

Introduction

► Objectif global du projet

- ▷ créer un dispositif portable et autonome
- ▷ intégrer un système de reconnaissance de la parole
 - * l'adapter aux besoins des personnes sourdes ou malentendantes
 - améliorer la communication entre les personnes sourdes et leur entourage
 - outil de socialisation et / ou d'intégration au lieu de travail
 - * l'adapter aux contraintes imposés par la solution embarquée
 - capacité mémoire & puissance de calcul

► Approche

- ▷ cibler seulement les personnes avec une bonne maîtrise du Français écrit
- ▷ faire des sacrifices par rapport à la taille de modèles de reconnaissance

Premier objectif : extraire des informations linguistiques pertinentes

- évaluer différentes unités linguistiques : mots, phonèmes, syllabes
 - les syllabes offrent des bonnes performances, malgré un lexique limité
- l'importance des mots pour la compréhension de la transcription par des personnes sourdes
- faire un compromis : combiner mots et syllabes dans un seul modèle de langage
 - ▷ assurer une reconnaissance correcte des mots les plus fréquents
 - ▷ proposer des suites de syllabes pour les segments hors vocabulaire

Création d'un modèle de langage hybride

- constituer un corpus d'apprentissage qui repose sur ces deux unités lexicales
- le vocabulaire est défini en sélectionnant
 - ▷ les mots les plus fréquents
 - ▷ les syllabes correspondant aux mots hors vocabulaire
- **Méthode pour définir les syllabes**
 - ▷ corpus d'apprentissage entièrement phonétisé (par alignement forcé)
 - * pour prendre en compte les événements de **liaison & réduction**
 - ▷ séquence de phonèmes traitée par l'outil de syllabation
 - ▷ règles de syllabation [Bigi et al, 2010]
 - * une syllabe contient une seule voyelle
 - * une pause désigne une frontière de syllabe

Exemples de règles de syllabation

Séquence de phonèmes	Position de coupure	Syllabes obtenues
VV	0	V V
VxV	0	V xV
VxxV	1	Vx xV
VxxxV	2	Vxx xV

► Exemple d'une transcription "mots & syllabes"

quel est le prix du **tournevis**
 quel est le prix du **t u r n s w a v i s** ← alignement forcé
 quel est le prix du **t . u r n . s w a . v i s** ← mots & syllabes

► imposer différents seuils minimaux sur la fréquence d'occurrence des mots

$$\theta \in \{3, 5, 10, 25, 50, 100, 300\}$$

- différentes transcriptions du corpus d'apprentissage
- différents lexiques et modèles de langage

Configuration

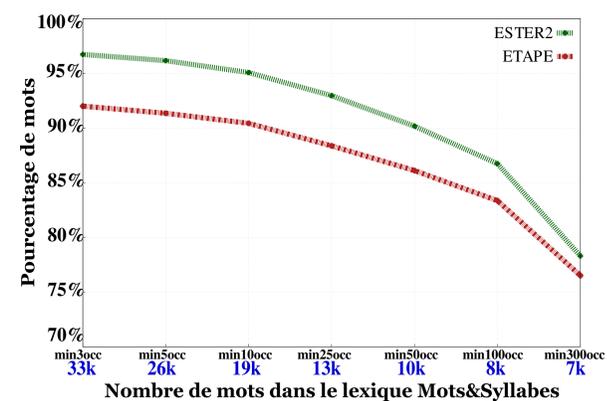
- Analyse acoustique MFCC
 - 12 paramètres MFCC et le logarithme de l'énergie par trame (+ Δ , $\Delta\Delta$)
 - fenêtre de 32 ms, décalage de 10 ms
- SRILM : apprentissage des modèles de langage
- Sphinx3 : apprentissage des modèles acoustiques
 - des modèles HMM (avec des mélanges de 64 gaussiennes)
 - adaptés homme/femme
- PocketSphinx : décodage & mesures de confiance (probabilité a posteriori)

