

UNIVERSITÉ D'AVIGNON
ET DES PAYS DE VAUCLUSE

Workshop ETAPE

Systèmes de segmentation et regroupement en locuteurs

Corinne Fredouille ♦

♦ Université d'Avignon, Laboratoire Informatique – CERI/LIA



LABORATOIRE
INFORMATIQUE
D'AVIGNON

Contexte

- Participation aux tâches SRL et SRL-x
- 3 systèmes soumis :
 - Système Baseline SRL
 - Système Baseline couplé à un système d'identification du locuteur
 - Document seul – SRL (primary)
 - Collection de documents - SRL-x

Système SRL Baseline

- Système “2000”
 - Thèse de Sylvain Meignier au LIA
 - Approche “top-down” (vs “bottom up”)
- 3 passes :
 1. Segmentation parole/non parole
 2. Phase de segmentation : Ajout itératif de locuteurs couplé à un décodage Viterbi & un réapprentissage des modèles locuteurs
 3. Phase de resegmentation : processus itératif couplant décodage Viterbi & réapprentissage des modèles à chaque itération

Système SRL Baseline : en détail

1. Segmentation parole/non parole

- HMM à 4 états : S, MS, T et NP (musique, silence)
- GMM à 64 gaussiennes
- Processus itératif : décodage Viterbi + réapprentissage des GMM par MAP
- 12 LFCC + énergie + Δ + $\Delta\Delta$
- Règles sur durée minimale des segments (T, S et MS: 2s, M: 1s)

2. Phase de segmentation

- E-HMM: ajout nouveaux locuteurs sur segments de 6s
- GMM à 16 gaussiennes
- Processus itératif : Ajout locuteurs + décodage Viterbi & réapprentissage des GMM par EM
- 20 LFCC + énergie
- Règles sur intervention minimale d'un locuteur : 6s

Système SRL Baseline : en détail

3. Phase de Resegmentation: affiner la segmentation

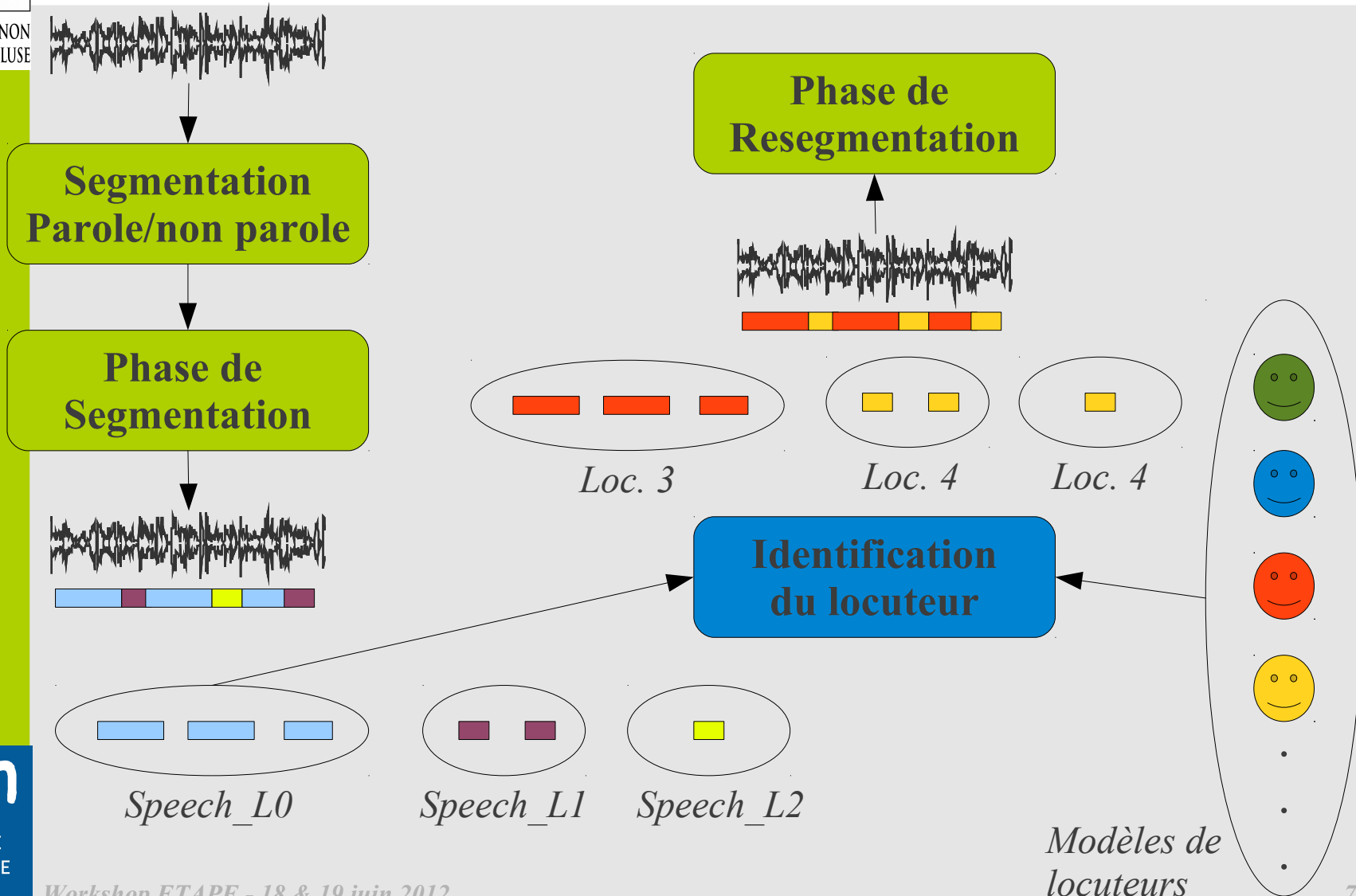
- Hmm à n états/locuteurs issus de la phase de segmentation
- GMM à 128 gaussiennes
- Processus itératif : décodage Viterbi & réapprentissage des GMM par MAP + suppression locuteurs (8s minimum)
- 20 LFCC + énergie
- Modèle du monde appris sur des données vidéo du web

