Structuration audiovisuelle par composantes primaires

Plate-forme AFIA

Atelier : Connaissance et Documents Temporels

Julien Pinquier et Régine André-Obrecht

Équipe SAMoVA

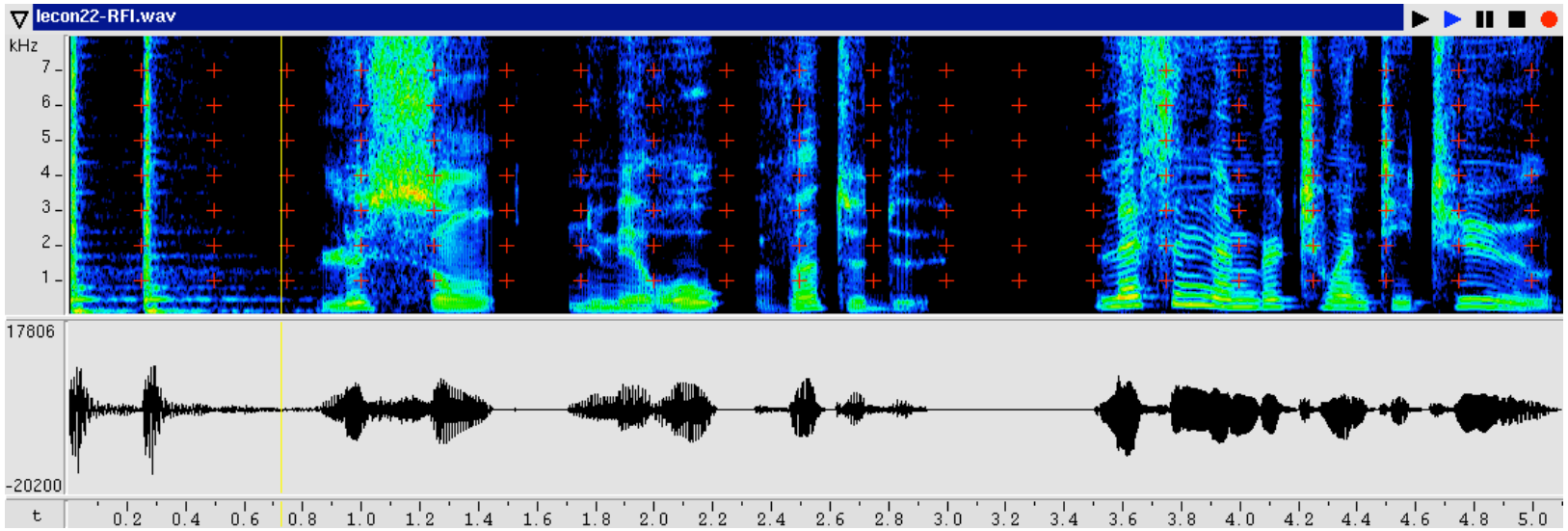
(**S**tructuration **A**nalyse et **Mo**délisation de la **V**idéo et de l'**A**udio)

