

Systeme de transcription orthographique

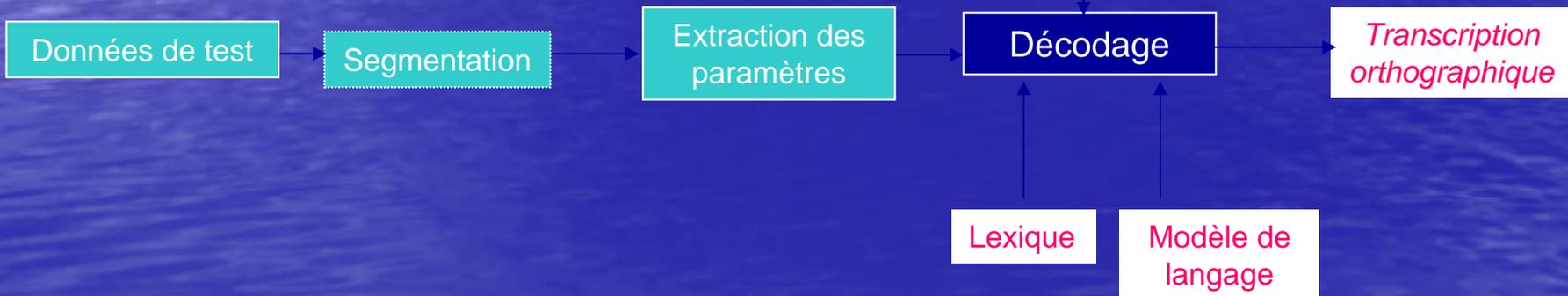
Leila ZOUARI
Gérard CHOLLET
TSI-ENST

Architecture Globale

1. Apprentissage



2. Décodage



- **Segmentation** (réalisée par l'IRISA)

Parole | musique | silence | parole + musique | silence | ..



Parole | parole + musique

- **Paramétrisation**

- 12 coefficients MFCC + delta + delta delta

- Spro 4.0

● Modèles acoustiques

- ❖ Base d'apprentissage : Ester phase1 (train + dev + test) + train phase2
- ❖ 38 modèles (hors contexte)
- ❖ HMMs à 3 états
- ❖ Estimation des paramètres
 - Alignement phonétique
 - Apprentissage modèles 64 gaussiennes
 - Augmentation de gaussiennes à 128
 - Ré-estimation des modèles 128 gaussiennes

- **Lexique**

- + 118 000 mots
- + le monde + corpus d'apprentissage
- + phonétisé

- **Modèle de langage**

- le monde + corpus d'apprentissage
- 4.25 M bi grammes
- 4.25 M trigrammes

- **Décodage**

- SI ROCCO

- Niveaux d'élagage :

- Hypothèses acoustiques
- Hypothèses de mots

- Critère d'élagage : Nombre maximal d'hypothèses simultanées

- 1 seule passe de décodage

Résultats

- **DEV phase2**

	Elagage	Wer
64 gaussiennes	50000 250	40.8
128 gaussiennes	50000 250	39.1

- **TEST phase2**

	Elagage	WER
128 gausiennes	30000 250	45.7

Systeme de suivi du locuteur

Najim DEHAK
Gérard CHOLLET
TSI-ENST

SVL train(1/3)

- **Segmentation** (parole / non parole) : réalisé par IRISA.

Parole | musique | silence | parole + musique | silence | ..



Parole

- **Paramétrisation :**
 - 12 coefficients MFCC + delta + delta delta
 - CMS
 - Spro 3.0

SVL train(2/3)

Modèle du monde UBM

- 1 seul UBM (GMM de 512 gaussiennes) indépendant du genre et des conditions d'acquisitions.
- Base d'apprentissage: 10heures extraite des données d'ester de la phase 1 (train, dev, test) + train phase 2
- Méthode d'apprentissage EM – ML
- Logiciel utilisé pour l'apprentissage: Becars

SVL train(3/3)

Modèle du locuteur

- Base d'apprentissage: les données des locuteurs sont extrait des données d'ester de la phase 1 (train, dev, test) + train phase 2.
- Adaptation MAP.
- Adaptation des moyennes des gaussiennes.
- Logiciel utilisé pour l'apprentissage: Becars

SVL test (1/2)

- Segmentation (parole / non parole) : réalisé par IRISA.
- Détection d'un changement du locuteur dans un segment de parole.
 1. Utilisation du critère BIC (3 passes)
 2. Passe 1 : fenêtre de 200 trames, décalage de 20 trames
passe 2 : fenêtre de 100 trames, décalage de 10 trames
 3. Seuil de décision de 1.5 fixé dans le Dev de la phase 2
 4. logiciel : Audioseg (IRISA)

SVL test (2/2)

- Calcul des scores
 - Le log du rapport de vraisemblance
 - Pas de normalisation
 - Becars
- Décision
 - Seuil unique fixé sur les données Dev de la phase 2
 - Choisir le locuteur avec le plus grand score

SVL résultat

- Seuil de décision pour le critère BIC : 1.5
- Seuil de décision pour la détection des locuteurs : 1
- Taux moyen sur la partie Test : 0.475

This document was created with Win2PDF available at <http://www.daneprairie.com>.
The unregistered version of Win2PDF is for evaluation or non-commercial use only.