

ESTER 2003

Les systèmes du LIA

- Personnels :
 - F. Béchet (TRS)
 - J.F Bonastre (SES, SRL)
 - C. Fredouille (SES, SRL)
 - G. Linares (TRS,TTR)
 - D. Matrouf (TRS,TTR)
 - P. Nocera (TRS,TTR)

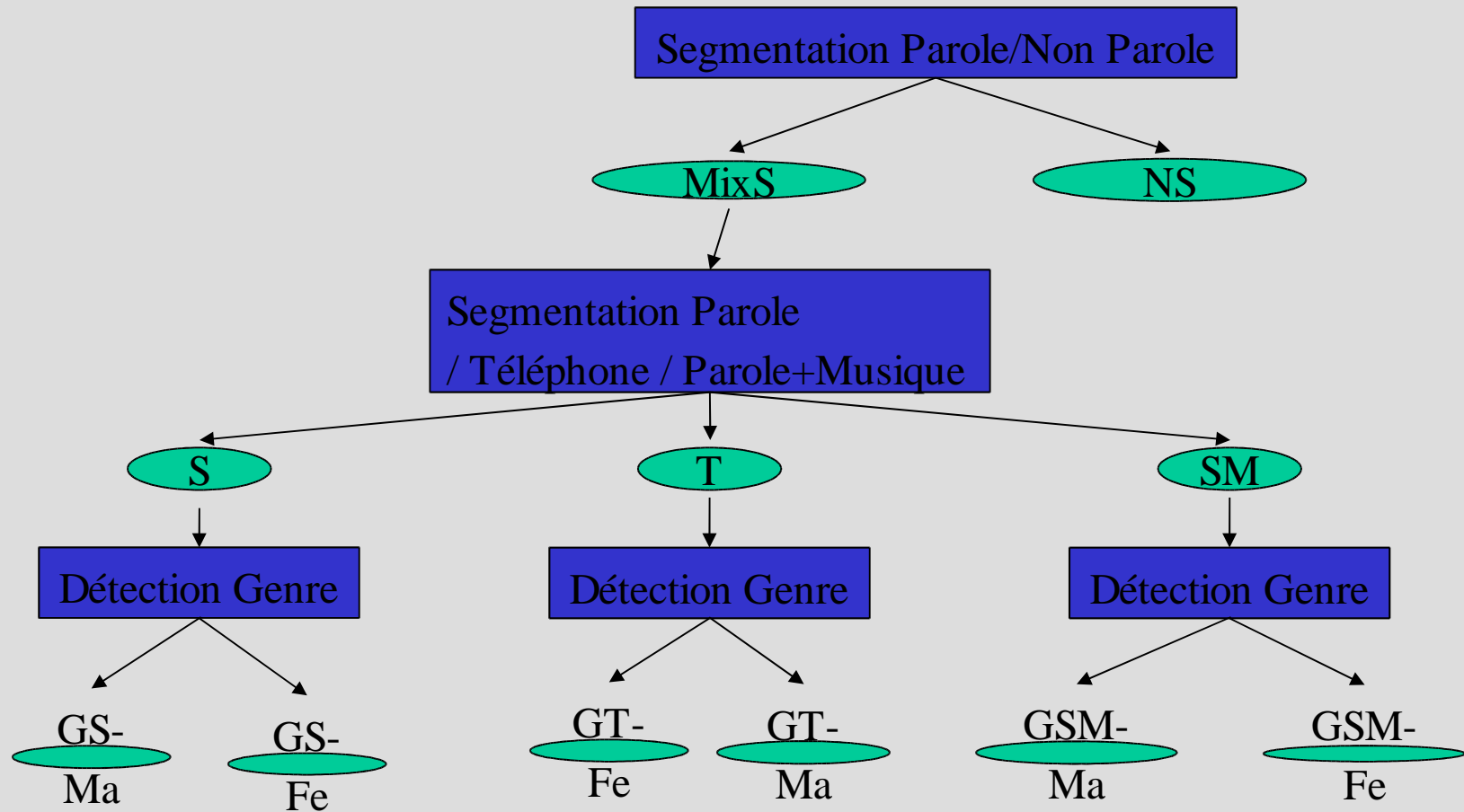
ESTER 2003

- Systèmes LIA phase 1:
 - Segmentation en événements sonores
 - Segmentation/regroupement en locuteurs
 - Transcription
- Résultats phase 1
- Travaux en cours & perspectives

Segmentation en événements sonores

- Méthode & modèles issus de RT-03
- Adaptation aux données ESTER (MAP)
- Principe : classification hiérarchique en macro-classes
- Outils : HTK