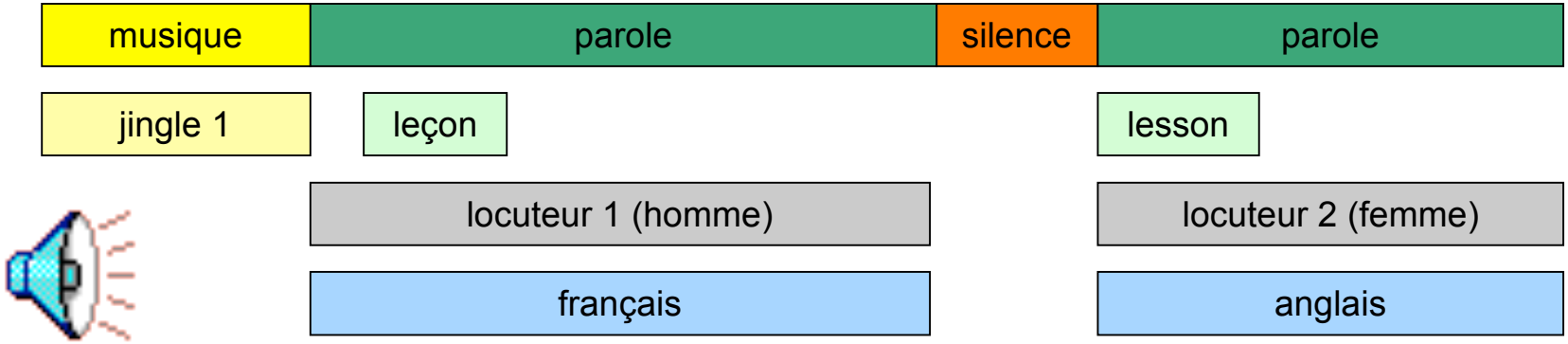
SAMoVA

IRIT CNRS INPT UPS UT1

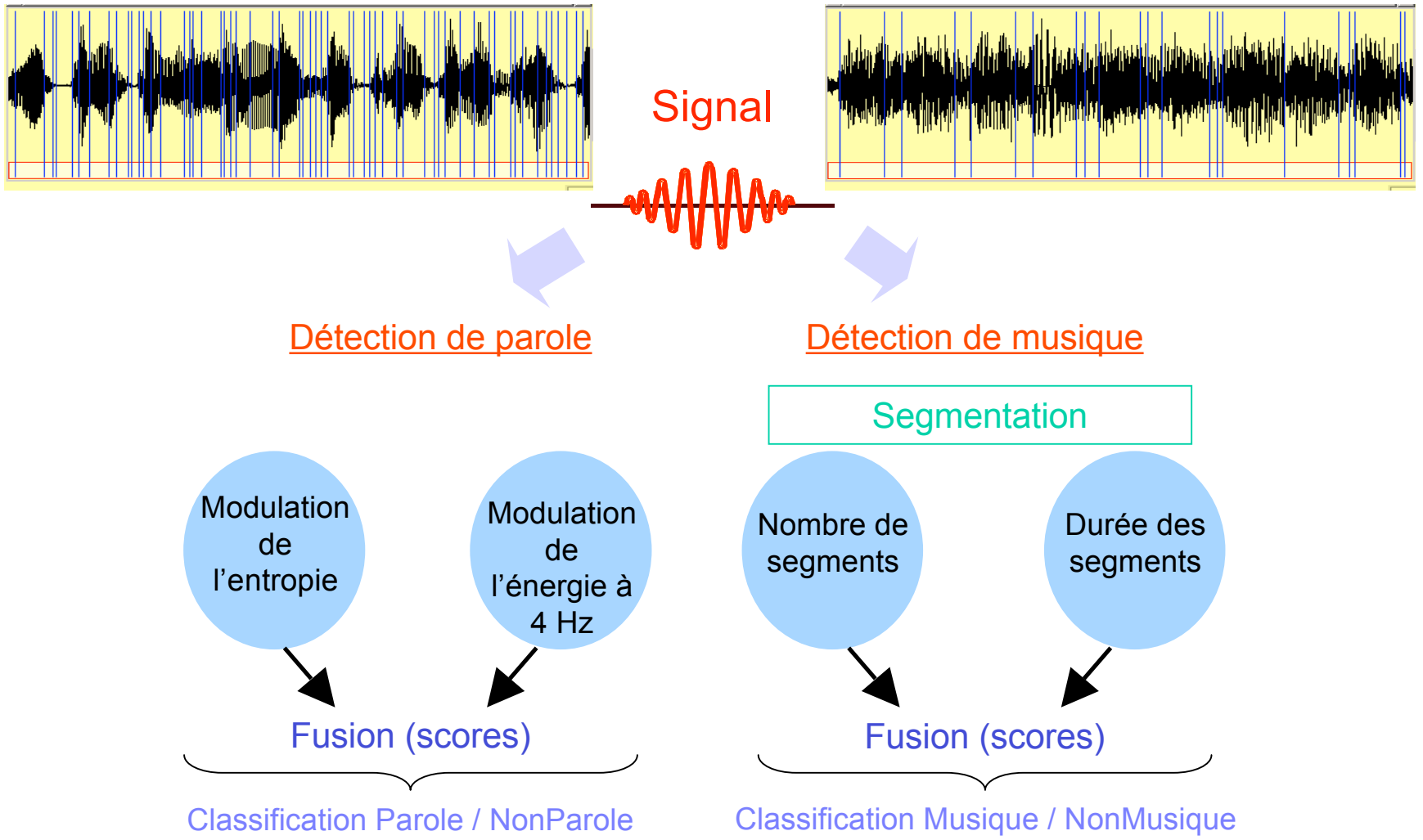
Institut de Recherche en Informatique de Toulouse

