Génération automatique de résumés audio musicaux

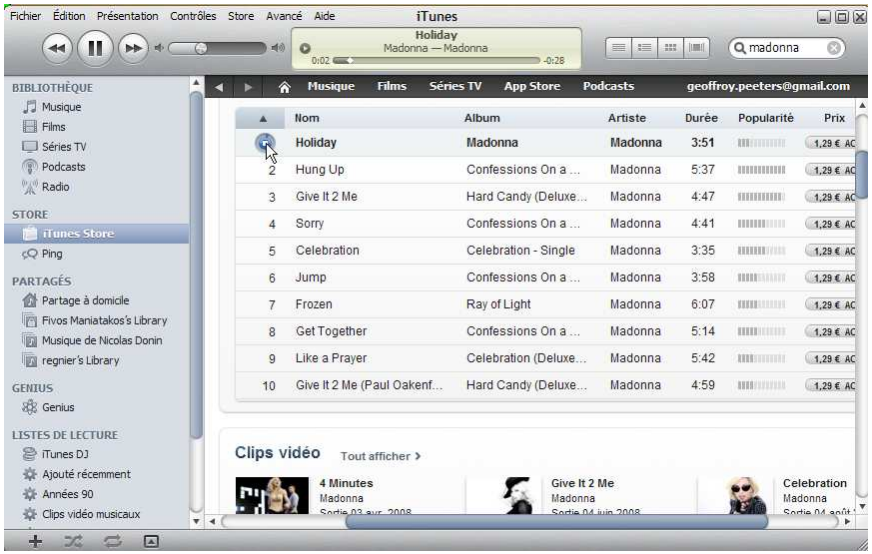
Journée GDR-ISIS, ATALA, AFCP
Résumé Automatique Multimédia
17 mars 2011

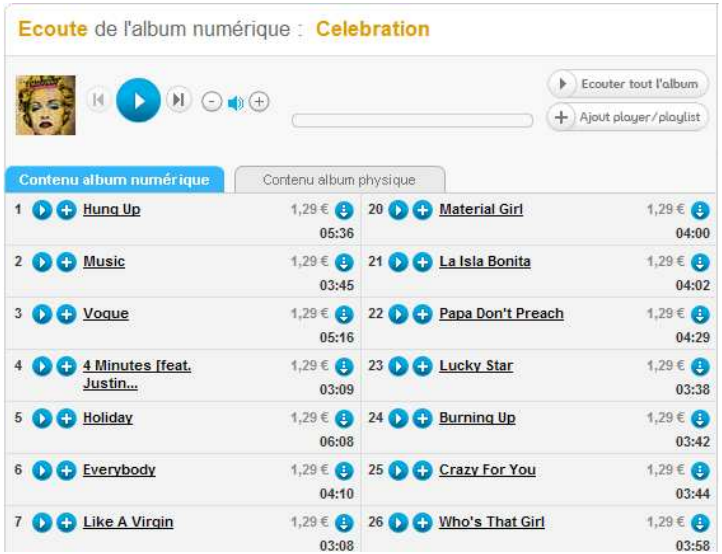
G. Peeters
A. Laburthe, F. Mislin, A. Wronecki, E. Deruty, X. Rodet

ircam
Centre Pompidou

Présentation des Use Cases

- Use Case 1: Résumé audio
  - Remplacement de l'extrait 30s continu de type iTunes ou FNAC



