

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](https://doi.org/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.

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Single-ended prediction of listening effort using deep neural networks

Rainer Huber¹, Melanie Krüger², Bernd T. Meyer¹

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

1 Abstract

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

Download English Version:

<https://daneshyari.com/en/article/8842415>

Download Persian Version:

<https://daneshyari.com/article/8842415>

[Daneshyari.com](https://daneshyari.com)