Système Baseline couplé à un système d'identification du locuteur

- Motivations : Répondre à la tâche SRL-x
 - Même étiquette pour un locuteur apparaissant dans deux documents différents
 - Impossible avec Système Baseline SRL => speech_L0, speech_L1, ..., speech_Lx
- Approche : Attribuer une identité à chaque cluster/locuteur
 - Deux clusters de deux documents différents appartenant au même locuteur => même identité !
 - *Deux clusters d'un même document appartenant au même locuteur => même identité*
 - Inspirée des modèles d'ancrage ([Reynolds et al.,01], [Collet et al., 05, 06, ...]) => simplifiés ici !

Système Baseline couplé à un système d'identification du locuteur

UNIVERSITÉ D'AVIGNON
ET DES PAYS DE VAUCLUSE



Système Baseline couplé à un système d'identification du locuteur : en détail

- Système baseline SRL inchangé
- Système d'identification du locuteur:
 - GMM/UBM état de l'art (ALIZE/SpkDet)
 - GMM à 512 gaussiennes
 - 19 LFCC + énergie + Δ + $\Delta\Delta$ normalisés
 - décision sur maximum de vraisemblance
 - appliqué sur les clusters de segments

Système Baseline couplé à un système d'identification du locuteur : en détail

- Système d'identification du locuteur (suite)
 - 235 modèles de locuteurs à disposition :
 - Données train+dev d'ETAPE – les émissions radio
 - Données dev du Défi REPERE (émissions TV identiques à ETAPE hors TV8)
 - Données JT
 - Importance de la couverture en terme d'identité des locuteurs ?
 - Aucune => (*modèles d'ancrage*)
 - Le système doit être le plus cohérent et stable possible à l'intérieur d'un même document et d'un document à l'autre (sous condition de pureté des clusters !)
=> *pas d'identification du locuteur in fine*

Résultats

La soumission : SRL

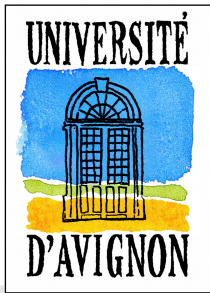
- Scoring NIST
- Recouvrement de la parole