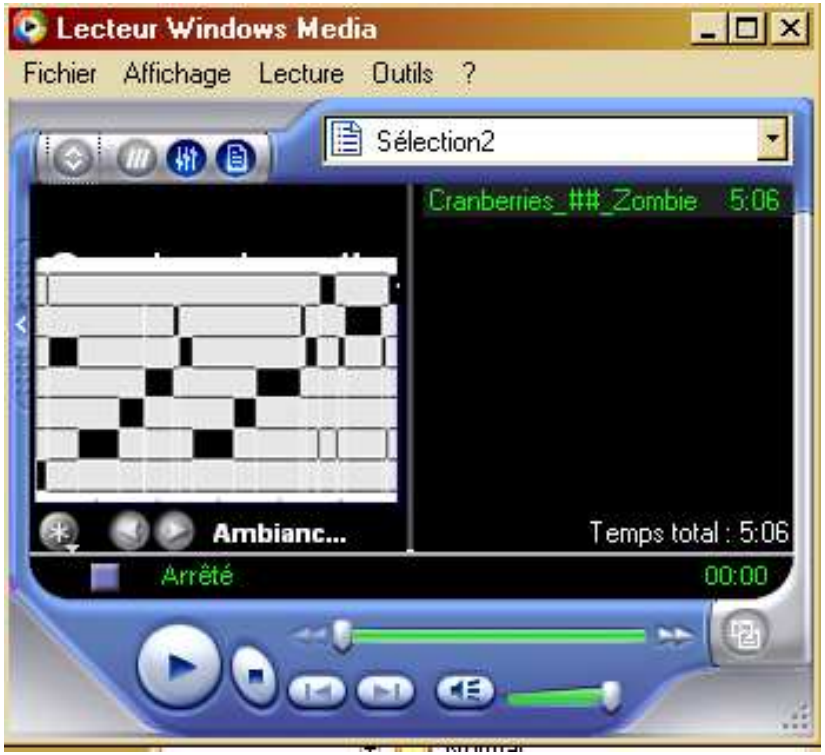


quaero

ircam Centre Pompidou

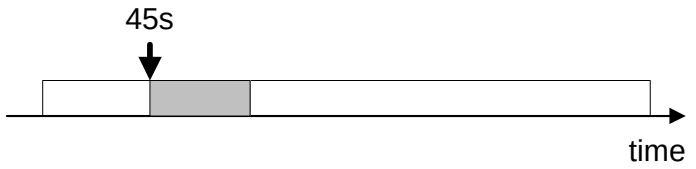
Présentation des Use Cases

- Use Case 2: Structure temporelle
  - Affichage et interaction avec la structure d'un morceau: navigation intra-document, chapitrage musicale
    - PAS AUJOURD'HUI

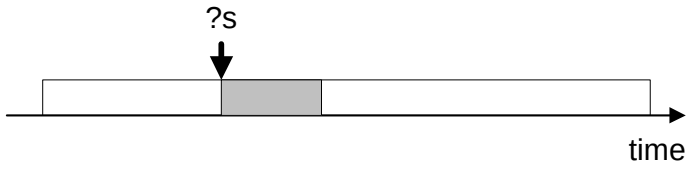


Différentes choix possibles pour le résumé

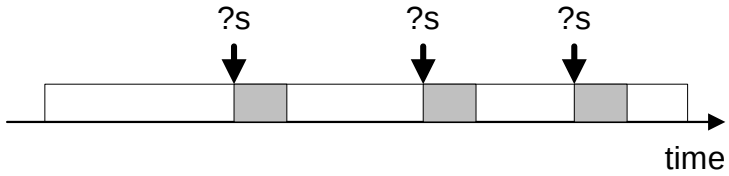
- Choix actuel sur les sites
  - 30s au début du morceau
  - 30s à partir de la 45èm sec.



- Meilleurs choix ?
  - Extrait unique, « le plus représentatif »



  - Suite d'extraits, « les plus représentatifs »

