EVALUATING AN AUDITORY MODEL AS PREDICTOR OF SPEECH UNDERSTANDING IN HEARING-IMPAIRED LISTENERS

Helia Relaño-Iborra[1*] **Johannes Zaar**[1,2] **Torsten Dau**[1]
[1] Hearing Systems, Department of Health Technology, Technical University of Denmark, Kgs. Lyngby, Denmark
[2] Eriksholm Research Centre, Snekkersten, Denmark

ABSTRACT

Speech intelligibility (SI) models that represent effects of hearing impairment (HI) in their processing stages may help understand the link between clinical measures of auditory dysfunction and daily-life challenges with speech-in-noise understanding experienced by listeners. Here, we present a thorough evaluation of a well-established normal-hearing (NH) SI model, the speech-based computational auditory signal processing and perception model [1], as a predictor of SI performance in HI listeners. By modelling three previously published datasets, we evaluated the model's predictive power of (i) the role of audibility in unaided speech reception thresholds, (ii) the masking release obtained in fluctuating (relative to stationary) maskers, and (iii) aided performance in conditions where amplification was provided. The model was evaluated both at the listener group level to assess differences between NH and HI populations, and as a predictor of individual listener performance within the HI population. For comparison, clinical estimates used to fit the model were also analyzed as individual SI predictors. We show that the current model accurately captures effects of audibility and represents the loss of masking release observed in the HI listeners. However, the model is still limited in accounting for data associated with supra-threshold auditory deficits.

Keywords: *auditory modeling, speech intelligibility, hearing loss*

*Corresponding author: heliaib@dtu.dk.

1. INTRODUCTION

Auditory models [1-4] provide insights into the inner workings of the auditory system and can be a useful tool when, e.g., analyzing novel audio-processing techniques and hearing-aid algorithms. More importantly, models provide a valuable framework to test hypotheses regarding potential effects of hearing deficits in the auditory system on relevant outcome measures. Models allow separate access to individual auditory processing stages, such that different types of impairment can be characterized and their effects on the overall system's outcomes can be evaluated. If auditory models are to be used to evaluate potential improvements in HI performance through diverse hearing loss compensation strategies, they must first be validated as predictors of un-aided hearing-impaired (HI) listeners' speech understanding.

This contribution focuses on the evaluation of the speech-based computational auditory signal processing and perception model (sCASP) [1] as a predictor of speech intelligibility for HI listeners. sCASP, originally developed as a psychoacoustic model (CASP) [5, 6] to predict behavioural data from both normal-hearing (NH) an HI listeners, was extended to account for speech intelligibility data of NH listeners, in a wide variety of listening conditions, including conditions considering effects of speech degradations as well as non-linear speech enhancement [1]. Here, sCASP was tested in several masking conditions to evaluate whether the model (i) captures the role of audibility in unaided speech reception thresholds (SRTs), (ii) whether it can reflect the reduced masking release experienced by HI listeners as compared to NH listeners in fluctuating (relative to stationary) maskers, and (iii) if its predictions of aided performance are accurate when amplification is provided to the listeners. In this extended abstract, selected results are shown, and the overall findings are discussed.

2. METHODS

2.1 Model structure and configuration

sCASP combines a non-linear auditory-inspired preprocessing chain, including a dual resonance non-linear filterbank (DRNL) [7], inner-hair cell (IHC) transduction, adaptation and modulation filter stages, with a backend based on the cross-correlation between the clean and the noisy speech representations in the modulation envelope domain.

The model is parametrized to the individual listener's auditory profile using estimates of cochlear compression and outer- and inner-hair cell loss. In this study, outer hair cell (OHC) loss and IHC loss parameters were based either in psychoacoustical estimates [8] or derived from the listener's pure tone audiogram [9, 10] assuming that the total audiometric loss can be divided into 2/3 OHC loss and 1/3 IHC loss [11, 12]. When the stimuli were presented diotically, model simulations were carried out for each ear separately, and the best (i.e., lowest) resulting SRT was selected as the final model prediction. Only one NH model configuration was considered, assuming homogeneity across the performance of NH listeners.

In order to transform model outputs (i.e., correlation values) into intelligibility scores comparable to human responses, a fitting condition was established to provide a mapping between the two domains. For each considered speech material, the model was fitted once to the condition of speech masked by a speech-shaped stationary noise (SSN), using the NH model configuration and normative NH data. The model was evaluated in terms of its predictive power of the average data across listener populations and as a predictor of the individual listeners' performance.

