

---

# Suivi du Locuteur

## *Systeme et résultats*

N. Scheffer, D. Istrate,  
C. Fredouille, J-F Bonastre

<http://www.lia.univ-avignon.fr/heberges/ALIZE/>

# Systeme utilise: LIA\_SpkDet

---

- *ALIZE* pour l'utilisateur
  - Outils de Détection du Locuteur basé sur la plateforme ALIZE
- Première validation sur *NIST* SRE 2004
  - Systeme *GMM* état de l'art en vérification.
- Nouvelle Fonctionnalité:
  - Gestion des segments temporelles pour le calcul de statistiques (EM, LK, Viterbi Path)
  - Adapté au Suivi de Locuteur
- Outils sous license GPL (LIA\_SpkDet) et LGPL (ALIZE)

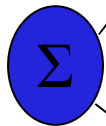
# Stratégies proposées

---

- Segmentation en macro-classes acoustiques en entrée: Téléphone, Studio / Homme, Femme.

- **Systemes contrastifs:**

Systeme Primaire



- BM Studio, adaptation donnés multi-canal, tests cross-channel, non cross-gender.
- BM Tél, adaptation donnés multi-canal, tests cross-channel, non cross-gender.
- Adaptation donnés mono-canal (canal majoritaire pour un locuteur), pas de tests cross-channel ni cross-gender.
- Pseudo Feature-Mapping.

# En détails (1)

---

- **Paramétrisation:** MFCC,  $16+\Delta$ .
- **Normalisation:**
  - (0,1) sur gender, pas de suppression de trames (!)
- **Modélisation:** (*macro-classes acoustiques*)
  - GMM 1024, Covariance Diag.
- **Adaptation au locuteur:** (*segmentation de référence*)
  - Moyenne seulement, Occupation Dependant (LIMSI/MIT), Regulation Factor 14
  - Limite durée adaptation 20 minutes

# En détails (2)

---

- **Tests:**
  - Sur les « top ten », LLR: moyenne non pondérée sur les trames
- **Normalisation des scores:**
  - **T-Norm:** La distribution imposteur est modélisée par les autres locuteurs à suivre.
- **Décision:**
  - Un segment = un locuteur (meilleure LK)
  - Un segment = un locuteur = 1 décision
    - Seuil indépendant du genre, réglé sur le DEV

# Résultats – DEV Phase II

Segmentation de référence  
(MDTM)

Source	C1	C2	C3	C4	P
20030418_1200_1300_FRANCEINFO_DGA	0.847	0.796	0.840	NC	0.841
20030418_1700_1800_FRANCEINFO_DGA	0.811	0.761	0.810	NC	0.804
20030418_0700_0800_FRANCEINTER_DGA	0.829	0.826	0.785	NC	0.843
20030418_0800_0900_FRANCEINTER_DGA	0.734	0.738	0.646	NC	0.743
20030508_1400_1500_RFI_ELDA	0.973	0.822	0.731	NC	0.919
20030509_1400_1500_RFI_ELDA	0.877	0.855	0.809	NC	0.878
20030717_0700_0715_RTM_ELDA	0.541	0.541	0.551	NC	0.541
20030717_1300_1320_RTM_ELDA	0.149	0.149	0.149	NC	0.149
20030717_2000_2020_RTM_ELDA	0.000	0.000	0.000	NC	0.000
20030717_2300_2315_RTM_ELDA	0.125	0.125	0.131	NC	0.125
20030719_0700_0715_RTM_ELDA	0.000	0.000	0.000	NC	0.000
20030719_1300_1320_RTM_ELDA	0.000	0.000	0.000	NC	0.000
20030719_2000_2015_RTM_ELDA	0.000	0.000	0.000	NC	0.000
20030719_2300_2310_RTM_ELDA	0.000	0.000	0.000	NC	0.000
F-mesure	0.629	0.597	0.572	NC	0.633

# Résultats – TEST Phase II

Segmentation du système SRL

Source	C1	C2	C3	C4	P
20041008_1800_1830_INFO_DGA	0.582	0.743	0.251	0.319	0.644
20041012_1800_1830_INFO_DGA	0.398	0.295	0.390	0.189	0.393
20041013_1700_1800_INFO_DGA	0.511	0.400	0.458	0.346	0.512
20041007_0800_0900_INTER_DGA	0.622	0.813	0.575	0.581	0.804
20041011_1300_1400_INTER_DGA	0.332	0.332	0.300	0.330	0.332
20041006_0700_0800_CLASSIQUE	0.000	0.000	0.000	0.000	0.000
20041006_0800_0900_CULTURE	0.000	0.000	0.000	0.000	.000
20041025_1930_2000_RFI_ELDA	0.730	0.251	0.113	0.505	0.826
20041026_1930_2000_RFI_ELDA	0.546	0.405	0.000	0.493	0.556
20041027_1230_1300_RFI_ELDA	0.000	0.000	0.000	0.000	0.000
20041124_1230_1300_RFI_ELDA	0.000	0.000	0.000	0.000	0.000
20041217_1300_1322_RTM_ELDA	0.749	0.748	0.786	0.839	0.748
20041218_1300_1314_RTM_ELDA	0.000	0.000	1.000	0.000	0.000
20041219_1300_1314_RTM_ELDA	0.000	0.000	0.000	0.000	0.000
20041220_1300_1314_RTM_ELDA	0.799	0.839	0.822	0.984	0.841
20041221_1300_1321_RTM_ELDA	0.747	0.778	0.322	0.804	0.747
20041222_1300_1320_RTM_ELDA	0.711	0.673	0.593	0.671	0.712
20041223_1300_1318_RTM_ELDA	0.681	0.646	0.000	0.588	0.681
F-mesure	0.605	0.585	0.413	0.507	0.659

# Discussion

---

- FR: 40%, FA:4%
  - La décision ne porte que sur un seul locuteur par segment
- Même stratégie d'adaptation pour tous les locuteurs
  - Volume d'apprentissage intra-locuteur variable !
- Robustesse:
  - Robuste aux erreurs de segmentation en canal
  - Ne pallie pas les erreurs de segmentation en genre



# Contraste: TEST Phase II

- Système primaire (P)
- Système primaire avec segmentation de référence
  - Segments
  - Genre

Source	P	Pref
20041008_1800_1830_INFO_DGA	0.644	0.964
20041012_1800_1830_INFO_DGA	0.393	0.386
20041013_1700_1800_INFO_DGA	0.512	0.423
20041007_0800_0900_INTER_DGA	0.804	0.797
20041011_1300_1400_INTER_DGA	0.332	0.844
20041006_0700_0800_CLASSIQUE	0.000	0.000
20041006_0800_0900_CULTURE	0.000	0.037
20041025_1930_2000_RFI_ELDA	0.826	0.891
20041026_1930_2000_RFI_ELDA	0.556	0.680
20041027_1230_1300_RFI_ELDA	0.000	0.000
20041124_1230_1300_RFI_ELDA	0.000	0.000
20041217_1300_1322_RTM_ELDA	0.748	0.802
20041218_1300_1314_RTM_ELDA	0.000	0.000
20041219_1300_1314_RTM_ELDA	0.000	0.000
20041220_1300_1314_RTM_ELDA	0.841	0.978
20041221_1300_1321_RTM_ELDA	0.747	0.890
20041222_1300_1320_RTM_ELDA	0.712	0.710
20041223_1300_1318_RTM_ELDA	0.681	0.741
F-mesure	0.659	0.752