

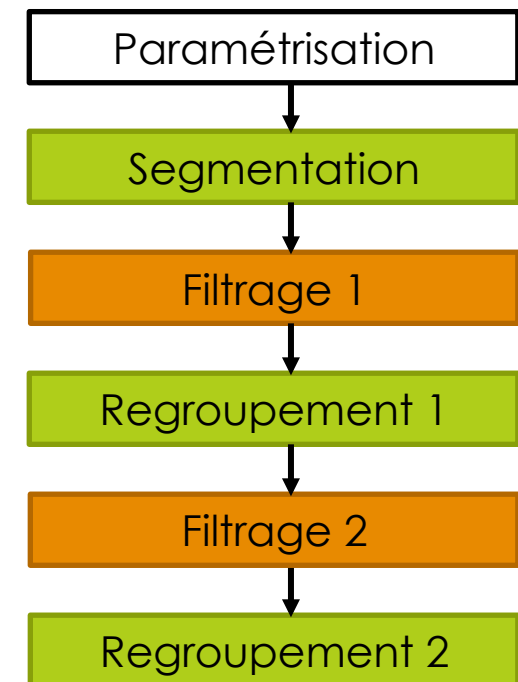
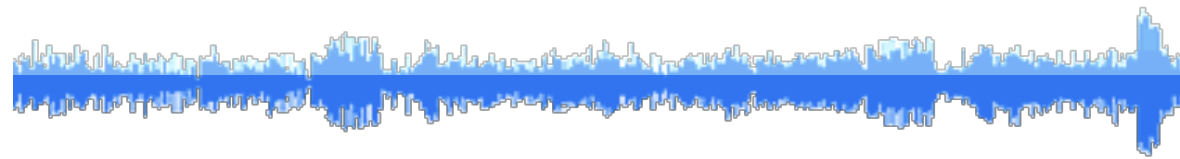


Speaker Diarization

Gregor Dupuy, Sylvain Meignier et Mickael Rouvier

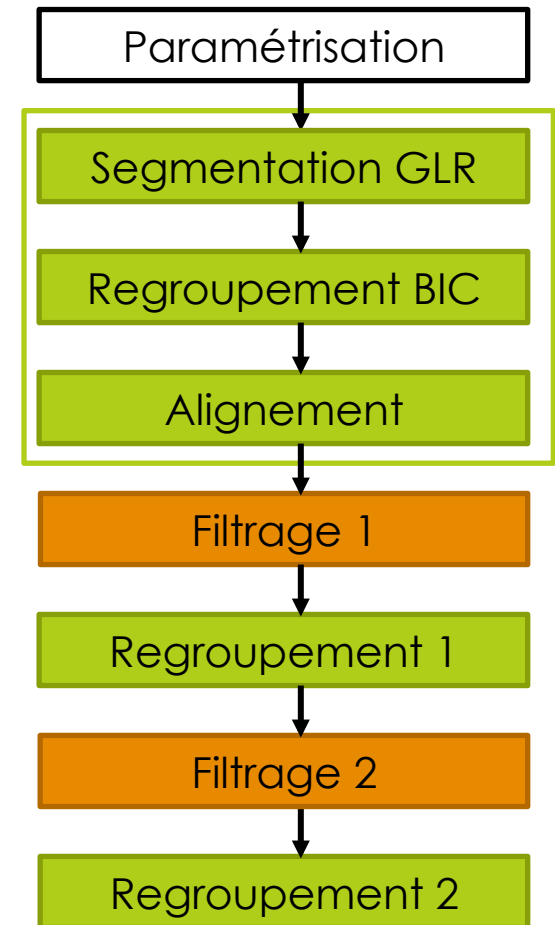
LIUM, University of Maine, France – Le Mans

