An Automatic Speech Recognition Android App for ALS patients

This paper describes *AllSpeak*, an Automatic Speech Recognition (ASR) Android Application developed for patients with *Amyotrophic Lateral Sclerosis* (ALS). It allows patients to record a number of basic sentences that are used in everyday life and, at a later stage, with the progression of their disease, to associate patient's degraded spoken sentences to a predefined internal vocabulary customized by the user. In this way, ALS patients will still be able to communicate their basic needs to their families or carers.

AllSpeak is based on Deep Neural Networks (DNNs). Although DNN-based ASR has achieved impressive results in the last 6 years [1], its accuracy significantly decreases in training-testing mismatched conditions [2]. That represents a challenge for the recognition of speakers whose acoustic characteristics are not sufficiently covered in the training dataset. Unfortunately, ALS patients with strong speech impairments and, more generally, speakers with dysarthria are poorly represented in both proprietary and publicly available datasets – with probably the most well-known, iconic dataset being the TORGO dataset [3].

Analysis of the tiny previous work on DNN-based ASR for dysarthric speech (e.g. [4], [5]) shows that even medium vocabulary robust ASR is not a viable solution, yet. As a consequence, *AllSpeak* aims at a very robust recognition of a limited number of spoken commands that are crucial for the patients to communicate their primary needs and feelings (e.g. "I am hungry", "I am thirsty", "I am cold", and so on). To increment robustness in speech recognition, we have explored several DNN adaptation strategies evaluating the resulting ASR performance. Specifically, we were interested to find out which adaptation strategy was more suitable for handling the dysarthric speech. As a result, we focus on adaptation of a four hidden-layers DNN model to speech control data performed by normal speakers captured by a Motorola Moto G4 Play microphone in a quiet environment. We show that significant performance gains can be obtained by employing the transfer-learning technique to compensate the mismatch between a clean speech-trained model and a small set of impaired speakers' data.

Another characteristic of *AllSpeak* consists in continuous "listening" and recognition that does not require any internet connection. That is achieved through a simple, computationally efficient, voice activity detection (VAD) system and trivial featherweight decoding strategy (similar to [6]).

AllSpeak is a hybrid App composed by a front-end developed with the *Ionic Framework* (ver. 1.X) and a multi-threaded Android Service that implements all the speech processing part. DNN were implemented with *TensorFlow* engine (version 1.1). *TensorFlow* is an open source software library released in 2015 by *Google Inc.* to make it easier for developers to design, build, and train deep learning models [7]. We chose to use it here because of its thoughtful design and ease of use, and because it can be successfully embedded in Android.

Preliminary results with 23 different recorded commands pronounced by 14 ALS patients and by 12 healthy controls show an accuracy rate of 80.5 ± 16 % and 98.3 ± 5 % respectively. They are primarily related to patient's vocabulary size and modality of speech (intelligible or degraded speech) depending on the extent of the disease in each patient at the time of this study.

Besides facilitating the manipulation for current users of mobile devices, the market is opened to a whole new range of people who might use these technologies, as has been shown by our study. By using our *AllSpeak* application, people with speech disorders would have the opportunity to participate in the technology present and feel the benefits of smartphones which are powerful devices able to mitigate their disabilities.

References

[1] Hinton, G., Deng, L., Yu, D., Dahl, G. E., Mohamed, A. R., Jaitly, N., ... & Kingsbury, B. (2012). Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *IEEE Signal Processing Magazine*, *29*(6), 82-97.

[2] Seltzer, M. L., Yu, D., & Wang, Y. (2013, May). An investigation of deep neural networks for noise robust speech recognition. In *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on* (pp. 7398-7402). IEEE.

[3] Rudzicz, F., Namasivayam, A. K., & Wolff, T. (2012). The TORGO database of acoustic and articulatory speech from speakers with dysarthria. *Language Resources and Evaluation*, *46*(4), 523-541.

[4] Espana-Bonet, C., & Fonollosa, J. A. (2016). Automatic speech recognition with deep neural networks for impaired speech. In *Advances in Speech and Language Technologies for Iberian Languages: Third International Conference, IberSPEECH 2016, Lisbon, Portugal, November 23-25, 2016, Proceedings 3* (pp. 97-107). Springer International Publishing.

[5] Joy, N. M., Umesh, S., & Abraham, B. (2017). On Improving Acoustic Models For TORGO Dysarthric Speech Database. *Proc. Interspeech 2017*, 2695-2699.

[6] Chen, G., Parada, C., & Heigold, G. (2014, May). Small-footprint keyword spotting using deep neural networks. In *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on* (pp. 4087-4091). IEEE.

[7] Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., ... & Kudlur, M. (2016, November). TensorFlow: A System for Large-Scale Machine Learning. In *OSDI* (Vol. 16, pp. 265-283).