Indexation sonore : que faire ?





Détection PMB

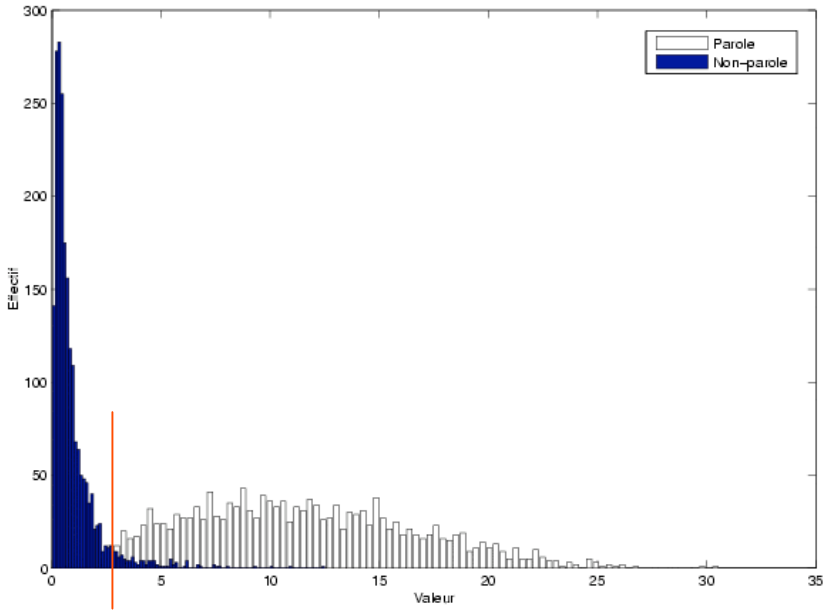


Détection PMB

- Les seuils
  - Parole : corpus MULTEXT [Campione98]
  - Musique : base personnelle

Exemple :

Modulation de l'énergie à 4 Hertz



Seuil

Détection PMB : résultats

Paramètres	Score

PAROLE

Modulation de l'énergie à 4 Hertz Modulation de l'entropie	87,3 % 87,5 %
Fusion (max)	90,5 %

CORPUS RFI (6 heures)

Campagne d'évaluation ESTER (2004-2005) :
1er rang sur 7 participants (tâche SES)

