

Segmentation et regroupement de locuteurs

Campagne ESTER

Le 12 mars 2004

Mikaël COLLET

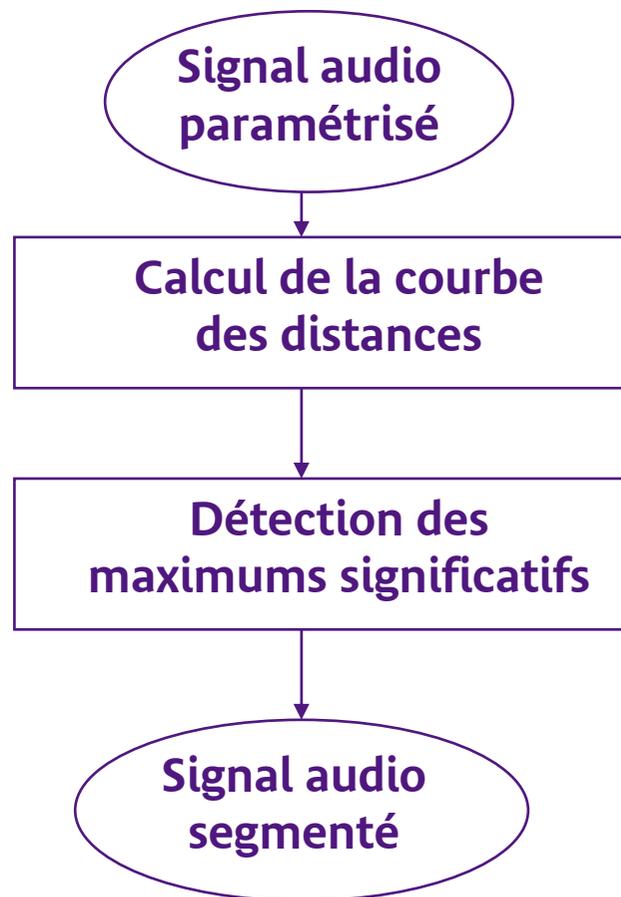
Le présent document contient des informations qui sont la propriété de France Télécom. L'acceptation de ce document par son destinataire implique, de la part de ce dernier, la reconnaissance du caractère confidentiel de son contenu et l'engagement de n'en faire aucune reproduction, aucune transmission à des tiers, aucune divulgation et aucune utilisation commerciale sans l'accord préalable écrit de France Télécom R&D

Segmentation et regroupement de locuteurs



- ▶ **Module de segmentation**
- ▶ **Module de regroupement**
- ▶ **Limitations du système**
- ▶ **Perpectives d'améliorations du système**

▶ Descriptif du module de segmentation

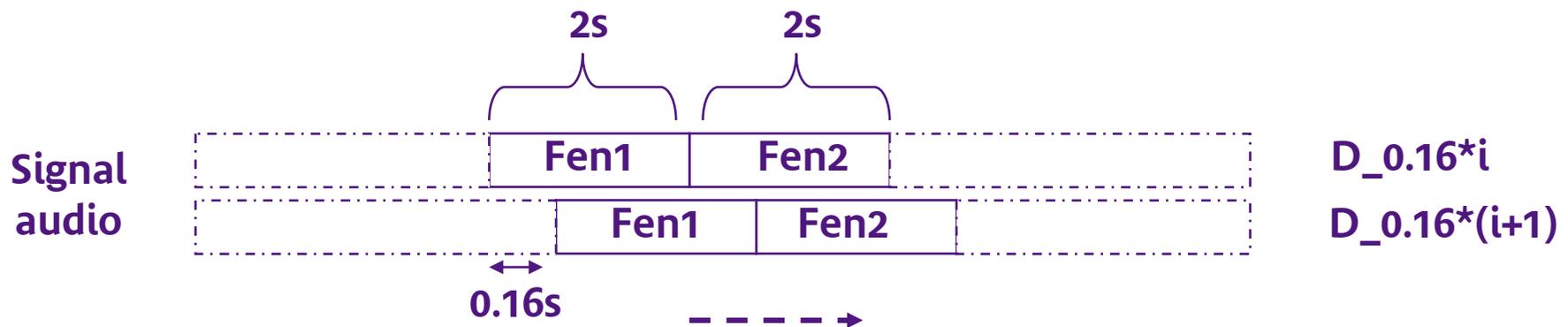


Segmentation et regroupement de locuteurs



▶ Descriptif du module de segmentation

- ▶ Analyse acoustique : $E + \Delta E + 13$ MFCC
- ▶ Calcul de la courbe des distances :
 - Calcul d'une distance entre deux fenêtres de signal audio paramétrisé
 - Longueur des fenêtres : 2 s
 - Décalage des fenêtres : 0.16 s
 - Métrique utilisée : Rapport de vraisemblance généralisé (GLR)

