

Reconnaissance de la parole pour l'aide à la communication pour les sourds et malentendants

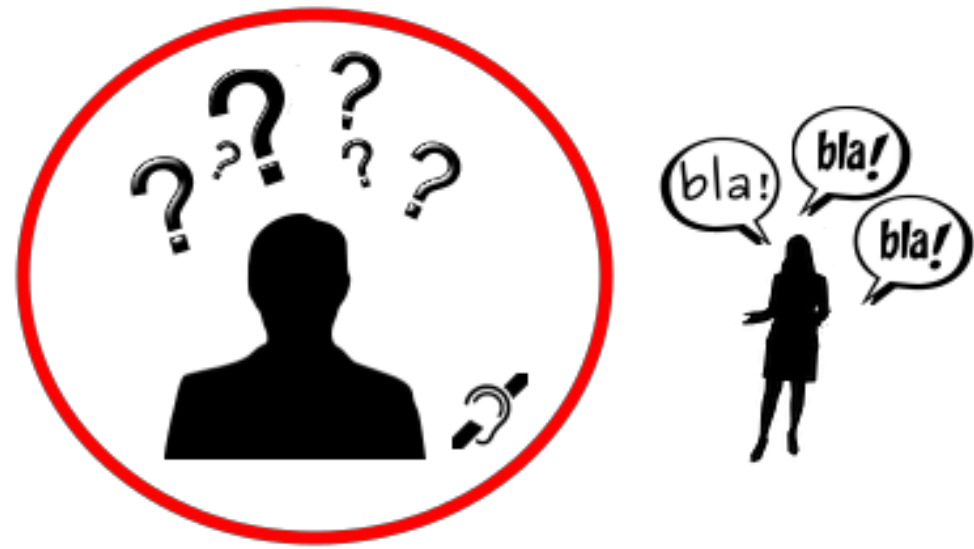
Luiza Orosanu
Denis Jouvét



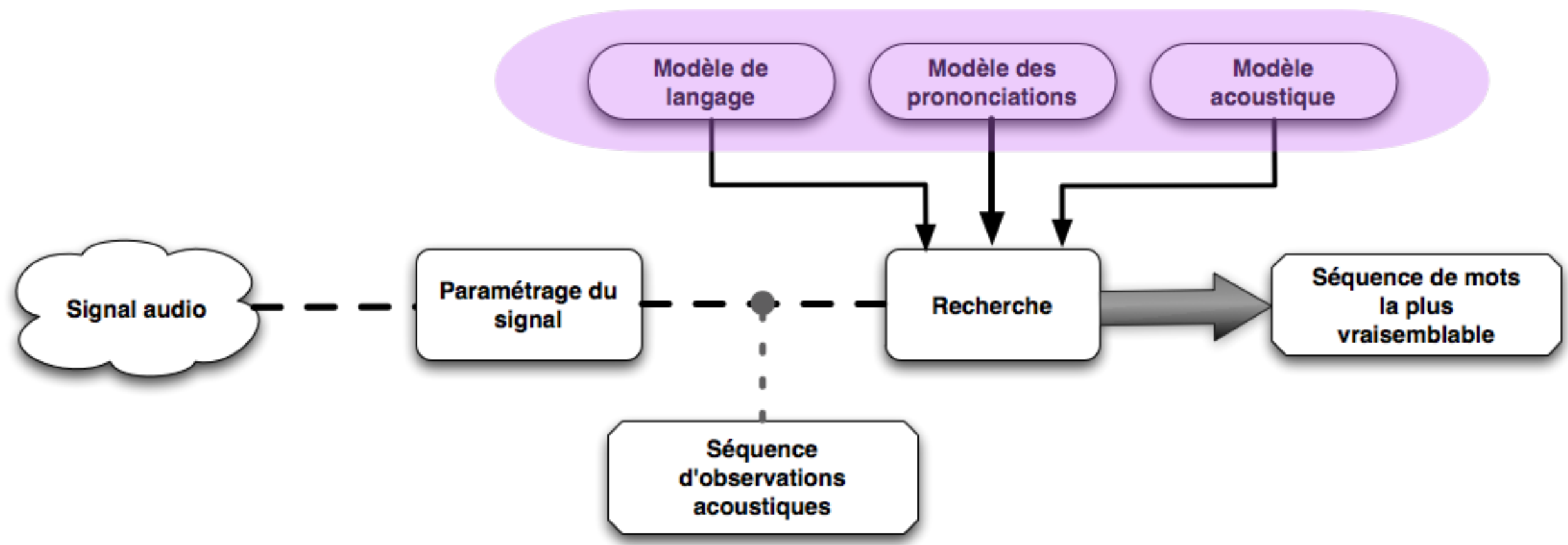
Équipe PAROLE, Loria
Nancy, France

Objectif global du projet RAPSODIE

Aider les **personnes sourdes ou malentendantes**
