

# Analyse par bandes de fréquences pour la détection des zones de «déviance» dans la parole dysarthrique

G. Pouchoulin<sup>1</sup>, C. Fredouille<sup>1</sup>, C. Fougeron<sup>3</sup>, A. Ghio<sup>2</sup>, N. Levêque<sup>3</sup>, C. Meunier<sup>2</sup>, O. Panseri<sup>3</sup>

<sup>1</sup>Université d'Avignon, CERI/LIA, Avignon (France)

<sup>2</sup>Laboratoire Parole et Langage, CNRS, Université Aix-Marseille, Aix en Provence (France)

<sup>3</sup>Laboratoire de Phonétique et Phonologie, UMR 7018 CNRS-Paris3/Sorbonne Nouvelle, Paris (France)

*alain.ghio@lpl-aix.fr; corinne.fredouille@univ-avignon.fr; cecile.fougeron@univ-paris3.fr*

Ce travail s'inscrit dans le projet ANR DesPho-APaDy dont l'un des objectifs vise à décrire les caractéristiques phonéto-acoustiques de la parole dysarthrique par le biais d'une approche combinant des procédures analytiques manuelles et des procédures issues du traitement automatique de la parole. Caractérisant un trouble de l'exécution motrice de la parole d'origine neurologique, la dysarthrie peut s'avérer comme un symptôme particulièrement important à évaluer dans le cadre du suivi de l'évolution d'une maladie neurologique ou d'essais thérapeutiques. En ce sens, ce papier s'intéresse à des personnes atteintes de maladie lysosomale *i.e.* pathologie neurodégénérative en rapport avec le dysfonctionnement d'une des enzymes contenues dans le lysosome. Liée à diverses atteintes dans le système nerveux, la dysarthrie associée peut être qualifiée de « mixte » dans la classification physiopathologique proposée par Darley *et al.* [1]. Dans cette étude, les auteurs décrivent une méthodologie originale utilisant des outils issus du traitement automatique de la parole et développés par le Laboratoire Informatique d'Avignon (LIA), pour détecter les portions « déviantes » dans la parole produite par des sujets atteints de cette forme de dysarthrie.

L'alignement automatique contraint par le texte consiste, à partir d'un signal de parole et de sa transcription en mots, à détecter les frontières dans le signal de parole de chacun des phonèmes présents dans le texte prononcé. Cette tâche repose sur un décodage du signal de parole par l'algorithme Viterbi [2] utilisant des modèles statistiques (37 modèles Hidden Markov Model indépendants du contexte appris sur le corpus d'émissions radiographiques ESTER associés à chacun des phonèmes recherchés [3]. Il en résulte une segmentation temporelle en phonèmes comprenant, pour chaque phonème, ses frontières de début et de fin au sein du signal de parole ainsi qu'un score de vraisemblance entre le segment de parole associé et le modèle statistique correspondant. Ces scores de vraisemblance sont normalisés avec les vraisemblances résultantes du décodage acoustico-phonétique (DAP) non-contraint afin de différencier les perturbations dues à la qualité du signal, d'une mauvaise prononciation du phonème. Ces derniers sont considérés comme des scores de confiance associés à chaque segment phonémique. En comparant les scores obtenus sur une population de locuteurs sains avec ceux obtenus sur une population de patients atteints de trouble de la parole, il est alors possible de mettre en place un système de détection de zones de déviance de la parole dysarthrique [4]. Le corpus utilisé est composé de 8 patients (4 femmes et 4 hommes) évalués à différents degrés de sévérité de dysarthrie suivant le stade d'évolution de leur maladie. Les productions de parole des locuteurs dysarthriques (texte « Tic Tac » de la batterie C. Chevrie-Müller) ont été transcrites orthographiquement selon une convention rigoureuse et les signaux de parole associés ont été alignés par le système automatique. Parallèlement, un expert humain a annoté chacun des phonèmes du corpus (par analyse et écoute du signal) en précisant pour chaque phonème la présence d'une déviance ou non ainsi que le type de déviance le cas échéant. Cette annotation est utilisée pour mesurer la qualité du système de détection automatique des zones de déviance.

La procédure de détection automatique de zones de déviance a été répétée 8 fois avec des modèles HMM de phonèmes appris selon des bandes de fréquences d'une largeur de 1kHz (de 0 à 8kHz). Les scores acoustiques issus du système automatique juxtaposés à l'alignement manuel peuvent ainsi être analysés et comparés afin de déterminer si des bandes fréquentielles sont plus pertinentes que d'autres pour la mise en évidence d'une variabilité en fonction de la sévérité de la dysarthrie.

- [1] F. L. Darley, A. E. Aronson, J. R. Brown, 1969. Clusters of deviant speech dimensions in the dysarthrias. *Journal of Speech and Hearing Research* 12, 462-496.
- [2] L. Rabiner, B. H. Juang, 1993. *Fundamentals of speech recognition*. Prentice-Hall.
- [3] F. Brugnara, D. Falavigna, M. Omologo, 1993. Automatic segmentation and labeling of speech based on hidden markov models. *Speech Communication*, vol. 12, p. 357-370.
- [4] P. Clément, C. Fredouille, N. Lévêque, 2009. Méthodes objectives appliquées à la dysarthrie. *Troisièmes Journées de Phonétique Clinique*, 4-5 décembre, Aix-en-Provence.