Les données utilisés lors de nos expériences

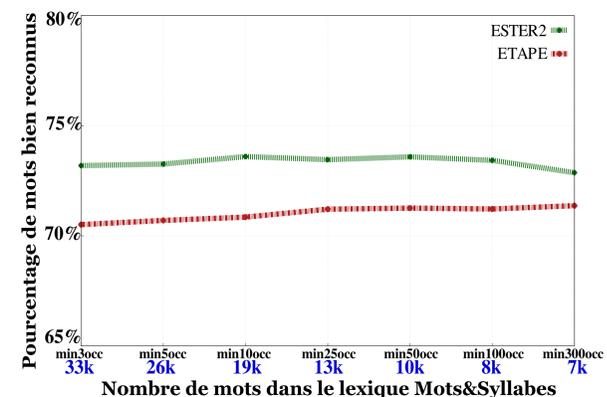
- Pour l'apprentissage des **modèles acoustiques phonétiques**
 - ▷ les données de l'ensemble d'apprentissage d'ESTER2 et d'ETAPE
 - ▷ les données transcrites du corpus EPAC
 - ▷ 300h de parole et 4 millions de mots
- Pour l'apprentissage des **modèles de langage hybrides**
 - ▷ les corpus des transcriptions d'ESTER2, d'ETAPE et d'EPAC
 - ▷ **après alignement forcé et transformation mots+syllabes**
- Pour les tests : les ensembles de développement d'ESTER2 et d'ETAPE

Résultats

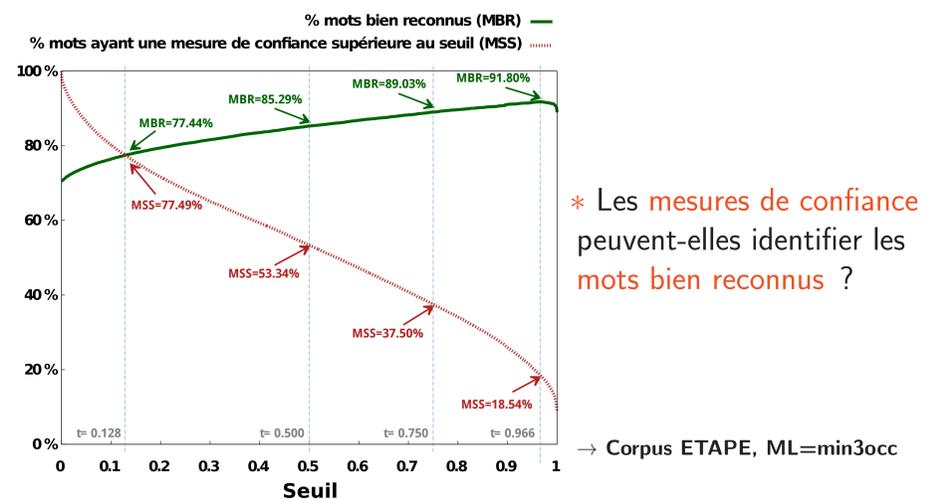
► But final : récupérer le message porté par la parole (par les mots)



* Combien de mots ont été produits par le décodeur ?



* Parmi ces mots, combien d'entre eux ont été bien reconnus ?



* Les mesures de confiance peuvent-elles identifier les mots bien reconnus ?

- La contribution des **mesures de confiance sur les syllabes**
 - pertinente seulement s'il existe une quantité relativement importante de syllabes dans le modèle de langage

Conclusions

- le modèle de langage hybride est un compromis efficace
- parmi les mots reconnus qui ont une mesure de confiance supérieure à 0.5, 85% d'entre eux sont corrects

Travaux futurs

- étudier d'autres solutions pour mieux modéliser les syllabes à l'intérieur d'un modèle hybride
- analyser les mesures de confiance sur les syllabes
- analyser les segments correspondants aux mots hors vocabulaire