

Modélisation et conception de système dédié à l'enregistrement de corpus de parole.

Un corpus est un ensemble de données organisées contenant des informations bas-niveau (par exemple signal sonore) et haut-niveau (par exemple POS tags). La plupart des domaines de recherche en intelligence artificielle utilisent de tels corpus, notamment la synthèse de la parole ou le traitement automatique des langues. Les informations caractéristiques d'un corpus de parole varient selon le domaine pour lequel il est conçu. Il contient généralement un texte associé à un signal de parole ainsi que des informations acoustiques, linguistiques, phonétiques appropriées.

Pour enregistrer un corpus de parole dédié à la synthèse de la parole, on a recours à un protocole simple, qui consiste à demander à un locuteur de prononcer un ensemble de phrases ou de mots défini suivant l'application souhaitée. Classiquement, le locuteur lit un texte écrit sur papier ou sur un support numérique couvrant l'ensemble des sons de la langue étudiée. La voix du locuteur est enregistrée dans de très bonnes conditions (microphone, cabine d'enregistrement, etc.) afin d'obtenir un signal de bonne qualité.

Actuellement, le signal de parole est segmenté en unités temporelles (phrases, mots, phonèmes) de manière manuelle ou automatique, puis chacun des segments est aligné avec la portion de texte correspondante. Les outils existants permettent de réaliser l'alignement une fois que tous les signaux de parole sont enregistrés. Ce processus est coûteux en terme de temps et de main d'oeuvre et peut introduire un certain nombre d'erreurs dans l'alignement.

L'objectif du projet est de proposer un système qui intègre les outils nécessaires à l'enregistrement d'un corpus pour la synthèse de parole et qui permet au locuteur d'interagir avec l'expérimentateur par le biais d'une interface automatisée. Les livrables attendus sont une analyse des besoins, la construction d'un cahier des charges et la modélisation en UML du système proposé.

References