2.2 Experimental data

The model's predictive power was evaluated using three different datasets [8,9,10]. Here, results from a subset of conditions are reported. The intelligibility of masked speech was modelled for unaided listeners in the presence of stationary and fluctuating interferers. SRTs were modelled for 13 HI-listeners and compared to data collected in [9]. The stimuli consisted of Danish sentences from the CLUE corpus presented diotically at 80 dB SPL in the presence of three different maskers: SSN, an 8-Hz modulated SSN (SAM) and the International Speech Test Signal [13], a nonsense-speech masker.

3. SELECTED RESULTS

Fig. 1 shows speech reception thresholds (SRTs) for NH listeners (diamonds) and HI listeners (boxplots). The human data are indicated by the open symbols while model predictions are shown using filled symbols. Each panel represents one masking condition. The NH data, measured in [14] and modelled in [1], are included for comparison, showing lower SRTs (i.e., better speech intelligibility) for speech that was presented in fluctuating maskers (i.e., SAM and ISTS) as compared to the stationary noise masker. This phenomenon, known as masking release, can be accounted for by sCASP). The HI data, measured by [9], show a reduced masking release as compared to the NH listeners with elevated SRTs in all listening conditions, although less markedly for the SSN interferer. sCASP captures the reduced masking release but does not capture the elevation of SRTs for the SSN masker. Furthermore, the across-listener variability observed in the HI listeners' data is not reflected in the HI model predictions, despite the individualized model configurations.

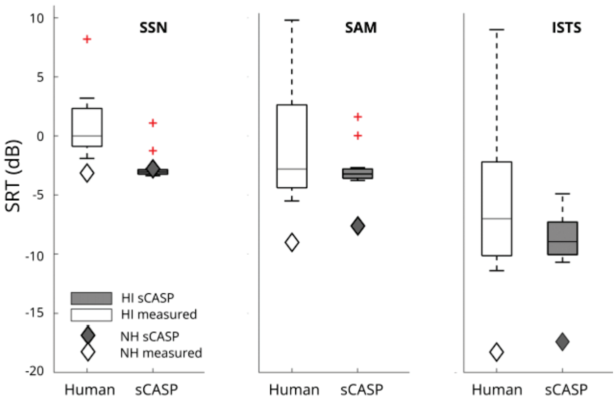

Figure 1: Measured (open symbols) and modelled (filled gray symbols) speech reception thresholds for normal-hearing (diamonds) and hearing-impaired listeners (boxplots). Each panel represents one interferer: Speech-shaped noise (SSN); amplitude-modulated speech-shaped noise (SAM); and the international speech test signal ISTS [13]. The HI data are from [9] and the reference NH data are from [14].

Fig. 2 shows measured (x-axis) vs modelled (y-axis) speech reception thresholds for the individual HI listeners. Each panel represents one masking condition. Significant correlations were found for the SSN and SAM, but not for the ISTS masker.

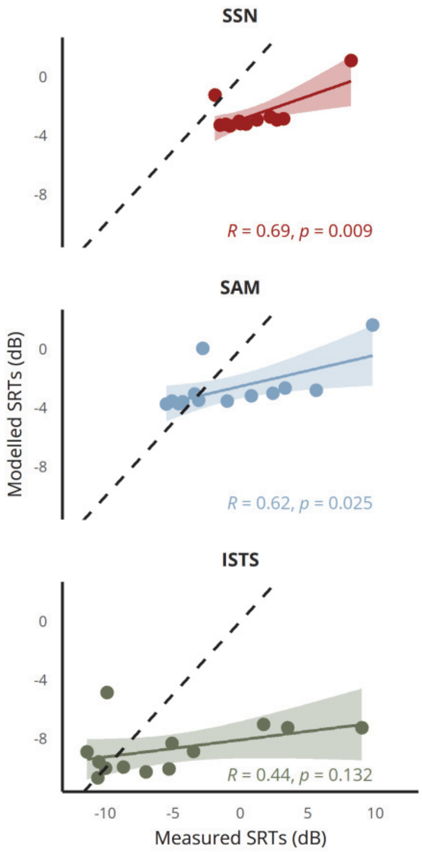

Figure 2: Measured (x-axis) vs modelled (y-axis) speech reception thresholds. Each panel represents one masking condition. SSN, SAM, and ISTS [13]. R indicates the correlation value and *p* the significance level. Human data were collected in [9].