Fichiers	Miss (%)	FA (%)	Baseline		Baseline+IAL		Delta
			Spk (%)	DER (%)	Spk (%)	DER (%)	
BFMStory_2011-05-31_175900	1,6	1,6	24,2	27,1	26,9	29,8	2,7
EST2BC_20100208_1000	3,8	2,8	11,2	17,9	8,8	15,5	-2,4
EST2BC_20100208_1750	0,7	17,6	34,1	52,4	36,1	54,3	2,0
EST2BC_20101007_2152	0,6	21,7	54,4	76,7	52,8	75,2	-1,5
EST2BC_20101014_2152	7,2	5,5	16,9	29,6	16,4	29,0	-0,5
EST2BC_20101018_0910	0,9	5,4	9,8	16,1	10,2	16,5	0,4
EST2BC_20101024_2004	5,0	5,8	15,3	26,1	19,4	30,2	4,1
CaVousRegarde_2011-05-12_235900	5,1	1,1	29,2	35,4	24,5	30,7	-4,7
EntreLesLignes_2011-05-06_192800	5,4	2,5	35,5	43,3	14,7	22,5	-20,8
EntreLesLignes_2011-05-13_192800	7,7	2,3	14,6	24,6	20,4	30,4	5,7
PileEtFace_2011-05-26_192800	12,1	0,7	14,2	27,0	10,7	23,5	-3,4
TopQuestions_2011-05-18_000400	0,1	12,4	0,6	13,0	31,2	43,7	30,7
TopQuestions_2011-05-25_213800	0,5	2,2	10,5	13,2	6,0	8,7	-4,5
TV8_Village_2011-05-03_201300	1,4	7,7	21,8	30,9	16,9	26,0	-4,9
TV8_Village_2011-05-12_172800	1,7	5,5	34,0	41,2	23,7	31,0	-10,2
Global	4,2	3,9	20,0	28,1	18,8	26,9	-1,2

La soumission : SRL

- Scoring NIST
- Recouvrement de la parole

Fichiers	Miss (%)	FA (%)	Baseline		Baseline+IAL		Delta
			Spk (%)	DER (%)	Spk (%)	DER (%)	
BFMStory_2011-05-31_175900	1,6	1,6	24,2	27,1	26,9	29,8	2,7
EST2BC_20100208_1000	3,8	2,8	11,2	17,9	8,8	15,5	-2,4
EST2BC_20100208_1750	0,7	17,6	34,1	52,4	36,1	54,3	2,0
EST2BC_20101007_2152	0,6	21,7	54,4	76,7	52,8	75,2	-1,5
EST2BC_20101014_2152	7,2	5,5	16,9	29,6	16,4	29,0	-0,5
EST2BC_20101018_0910	0,9	5,4	9,8	16,1	10,2	16,5	0,4
EST2BC_20101024_2004	5,0	5,8	15,3	26,1	19,4	30,2	4,1
CaVousRegarde_2011-05-12_235900	5,1	1,1	29,2	35,4	24,5	30,7	-4,7
EntreLesLignes_2011-05-06_192800	5,4	2,5	35,5	43,3	14,7	22,5	-20,8
EntreLesLignes_2011-05-13_192800	7,7	2,3	14,6	24,6	20,4	30,4	5,7
PileEtFace_2011-05-26_192800	12,1	0,7	14,2	27,0	10,7	23,5	-3,4
TopQuestions_2011-05-18_000400	0,1	12,4	0,6	13,0	31,2	43,7	30,7
TopQuestions_2011-05-25_213800	0,5	2,2	10,5	13,2	6,0	8,7	-4,5
TV8_Village_2011-05-03_201300	1,4	7,7	21,8	30,9	16,9	26,0	-4,9
TV8_Village_2011-05-12_172800	1,7	5,5	34,0	41,2	23,7	31,0	-10,2
Global	4,2	3,9	20,0	28,1	18,8	26,9	-1,2



La soumission : SRL-x

- Scoring ETAPE
- Recouvrement de la parole
- Résultats:
 - Baseline: 57%
 - Baseline+IAL: 37,5%



LABORATOIRE
INFORMATIQUE
D'AVIGNON

La soumission : SRL – un peu revue et "corrigée"

- Le système tient compte des UEM (zones non transcrites) !
- Scoring NIST, Recouvrement de la parole

