

Speech Intelligibility Prediction for Hearing-Impaired Listeners with Phoneme Classifiers based on Deep Learning

Jana Roßbach

Rainer Huber, Saskia Röttges, Christopher F. Hauth, Thomas Biberger,
Thomas Brand, Bernd T. Meyer, Jan RENNIES

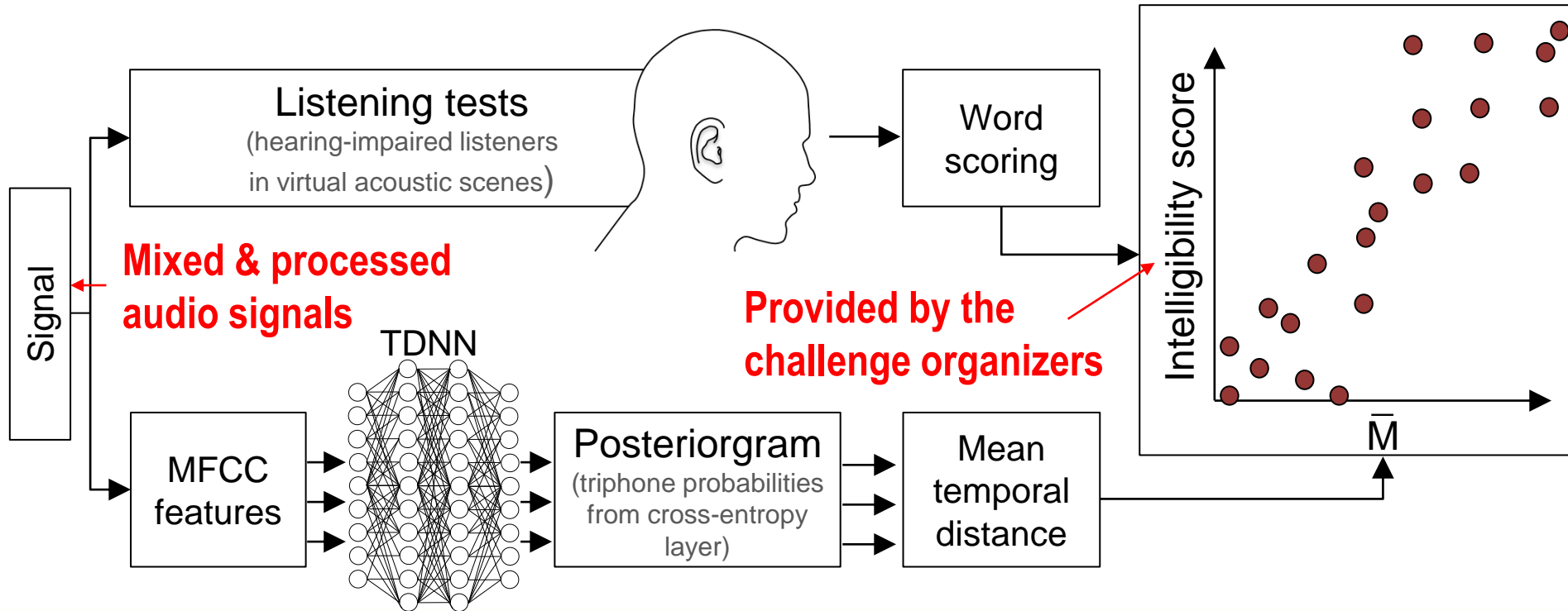
6/29/2022

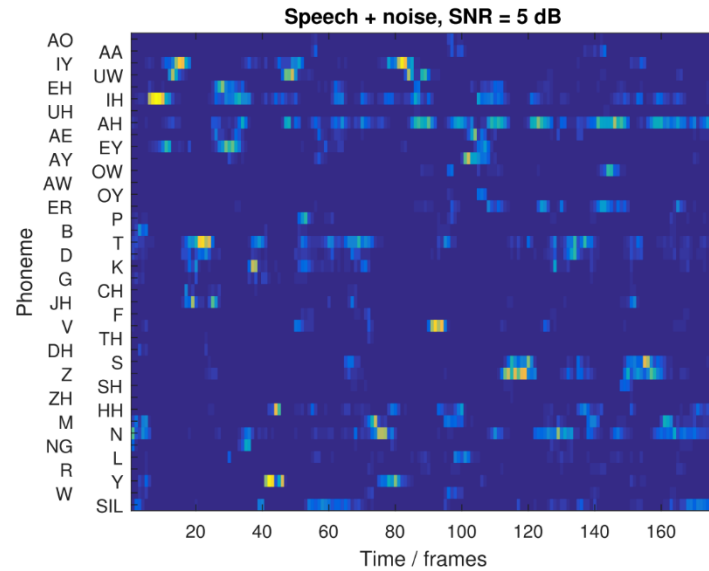
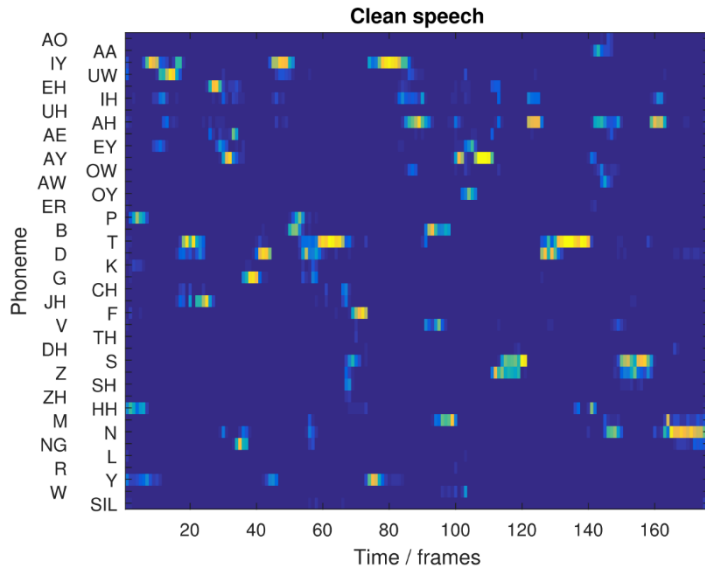
- Why are speech intelligibility (SI) models important?
 - are significantly less time and cost intensive than SI measurements
 - can give us a better understanding of the auditory system

- What can SI models be used for?
 - Optimization of speech enhancement algorithms