MUSIQUE

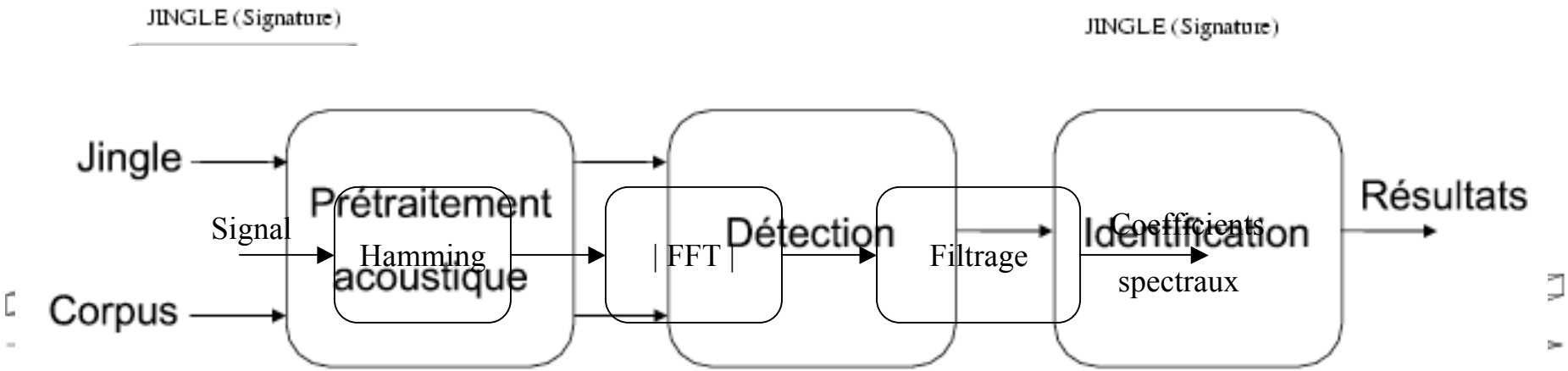
Nombre de segments Durée des segments	86,4 % 78,1 %
Fusion (max)	89 %

