

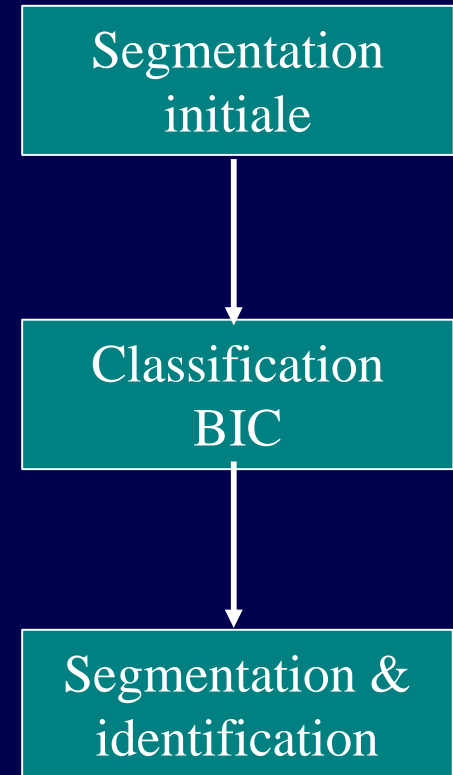


# ESTER DryRun SRL - SVL

Sylvain Meignier, Xuan Zhu,  
Claude Barras, Jean-Luc Gauvain

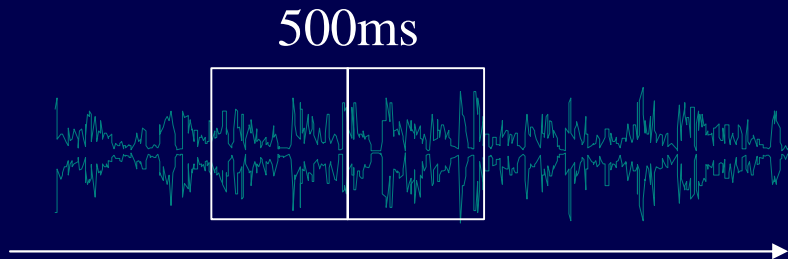
# SRL

- SRL en 3 phases
  - Segmenter en unités mono locuteur
  - Regrouper les segments
  - Affiner les frontières des segments et identifier le genre et le type



# SRL : Segmentation initiale

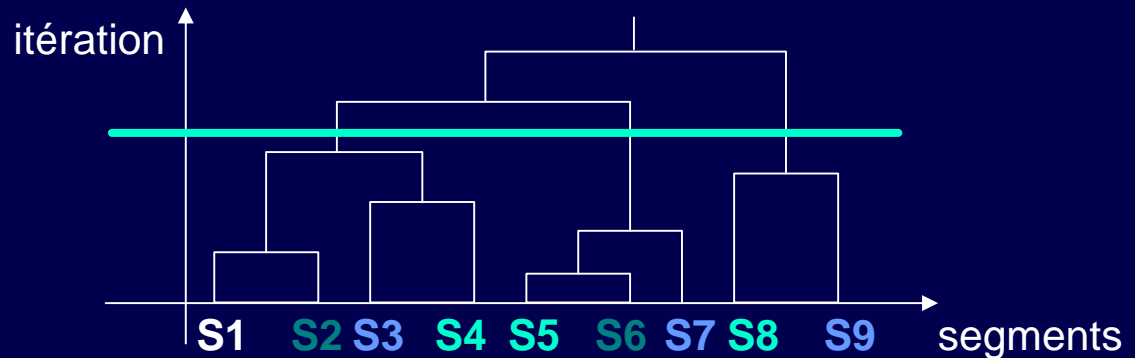
- Paramètre acoustique
  - $12\text{plp} + ? + ?? + ?E + ??E = 38$
  - Segmentation en classes acoustiques
  - Musique, Silence, Parole
  - 6x64 GMM diag, hub4
- Détection de ruptures
  - Divergence gaussienne  $D(x, y) = (\mu_x - \mu_y)^2 / (\sigma_x \sigma_y)$
  - Fenêtre glissante
    - $D(x, y) > 0.1$



- Segmentation: EM-ML+ Viterbi
  - $n \times 8$  GMM, 1 par segment, 38 paramètres

# SRL : Classification hiérarchique

- Paramètre acoustique
  - 12plp+E, pas de CMS...
  - Décalage de 10ms, fenêtre de 30ms
- Modèle
  - Mono gaussien plein
- Mesure
  - BIC
- Critère d'arrêt : BIC
  - $\lambda = 5.0$



# SRL : segmentation & identification

- Segmentation : EM + Viterbi
  - nx8 GMM, 1 par segment, 38 paramètres
- Segmentation : Viterbi
  - Frontières contraintes par l'énergie
- Identification Studio / Téléphone des segments
  - 38 paramètres, 12 plp
    - bande large + bande étroite
  - 2x64 GMM diag
- Identification du genre des segments
  - 2x64 GMM diag + 2x64 GMM diag
  - 39 paramètres, 12 plp

# SRL : Résultats

- DryRun RT'03 , 6 show de 10m sans pub sur 5 fichiers = 10.88%
  - 13.15, 2.44, 0.69, 17.37, 4.20, 29.43 (%)
- Train RT'03, 6 shows de 20m avec pub = 15.52 %
  - 22.79, 15.54, 10.48, 19.86, 16.19, 7.09 (%)
- RT'03 Spring + Fall , 6 shows de 20m avec pub = 12.50 %
  - 14.71, 10.90, 5.06, 18.04, 25.95, 3.28 (%)
- Dev Ester = 13.34 %, 6 shows
  - 6.51, 16.98, 8.22, 7.33, 11.31, 22.48 (%)
- ALL = 13.42 %
  - 6 meilleurs = 4.88 %
  - 6 + mauvais = 19.98 %
- DryRun Ester = 13.8 %

# SVL

- Système de RAL
- Paramètre identique au système 1sp NIST 2003
  - Cf. ICASSP 2003
  - Sauf
    - Bande large
    - Pas de TNorm
- Principe

# SVL : Train

- 15 plp + delta, CMS + Var sur les classes
- UBM
  - Données
    - Hub4 BN96, hub4 spanish, hub4 arabe
    - 2x1024 GMM (dep genre)
  - EM-ML
- Locuteur
  - MAP, 5 itérations MAP
  - moyennes



# SVL : Test

- Score
  - Top 5 gaussiennes
  - Rapport de vraisemblance
  - Pas de normalisation ZNorm ou TNorm
- Seuil
  - Fixé pour une autre application