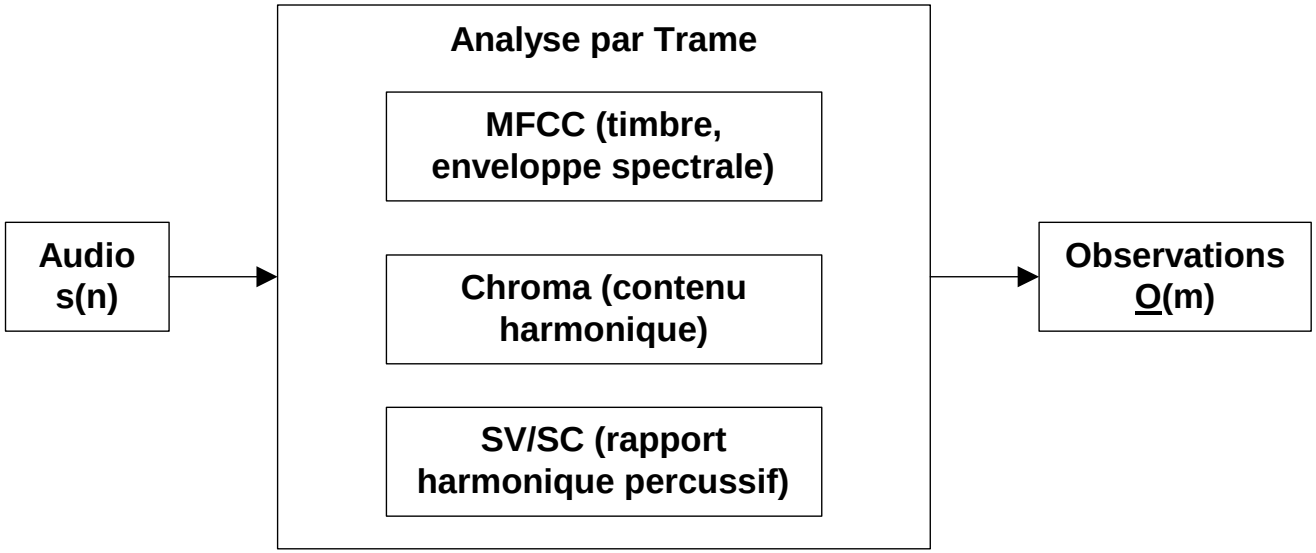


Comment ça marche ?
Détection de répétitions

- Principe de base:
  - Détection de répétitions au cours du temps dans un morceau de musique (couplet, refrain, …)
  - Utilisation pour
    - Estimation du/des points clefs sur base de répétitions
    - Estimation de la structure sur base de répétitions

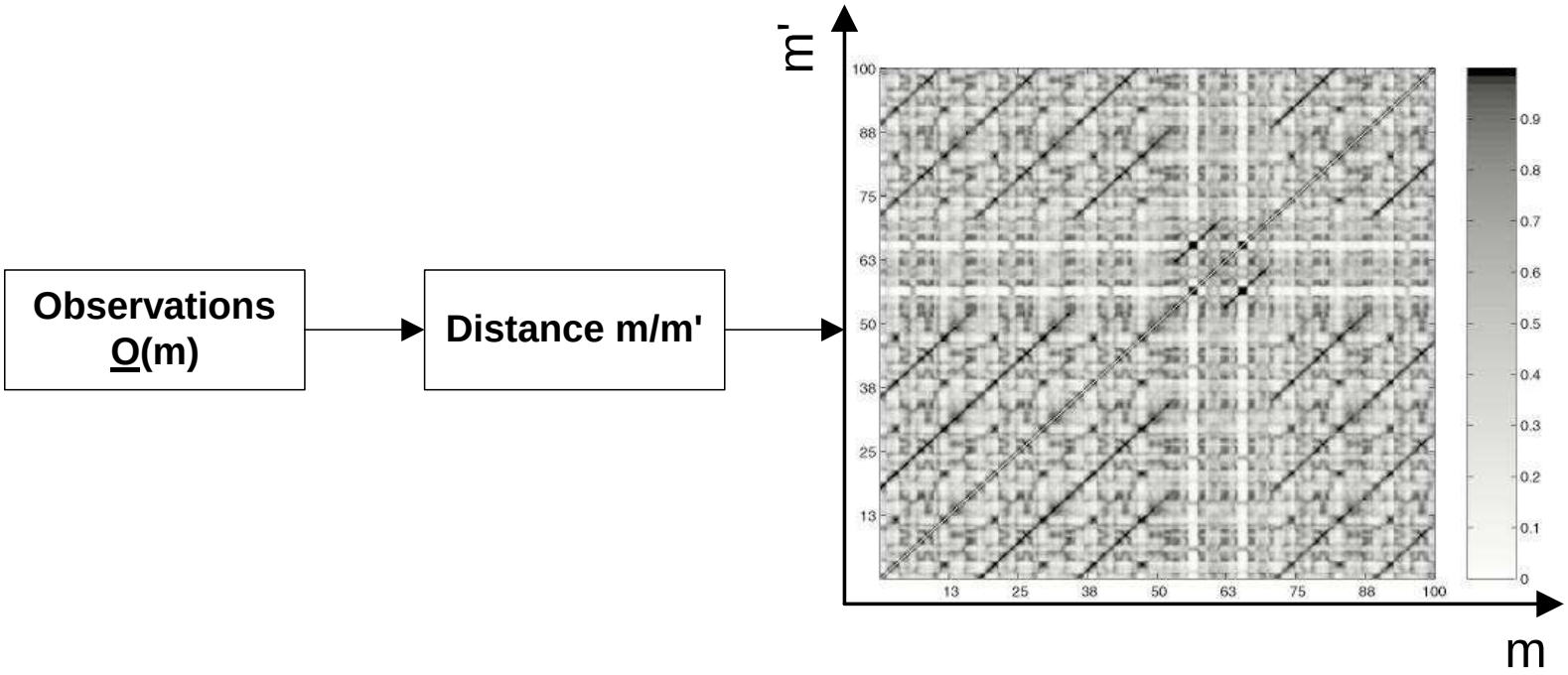
Comment ça marche ?
Détection de répétitions