(améliorer la communication entre les sourds et leur entourage)



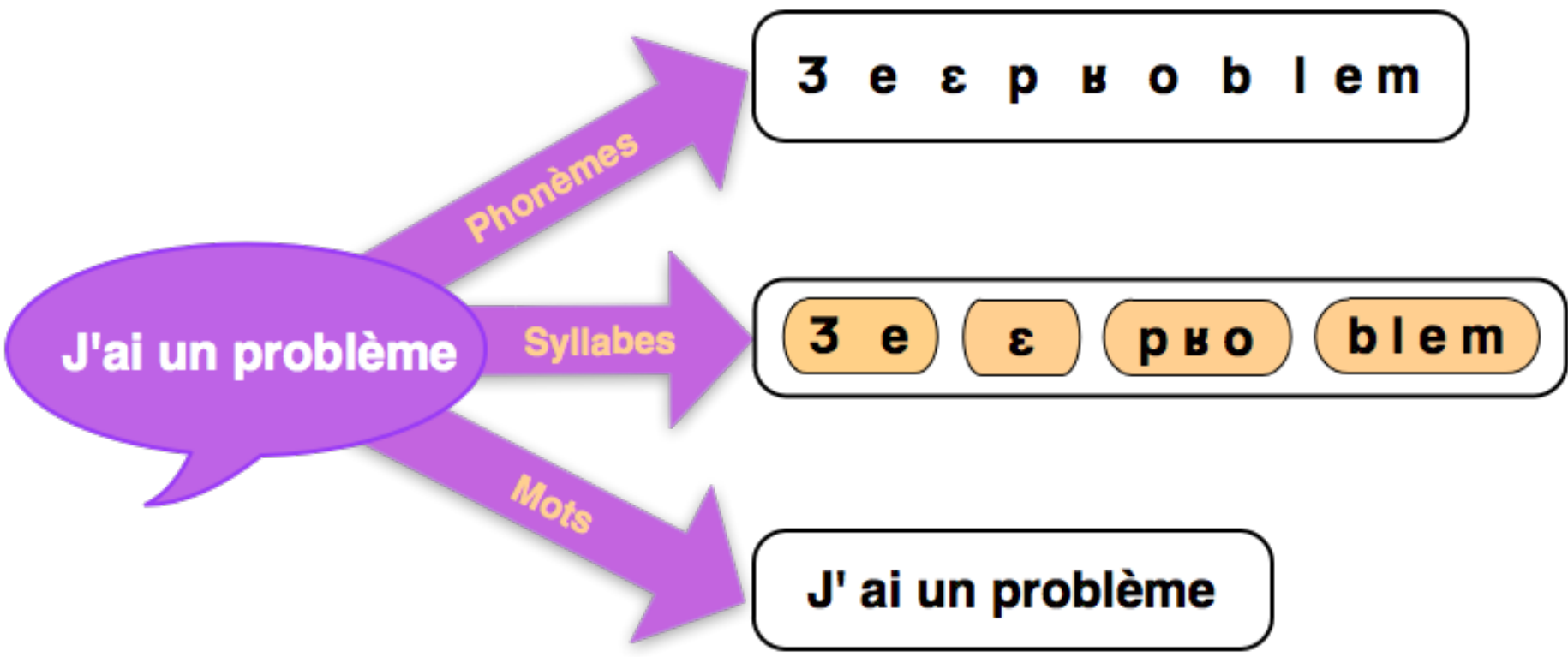
Créer un système automatique de **transcription de la parole**



adapté aux contraintes imposés par une **solution embarquée** :
capacité mémoire & puissance de calcul limitées

