Le système français de dictée vocale du LIMSI: développements et évaluation

M. Adda-Decker, G. Adda, L. Lamel, J.L. Gauvain

LIMSI - CNRS, B.P. 133, 91403 Orsay, France
{madda,gadda,lamel,gauvain}@limsi.fr

Nous avons développé au LIMSI un système de dictée vocale en français, anglais et allemand ([1], [2]). Ce système est indépendant du locuteur, capable de traiter de grands vocabulaires (jusqu'à 65.000 mots), en parole lue. La version anglaise de ce sytème a été évaluée avec succès lors des derniers tests ARPA américains ([3]). Une évaluation multilingue a été organisée en Europe au printemps 95 dans le cadre du projet LRE-SQALE. Lors de cette évaluation des lexiques de 20.000 mots ont été utilisés pour le français, l'anglais américain et l'anglais britannique, alors que le système allemand a fait appel à un lexique de 64.000 mots. Dans le cadre du projet de *Dictée Vocale* de l'ARC-B1 de l'AUPELF nous élargissons notre système français à 65.000 mots.

Notre système de reconnaissance repose sur des principes de modélisation markovienne tant au niveau acoustique, qu'au niveau linguistique. Les connaissances spécifiques à la langue française sont apportées par les bases de données (parole et texte), le choix d'un ensemble de phonèmes, et un lexique donnant pour chaque mot graphémique une ou plusieurs transcriptions phonétiques utilisant les symboles phonétiques fixés auparavant. Le système de reconnaissance utilise des modèles acoustiques de phones en contexte, chaque phone en contexte étant représenté par une chaîne de Markov cachée à 3 états avec des densités d'observations multi-gaussiennes. Les modèles de langage utilisés sont des statistiques N-grammes (bigrammes et trigrammes) estimés sur des textes de journaux.

Nous proposons de décrire dans cet article nos travaux de développement du système de reconnaissance concernant le choix du lexique, le choix du modèle de langage, les transcriptions phonétiques et les modèles acoustiques. Ces travaux sur la modélisation du français motivés par une évaluation des divers systèmes francophones existants permettra d'améliorer à la fois les performances du système et notre compréhension des différents problèmes posés.

Concernant la langue française des problèmes spécifiques ont d'ores et déjà pu être mis en évidence. Ainsi des sources d'erreurs de reconnaissance pour le français sont par exemple le nombre élevé de mots courts (à un ou deux phonèmes), le nombre élevé de homophones, l'occurrence non systématique de liaisons et de e-muet à la frontière de mot. Pour une même taille de lexique on obtient une couverture lexicale plus faible en français qu'en anglais, mais plus élevée qu'en allemand. Les résultats seront analysés afin de préciser les problèmes spécifiques au français.

(1) L.F. Lamel, M. Adda-Decker, G. Adda, J.L. Gauvain, "Reconnaissance Multilingue de Grands Vocabulaires," école d'été GDR-PRC CHM, Marseille, juillet 1995 (à paraître dans *livre GDR-AUPELF 1996*).

(2) J.L. Gauvain, L.F. Lamel, G. Adda, M. Adda-Decker, "Speaker Independent Continuous Speech Dictation," Speech Communication, Oct. 1994.

(3) J.L Gauvain, L.F. Lamel, G. Adda, D. Matrouf, "Developments in Continuous Speech Dictation using the 1995 ARPA NAB News Task," Proceedings ICASSP, 1996.