

# Etape

# Speaker Diarization

Orange Labs

CHARLET Delphine, Recherche & Développement  
18/06/2012 Workshop Etape, Rennes



diffusion libre



# Système de base

- Segmentation
- Regroupement BIC
- Viterbi et Regroupement CLR

# Segmentation

- Segmentation parole/non-parole
  - segmentation en locuteur des segments de parole
    - MFCC+E, delta, delta-delta
    - critère BIC en 2 passes:
      - 1ere passe, distance entre 2 fenêtres glissantes de taille fixe
      - 2eme passe, entre les segments consécutifs obtenus à la première passe
- => 2 seuils (critères d'arrêt bic pour chaque passe)

# Regroupement

- Regroupement BIC
    - MFCC+E, delta, delta-delta + warping
    - Regroupement hiérarchique ascendant critère BIC
      - permet d'initialiser des modèles GMM (64 gaussiennes) de locuteurs (adaptation MAP modèle UBM de l'émission)
  - Regroupement CLR avec décodage Viterbi:
    - regroupement itératif:
      - décodage de Viterbi
      - re-estimation des modèles GMM
      - regroupement CLR (post-soumission: c'est mieux de faire NCLR!)
- => 2 seuils (critère d'arrêt BIC et critère d'arrêt CLR)

# Prise en compte de la parole superposée

- Constat

- le système de base crée souvent un cluster de parole superposée. Dans les évaluations excluant la parole superposée, cela n'a pas d'incidence. Dans les évaluations incluant la parole superposée, les données attribuées à ce cluster sont considérées plusieurs fois comme de l'erreur !!!

ex: données du segment de parole superposée A et B, attribuée au cluster Brouhaha: confusion A->Brouhaha, B->Brouhaha

- Principe

- détection de la parole superposée
- segmentation/regroupement en locuteurs sans parole superposée
- post-traitement d'attribution des segments de parole superposée: heuristique

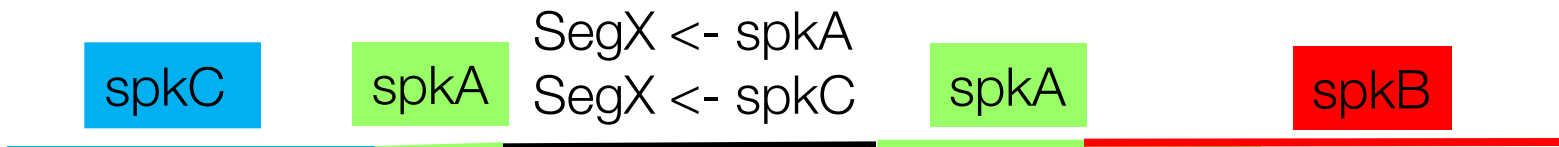
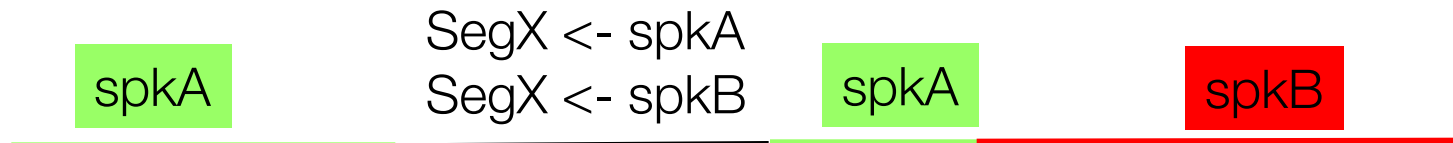
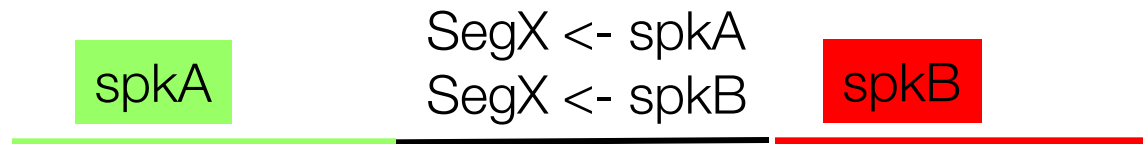
# Détection de la parole superposée