Premier objectif : extraire des informations linguistiques

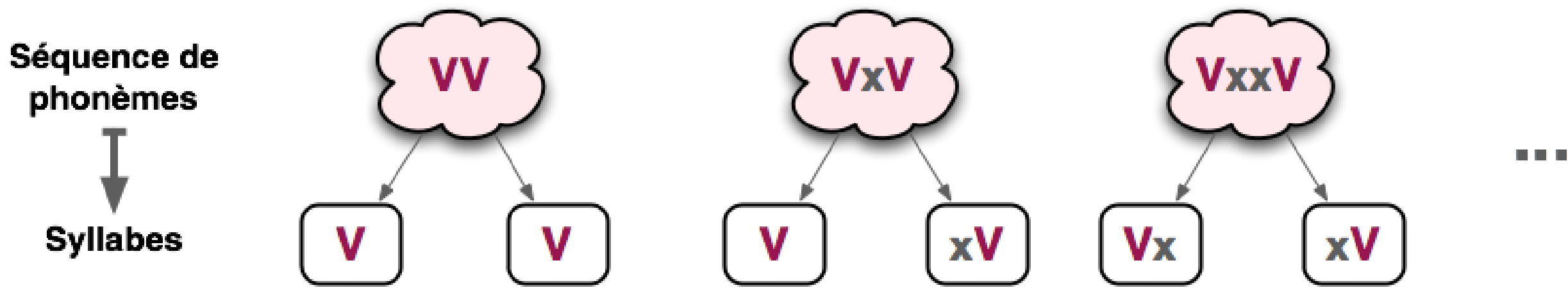
► différentes unités linguistiques évaluées : mots, phonèmes, syllabes



- avec un lexique limité, les syllabes offrent des bonnes performances
- à prendre en compte
 - ▷ l'importance des mots pour la compréhension du message pour les sourds
 - ▷ la problématique des mots hors de vocabulaire
- 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
 - ▷ conserver les mots les plus fréquents
 - ▷ décomposer en syllabes les mots hors vocabulaire
- **Méthode pour définir les syllabes**
 - ▷ corpus d'apprentissage entièrement phonétisé
 - * pour prendre en compte les événements de **liaison & réduction**
 - ▷ séquence de phonèmes traitée par des règles de syllabation [Bigi et al, 2010]
 - * une syllabe contient une seule voyelle
 - * une pause désigne une frontière de syllabe



Exemple d'une transcription "mots & syllabes"

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

Résultats

- le modèle de langage hybride est un compromis efficace
- parmi les 69 à 96% de mots qui sont reconnus par le système, environ 70% d'entre eux sont effectivement correctement reconnus
- parmi les mots reconnus qui ont une mesure de confiance supérieure à 0.5, 85% d'entre eux sont corrects

Deuxième objectif : extraire des informations complémentaires

- détecter la modalité des énoncés : **question ou affirmation**
- dans un contexte interactif il est important de savoir si un message correspond à une question
- deux types de questions
 - ▷ exprimées par des tournures interrogatives
 - ▷ perçues comme des interrogations qu'au travers de la prosodie

