



MINISTÈRE DE L'ÉDUCATION
NATIONALE, DE L'ENSEIGNEMENT
SUPÉRIEUR ET DE LA RECHERCHE

>>>

AVIS DE SOUTENANCE DE THESE DE DOCTORAT

Monsieur Loïc BARRAULT soutiendra une thèse
le 18 juillet 2008 à 14h 30

L.I.A.

Amphi B. Pascal

SPÉCIALITÉ : INFORMATIQUE ED 166

Titre de la thèse : diagnostic pour la combinaison de systèmes de reconnaissance automatique de la parole.

Membres du jury :

ANDRE-OBRECHT Régine PR Informatique Université Paul Sabatier,
LAFACE Pietro PR Informatique Politecnico di Torino, Italie
MÉLONI Henri PR Informatique Université d'Avignon et des Pays de Vaucluse
ADDA-DECKER Martine CR1-CNRS HDR Informatique Université de Paris-Sud
GRAVIER Guillaume CR Informatique INRIA Rennes
DE MORI Renato PR Informatique Université d'Avignon et des Pays de Vaucluse
MATROUF Driss MCF Informatique Université d'Avignon et des Pays de Vaucluse

Résumé de la thèse :

La Reconnaissance Automatique de la Parole (RAP) est affectée par les nombreuses variabilités présentes dans le signal de parole. En dépit de l'utilisation de techniques sophistiquées, un système de RAP seul n'est généralement pas en mesure de prendre en compte l'ensemble de ces variabilités. Nous proposons l'utilisation de diverses sources d'information acoustique pour augmenter la robustesse des systèmes de reconnaissance. La combinaison de différents jeux de paramètres acoustiques repose sur l'idée que certaines caractéristiques du signal de parole sont davantage mises en avant par certains jeux de paramètres que par d'autres. L'intérêt est donc d'exploiter les points forts de chacun. Par ailleurs, les différentes partitions de l'espace acoustique opérées par les modèles acoustiques peuvent être mises à profit dans des techniques de combinaison bénéficiant de leur éventuelle complémentarité.

Le diagnostic est au cœur de ce travail. L'analyse des performances de chaque jeu de paramètres permet la mise en évidence de contextes spécifiques dans lesquels la prédiction du résultat de reconnaissance est possible. Nous présentons une architecture de diagnostic dans laquelle le système de RAP est vu comme un « canal de transmission » dont l'entrée correspond aux phonèmes contenus dans le signal de parole et la sortie au résultat de reconnaissance. Cette architecture permet de séparer les différentes sources d'ambiguïté à l'intérieur

du système de reconnaissance. Les analyses ont permis d'intégrer des stratégies de combinaison post-décodage à un niveau segmental élevé (phonème ou mot).

Des techniques de combinaison des probabilités a posteriori des états d'un modèle de Markov caché connaissant un vecteur de paramètres acoustiques sont également proposées. Afin d'améliorer l'estimation des probabilités a posteriori, les probabilités obtenues avec différents modèles acoustiques sont fusionnées. Pour combiner les probabilités de manière cohérente, les modèles acoustiques doivent avoir la même topologie. Par conséquent, nous avons développé un protocole permettant d'entraîner des modèles de même topologie avec

des paramètres acoustiques différents. Plusieurs méthodes pour estimer des facteurs de pondération et pour générer des modèles acoustiques complémentaires sont également présentées.