

Confusion phonémique dans la reconnaissance automatique de la parole dysarthrique

Maxime Manos¹ Imed Laaridh^{1,2} Corinne Fredouille¹

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

(2) Université d'Aix Marseille, CNRS, LPL UMR 7309, 13100, Aix-en-Provence, France

{maxime.manos, imed.laaridh}@alumni.univ-avignon.fr,

corinne.fredouille@univ-avignon.fr

MOTS-CLÉS : Dysarthrie, troubles de parole, alignement automatique, confusion phonémique.

KEYWORDS: Dysarthria, speech disorders, automatic phone alignment, phonemic confusion.

1 Introduction

Malgré le niveau avancé atteint par les outils de traitement automatique de la parole, en particulier dans les applications destinées au grand public, ces derniers éprouvent de grandes difficultés face à des cas de parole non standard, comme la parole produite par des enfants, des locuteur du L2 ou des locuteurs atteints de trouble de la parole. Nous nous intéressons dans ce travail à la parole dysarthrique.

La dysarthrie est un trouble de la réalisation motrice de la parole résultant d'une lésion du système nerveux central ou périphérique. Elle peut être associée à plusieurs pathologies et affecter différents niveaux de production de la parole.

Le travail présenté ici s'inscrit dans le cadre de l'étude de l'apport des outils de traitement automatique face à la parole pathologique (Laaridh *et al.*, 2015, 2016) et a pour objectif l'observation et l'analyse du comportement d'un système d'alignement automatique de la parole face à la variabilité observée dans la parole dysarthrique. L'alignement automatique de la parole consiste à retrouver, à partir d'un signal de parole et de sa transcription en mots, les frontières de début et de fin de chaque phonème prononcé. Pour chaque segment y_p résultant de cet alignement, un calcul de vraisemblance représentant la ressemblance entre le segment et le modèle de chaque phonème est réalisé. Le tri de ces scores de vraisemblance permet de détecter et d'étudier la distance et l'altération d'un phonème "atypique" p par rapport à son modèle générique et son possible rapprochement du modèle d'un autre phonème p' . Cette tâche est basée sur un décodage Viterbi (Brugnara *et al.*, 1993) utilisant des modèles acoustiques (Modèles de Markov Cachée - HMM) représentant les 37 phonèmes du Français appris sur des enregistrements radiophoniques (Galliano *et al.*, 2005).

Le corpus de données utilisé dans cette étude est le corpus TYPALOC (Meunier *et al.*, 2016) composé de 40 enregistrements réalisés par 28 patients dysarthriques et 12 contrôles sains. Tous les locuteurs ont lu le même texte : "tic tac" de la batterie C. Chevie-Müller. Les patients sont répartis sur 3 pathologies différentes : la maladie de Parkinson (Park.), la Sclérose Latérale Amyotrophique (SLA) et une ataxie cérébelleuse (Céréb.). Ils présentent différents grades de sévérité de dysarthrie. Tous ces enregistrements ont été segmentés manuellement (moyennant l'outil Praat) afin de délimiter les

frontières de début et de fin de chaque phonème. Cette segmentation manuelle nous a permis d'évaluer la qualité de notre alignement automatique et de limiter l'étude aux phonèmes bien alignés (décalages entre les frontières de début et fin manuelles et automatiques inférieures à 20 ms)

2 Discussion

Considérant les différentes populations, on observe que les contrôles présentent le plus grand nombre de phonèmes bien reconnus (sans confusion) comparés aux patients (81% pour les contrôles contre 67% pour les Park. et Céréb. et 58% pour les SLA). Cette variabilité inter-pathologique est vraisemblablement due à la différence de grade de sévérité de la dysarthrie plus important chez les patients atteints de SLA par rapport aux autres pathologies.

De plus, les taux de reconnaissance (TR) de phonèmes présentent une large variabilité inter-phonème sur toutes les populations. Comparons occlusives et fricatives, on trouve que contrairement aux contrôles et SLA pour lesquelles les deux catégories présentent des TR globaux comparables, l'outil d'alignement réalise beaucoup plus de confusions sur les fricatives que les occlusives pour les Park. et les Céréb. ($TR_{occlusives}$; $TR_{fricatives}$) de (70% ;45%) et (65% ;38%) respectivement). Il est intéressant de constater qu'une large majorité de confusions observées sur les fricatives pour les contrôles et même les SLA sont avec d'autres fricatives contrairement aux Park. pour lesquelles il y a plus de confusions de fricatives avec des occlusives qu'avec d'autres fricatives.

Observant les voyelles, on observe que pour les contrôles, Park. et Céréb., les voyelles orales sont généralement bien reconnues (TR de 81%, 71% et 76% respectivement) et leurs taux de confusion avec les voyelles nasales ne dépassent pas 6% des phonèmes. Par contre, sur les SLA, les voyelles orales montrent un TR de 61% seulement avec 14% des phonèmes confondus avec des voyelles nasales. Cela peut s'expliquer par le phénomène d'hypernasalisation caractéristique de la SLA (dysarthrie mixte) (Auzou *et al.*, 2007). Cette tendance est confirmée en observant les confusions faites par le système sur les occlusives sonores pour les différentes populations. En effet, 19% des phonèmes de cette classe sont confondus avec des consonnes nasales pour les SLA alors que ce taux est nul pour les contrôles et ne représente que 4% et 1% pour les Park. et les Céréb. respectivement.

Un futur travail consistera à élaborer une mesure de robustesse/vulnérabilité des phonèmes sur la base des observations des confusions issues de l'alignement.

Références

- AUZOU P., ROLLAND-MONNOURY V., PINTO S. & ÖZSANCAK C. (2007). *Les dysarthries*. Groupe de Boeck.
- BRUGNARA F., FALAVIGNA D. & OMOLOGO M. (1993). Automatic segmentation and labeling of speech based on hidden markov models. *Speech Communication*, **12**(4), 357–370.
- GALLIANO S., GEOFFROIS E., MOSTEFA D., CHOUKRI K., BONASTRE J.-F. & GRAVIER G. (2005). ESTER phase II evaluation campaign for the rich transcription of French broadcast news. In *Proceedings of Interspeech'05*, p. 1149–1152.
- LAARIDH I., FREDOUILLE C. & MEUNIER C. (2015). Automatic detection of phone-based anomalies in dysarthric speech. *ACM Transactions on accessible computing*, **6**(3), 9 :1–9 :24.

LAARIDH I., FREDOUILLE C. & MEUNIER C. (2016). Automatic anomaly detection for dysarthria across two speech styles : read vs spontaneous speech. In *Proceedings of the International Conference on Language Resources and Evaluation (LREC'16)*, Portoroz.

MEUNIER C., FOUGERON C., FREDOUILLE C., BIGI B., CREVIER L., DELAIS-ROUSSARIE E., GEORGETON L., GHIO A., LAARIDH I., LEGOU T., PILLOT-LOISEAU C. & POUCHOULIN G. (2016). The tupaloc corpus : A collection of various dysarthric speech in read and spontaneous speech. In *Proceedings of the International Conference on Language Resources and Evaluation (LREC'16)*, Portoroz.