Fichiers	Baseline+IAL				Baseline+IAL sur UEM				Delta
	Miss (%)	FA (%)	Spk (%)	DER (%)	Miss (%)	FA (%)	Spk (%)	DER (%)	
BFMStory_2011-05-31_175900	1,4	1,6	26,9	29,8	1,4	1,5	26,7	29,6	-0,2
EST2BC_20100208_1000	3,8	2,8	8,8	15,5	3,9	2,8	8,6	15,3	-0,2
EST2BC_20100208_1750	0,7	17,6	36,1	54,3	0,7	17,6	36,1	54,3	0,0
EST2BC_20101007_2152	0,6	21,7	52,8	75,2	26,4	4,6	44,5	75,5	0,3
EST2BC_20101014_2152	7,2	5,5	16,4	29,0	0,2	16,5	23,3	40,1	11,0
EST2BC_20101018_0910	0,9	5,4	10,2	16,5	0,4	6,5	1,2	8,1	-8,5
EST2BC_20101024_2004	5	5,8	19,4	30,2	5	5,8	3,1	13,9	-16,3
CaVousRegarde_2011-05-12_235900	5,1	1,1	24,5	30,7	5,1	1,1	19	25,2	-5,5
EntreLesLignes_2011-05-06_192800	5,4	2,5	14,7	22,5	5,4	2,5	4,3	12,2	-10,3
EntreLesLignes_2011-05-13_192800	7,7	2,3	20,4	30,4	7,7	2,3	5,5	15,6	-14,8
PileEtFace_2011-05-26_192800	12,1	0,7	10,7	23,5	12,1	0,9	11,5	24,5	1,0
TopQuestions_2011-05-18_000400	0,1	12,4	31,2	43,7	0,1	12,4	31,2	43,7	0,0
TopQuestions_2011-05-25_213800	0,5	2,2	6	8,7	0,3	2,4	6,2	8,9	0,2
TV8_Village_2011-05-03_201300	1,4	7,7	16,9	26,0	1,6	7,5	6,8	15,9	-10,1
TV8_Village_2011-05-12_172800	1,7	5,5	23,7	31,0	1,7	5	11,3	18,1	-12,9
Global	4,2	3,9	18,8	26,9	4,4	3,9	12,5	20,9	-6,0

La soumission : SRL – un peu revue et "corrigée"

- Le système tient compte des UEM (zones non transcrites) !
- Scoring NIST, Recouvrement de la parole

Fichiers	Baseline+IAL				Baseline+IAL sur UEM				Delta
	Miss (%)	FA (%)	Spk (%)	DER (%)	Miss (%)	FA (%)	Spk (%)	DER (%)	
BFMStory_2011-05-31_175900	1,4	1,6	26,9	29,8	1,4	1,5	26,7	29,6	-0,2
EST2BC_20100208_1000	3,8	2,8	8,8	15,5	3,9	2,8	8,6	15,3	-0,2
EST2BC_20100208_1750	0,7	17,6	36,1	54,3	0,7	17,6	36,1	54,3	0,0
EST2BC_20101007_2152	0,6	21,7	52,8	75,2	26,4	4,6	44,5	75,5	0,3
EST2BC_20101014_2152	7,2	5,5	16,4	29,0	0,2	16,5	23,3	40,1	11,0
EST2BC_20101018_0910	0,9	5,4	10,2	16,5	0,4	6,5	1,2	8,1	-8,5
EST2BC_20101024_2004	5	5,8	19,4	30,2	5	5,8	3,1	13,9	-16,3
CaVousRegarde_2011-05-12_235900	5,1	1,1	24,5	30,7	5,1	1,1	19	25,2	-5,5
EntreLesLignes_2011-05-06_192800	5,4	2,5	14,7	22,5	5,4	2,5	4,3	12,2	-10,3
EntreLesLignes_2011-05-13_192800	7,7	2,3	20,4	30,4	7,7	2,3	5,5	15,6	-14,8
PileEtFace_2011-05-26_192800	12,1	0,7	10,7	23,5	12,1	0,9	11,5	24,5	1,0
TopQuestions_2011-05-18_000400	0,1	12,4	31,2	43,7	0,1	12,4	31,2	43,7	0,0
TopQuestions_2011-05-25_213800	0,5	2,2	6	8,7	0,3	2,4	6,2	8,9	0,2
TV8_Village_2011-05-03_201300	1,4	7,7	16,9	26,0	1,6	7,5	6,8	15,9	-10,1
TV8_Village_2011-05-12_172800	1,7	5,5	23,7	31,0	1,7	5	11,3	18,1	-12,9
Global	4,2	3,9	18,8	26,9	4,4	3,9	12,5	20,9	-6,0