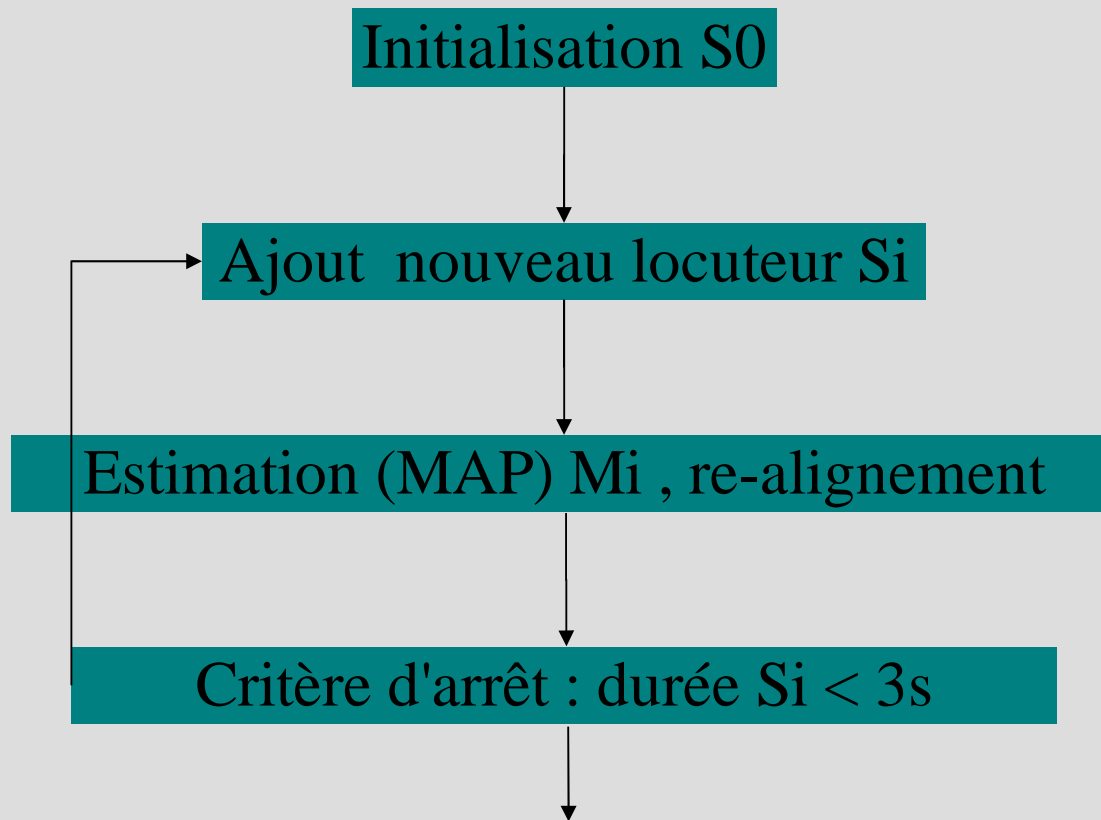
Segmentation en événements sonores



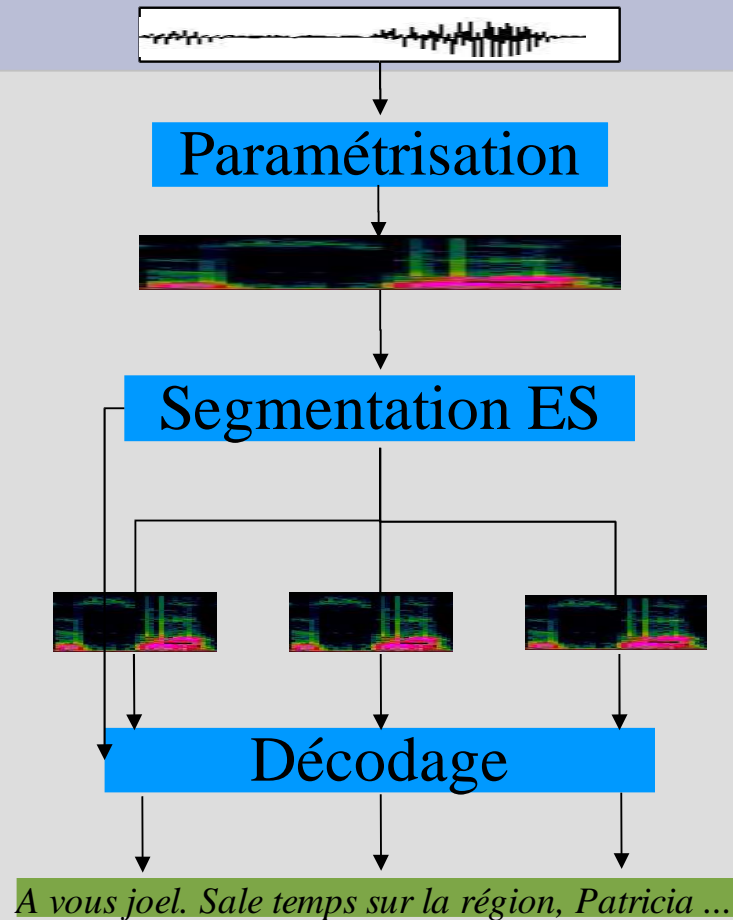
Segmentation en locuteurs

- Méthode issue de RT-03
- Apprentissage sur ESTER
- Post traitement de SES
- Outils : LIA

