Résumé

Le domaine de la reconnaissance automatique du locuteur regroupe les applications pour lesquelles on désire identifier une personne à partir de sa voix. Le champ d'application couvre de nombreux secteurs tels que l'accès sécurisé, les transactions téléphoniques, la surveillance, l'indexation audio ou encore l'expertise judiciaire.

Dans ce travail, nous nous intéressons à une méthode de modélisation statique basée sur les mélanges gaussiens GMM (Gaussian Mixture Model), qui est devenue, dans les dix dernières années, l'approche la plus performante et la plus répandue pour les systèmes de reconnaissance du locuteur indépendante du texte. Chaque locuteur est représenté par un modèle GMM. Ce dernier est une somme pondérée de M distributions gaussiennes multidimensionnelles.

Dans notre travail, la base de données VoxForge est utilisée pour l'extraction des paramètres pertinents caractérisant les locuteurs. La technique utilisée est celle de l'analyse cepstrale, dont les résultats sont les coefficients MFCC couramment utilisés en reconnaissance du locuteur.

Durant la phase d'apprentissage, la phase de construction des modèles des locuteurs, les paramètres du mélange de gaussiennes sont estimés par l'algorithme EM/ML. Deux phases d'apprentissage sont mises en œuvre pour pallier le manque de données d'apprentissage disponibles pour chaque locuteur:

1. Apprentissage d'un modèle générique (aussi appelé « modèle du monde ») estimé par l'algorithme EM/ML sur une grande quantité de données (population de locuteurs);
2. apprentissage du modèle locuteur dérivé du modèle du monde par application d'une technique d'adaptation (par exemple par Maximum a Posteriori).

Nous avons considéré deux méthodes pour la phase de décision; la première est celle du rapport de la maximum de la log-vraisemblance (LLR : Logarithm of the Likelihood Ratio), la deuxième approche considérée est basée sur les SVM. Les performances des deux systèmes sont comparées.

Notre système de reconnaissance du locuteur réalise deux tâches qui sont l'identification du locuteur et la vérification du locuteur. Le taux d'égale erreur est un indice de performance offert par le système de vérification, tandis que le taux d'identification correcte est un indice offert par le système d'identification. Les résultats obtenus en termes de taux d'identification correcte sont de 91.1%, et le taux d'égale erreur EER est de 7.4%.

Mots-clés: Identification du locuteur, vérification du locuteur, l'approche GMM-UBM. SVM.