 - SI monitoring
 - prediction of the benefit of hearing aids

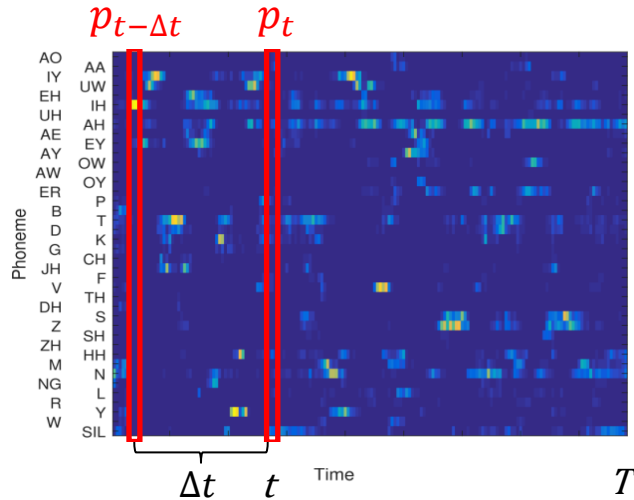
- An existing model was used because accurate predictions would indicate that the model can be used in other acoustic situations as well
- SI and listening effort (LE) are closely related
 - We used an LE model with a mapping from model output to intelligibility scores
- It is a monaural model, but the signals does not contain many binaural cues
 - we used better ear listening

Listening Effort prediction from Acoustic Parameters (LEAP) (Huber et al., 2018)



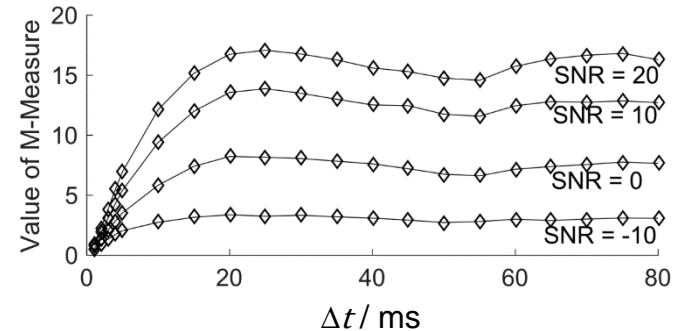


Mean temporal distance (Hermansky et al., 2013)