Sorties outils : Format FDL

Détection de jingles

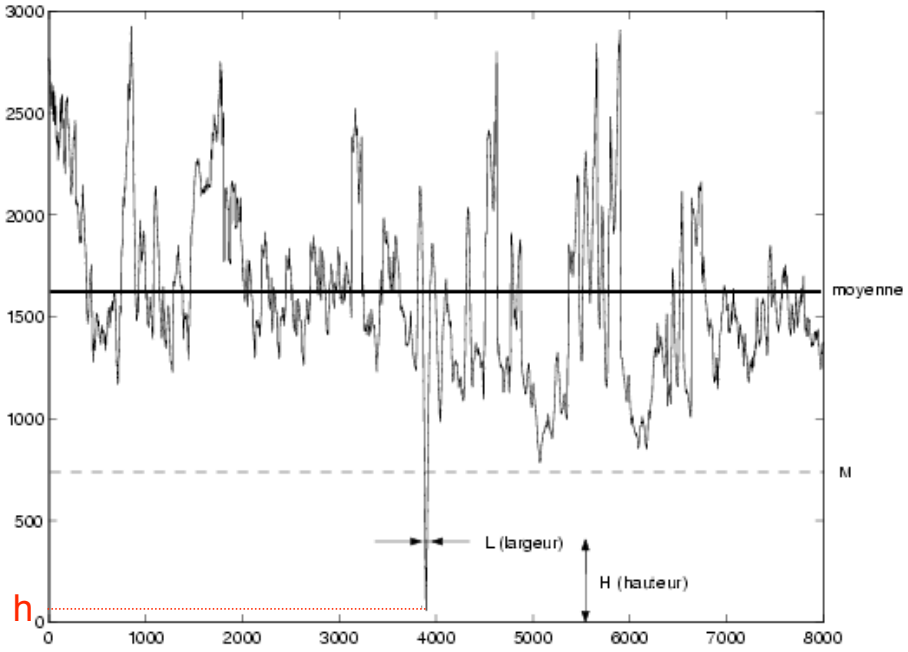
- Extrait sonore

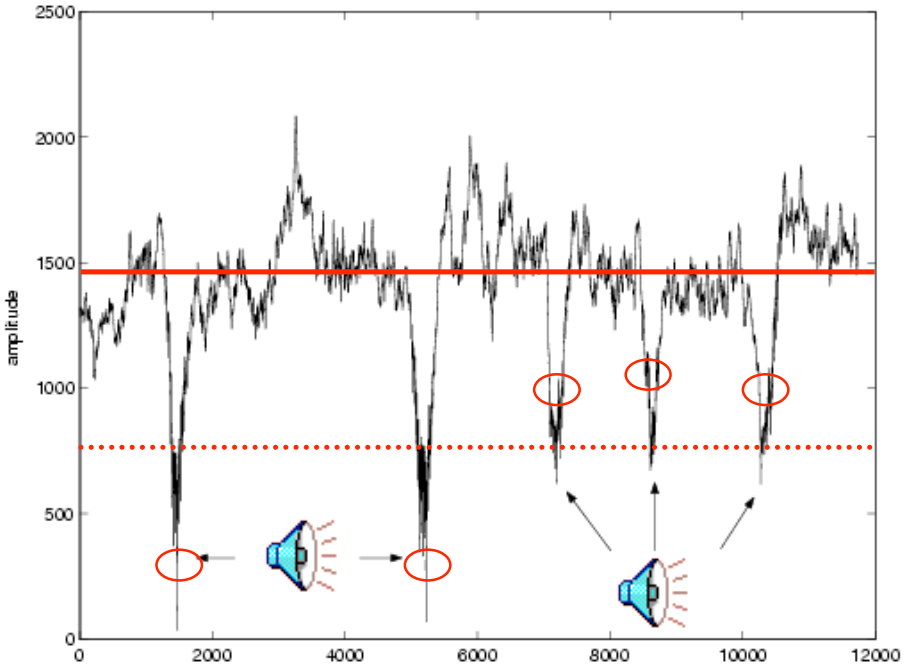
- Système
  - Analyse spectrale (29 coefficients)
  - Comparaison (distance Euclidienne)
  - Analyse des « pics »