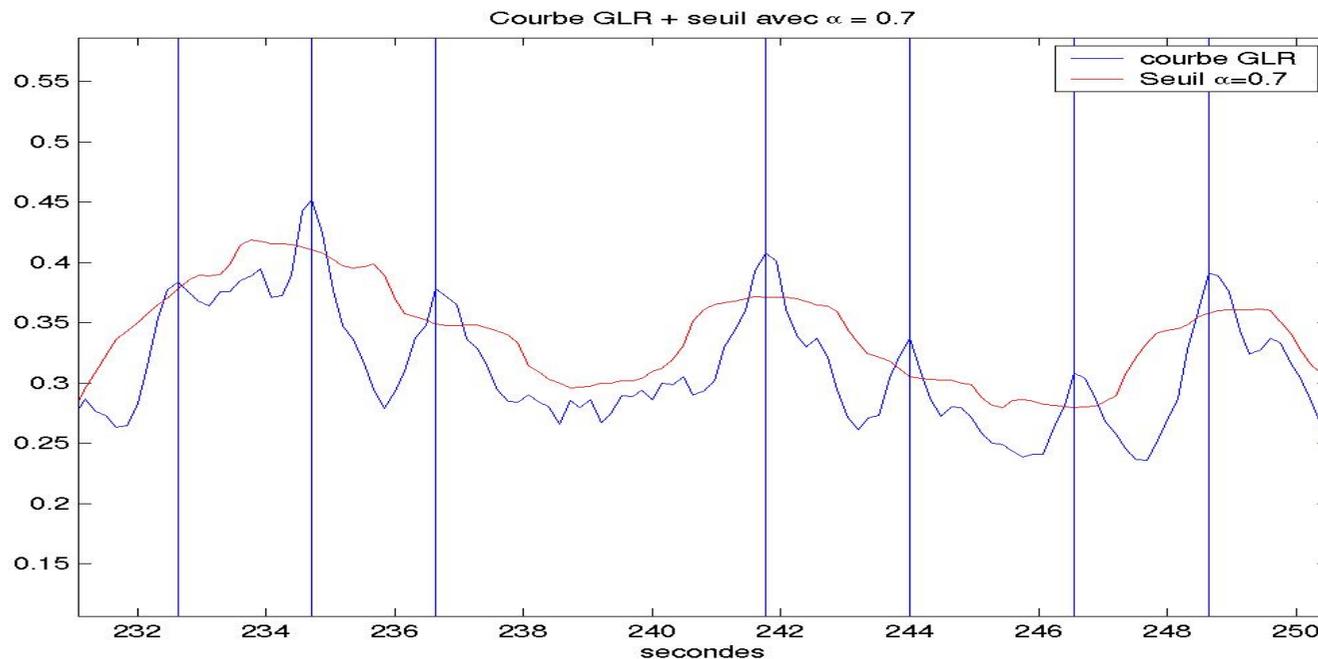


Segmentation et regroupement de locuteurs



▶ Descriptif du module de segmentation

- ▶ Détection des maximums significatifs :
 - Lissage de la courbe : moyenne sur une fenêtre de 3 échantillons.
 - Calcul d'un seuil : $\text{seuil} = \text{moyenne sur 15 points} + \alpha \times \text{écart type}$.
 - Changement de locuteur détecté lorsqu'un maximum se trouve au dessus du seuil