Méthodologie pour la détection de questions

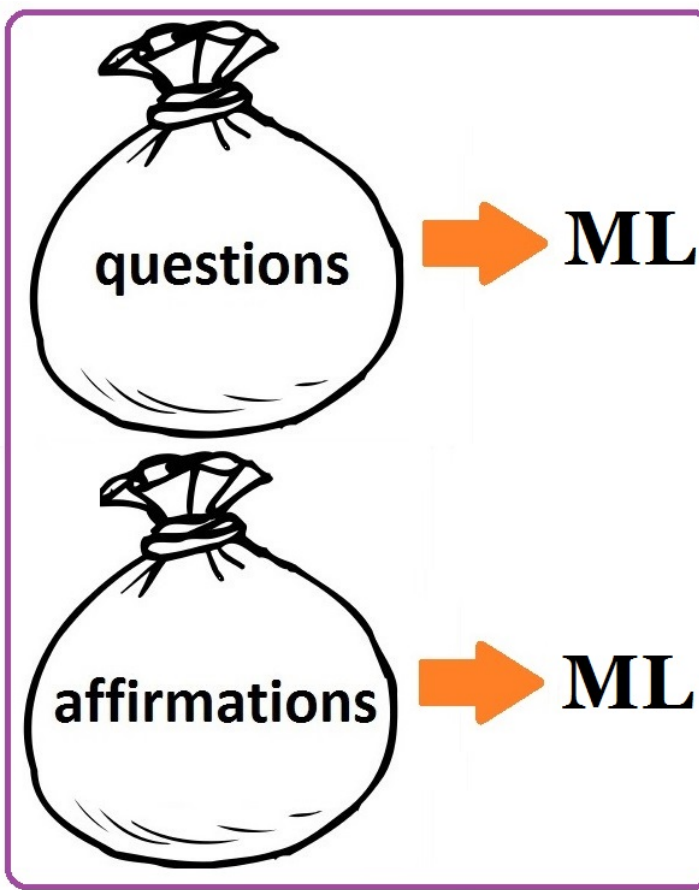
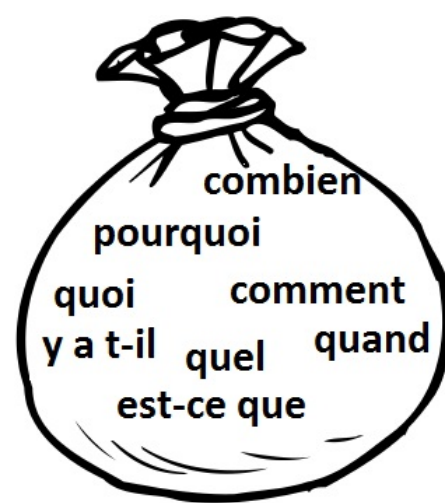
- évaluer plusieurs approches
 - ▷ un classifieur prosodique : utilise la prononciation
 - ▷ un classifieur linguistique : utilise les mots qui compose la phrase
 - ▷ un classifieur combiné : utilise les informations prosodique et linguistique

Les paramètres linguistiques

- ▷ les tournures interrogatives : pour indiquer la présence ou l'absence d'une tournure interrogative dans la phrase
- ▷ la probabilité que la phrase soit une question
 - * par rapport à deux modèles de langage de référence