Détection de jingles

✧ Méthode d'analyse des pics

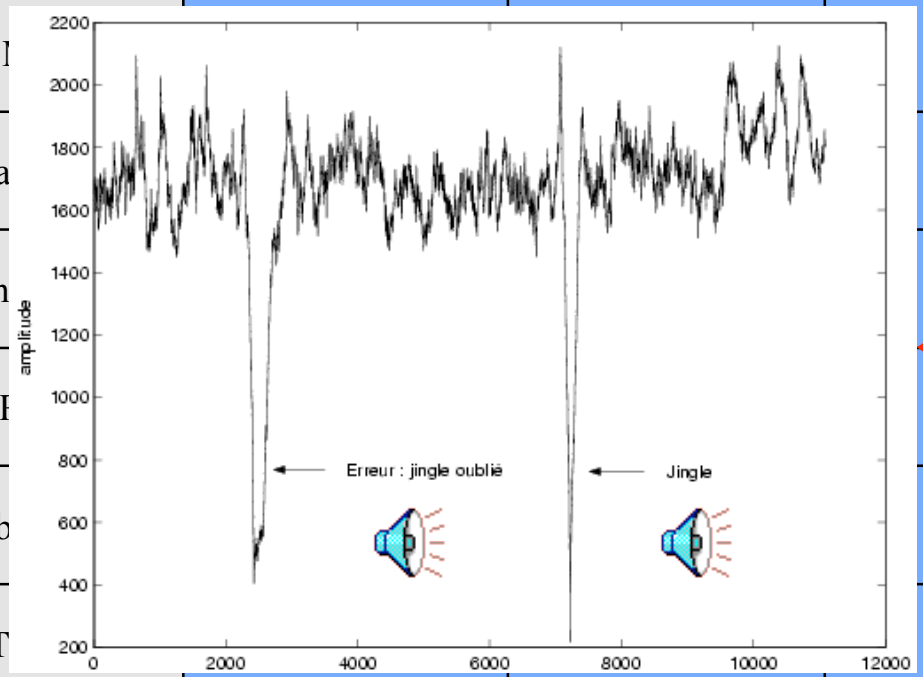




Détection de jingles

✧ Résultats

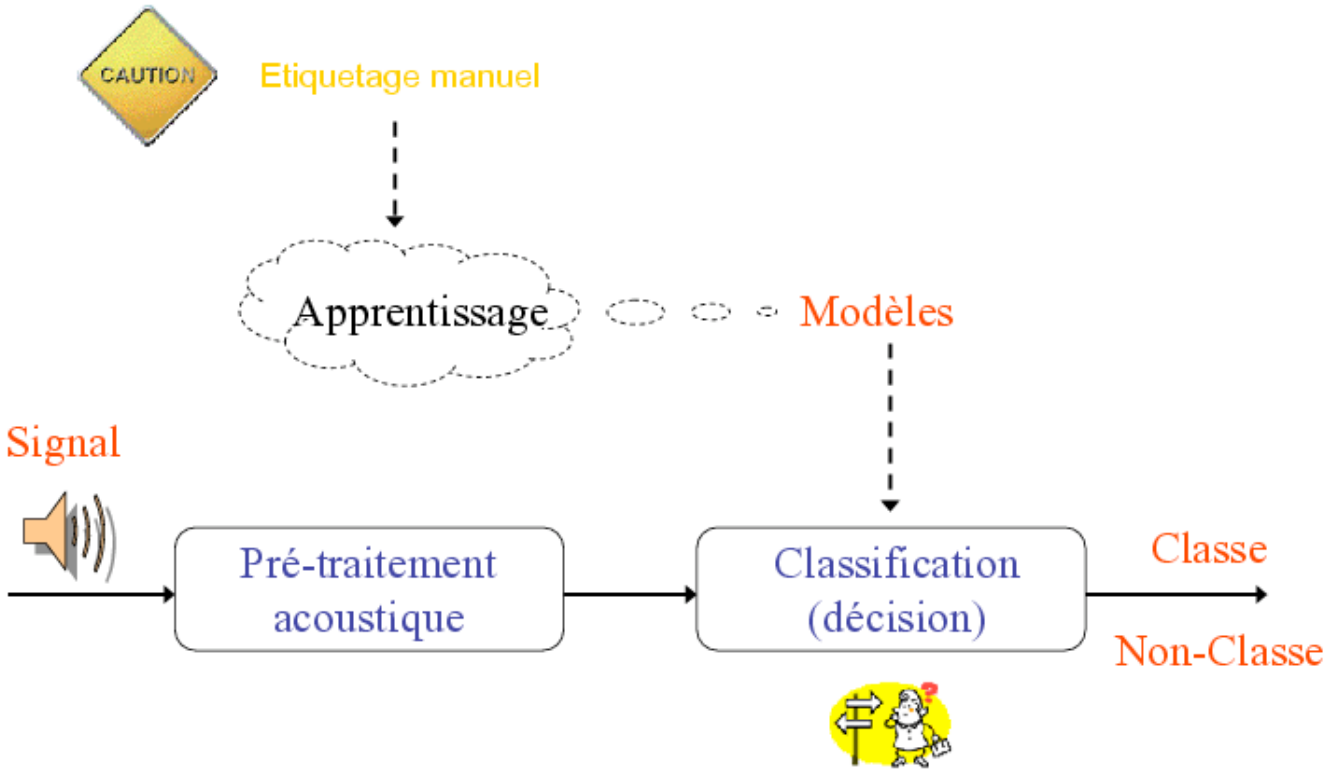
Corpus	Durée	Jingles	Détection manuelle
France 3	15 min	1	4
I			16
Ca			6
Fran			12
F			60
Pub			34
T			**132**