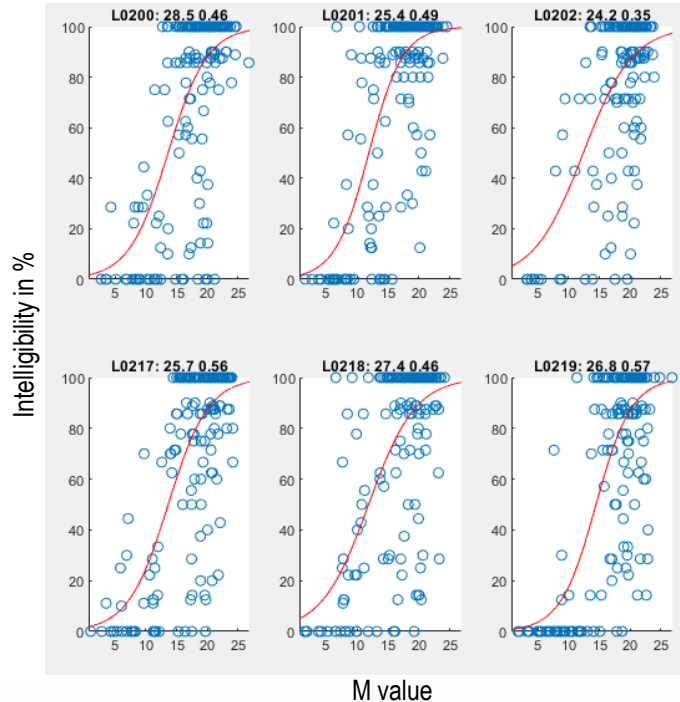
- Much noise $\rightarrow p_{t-\Delta t} \approx p_t \rightarrow$ small M
- No noise $\rightarrow p_{t-\Delta t} \neq p_t \rightarrow$ large M

$$M(\Delta t) = \frac{1}{T - \Delta t} \sum_{t=\Delta t}^T D(p_{t-\Delta t}, p_t),$$

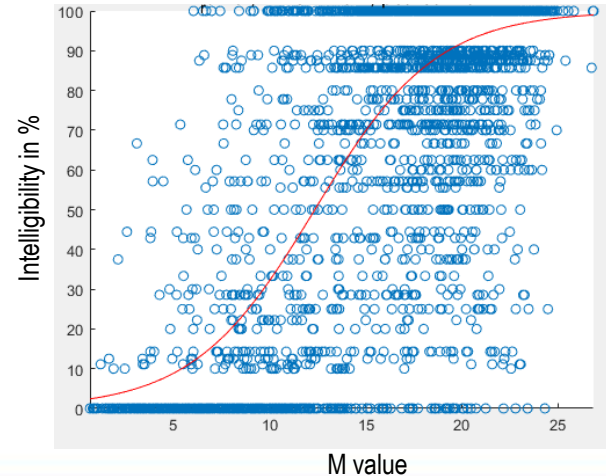


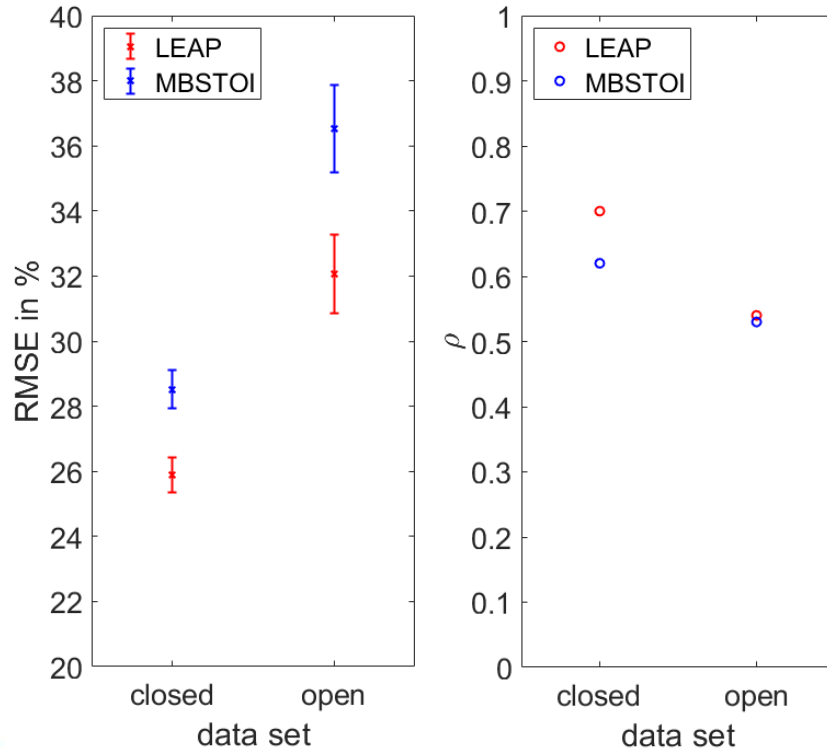
Final predictor: $\bar{M} := \frac{1}{10} \sum_{n=1}^{10} M(300\text{ms} + n \cdot 50\text{ms})$