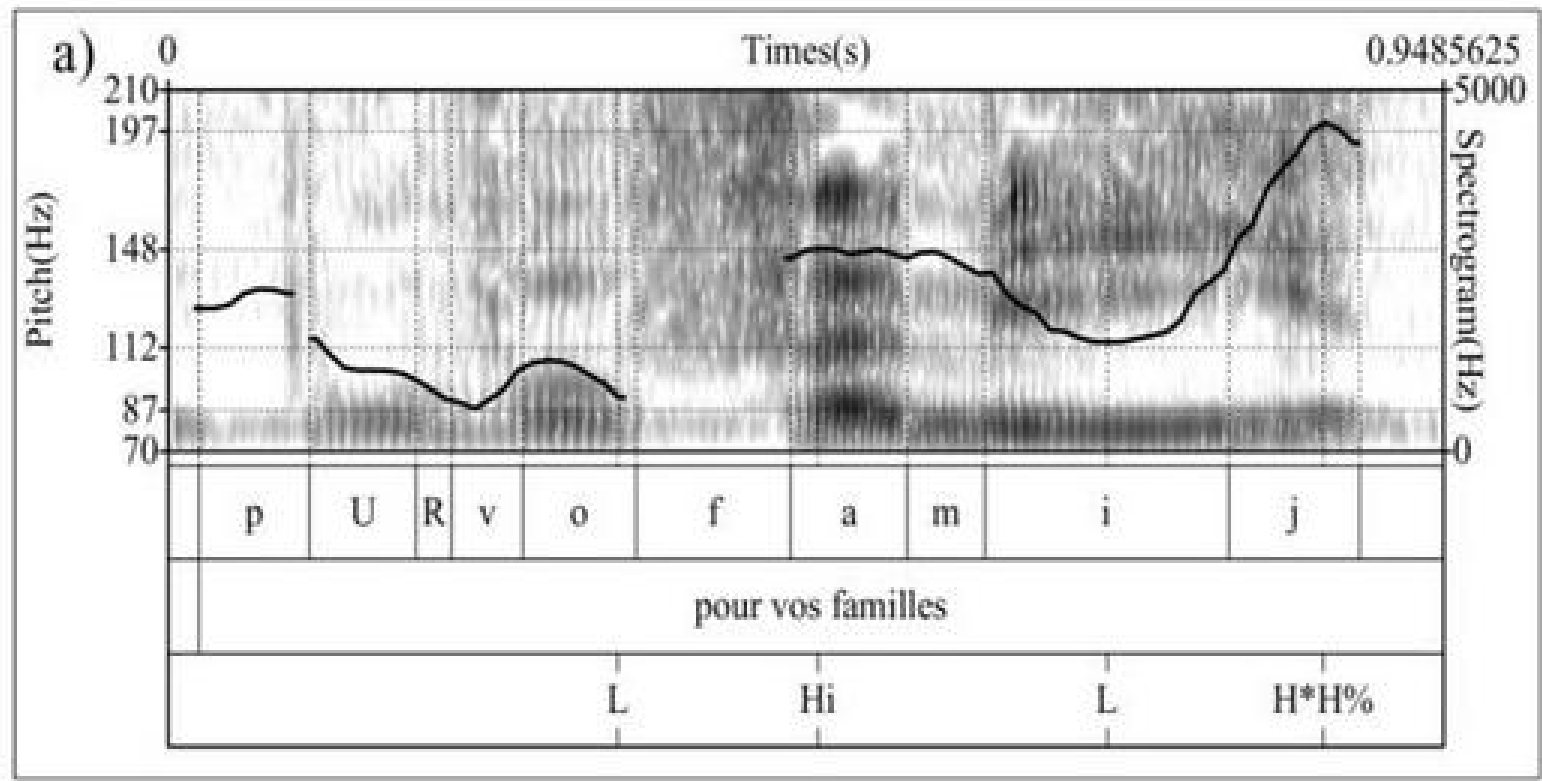
$$LL(\text{phrase}) = \text{Log} \left(\frac{P(\text{phrase}|\text{ML-question})}{P(\text{phrase}|\text{ML-affirmation})} \right)$$

- * $LL \geq 0 \rightarrow$ susceptible d'être une question
- * $LL < 0 \rightarrow$ susceptible d'être une affirmation



Les paramètres prosodiques

- ▷ l'intonation d'une question comporte en général un pitch final montant
- ▷ 10 paramètres prosodiques qui prennent en compte la durée, l'énergie et le pitch du dernier groupe prosodique de la phrase



- **L'évaluation** est effectuée en utilisant des paramètres linguistiques issus des transcriptions **manuelles** et **automatiques**

Résultats

- l'essentiel de l'information pour la détection de la modalité des énoncés provient du contenu linguistique de l'énoncé
- lorsque l'on se place dans un contexte applicatif (transcription automatique), les performances du détecteur linguistique baissent, et la prosodie apporte un complément d'information significatif

Travaux futurs

- étudier d'autres solutions pour mieux modéliser les syllabes à l'intérieur d'un modèle hybride
- analyser les segments correspondants aux mots hors vocabulaire
- étudier plus les caractéristiques prosodiques et linguistiques
- prendre en compte les mesures de confiance dans le calcul des paramètres linguistiques