France Info

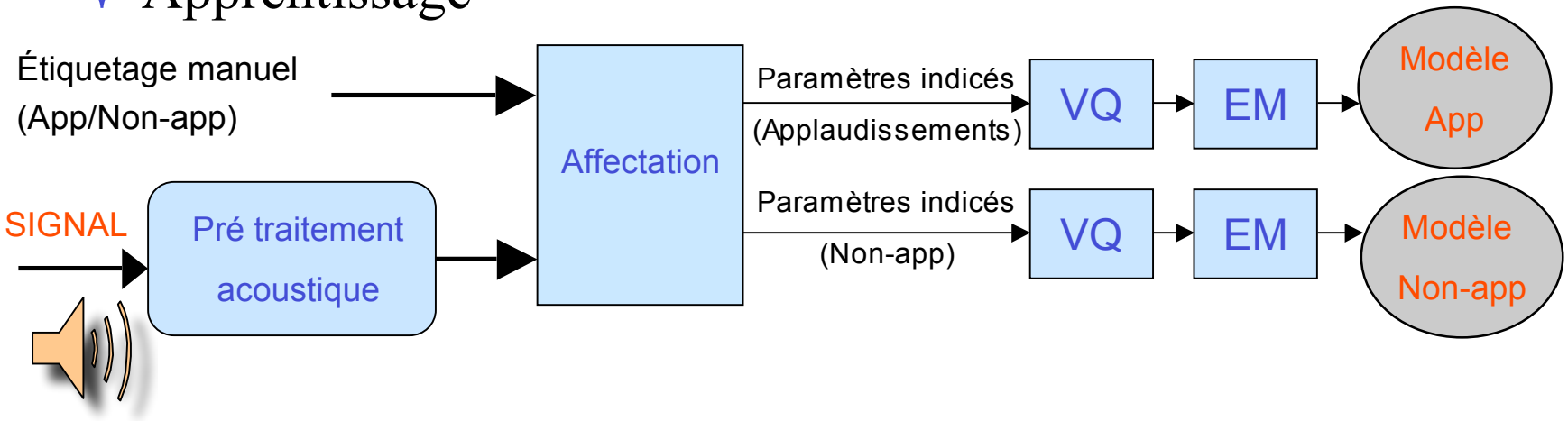
Détection des applaudissements

- Système



Détection des applaudissements

✧ Apprentissage



✧ Corpus : « Le Grand Échiquier », projet FERIA

- Apprentissage : 1 émission
- Reconnaissance : 1 émission

✧ Résultats

- Problème du critère évaluation : taux de reconnaissance > 98 %
- Segments significatifs : excellents

Structuration sonore

- Détection de motif dans une collection d'émissions
  - « Le grand Échiquier » 54 émissions de 3h
  - Motif :
    → présentateur / [APP] / spectacle / [APP/spectacle] / APP / présentateur
  - Détections automatiques :
    → détection de musique (chansons, spectacle)
    → détection de parole, puis du présentateur
    → détection des applaudissements
  - Apprentissage : 3 min (présentateur) et 6 min (applaudissements)
  - Résultats : une émission → détection de 10 motifs
  - Application aux autres émissions de la collection (vérité terrain)

Structuration sonore

- Macrosegmentation automatique (exemple du motif)
  - Annotations automatiques
  - Recherche de suites récurrentes [Haidar04]
  - Inférence d'un motif
  - Structuration

