



Open Archive Toulouse Archive Ouverte

OATAO is an open access repository that collects the work of Toulouse researchers and makes it freely available over the web where possible

This is an author's version published in:
<http://oatao.univ-toulouse.fr/24950>

To cite this version: Farinas, Jérôme and Pellegrini, Thomas and Pinquier, Julien *Comparaison de systèmes automatiques de reconnaissance grand vocabulaire appliqué à de la parole pathologique*. (2019) In: 8e Journées de Phonetique Clinique (JPC 2019), 14 May 2019 - 16 May 2019 (Mons, Belgium).

Any correspondence concerning this service should be sent to the repository administrator: tech-oatao@listes-diff.inp-toulouse.fr

Comparaison de systèmes automatiques de reconnaissance grand vocabulaire appliqué à de la parole pathologique

Jérôme FARINAS¹
Thomas PELLEGRINI¹
Julien PINQUIER¹

¹Institut de Recherche en Informatique de Toulouse,
UMR 5505 CNRS,
Équipe SAMOVA,
Université Toulouse

Les performances actuelles des systèmes automatiques de reconnaissance de la parole grand vocabulaire permettent d'envisager des applications dans le domaine de la santé. Cela permettrait d'envisager des automatisations de divers tests (par exemple la fluence verbale) mais également d'apporter des informations objectives d'assez haut niveau issues de la voix (par exemple des mesures d'intelligibilité). Mais comment se comportent ces systèmes automatiques de reconnaissance de la parole sur des voix pathologiques ? Une solution entièrement automatique est-elle envisageable ?

Dans le cadre d'une étude financée par la Société d'Accélération et de Transfert Technologique Toulouse Tech Transfert, une évaluation de systèmes de transcription académiques et industriels a été menée sur un corpus de parole de 385 minutes. Les données sont issues d'enregistrements produits dans différentes conditions : différents styles de parole, environnements bruités, locuteurs avec accents régionaux, personnes atteintes de cancers des voix aériennes supérieures présentant différents degrés de sévérité (extraits du corpus PARALOTHEQUE/C2SI (Astesano 2018)) et également des enregistrements de parole simulant différents degrés de presbyacousie (projet ARCHEAN/Projet AGILE IT (Fontan 2017)). Dix systèmes ont ainsi été évalués : Authôt (société française), Bing de Microsoft, Google, IBM ViaVoice, Nuance, Speechmatics, Sphinx, Wit ainsi que les laboratoires de recherche LIA et IRIT. Aucune adaptation particulière n'a été effectuée sur ce type de données. En effet, une phase d'adaptation permettrait de mettre en meilleure adéquation les modèles (acoustiques et de langage) utilisés par les systèmes de reconnaissance et les enregistrements qui leur sont soumis. Les performances ne sont bien évidemment pas aussi bonnes que celles obtenues sur des enregistrements

de parole en conditions normales : environ 94 % de bonnes reconnaissances sur un corpus de 12.500 h d'entraînement (Chiu, 2018). Par exemple, le meilleur système atteint seulement 38 % de taux de reconnaissance de mots sur des voix cancer. Les résultats présentés ici sont donc « bruts » mais permettent d'avoir une vue sur les performances que nous pouvons obtenir directement en utilisant ces services/systèmes. Cela permet également de mesurer l'effort à fournir pour collecter et annoter des données en quantité suffisante pour adapter et rendre pleinement utilisables de tels systèmes afin de traiter des données de voix pathologique.

Références bibliographiques

- ASTESANO, C. *et al.*, Carcinologic Speech Severity Index Project: A Database of Speech Disorders Productions to Assess Quality of Life Related to Speech After Cancer (regular paper), *Language Resources and Evaluation Conference (LREC 2018)*, European Language Resources Association (ELRA), , 7-12 mai 2018, Miyazak, Japon, 4265-4271.
- CHIU, Chung-Cheng, SAINATH, T. N., WU, Y., *et al.*, State-of-the-art speech recognition with sequence-to-sequence models, in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2018, 4774-4778.
- FONTAN, L. *et al.*, Automatic speech recognition predicts speech intelligibility and comprehension for listeners with simulated age-related hearing loss, *Journal of Speech, Language and Hearing Research*, 2017, Vol. **60**, American Speech-Language-Hearing Association (ASLHA), USA, septembre 2017, 2394-2405.