Speaker Diarization



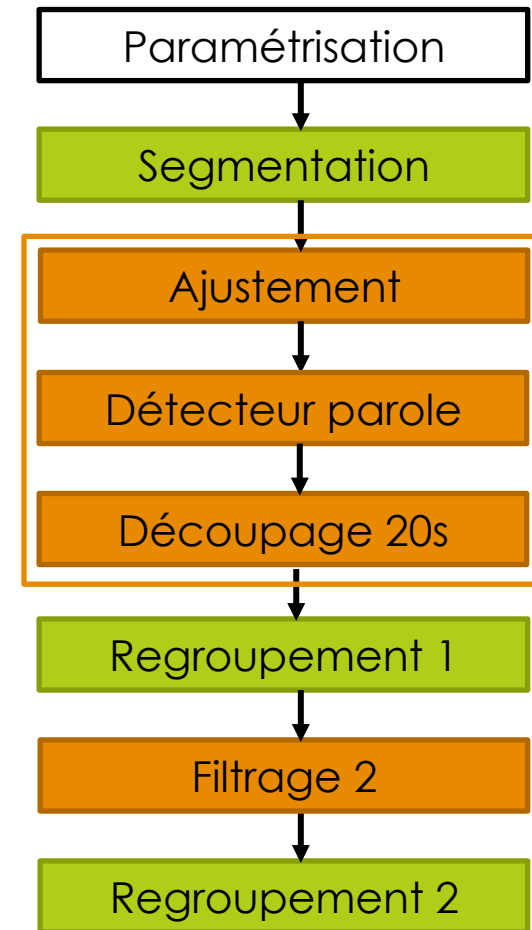
Speaker Diarization : Segmentation

- Segmentation GLR
- Regroupement BIC
 - Contraint : sur les segments contigus
 - Non contraint : sur l'ensemble des segments
- Alignement viterbi
 - HMM, 1 état = 1 classe = 1 locuteur
 - 8-GMM



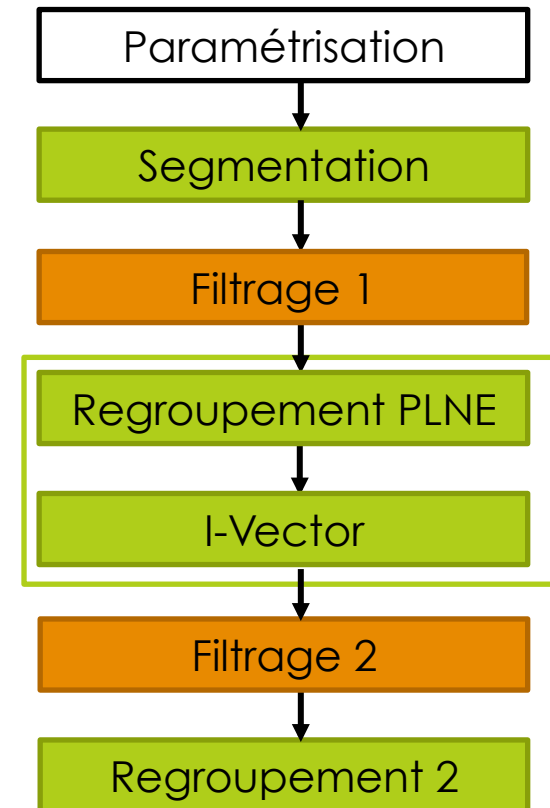
Speaker Diarization : Filtrage 1

- Ajustement des frontières en utilisant l'énergie
- Détection des zones de parole
 - 8 GMM à 64 gaussiennes diagonales
 - 12MFCC + Δ
 - Alignement + Filtrage contraint
 - Rogner les silences : – 25 frames
 - Pas de silence de < 25 frames
 - Pas de segment < 150 frames
- Découpage en segments de 20s maxi.
 - Utilise un détecteur de silence



Speaker Diarization : Regroupement 1

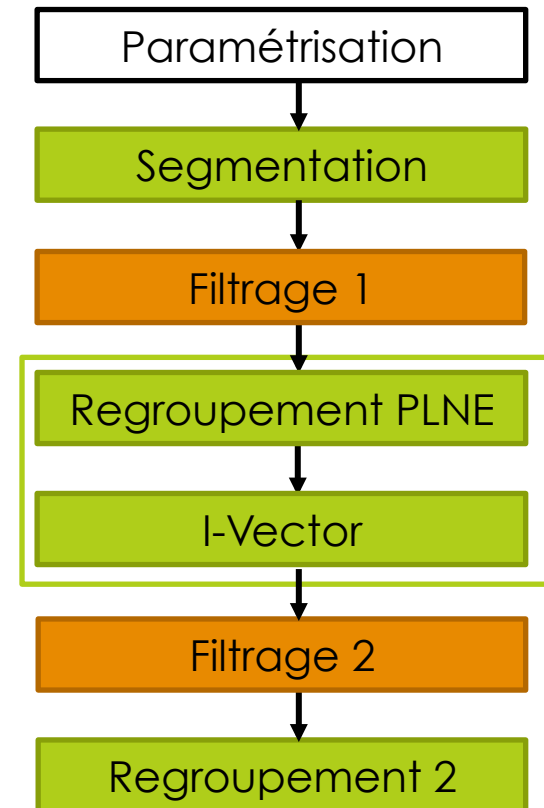
- Problème : La nature gloutonne de l'algorithme bottom-up
- Problème de **partitionnement** exprimé sous forme de **PLNE**
- Résolution grâce à l'algorithme branch-and-bound
 - Explore l'ensemble des solutions possibles
 - Sélectionne celle qui optimise la fonction objectif