- Observations du signal audio ?
  - Observations instantanées (60ms/20ms)



Comment ça marche ?
Détection de répétitions

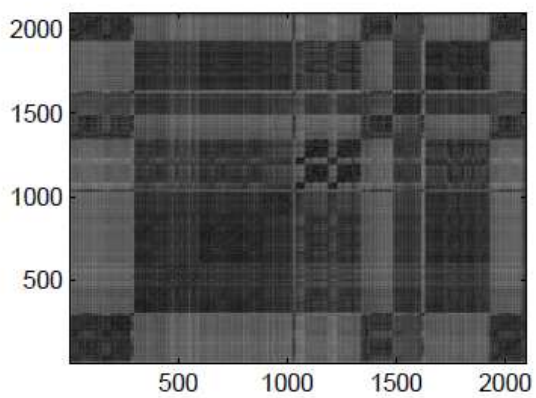
- Détection de répétitions ?
  - Calcul de la similarité entre deux instants m et m'
  - Distance cosinusoidale entre les observations à l'instant m et m':
    - d (O(m), O(m'))
  - Stockage dans une matrice d(m,m') :
    - matrice de similarité ou de co-occurrence

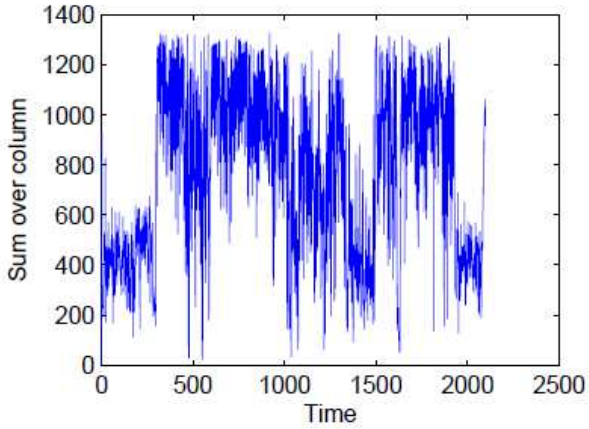


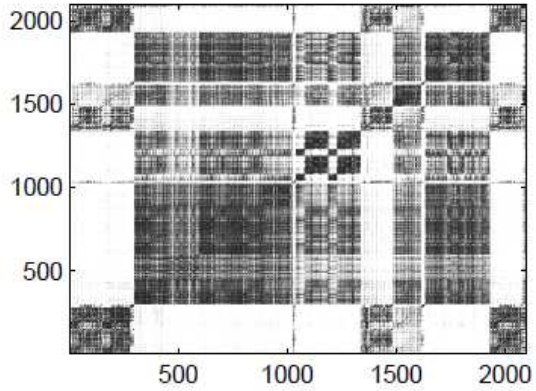
Extrait unique,
« le plus représentatif »