→ Important : difficile manuellement

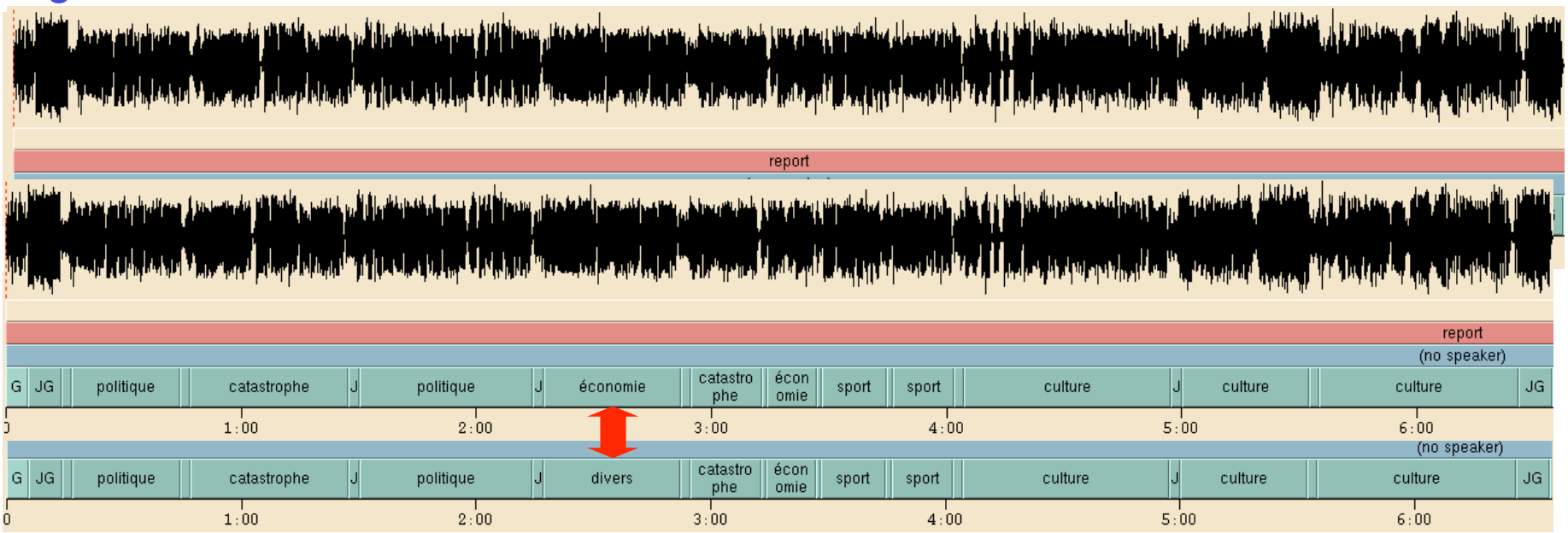
Conclusion

- Indexation sonore : étude de composantes primaires
- « Unités communes »
  - Parole et musique : → robustesse (plus d'apprentissage)
- « Unités exotiques »
  - Jingles : résultats excellents → 1 occurrence
  - Applaudissements : résultats très bons → universel
- Étude de structuration sonore
  - Détection d'un motif → très intéressante

Bonus : une autre étude de structuration

- Structuration d'un journal télévisé (« 6 minutes » de M6)
  - Détection de jingles (J et JG)
  - Détections de parole et de musique
  - Détection de mots clés → thème du reportage
  - 1 erreur

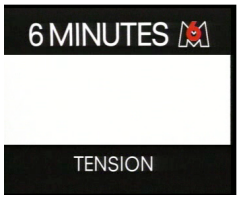
J

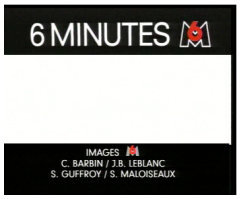


Perspectives

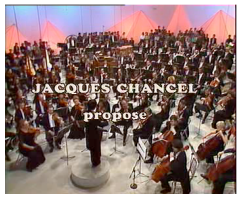
✧ Apport de la vidéo

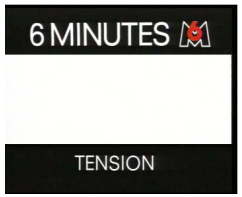
- Détection de logos

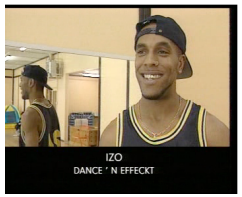




- Extraction de texte







- Reconnaissance de l'intervenant

⇨ Chanteur ?

Perspectives

- Caractérisation de l'intervenant
  - Voix IN, OUT, OFF
  - Vidéo : mouvements

    → Corps, tête, lèvres
  - Audio : conditions enregistrements