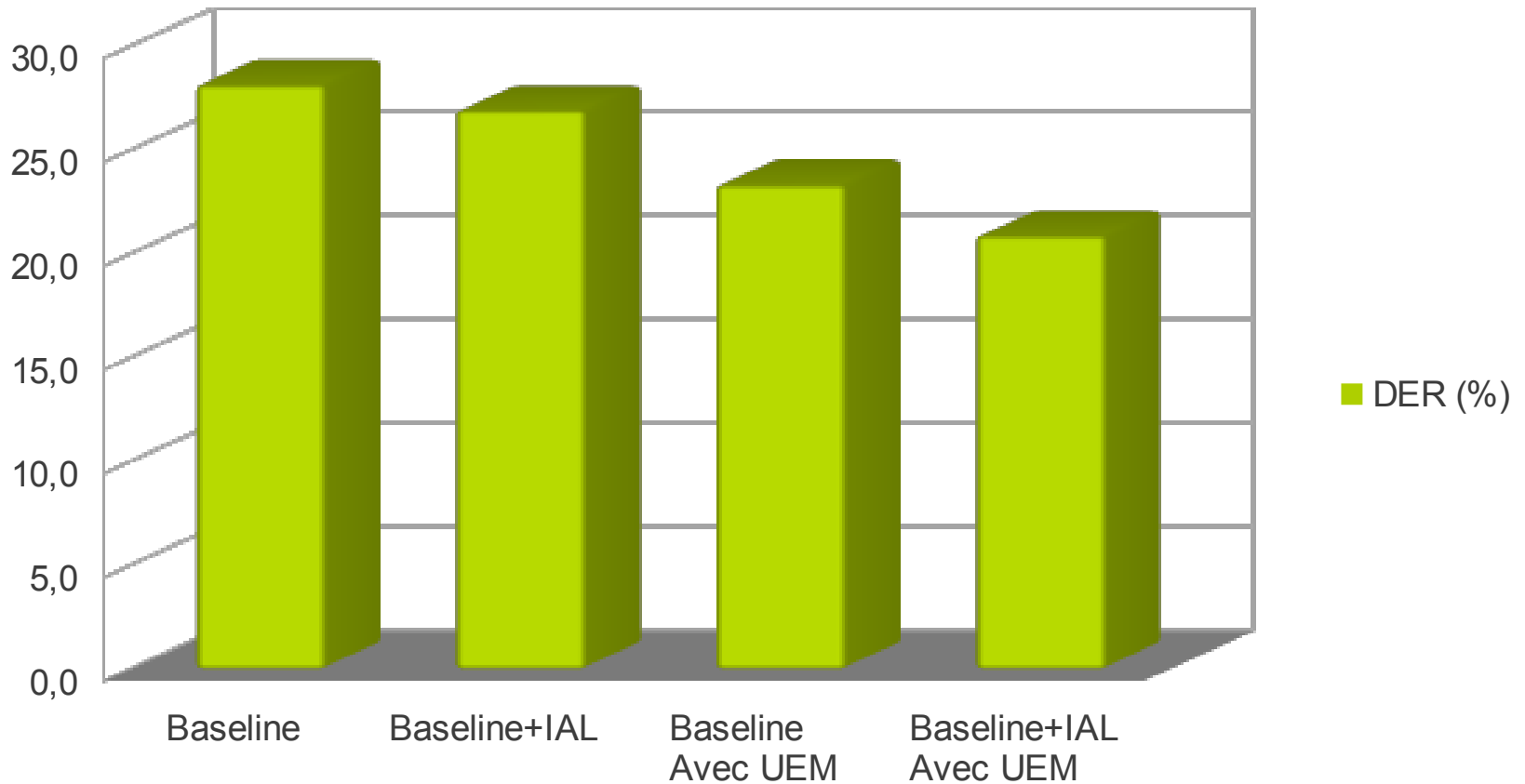
La soumission : SRL-x – un peu revue et "corrigée"