Speaker Diarization : Regroupement 1

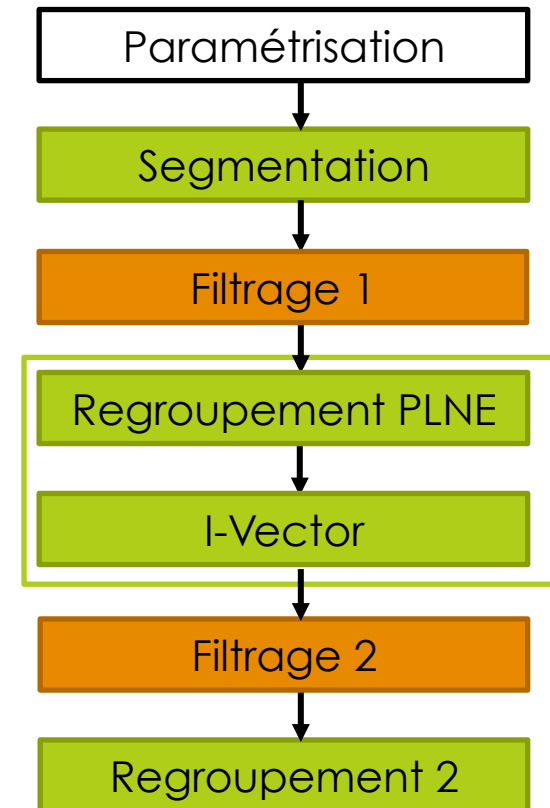
- Fonction objectif
 - Minimiser le nombre de classe
 - Minimiser la dispersion intra-classe
- Contrainte
 - Un segment peut appartenir qu'à une seule classe
 - Un segment peut appartenir à une classe si la distance < seuil

$$\begin{aligned}
 \text{Minimize:} \quad & \sum_{i=1}^N x_{i,i} + \frac{1}{F} \sum_{i=1}^N \sum_{j=1}^N d(w_i, w_j) x_{i,j} \\
 \text{Subject To:} \quad & \sum_{i=1}^N x_{i,j} = 1, \quad j \\
 & d(w_i, w_j) x_{i,j} \leq \epsilon, \quad i, j \\
 & x_{i,j} \in \{0, 1\}, \quad i, j
 \end{aligned}$$



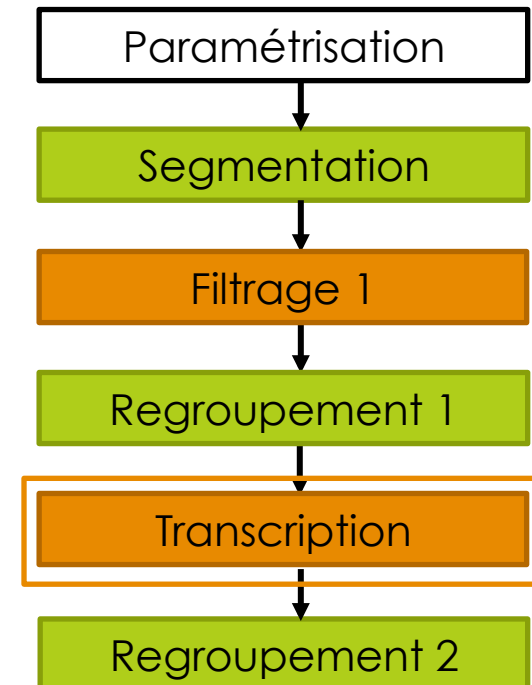
Speaker Diarization : Regroupement 1

- Problème : Les GMM modélisent l'information utile mais aussi l'information inutile
- i-vector : Transformer un supervecteur en un vecteur de plus faible dimension où toute l'information du locuteur est conservé
 - 256 composantes
 - Appris sur les données : ETAPE et REPERE
- Comparaison entre deux modèles/i-vecteurs
 - Distance de Mahalanobis