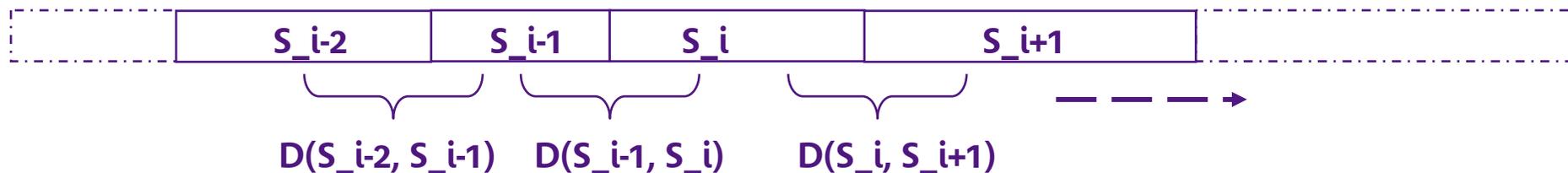
▶ Limitations du module de segmentation

- ▶ Taux de fausses alarmes très élevé
 - Nombre de segments très important
 - Longueur moyenne des segments : 2 s
- ▶ Conséquences :
 - Les mesures de similarité sont moins fiables sur des segments très courts.
 - Le regroupement de locuteurs est moins performant lorsque le nombre de segment est important.
- ▶ Solution : diminuer le nombre de fausses alarmes en regroupant les segments voisins qui semblent avoir été prononcés par le même locuteur.

▶ Diminution du nombre de fausses alarmes

- ▶ Regroupement des segments voisins qui semblent avoir été prononcés par le même locuteur.
 - Même méthode que précédemment : calcul d'une courbe des distances + détections des maximums significatifs.

Signal audio segmenté



► Descriptif du module de regroupement

- Utilisation de la technique des modèles d'ancrage :
 - Un segment de parole X de longueur N est représenté par un vecteur dans un espace de locuteurs de référence ($E = 250$ locuteurs).
 - Le vecteur de coordonnées est défini comme l'ensemble des scores de vraisemblance entre le segment de parole et les modèles des locuteurs de références $\bar{\lambda}_e$ (modèles d'ancrage).

$$w = \begin{bmatrix} \tilde{p}(X | \bar{\lambda}_1) \\ \tilde{p}(X | \bar{\lambda}_2) \\ \vdots \\ \tilde{p}(X | \bar{\lambda}_E) \end{bmatrix}$$

$$\tilde{p}(X | \bar{\lambda}_e) = \frac{1}{N} \log \frac{p(X | \bar{\lambda}_e)}{p(X | \lambda_{UBM})}$$

▶ Descriptif du module de regroupement

▶ Comparaison entre deux segments de parole (X et Y) par la technique des modèles d'ancrage.

- Distance euclidienne sur les vecteurs de coordonnées.

$$d(X, Y) = \sqrt{\sum_{i=1}^E (x_i - y_i)^2}$$

- Distance angulaire sur les vecteurs de coordonnées.

$$\delta(X, Y) = \arccos \left[\frac{x^T y}{\sqrt{x^T x \cdot y^T y}} \right]$$

▶ Descriptif du module de regroupement

▶ Regroupement hiérarchique ascendant

- Critère de regroupement de 2 classes : minimisation de la dispersion intra-classe.

$$I_C = \frac{1}{K} \sum_{i=1}^K D(X_i, G)$$

- . G est le centre de gravité de la classe C de K éléments.
 - . X un élément de la classe.
 - . D(X,G) est la distance entre l'élément X et le centre de gravité.
- Critère d'arrêt : regroupement des segments en un seule classe.
 - Critère d'élagage du dendrogramme : dispersion intra-classe supérieur à un seuil.

Segmentation et regroupement de locuteurs



▶ Résultats

Fichier	1	2	3	4	5	6	Total
Taux d'erreur	29,13	32,79	29,22	15,20	32,50	35,41	30,76

- ▶ Fichier 1 : 19981218_0700_0800_inter_fm_dga
- ▶ Fichier 2 : 19981218_0800_0900_inter_fm_dga
- ▶ Fichier 3 : 19980624_1900_1920_inter_fm_dga
- ▶ Fichier 4 : 19980625_1900_1920_inter_fm_dga
- ▶ Fichier 5 : 20000908_0930_1030_inter_fm_dga
- ▶ Fichier 6 : 20000908_1130_1230_inter_fm_dga



Limitations du système

- Problème lié à la normalisation des mesures de similarité.
 - Seuil de décision des changements de locuteur du module de segmentation.
 - Seuil de décision pour l'élagage du dendrogramme du module de regroupement.
- Les mesures de similarité ne sont pas fiables sur les segments courts (< 2 s).
- A l'heure actuelle, les modèles des locuteurs de référence sont appris à partir de données téléphoniques.



Perspectives d'amélioration du système

- Normalisation des mesures de similarité.
- Etude d'une nouvelle méthode de détections des changements de locuteurs permettant de diminuer le taux de fausses alarmes.
- Modification du critère d'élagage du dendrogramme.
- Création d'un nouvel espace de locuteurs de référence à partir de données radiophoniques.