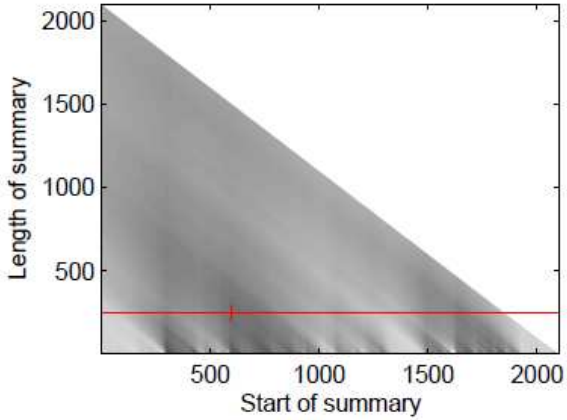
Comment ça marche ?
Choix de l'extrait le plus représentatif

- On cherche la tranche de la matrice
  telle que ses instants soient les plus répétés au cours du temps









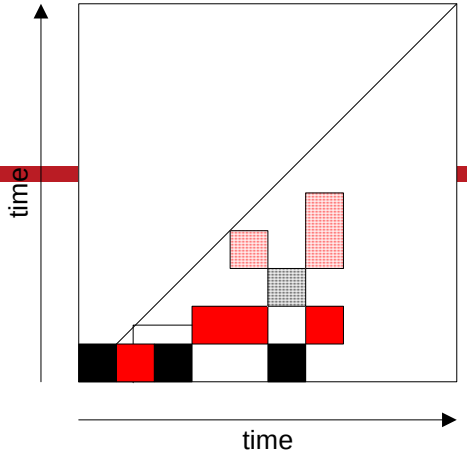
quaero

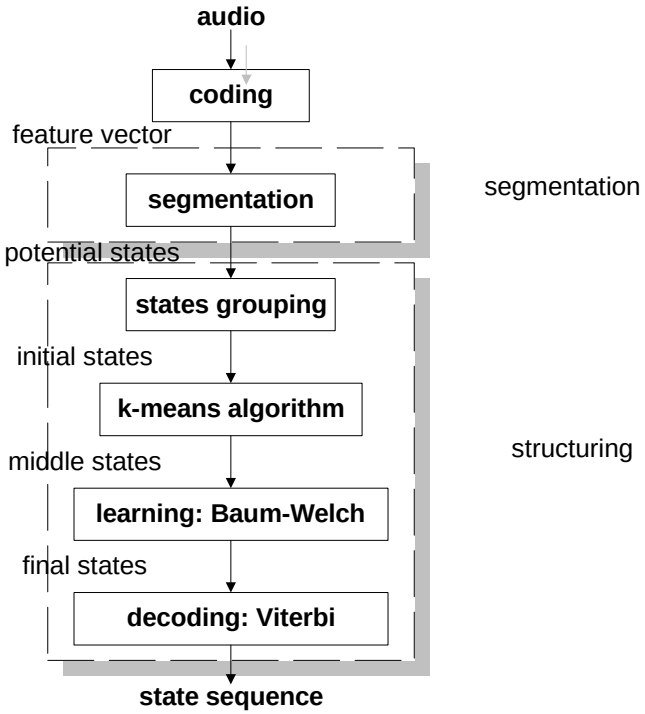
ircam Centre Pompidou

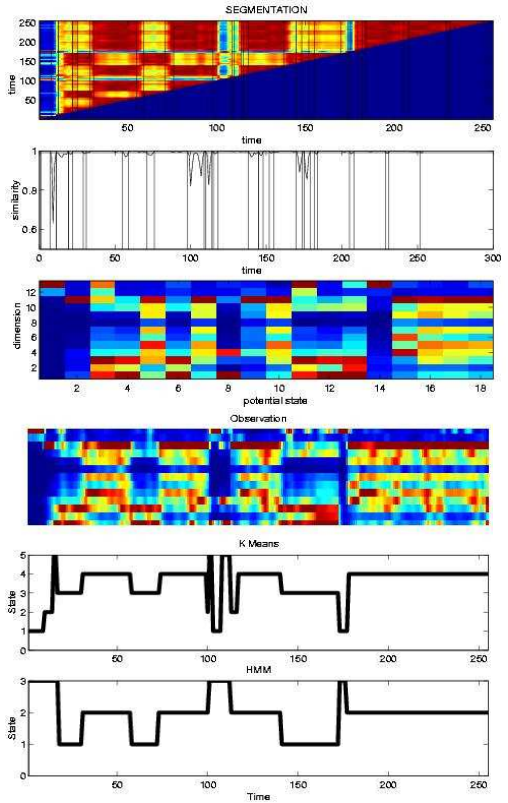
Suite d’extraits,
« les plus représentatifs »

