Predicting speech intelligibility across acoustic conditions and hearing status using a physiologically inspired auditory model

Johannes Zaar*[†1,2] and Laurel Carney[3]

[1]Eriksholm Research Centre, DK-3070 Snekkersten, Denmark – Denmark
[2]Hearing Systems Section, Department of Health Technology, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark – Denmark
[3]Departments of Biomedical Engineering and Neuroscience, University of Rochester, Rochester, NY, 14642 USA – United States

Abstract

Various computational speech-intelligibility (SI) prediction models have been developed, but most focus on normal-hearing (NH) listeners in a large range of acoustic conditions. Only a few SI-prediction models aim to account for effects of hearing impairment on SI. It remains a challenge to faithfully predict speech-test outcomes for groups of hearing-impaired (HI) listeners, and even more so for individual HI listeners, especially when stimuli are presented at supra-threshold levels. The current study presents a major update and full evaluation of a previously introduced SI model. Inputs to the model are the noisy speech stimulus and the noise alone, as a reference signal. The two signals are processed through a physiologically inspired nonlinear model of the auditory periphery, for a range of characteristic frequencies (CFs), followed by a modulation analysis in the range of the fundamental frequency (F_0) of speech. The decision metric of the model is the mean of a series of short-term across-CF correlations between population responses to noisy speech and noise alone, with a sensitivity limitation process imposed on it. This correlation is assumed to be inversely related to SI. The model was evaluated in a number of speech-in-noise conditions with stationary, fluctuating, and speech-like interferers using data previously obtained in NH as well as HI listeners. A conversion function between the model's decision metric and SI (in percent correct) was computed based on predictions and data obtained in the stationary-noise condition in NH listeners; this conversion was then used for all other tests, irrespective of acoustic condition or hearing status. Speech reception thresholds (SRTs) were derived from predicted psychometric functions, for comparison to measured SRTs. The model made accurate predictions of NH listeners' SRTs across the different acoustic conditions. Furthermore, the model predicted plausible effects in response to changes in presentation level, accounting both for inaudible stimuli (0% correct at very low levels) and overly loud stimuli (elevated SRTs at very high levels – "roll-over"). Crucially, the model also accounted for effects of hearing impairment on the SRTs when simply adjusting the front-end processing based on individual audiograms using standard assumptions of inner-hair-cell (IHC) and outer-hair-cell (OHC) loss. HI predictions were accurate at the group level and also captured the large inter-individual variability across HI listeners. An evaluation of the model's performance

---

*Speaker
†Corresponding author: jozr@eriksholm.com

with a more fine-grained adjustment of IHC and OHC loss will be discussed. Furthermore, the model's predictive power will be further evaluated using additional NH and HI data sets. Overall, the present model provides a useful tool to accurately predict speech-in-noise outcomes in NH and HI listeners, and – perhaps more importantly – to facilitate important insights into processes that are crucial for speech understanding.