- Scoring ETAPE
- Recouvrement de la parole
- Résultats:
 - Baseline: 55,7% (vs 57%)
 - Baseline+IAL: 37% (vs 37,5%)

Synthèse

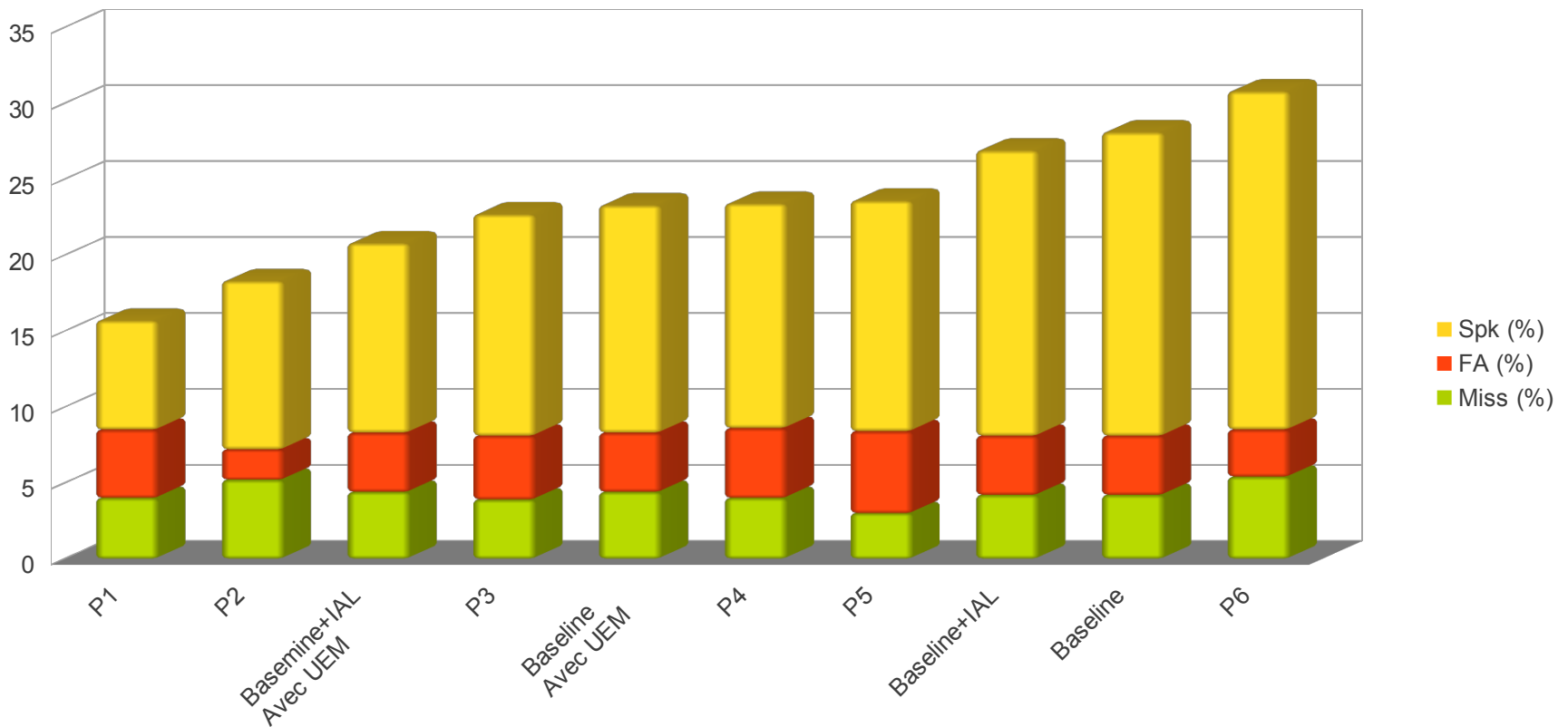
UNIVERSITÉ D'AVIGNON
ET DES PAYS DE VAUCLUSE

- Scoring NIST
- Recouvrement de la parole



Synthèse

- Scoring NIST
- Recouvrement de la parole



Un début d'analyse...