Comment ça marche ? Estimation de la structure

- Approche par état
  - Cadre non-supervisé
  - [Peeters 2002]: Fuzzy-KMeans / hidden Markov model

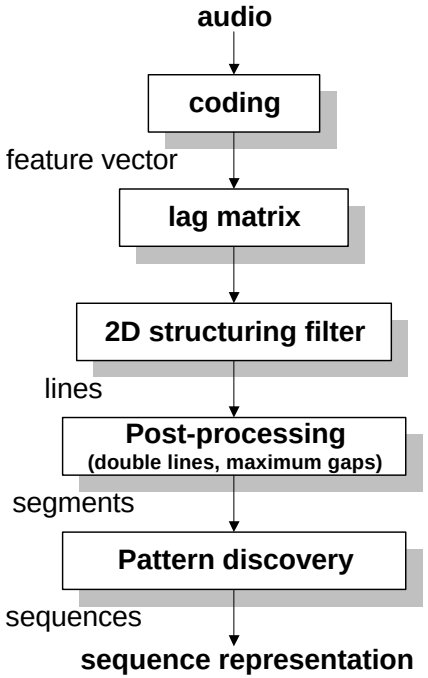


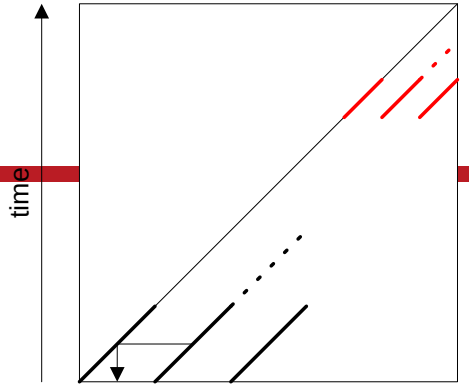


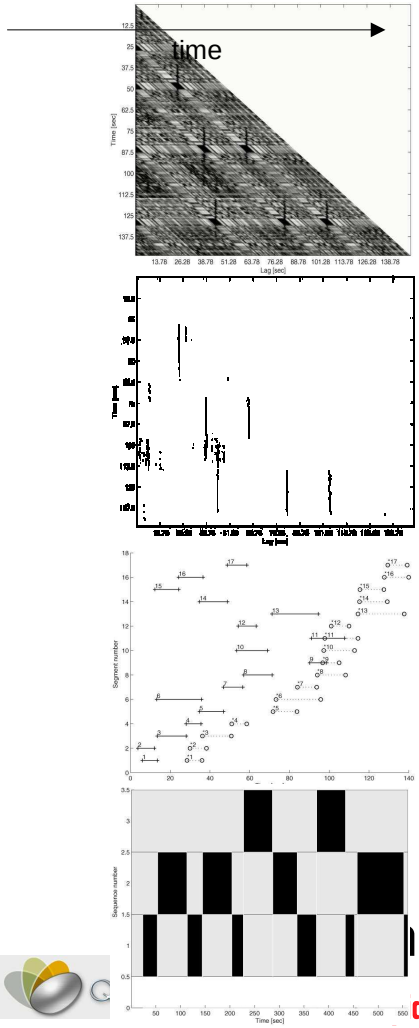


Comment ça marche ? Estimation de la structure

- Approche par séquence
  - Cadre non-supervisé
  - [Peeters 2003]: Analyse image de la matrice de similarité, heuristiques pour l'appariement des diagonales trouvées
  - [Peeters 2007]: Matrice de similarité d'ordre supérieur, approche par maximum de vraisemblance



