

Résumé

Cette thèse porte sur l'étude d'une méthode de recherche des N meilleures solutions, puis son adaptation au système de reconnaissance automatique de la parole du CNET et sur le développement de post-traitements segmentaux ou syntaxiques pour retrouver la "meilleure" solution parmi les N proposées.

Le système réalisé comporte deux modules. Le module markovien et le module de post-traitement.

Le module markovien se charge de générer les N meilleures solutions pour un mot prononcé (ou une phrase). La méthode employée consiste à utiliser l'algorithme de Viterbi dans la phase aller (effectuée de manière synchrone) et l'algorithme A* dans la phase retour (effectuée de manière asynchrone). Cette recherche est introduite au niveau le plus bas de la modélisation, c'est-à-dire au niveau acoustique. Nous montrons en fin de compte que le temps de recherche des N meilleures solutions n'est qu'une fonction affine du nombre N , pour N petit.

Le module de post-traitement se charge de retrouver la solution "correcte" dans la liste proposée. Pour cela deux approches sont étudiées : le post-traitement syntaxique et le post-traitement segmental statistique.

Beaucoup de syntaxes sont trop complexes pour être traitées directement. Dans le cas des villes épelées l'information syntaxique est la liste des villes possibles. Cependant cette liste est difficile à intégrer dans l'algorithme de décodage. Une solution consiste à rechercher les N meilleures solutions pour chaque ville épelée et à utiliser par la suite un post-traitement syntaxique pour retrouver parmi les N solutions proposées la première qui se trouve dans la liste des villes. Les résultats obtenus pour la reconnaissance de villes épelées dans une liste de 30000 villes, ont permis de montrer que le taux de substitution était très faible 3% avec un taux de rejet de 13%.

Cependant dans le cas où il n'y a pas de contraintes syntaxiques (reconnaissance de mots isolés par exemple) le post-traitement syntaxique n'est plus adapté. Dans ce cas nous utilisons un post-traitement segmental. Le post-traitement segmental consiste à calculer pour chaque solution, un score segmental qui est combiné par la suite avec le score markovien. La solution fournie par le système est celle ayant le score de combinaison le plus élevé.

Alors que la méthode généralement utilisée pour calculer le score segmental d'une solution utilise des réseaux de neurones, nous proposons dans le cadre de cette thèse d'utiliser une approche statistique. Cette approche repose sur une nouvelle technique dans la représentation statistique des segments. Deux modèles sont associés à chaque segment : le premier représente les statistiques d'une segmentation correcte et le second représente les statistiques d'une segmentation incorrecte. Cette approche a été testée sur différents corpus de données et a permis une réduction du taux d'erreur de l'ordre de 15 à 25 % par rapport à l'utilisation du modèle de Markov seul.

