 

Master Thesis Proposal

Title of the project:

Representation learning strategies to model pathological speech: effect of multiple spectral resolutions.

Supervisors from LME:

Prof. Dr. Ing. Elmar Nöth [elmar.noeth@fau.de](mailto:elmar.noeth@fau.de)

Msc. Juan Camilo Vasquez-Correa [juan.vasquez@fau.de](mailto:juan.vasquez@fau.de)

Description:

Speech signals contain paralinguistic information with speciﬁc cues about a given speaker including the presence of diseases that may alter their communication capabilities. The automatic classiﬁcation of paralinguistic aspects has many potential applications, and has received a good deal of attention by the research community [1-3]. In particular researchers look at clinical observations in the speech of patients and try to objectively and automatically measure two main aspects of a given disease: (1) the presence of a disease via classification of healthy control (HC) subjects and patients, and (2) the level of degradation of the speech of patients according to a specific clinical scale [4]. These aspects are evaluated using computer aided methods supported in signal processing and pattern recognition methods.

At the center of these computer aided methods and something that has been developed over the years to continually improve the diagnosis and the assessment of severity of different pathological diseases is the particular feature set and extraction method used [5-7]. Many recent studies focused on extracting features for assessment of pathological speech rely on deep learning strategies [3].

In this project we consider one such approach that uses a parallel representation learning strategy to model speech signals from patients with different speech disorders [8]. The model uses two types of autoencoders, a convolutional autoencoder (CAE) and recurrent autoencoder (RAE). Both take as input a spectrogram and output features derived from a hidden representation in the bottleneck space (i.e. a compressed representation of the input). In addition, the reconstruction error of the autoencoder in different spectral components of the speech signal is considered as a feature set.

The aim of this project is to evaluate the performance of the parallel representation learning strategy using different parametrized representations of the spectrogram (e.g. comparing broadband and narrowband spectral representations) as well as a wavelet representation to quantify the information loss for each representation, and the benefit of using all of them together as a multiple input channel. Methods for quantification include the overall ability of the proposed model to classify different pathologies and the associated level of degradation of a given patient’s speech, and also comparing the input and reconstructed speech signals using contours of phonological posteriors [9]. The aim is to evaluate which group of phonemes are more affected due to the compression of the autoencoders using the different spectral resolutions and their combinations.

References:

[1] Schuller, B., Batliner, A., 2013. Computational Paralinguistics: Emotion, Aﬀect and Personality in Speech and Language Processing. John Wiley & Sons.

[2] Schuller, B., et al., 2019. Aﬀective and Behavioural Computing: Lessons Learnt from the First Computational Paralinguistics Challenge. Computer Speech & Language 53, 156–180.

[3] Cummins, N., Baird, A., Schuller, B., 2018. Speech Analysis for Health: Current State-of-the-Art and the Increasing Impact of Deep Learning Methods.

[4] Orozco-Arroyave, J.R., et al., 2015. Characterization Methods for the Detection of Multiple Voice Disorders: Neurological, Functional, and Laryngeal Diseases. IEEE Journal of Biomedical and Health Informatics 19, 1820–1828.

[5] Dimauro, G., Di-Nicola, V., et al., 2017. Assessment of Speech Intelligibility in Parkinson’s Disease Using a Speech-to-Text System. IEEE Access 5, 22199–22208.

[6] Orozco-Arroyave, J.R., Vasquez-Correa, J.C., et al., 2016. Towards an Automatic Monitoring of the Neurological State of the Parkinson’s Patients from Speech, in: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 6490–6494.

[7] Schuller, B., Steidl, S., Batliner, A., Hantke, S., Hönig, F., Orozco-Arroyave, J.R., Nöth, E., Zhang, Y., Weninger, F., 2015. The INTERSPEECH 2015 Computational Paralinguistics Challenge: Nativeness, Parkinson’s & Eating Condition, in: Proceedings of INTERSPEECH, pp. 478–482.

[8] Vásquez-Correa, Juan Camilo et al. “Parallel Representation Learning for the Classification of Pathological Speech: Studies on Parkinson’s Disease and Cleft Lip and Palate” *Under Review* (2020).

[9] Vásquez-Correa, Juan Camilo et al. “Phonet: A Tool Based on Gated Recurrent Neural Networks to Extract Phonological Posteriors from Speech.”, in: INTERSPEECH (2019).