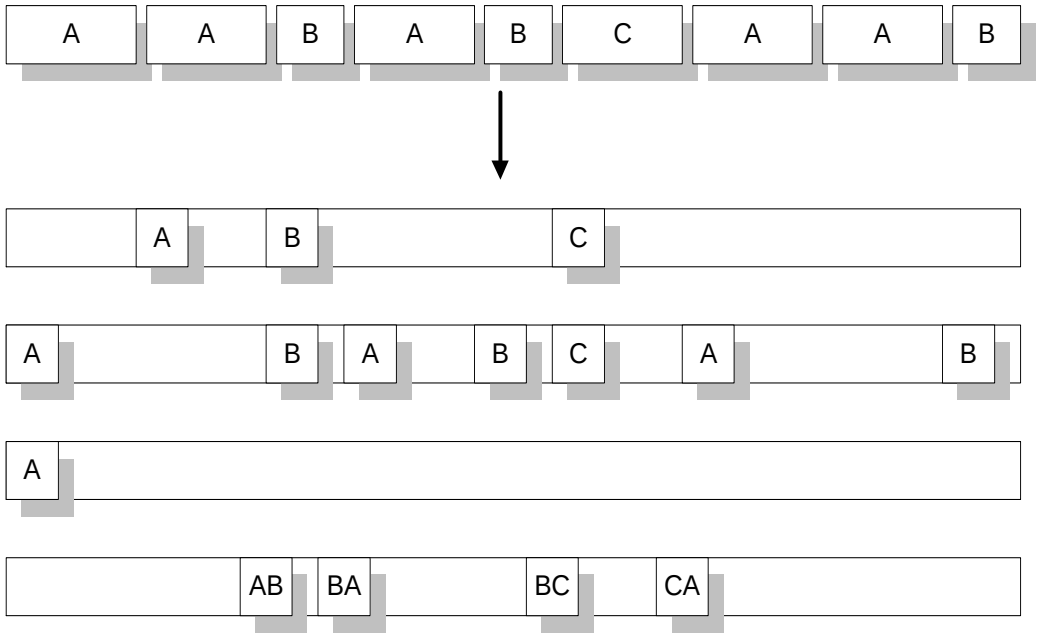




Création du résumé à partir de la structure

Choix des extraits les plus représentatifs

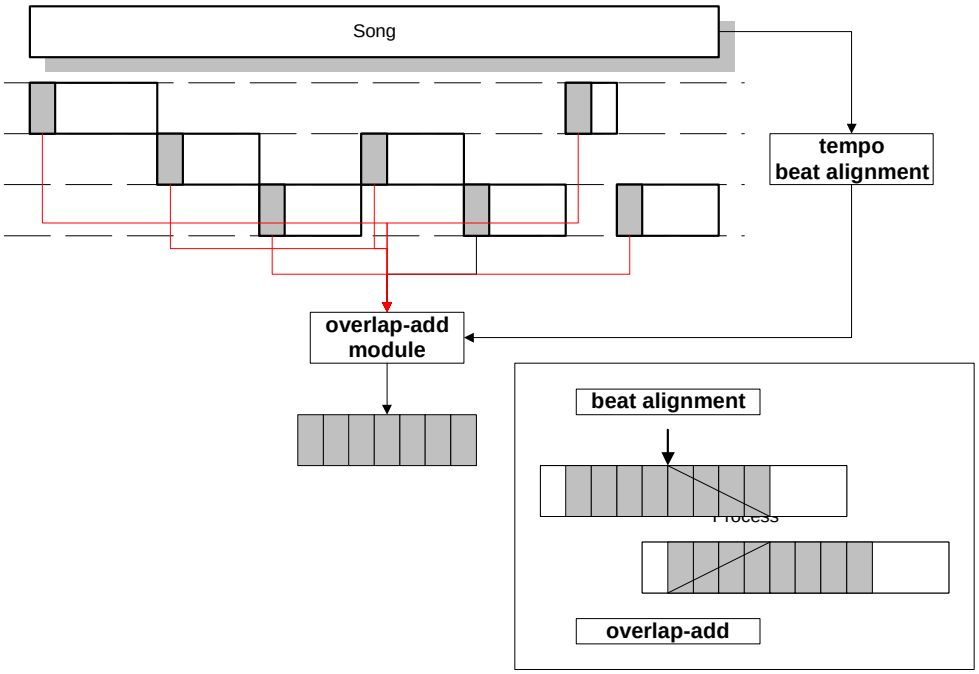
- Choix des extraits



- Un pour chaque partie
- Toutes les parties
- La partie la plus représentative
- Les transitions

Choix des extraits les plus représentatifs

- Techniques de montages
  - Cross-fade down-beat synchrone
    - Musique: structuration temporelle suit les battements, 1er temps, suite d'accords, …



Comment évaluer ?