- segmentation parole/non-parole inchangée
- détection de la parole superposée dans les segments de parole:
  - Modèle HMM 3 états: parole simple homme, parole simple femme, parole-superposée:
  - modélisation GMM (256 gaussiennes par classe)
  - MFCC+E, delta, delta-delta
  - critère de durée minimale dans les états parole simple: 2s, parole superposée: 0.5s
- Décodage de Viterbi (sur segment parole uniquement)

# Post-traitement d'attribution de la parole superposée

- Principe: réattribution aux locuteurs les plus proches.
  - pour chaque segment  $seg_X$  classé "parole superposée": recherche des 2 locuteurs les plus proches temporellement de ce segment
    - $distance\_temporelle(loc_i, seg_X) = \min(\text{délai entre fin du dernier segment du } loc_i \text{ précédent } seg_X \text{ et début de } seg_X, \text{délai entre début du premier segment du } loc_i \text{ suivant } seg_X \text{ et fin de } seg_X)$
    - attribution aux 2 plus proches locuteurs  $loc_i$  et  $loc_j$  si  $distance\_temporelle(loc_i, seg_X) < \text{seuil}$ .

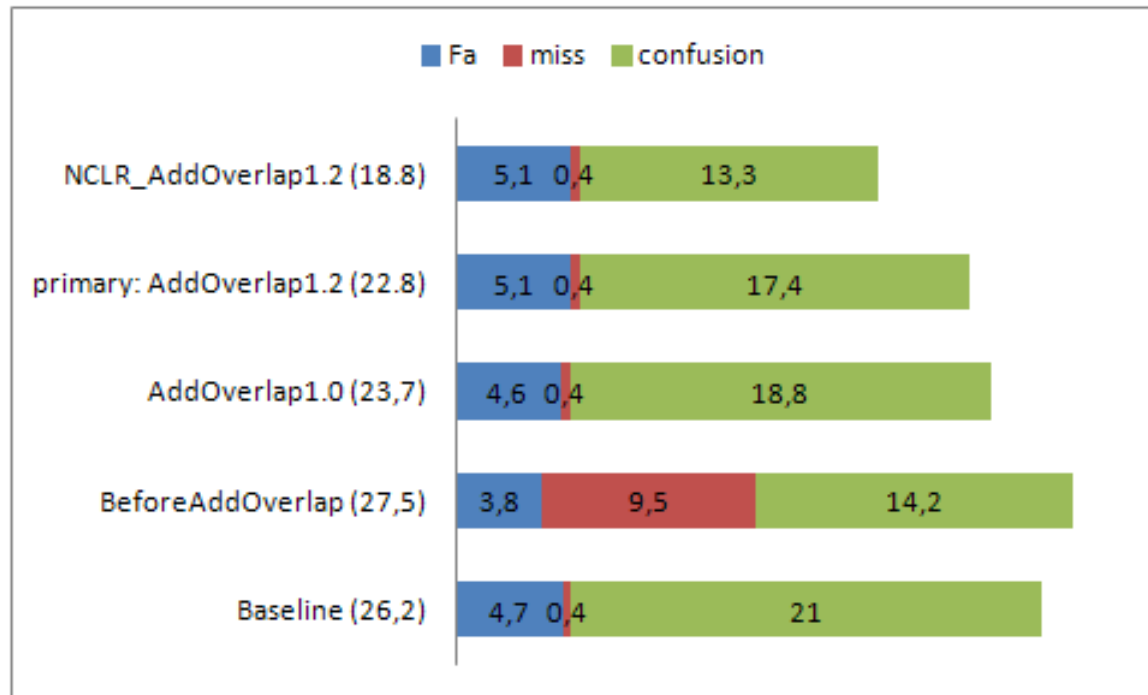
# Exemples



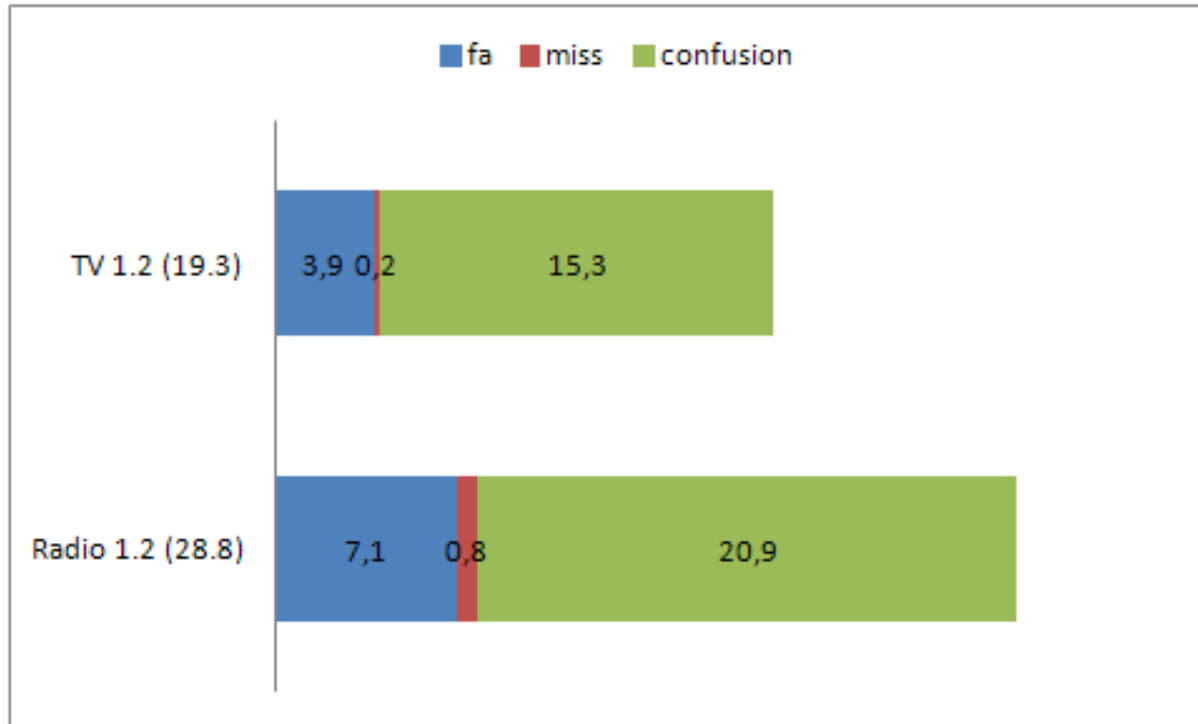


# Résultats TEST Etape

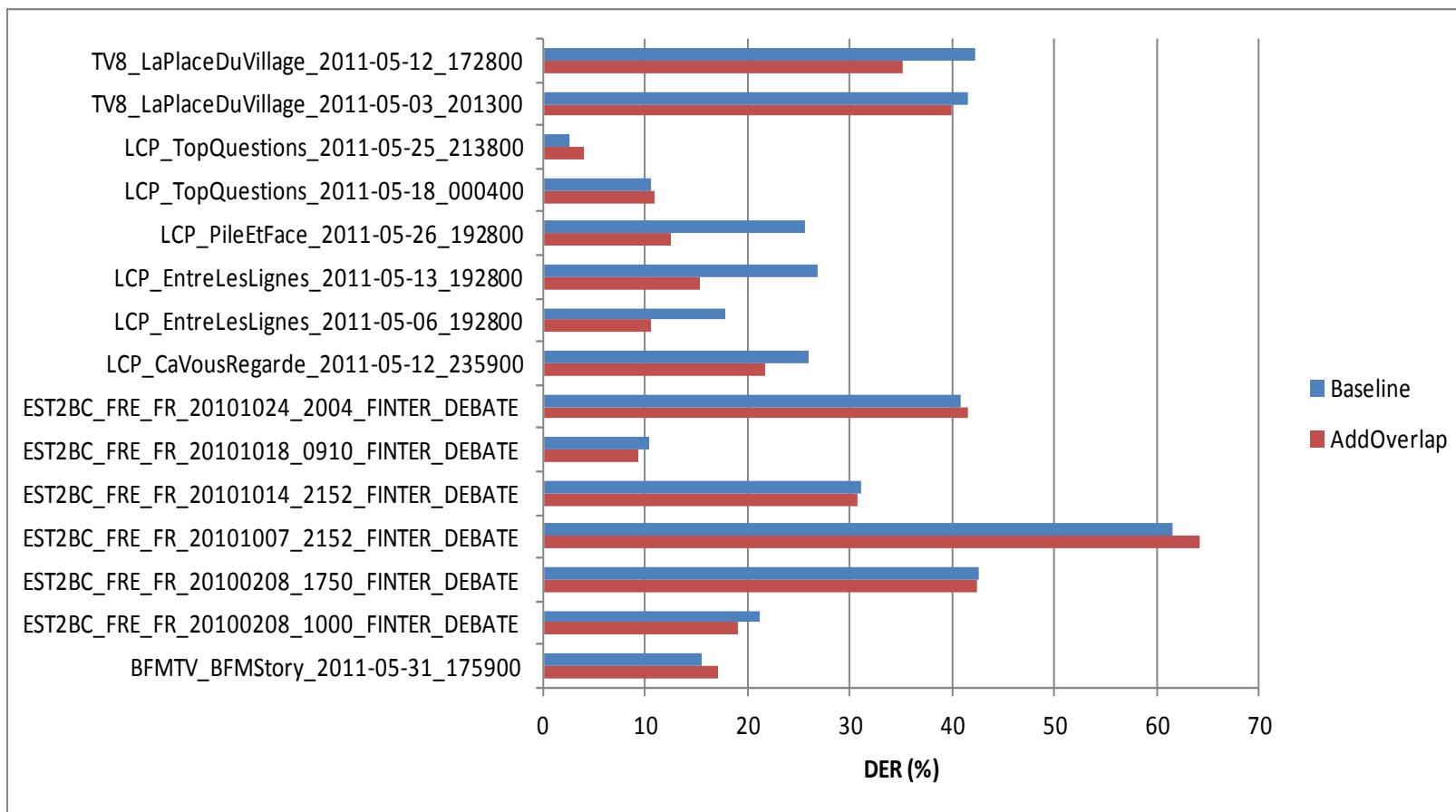
- **Baseline:** système initial sans détection d'overlap
- **BeforeAddOverlap:** système avec détection des zones de parole superposée, et clustering BIC/CLR en excluant les zones d'overlap
- **AddOverlap:** sortie de BeforeAddOverlap enrichie des réattributions des segments détectés parole superposée
  - 1.0: réattribution d'un seul locuteur (le plus proche)
  - 1.2: réattribution du plus proche locuteur puis d'un deuxième si "assez" proche ( $\text{distance\_temporelle}(\text{loc}_i, \text{seg}_x) < \text{seuil}$ )



# Résultats Etape TEST: Radio vs TV



# Résultats par émission



# Conclusions

- la stratégie de traiter la parole superposée spécifiquement s'avère payante
  - permet d'obtenir des clusters moins bruités (moins de confusion)
  - stratégie de réattribution actuelle : ok, au moins pour le premier locuteur
  - approfondir/évaluer la phase de détection de la parole superposée
- mais ne compense pas les problèmes intrinsèques de la sensibilité à certains seuils, pour déterminer le nombre de clusters (mauvais résultats sur la radio dû à un clustering trop fort d'une émission du Masque et la Plume)