4. DISCUSSION AND OUTLOOK

Overall, the predictions obtained with sCASP reflect the general decrease in performance observed for the HI listeners as compared to results from NH listeners. Furthermore, the model correctly predicts masker-type effects in the SRTs. The model accounts well for the trends observed at a group level, whereas significant correlations between the measured and predicted performance across the individual listeners are only found for a subset of the maskers. The fact that the model fails to account for the elevation of the SRTs in the SSN noise, as compared to the fluctuating maskers, suggests that the model is better suited to predict speech intelligibility in conditions where the audibility of the speech (e.g., in the dips of a fluctuating masker) might be the limiting factor of the listeners' performance. Additional simulations of two other datasets (not shown here) further support the hypothesis that, while the model accurately accounts for decreased speech intelligibility performance caused by a loss of sensitivity (i.e., audibility), it struggles to predict potential effects of suprathreshold hearing deficits on speech intelligibility. The results of this study encourage further investigations towards the prediction of individual performance.

5. ACKNOWLEDGMENTS

This research was supported by the Centre for Applied Hearing research (CAHR).

6. REFERENCES

[1] Relaño-Iborra, H., Zaar, J., & Dau, T. (2019). A speech-based computational auditory signal processing and perception model. The Journal of the Acoustical Society of America, 146(5), 3306–3317. https://doi.org/10.1121/1.5129114CASP 2008

[2] Relaño-Iborra, H., May, T., Zaar, J., Scheidiger, C., & Dau, T. (2016). Predicting speech intelligibility based on a correlation metric in the envelope power spectrum domain. The Journal of the Acoustical Society of America, 140(4), 2670–2679. https://doi.org/10.1121/1.4964505

[3] Zaar, J., & Carney, L. H. (2022). Predicting speech intelligibility in hearing-impaired listeners using a physiologically inspired auditory model. Hearing Research, 108553. https://doi.org/10.1016/j.heares.2022.108553

[4] Relaño-Iborra, H., & Dau, T. (2022). Speech intelligibility prediction based on modulation frequency-selective processing. Hearing Research, 108610. https://doi.org/10.1016/j.heares.2022.108610

[5] Jepsen, M. L., Ewert, S. D., & Dau, T. (2008). A computational model of human auditory signal processing and perception. The Journal of the

Acoustical Society of America, 124(1), 422–438. https://doi.org/10.1121/1.2924135

[6] Jepsen, M. L., & Dau, T. (2011). Characterizing auditory processing and perception in individual listeners with sensorineural hearing loss. The Journal of the Acoustical Society of America, 129(1), 262–281. https://doi.org/10.1121/1.3518768

[7] Lopez-Poveda, E. A., & Meddis, R. (2001). A human nonlinear cochlear filterbank. The Journal of the Acoustical Society of America, 110(6), 3107–3118. https://doi.org/10.1121/1.1416197

[8] Johannesen, P. T., Pérez-González, P., Kalluri, S., Blanco, J. L., & Lopez-Poveda, E. A. (2016). The Influence of Cochlear Mechanical Dysfunction, Temporal Processing Deficits, and Age on the Intelligibility of Audible Speech in Noise for Hearing-Impaired Listeners. Trends in Hearing, 20, 1–14. https://doi.org/10.1177/2331216516641055

[9] Christiansen, C., & Dau, T. (2012). Relationship between masking release in fluctuating maskers and speech reception thresholds in stationary noise. The Journal of the Acoustical Society of America, 132(3), 1655–1666. https://doi.org/10.1121/1.4742732

[10] Wardenga, N., Batsoulis, C., Wagener, K. C., Brand, T., Lenarz, T., & Maier, H. (2015). Do you hear the noise? the German matrix sentence test with a fixed noise level in subjects with normal hearing and hearing impairment. International Journal of Audiology, 54, 71–79. https://doi.org/10.3109/14992027.2015.1079929

[11] Moore, B. C. J., & Glasberg, B. R. (1997). A model of loudness perception applied to cochlear hearing loss. Auditory Neuroscience, 3(3), 289–311. https://doi.org/10.1016/S0378-5955(03)00347-2

[12] Plack, C. J., Drga, V., & Lopez-Poveda, E. A. (2004). Inferred basilar-membrane response functions for listeners with mild to moderate sensorineural hearing loss. The Journal of the Acoustical Society of America, 115(4), 1684–1695. https://doi.org/10.1121/1.1675812

[13] Holube, I., Fredelake, S., Vlaming, M. S. M. G., & Kollmeier, B. (2010). Development and analysis of an International Speech Test Signal (ISTS). International Journal of Audiology, 49, 891–903. https://doi.org/10.3109/14992027.2010.506889

[14] Jørgensen, S., Ewert, S. D., & Dau, T. (2013). A multi-resolution envelope-power based model for speech intelligibility. The Journal of the Acoustical Society of America, 134(1), 436–446. https://doi.org/10.1121/1.4807563