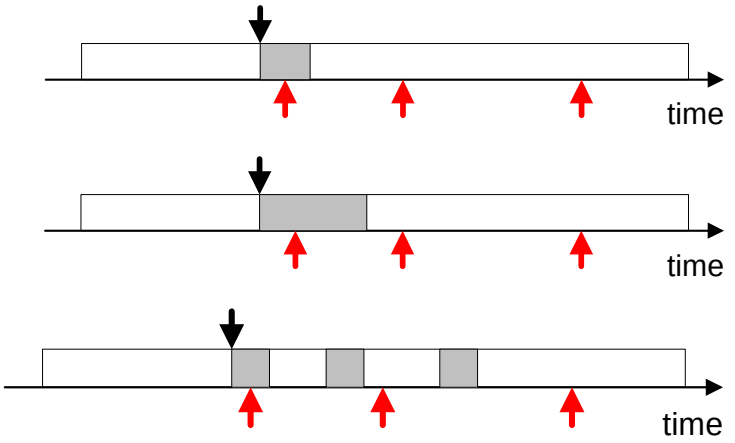
Evaluation

- Evaluation structure:
  - Comparaison:
    - Annotation/Estimation de la structure
  - Mesures de segmentation:
    - Recall/Precision/F-Measure
  - Labeling:
    - Frame Pair-Wise Clustering, Normalized Entropy (Sover, Sunder)

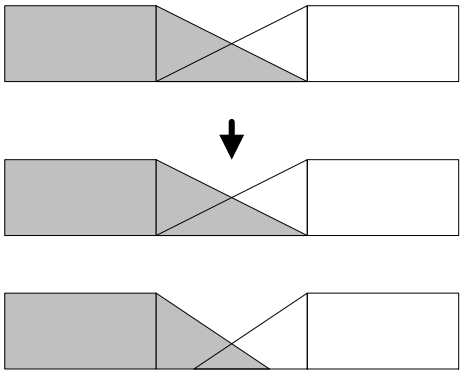
- Evaluation résumé audio
  - Contenu informatif:
    - est-ce que le chorus/titre du morceau est dans le résumé
  - Qualité de la présentation:
    - est-ce que le cross-fade dérange l'utilisateur

Evaluation

- Evaluation résumé audio
  - Contenu informatif:
    - est-ce que le chorus/titre du morceau est dans le résumé ?
    - Test-Set 150 popular top-ten hits
  - Résultats
    - 15s: 77%
      15s=snapshot summary
    - 30s: 90%
    - 3x10s: 95%



  - Qualité de la présentation (*)
    - Cross-fade total downbeat synchrone:
      - … Certains montages ressemblent de trop à un vrai morceau
    - Idem + Insertion de beep aux transitions
      - … Dérangeant
    - Idem+ seulement cross-fade partiel
      - … Mieux
    - Cross-fade partiel avec suivi visuel



(*) Résultats complets voir Orange-Labs (M. Vian, V. Botherel)

References

- G. Peeters, « Sequence representation of music structure using higher-order similarity matrix and maximum-likelihood approach », ISMIR (International Symposium for Music Information Retrieval) - NaN Vienna, Austria 2007 paper presentation

- G. Peeters, « Deriving Musical Structures from Signal Analysis for Music Audio Summary Generation: Sequence and State Approach » Springer Verlag - Lectures Notes in Computer Science, 2004, Volume 2771/2004 U. K. Wiil, Springer-Verlag Berlin Heidelberg 2004 paper

- G. Peeters « Audio summary » Patent - FR04/01493, 2004/06/16, 2003

- G. Peeters, A. Laburthe, X. Rodet « Toward Automatic Music Audio Summary Generation from Signal Analysis » ISMIR (International Symposium for Music Information Retrieval) - NaN Paris, France 2002 paper