

# MASTER LANGUE ET INFORMATIQUE

## Reconnaissance de la parole

### Projet

#### Apprentissage de Modèles de Markov cachées et détection de mots clés

La détection de mots-clés (word spotting) dans la parole continue consiste à reconnaître et à localiser toutes les occurrences des mots d'une liste de mots-clés dans un continuum de parole donné. L'objectif de ce projet individuel est de mettre en œuvre une telle détection à l'aide de modèles de Markov cachées phonétiques.

Le logiciel **HTK** (Hidden Markov model Tool Kit) regroupe un ensemble d'outils de développement de système de reconnaissance markovien; c'est à partir de cette plateforme que votre système sera développé. HTK vous sera présenté aux ateliers 4, 5 et 6. Il vous est cependant conseillé de consulter le site [htk.eng.cam.ac.uk](http://htk.eng.cam.ac.uk) au préalable.

Les différentes étapes de développement sont les suivantes :

#### 1) ELABORATION DES RESSOURCES

Chaque auteur de projet a transcrit les différents tours de parole correspondant au fichier audio *synthRadiox*. Les tours de parole des deux premières minutes seront découpés en autant de fichiers audio. Les modèles de Markov cachées seront appris sur ces fichiers audio qui constitueront le corpus d'apprentissage. Les tests auront lieu sur la dernière minute. Pour les besoins du processus d'apprentissage, la transcription phonétique de chaque tour de parole est indispensable et vous devrez phonétiser les tours de parole avec le codage *SAMPA* vu au cours 5 et à l'atelier 5.

#### 2) SEGMENTATION ET APPRENTISSAGE DES MACHINES DE MARKOV PHONÉTIQUES

Il est indispensable pour apprendre des modèles de Markov cachés d'avoir une segmentation du corpus d'apprentissage. Cette segmentation peut être faite manuellement ou automatiquement à l'aide des modèles de Markov appris sur un autre corpus (cf. cours 5). Chacun des phonèmes sera appris via une machine de Markov spécifique à ce phonème. L'apprentissage consiste à apprendre les paramètres de la machine à partir du corpus d'apprentissage. En cas de données insuffisantes, vous pouvez compléter ce corpus par des données provenant d'autres corpus (contacter-moi dans ce cas).

#### 3) DECODAGE ACOUSTICO-PHONÉTIQUE ET DETECTION DE MOTS-CLES

Le décodage acoustico-phonétique consiste à extraire d'un flux audio, la suite des phonèmes reconnus. En utilisant les modèles de Markov appris à l'étape précédente et le réseau de décodage vu en atelier, vous devez décoder phonétiquement le corpus de test. Quel est le taux d'erreur ? Vous construirez ensuite un réseau de décodage permettant de détecter une liste choisie de mots-clés du corpus de test. Le but sera de maximiser la Figure Of Merit (FOM) avec un système de récompenses et de pénalités bien choisi (cf. atelier 6). Donner les résultats de détection ?

#### DEADLINES

- Le rapport final (au minimum 10 pages) est individuel. Il comprendra, entre autres, la méthodologie de construction des modèles de Markov, la segmentation phonétique des tours de parole, les réseaux de décodage, les résultats de reconnaissance et leurs analyses, une introduction, une conclusion et les références bibliographiques employées.
- La recette de votre projet comprenant votre rapport, l'arborescence complète de votre système de reconnaissance devra être envoyée par mail à [Claude.Montacie@sorbonne-universite.fr](mailto:Claude.Montacie@sorbonne-universite.fr)