Segmentation en locuteurs



Transcription : chaîne de traitement



Transcription : paramétrisation

- MFCC 12 Coefs + nrj
- Delta+Delta2
- Paramètres centrés-réduits par segment
- Outils LIA

Transcription : modèles acoustiques

- Modèles initiaux issus de BREF120
- 580 diphones droits
- Dépendants du genre
- Adaptation aux données ESTER
- Outils : LIA

Transcription : modèles acoustiques

- Partitionnement du corpus ESTER
 - basé sur la segmentation en événements sonores
 - Genre (H/F)
 - Téléphone / Fond musical / Propre
- Adaptation 2 passes:
 - MAP
 - Réalignement

Transcription : modèles de langage

- Lexique 40K mots
 - le monde 1987-2002 (300 M mots)
 - + ester (300 K mots)
 - Phonétisation : LIA-PHON
- Modèles de langage :
 - combinaison
 - Le monde 1998-2004 (poids .68)
 - Ester (poids 0.32)
 - CMU Toolkit

Transcription : moteur de reconnaissance

- Algorithme dérivé de A^*
- Basé sur un treillis de phonèmes
- Coupures :
 - En largeur
 - En profondeur

Résultats SES

Fichiers Dev	Music			Speech		
	% Miss	% Fa	% Err.	% Miss	% Fa	% Err.
19981217_0700_0800_inter_dga	5.1	0.0	5.1	4.6	0.0	4.6
19981217_0800_0900_inter_dga	0.3	0.0	0.3	2.8	0.0	2.8
19990622_1900_1920_inter_dga	0.0	0.0	0.0	2.9	0.0	2.9
19990623_1900_1920_inter_dga	0.0	0.0	0.0	2.0	0.0	2.0
20000907_0930_1030_rfi_dga	46.1	0.0	46.1	10.0	0.0	10.0
20000907_1130_1230_rfi_dga	19.8	0.0	19.8	8.7	0.0	8.7
Fichiers Test	% Miss	% Fa	% Err.	% Miss	% Fa	% Err.
19981218_0700_0800_inter_dga	7.8	0.0	7.8	1.6	0.0	1.6
19981218_0800_0900_inter_dga	15.6	0.0	15.6	3.4	0.0	3.4
19990624_1900_1920_inter_dga	0.0	0.0	0.0	4.1	0.0	4.1
19990625_1900_1920_inter_dga	0.1	0.0	0.1	3.6	0.0	3.6
20000908_0930_1030_rfi_dga	23.1	0.0	23.1	8.0	0.0	8.0
20000908_1130_1230_r fi_dga	44.0	0.0	44.0	12.8	0.0	12.8

Résultats SL

	Segmentation Locuteur			
Fichiers Dev	% Miss Sp.	% Fa Sp.	% Err. Sp.	% All
19981217_0700_0800_inter_dga	0.5	0.0	28.5	29.0
19981217_0800_0900_inter_dga	0.4	0.0	9.3	9.7
19990622_1900_1920_inter_dga	0.7	0.0	10.8	11.5
19990623_1900_1920_inter_dga	1.4	0.0	7.3	8.7
20000907_0930_1030_rfi_dga	9.3	0.0	23.6	32.9
20000907_1130_1230_rfi_dga	7.3	0.0	24.4	31.7
Fichiers Test	% Miss Sp.	% Fa Sp.	% Err. Sp.	% All
19981218_0700_0800_inter_dga	0.2	0.0	18.6	18.8
19981218_0800_0900_inter_dga	1.0	0.0	8.0	9.1
19990624_1900_1920_inter_dga	2.1	0.0	14.1	16.2
19990625_1900_1920_inter_dga	1.5	0.0	1.7	3.2
20000908_0930_1030_rfi_dga	6.0	0.0	21.7	27.7
20000908_1130_1230_rfi_dga	9.8	0.0	23.5	33.3

Résultats TRS

Session	WER
19981218_0700_0800_inter	23.60%
19981218_0800_0900_inter	34.20%
19990624_1900_1920_inter	39.50%
19990625_1900_1920_inter	34.30%
20000908_0930_1030_rfi	45.50%
20000908_1130_1230_rfi	48.30%

Perspectives

- Segmentation ES: travail spécifique sur ESTER
 - Re-estimation des modèles
 - Réglage du système
- Suivi de locuteurs (à faire)

Perspectives

- Transcription :
 - Acoustique :
 - Autres stratégies pour l'estimation des HMMs
 - Adaptation
 - Linguistique : modèles spécifiques, re-segmentation
 - Stratégies de décodage :
 - ROVER
 - Temps Réel