Fichiers	Nb Loc. Ref.	Baseline	Baseline Avec UEM	Baseline+IAL	Baseline+IAL Avec UEM
BFMStory_2011-05-31_175900	21	36 (24)	21 (20)	23 (27)	16 (27)
EST2BC_20100208_1000	29	18 (11)	17 (11)	15 (9)	15 (9)
EST2BC_20100208_1750	10	7 (34)	7 (34)	6 (36)	6 (36)
EST2BC_20101007_2152	11	9 (54)	3 (49)	9 (53)	3 (44)
EST2BC_20101014_2152	9	9 (17)	6 (24)	9 (16)	6 (23)
EST2BC_20101018_0910	7	13 (10)	4 (1)	12 (10)	4 (1)
EST2BC_20101024_2004	6	10 (15)	9 (5)	7 (19)	7 (3)
CaVousRegarde_2011-05-12_235900	19	29 (30)	27 (18)	21 (24)	19 (19)
EntreLesLignes_2011-05-06_192800	5	16 (35)	10 (30)	12 (15)	6 (4)
EntreLesLignes_2011-05-13_192800	4	18 (15)	9 (13)	11 (20)	4 (5)
PileEtFace_2011-05-26_192800	3	14 (14)	7 (16)	12 (11)	4 (11)
TopQuestions_2011-05-18_000400	8	11 (0)	8 (5)	9 (31)	6 (31)
TopQuestions_2011-05-25_213800	7	13 (10)	7 (10)	11 (6)	5 (6)
TV8_Village_2011-05-03_201300	7	16 (22)	9 (18)	13 (17)	5 (7)
TV8_Village_2011-05-12_172800	9	19 (34)	11 (19)	13 (24)	6 (11)
Global	155	238	155	183	112

UEM : Moins de "bruit" =>
Moins de locuteurs !

Un début d'analyse...

UN
ET

Fichiers	Nb Loc. Ref.	Baseline	Baseline Avec UEM	Baseline+IAL	Baseline+IAL Avec UEM
BFMStory_2011-05-31_175900	21	36 (24)	21 (20)	23 (27)	16 (27)
EST2BC_20100208_1000	29	18 (11)	17 (11)	15 (9)	15 (9)
EST2BC_20100208_1750	10	7 (34)	7 (34)	6 (36)	6 (36)
EST2BC_20101007_2152	11	9 (54)	3 (49)	9 (53)	3 (44)
EST2BC_20101014_2152	9	9 (17)	6 (24)	9 (16)	6 (23)
EST2BC_20101018_0910	7	13 (10)	4 (1)	12 (10)	4 (1)
EST2BC_20101024_2004	6	10 (15)	9 (5)	7 (19)	7 (3)
CaVousRegarde_2011-05-12_235900	19	29 (30)	27 (18)	21 (24)	19 (19)
EntreLesLignes_2011-05-06_192800	5	16 (35)	10 (30)	12 (15)	6 (4)
EntreLesLignes_2011-05-13_192800	4	18 (15)	9 (13)	11 (20)	4 (5)
PileEtFace_2011-05-26_192800	3	14 (14)	7 (16)	12 (11)	4 (11)
TopQuestions_2011-05-18_000400	8	11 (0)	8 (5)	9 (31)	6 (31)
TopQuestions_2011-05-25_213800	7	13 (10)	7 (10)	11 (6)	5 (6)
TV8_Village_2011-05-03_201300	7	16 (22)	9 (18)	13 (17)	5 (7)
TV8_Village_2011-05-12_172800	9	19 (34)	11 (19)	13 (24)	6 (11)
Global	155	238	155	183	112

IAL: Regroupement de clusters (même identité) => moins de locuteurs. Attention aux erreurs d'identification !

Conclusion

- IAL, Piste prometteuse, à explorer davantage
 - Baseline: $0,76 \times RT$, Baseline+IAL: $1 \times RT$
- Sur corpus dev ETAPE :
 - SRL : DER sans recouvrement parole => baseline+IAL : 13,8% vs baseline : 19,6%
 - SLR-x : 25,3%
 - Biais possible sur l'identification pouvant aider le système !
- A considérer, le problème des UEM et du "bruit" potentiel amené par certaines zones de parole dans une émission
 - Approche bottom-up sensible également ?