Closed data set: individual mapping

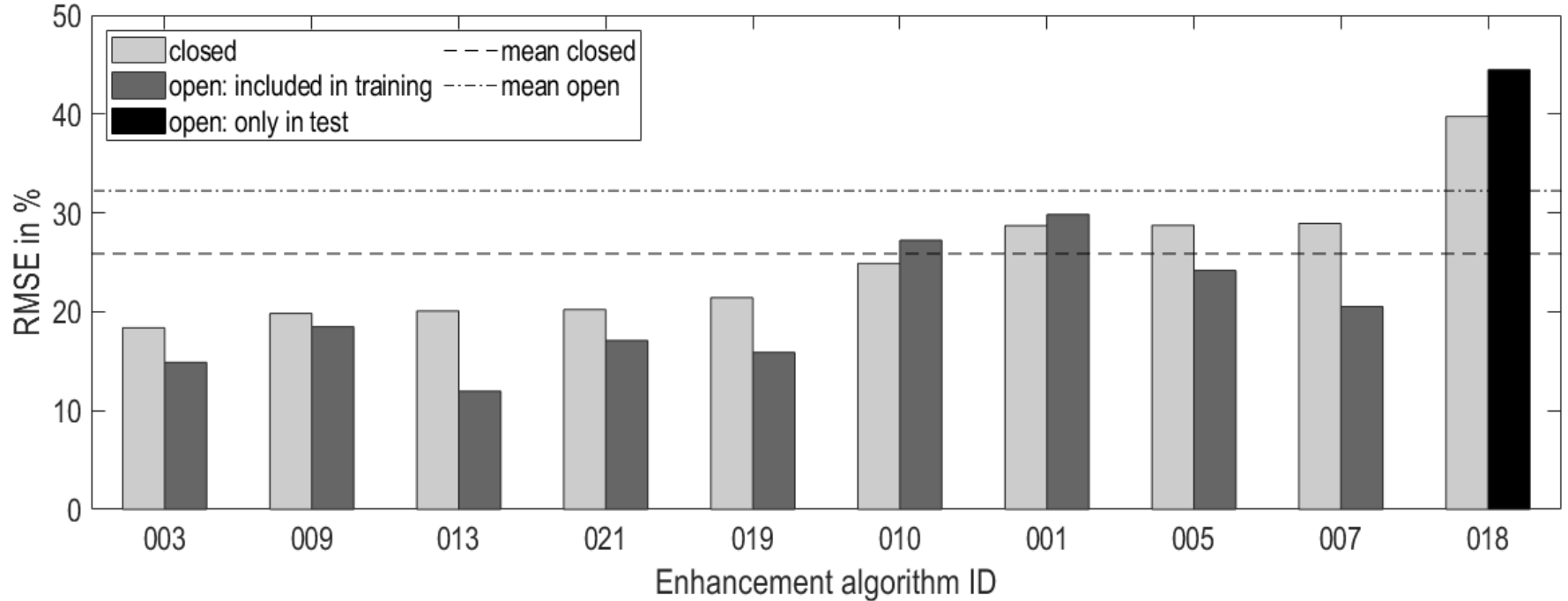


open data set: general mapping





- Possible reasons for difference between open and closed:
 - Mapping: individual for closed and general for open
 - Data: open test set contains five listeners and one algorithm excluded from training



- The non-intrusive LEAP model outperforms the intrusive MBSTOI for both data sets
 - LEAP was not trained for this challenge:
 - Trained with German, tested with English
 - Trained without spatial information, tested with reverberation
- The model generalizes and may also be used in other acoustic situations

Thank you for your attention!

- Jon Barker, Michael Akeroyd, Trevor J. Cox, John F. Culling, Jennifer Firth, Simone Graetzer, Holly Griffiths, Lara Harris, Graham Naylor, Zuzanna Podwinska, Eszter Porter and Rhoddy Viveros Munoz, "The 1st Clarity Prediction Challenge: A machine learning challenge for hearing aid intelligibility prediction", In Proceedings of the 23rd Annual Conference of the International Speech Communication Association (INTERSPEECH 2022). Incheon, Korea.
- S. Graetzer, J. Barker, T. J. Cox, M. Akeroyd, J. F. Culling, G. Naylor, E. Porter, and R. V. Muñoz, "Clarity-2021 challenges: Machine learning challenges for advancing hearing aid processing," Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH, vol. 2, pp. 1181–1185, 2021.
- H. Hermansky, E. Variani, and V. Peddinti, "Mean temporal distance: Predicting ASR error from temporal properties of speech signal," ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings, pp. 7423–7426, 2013.
- R. Huber, A. Pusch, N. Moritz, J. Rannies, H. Schepker, and B. T. Meyer, "Objective assessment of a speech enhancement scheme with an automatic speech recognition-based system," Speech Communication; 13th ITG-Symposium, Oldenburg, Germany, p. 86–90, 2018.