

THESE DE DOCTORAT

Présentée par

Jamal Eddine Rougui

Titre: Indexation de documents audio : Cas des grands volumes de données

Discipline : Sciences de l'ingénieur.

Spécialité : Informatique et Télécommunications.

U.F.R : Informatique et Télécommunications.

Période d'accréditation : 2005-2008.

Directeur de l'UFR : Prof. Driss Aboutajdine.

Directeurs de thèse : Prof. Driss Aboutajdine / Nouredine Mouaddib.

Soutenance :

Date : 16 Juillet 2008.

Heure : 15h.

Lieu : Amphi A.

Devant le jury :

Président :

D. Aboutajdine, Professeur à la Faculté des Sciences de Rabat.

Examineurs :

N. Mouaddib, Professeur à l'Ecole Polytechnique, Nantes.

EL. Mouaddib, Professeur à l'Université de Picardie Jules Vernes, Amiens.

D. Mammass, Professeur à Faculté des Sciences Ibn zouhr, Agadir.

K. Daoudi, Chargé de recherche 1ère classe au CNRS, Toulouse.

M.B García, professeur, University de Deusto, Bilbao, Espagne.

L. Gelgon, Maître de conférence HDR à l'Ecole Polytechnique, Nantes.

M. Rziza, Professeur Assistant à la Faculté des Sciences, Rabat.

Résumé : Cette thèse est consacrée à l'élaboration et l'évaluation des techniques visant à renforcer la robustesse des systèmes d'indexation de documents audio au sens du locuteur. L'indexation audio au sens du locuteur consiste à reconnaître l'identité des locuteurs ainsi que leurs interventions dans un flux continu audio ou dans une base de données d'archives audio, ne contenant que la parole. Dans ce cadre nous avons choisi de structurer les documents audio (restreints à des journaux radiodiffusés) selon une classification en locuteurs. La technique utilisée repose sur l'extraction des mel-cepstrales coefficients, suivi par l'apprentissage statistique de modèles de mélange de gaussiennes (MMG) et sur la détection des changements de locuteur au moyen de tests d'hypothèses Bayésiens. Le processus est incrémental : au fur et à mesure que de nouveaux locuteurs sont détectés, ils sont identifiés à ceux de la base de données ou bien, le cas échéant, de nouvelles entrées sont créées dans la base.

L'organisation des MMG des locuteurs en structure arborescente est un choix qui permet de réduire considérablement la complexité du traitement lors du passage à l'échelle incrémental. Deux principales contributions ont été réalisées dans ce travail de thèse :

- ✓ le calcul de la mesure de similarité entre les modèles des locuteurs ne fait intervenir que les paramètres des MMG. Différentes approximations de la divergence de Kullback-Leibler (KL) ont été adaptées au MMG. En effet, les nouvelles expressions de KL (KLm, KLut et KLut_m) offrent ainsi une meilleure discrimination entre les modèles des locuteurs à moindre coût ;
- ✓ grâce au développement des techniques de mesure de similarité entre GMM, des algorithmes d'organisation hiérarchique des MMG en structure (binaire, n'aire, Treillis) ont été réalisés et testés à la fois sur un lot de GMM ou directement sur un flux en continu. Des expériences sur un flux en continu montre la robustesse d'une telle organisation hiérarchique des MMG à un taux de plus de 88.5% moyennant une recherche non exhaustive avec une complexité logarithmique au lieu de linéaire, voire même exponentielle.

Comme toute structure de données adaptée au problème incrémental, notre système d'indexation permet d'effectuer la mise à jour des modèles MMG des locuteurs à l'aide de l'algorithme fusion des MMG. Cet algorithme a été conçu à la fois pour créer une structure ascendante en regroupant deux à deux les modèles GMM jugés similaires.

Enfin, Au travers de deux expérimentations utilisant des structures arborescentes binaires ou n'aires, une réflexion est conduite afin de trouver une structure ordonnée et adaptée au problème incrémental. Quelques pistes de réflexions sur l'apport de l'analyse vidéo sont discutées et les besoins futurs sont explorés.

Mots-clefs (9): Reconnaissance automatique de locuteurs, bases de données multimédias, structuration audiovisuelle, classification hiérarchique, modèle de mélange de gaussiennes, divergence de Kullback-Leibler, architecture arborescente, structure incrémentale, Archivage audio.