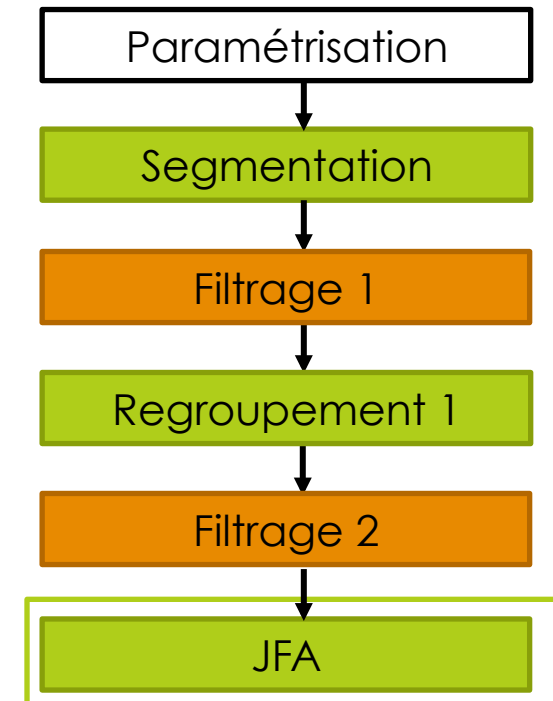
Speaker Diarization : Filtrage 2

- But : les fillers d'une transcription ASR peuvent fournir une information sur les zones non-paroles
- Transcription
 - Transcription : LIUM Sphinx - P1
 - Supprime dans la segmentation tous les fillers > 50 trames



Speaker Diarization : Regroupement 2

- But : Regrouper certaines classes en faisant de l'identification du locuteur
- Identification du locuteur : JFA
 - Repose sur le toolkit ALIZE
 - 252 Modèles de locuteurs
 - Journaliste
 - Présentateur
 - Homme/Femme politique
 - Appris sur les données ESTER 1 & 2 + ETAPE
 - UBM 1024 gaussienne indép. sexe et canal.

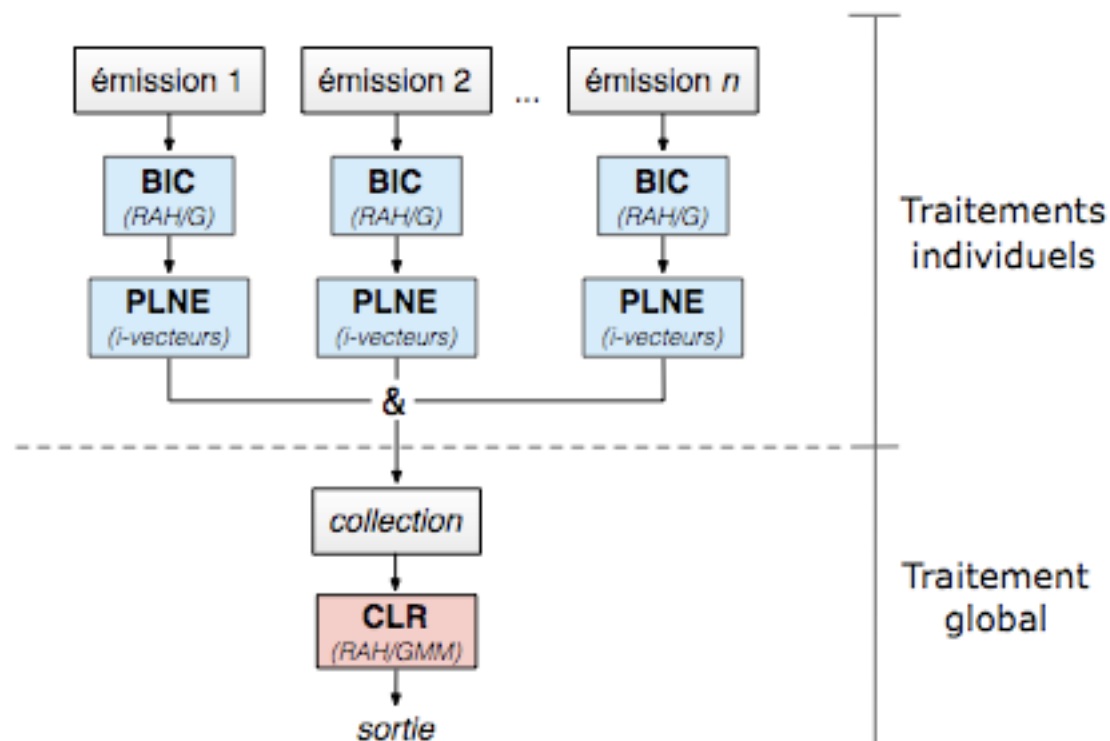


Speaker Diarization : Résultat

- **Dev** : Uniquement évalué sur les émissions BFM-TV, LCP et TV-8; problème d'overlap sur les émissions France Inter
- **Test** : Evalué en utilisant les MDTM du 14/06/2012

	Dev	Test
Regroupement 1	15.43	21.45
Filtrage 2	14.26	19.74
Regroupement 2	12.66	19.01

Speaker Diarization Cross-Show



SRL-X : Résultat

- **Dev** : Uniquement évalué sur les émissions BFM-TV, LCP et TV-8; problème d'overlap sur les émissions France Inter
- **Test** : Evalué en utilisant les mdtn du 14/06/2012

	Dev	Test
SRL-X	12.16	20.26