Accepted Manuscript

Single-ended prediction of listening effort using deep neural networks

Rainer Huber, Melanie Krüger, Bernd T. Meyer

PII: S0378-5955(17)30495-1

DOI: 10.1016/j.heares.2017.12.014

Reference: HEARES 7472

To appear in: *Hearing Research*

Received Date: 11 October 2017

Revised Date: 3 December 2017

Accepted Date: 23 December 2017

Please cite this article as: Huber, R., Krüger, M., Meyer, B.T., Single-ended prediction of listening effort using deep neural networks, *Hearing Research* (2018), doi: 10.1016/j.heares.2017.12.014.

Single-ended prediction of listening effort using deep neural networks

Rainer Huber[1], Melanie Krüger[2], Bernd T. Meyer[1]

[1] Medizinische Physik and Cluster of Excellence Hearing4all, Carl-von-Ossietzky Universität Oldenburg, Carl-von-Ossietzky-Str. 9-11, 26129 Oldenburg, Germany

[2] Hörzentrum Oldenburg, Marie-Curie-Str. 2, 26129 Oldenburg, Germany

Rainer.Huber@uni-oldenburg.de, Melanie.Krueger@hoerzentrum-oldenburg.de, Bernd.Meyer@uni-oldenburg.de

Corresponding author: Rainer Huber

Abstract

The effort required to listen to and understand noisy speech is an important factor in the evaluation of noise reduction schemes. This paper introduces a model for Listening Effort prediction from Acoustic Parameters (LEAP). The model is based on methods from automatic speech recognition, specifically on performance measures that quantify the degradation of phoneme posteriorgrams produced by a deep neural net: Noise or artefacts introduced by speech enhancement often result in a temporal smearing of phoneme representations, which is measured by comparison of phoneme vectors. This procedure does not require a priori knowledge about the processed speech, and is therefore single-ended. The proposed model was evaluated using three datasets of noisy speech signals with listening effort ratings obtained from normal hearing and hearing impaired subjects. The prediction quality was compared to several baseline models such as the ITU-T standard P.563 for single-ended speech quality assessment, the American National Standard ANIQUE+ for single-ended speech quality assessment, and a single-ended SNR estimator. In all three datasets, the proposed new model achieved clearly better prediction accuracies than the baseline models; correlations with subjective ratings were above 0.9. So far, the model is trained on the specific noise types used in the evaluation. Future work will be concerned with overcoming this limitation by training the model on a variety of different noise types in a multi-condition way in order to make it generalize to unknown noise types.