

Fouille de données audio pour la classification automatique de mots homophones

Rena NEMOTO*, Martine Adda-Decker*
Ioana Vasilescu*

*LIMSI-CNRS B.P. 133 91403 Orsay Cedex France
{nemoto, madda, ioana}@limsi.fr
<http://www.limsi.fr>

Résumé. Cet article présente une contribution à la modélisation acoustique des mots à partir de grands corpus oraux, faisant appel aux techniques de fouilles de données. En transcription automatique, de nombreuses erreurs concernent des mots fréquents homophones. Deux paires de mots (quasi-)homophones *à/a* et *et/est* sont sélectionnées dans les corpus, pour lesquels sont définis et examinés 41 descripteurs acoustiques permettant potentiellement de les distinguer. 17 algorithmes de classification, mis à l'épreuve pour la discrimination automatique de ces deux paires de mots, donnent en moyenne 77% de classification correcte sur les 5 meilleurs algorithmes. En réduisant le nombre de descripteurs à 10 (sélectionnés par l'algorithme le plus performant), les résultats de classification restent proches du résultat obtenu avec 41 attributs. Cette comparaison met en évidence le caractère discriminant de certains attributs, qui pourront venir enrichir à la fois la modélisation acoustique et nos connaissances des prononciations de l'oral.

1 Introduction

En transcription automatique de la parole, de grands corpus audio (incluant généralement des centaines d'heures de parole) servent à estimer des modèles acoustiques précis de phonèmes contextuels. Ces modèles de sons élémentaires sont ensuite concaténés pour aboutir à des modèles de mots en s'appuyant sur la connaissance de leur prononciation. Cette connaissance est incomplète à l'heure actuelle et une partie importante de l'information caractérisant les variantes de prononciations se trouve encodée implicitement dans les modèles acoustiques. L'objectif de ce travail est de s'appuyer sur les techniques de fouille de données afin d'extraire des connaissances relatives aux spécificités acoustiques et prosodiques caractérisant les prononciations. Cette approche a déjà pu montrer son intérêt pour la caractérisation des accents étrangers (Vieru-Dimulescu et al., 2007). Nous nous intéresserons ici aux mots considérés comme homophones, i.e. phonémiquement pareils, et qui sont de ce fait sujets à de nombreuses erreurs de confusion lors de la transcription automatique. Partant de ces constats, nous nous sommes interrogés si les mots homophones ne déploieraient pas de particularités acoustiques/prosodiques qui n'ont été prises en compte ni par les paramètres acoustiques classiques (vecteurs de cepstres), ni par les modèles acoustiques (Modèles de Markov Cachés à trois états) et qui permettrait leur discrimination. Nous faisons ainsi l'hypothèse que des informations prosodiques (concernant durée, fréquence fondamentale notée f_0 , cooccurrence avec des pauses, etc.) puissent contribuer à lever certains types d'homophonie, en particulier s'il s'agit d'homophones issus de classes syntaxiques différentes (hétéro-syntaxiques). Nous avons fait appel aux techniques de fouille de données afin de classer automatiquement ces