Prediction of speech intelligibility based on a correlation metric in the envelope power spectrum domain

Iborra, Helia Relano; May, Tobias; Zaar, Johannes; Scheidiger, Christoph; Dau, Torsten

Publication date: 2017

Document Version
Publisher's PDF, also known as Version of record

Link back to DTU Orbit

Citation (APA):
Prediction of speech intelligibility based on a correlation metric in the envelope power spectrum domain

Helia Relaño Iborra\textsuperscript{a),} Tobias May, Johannes Zaar, Christoph Scheidiger and Torsten Dau

Hearing Systems Group, Department of Electrical Engineering, Technical University of Denmark, DK-2800, Kgs. Lyngby, Denmark.

Introduction

A powerful tool to investigate speech perception is the use of speech intelligibility prediction models. Recently, a model was presented, termed correlation-based speech-based envelope power spectrum model (sEPSM\textsuperscript{corr})\cite{Jepsen2008}, based on the auditory processing of the multi-resolution speech-based Envelope Power Spectrum Model (mr-sEPSM)\cite{Jørgensen2011}, combined with the correlation back-end of the Short-Time Objective Intelligibility measure (STOI)\cite{Søndergaard2001}. The sEPSM\textsuperscript{corr} model can accurately predict HI data for a broad range of listening conditions, e.g., additive noise, phase jitter and ideal binary mask processing.

The sEPSM\textsuperscript{corr} model includes auditory thresholds, such that sensitivity loss can be incorporated based on the audiogram, but other types of hearing impairment (HI) cannot be simulated using this framework. However, speech perception can vary greatly among listeners even when hearing sensitivity is similar. Therefore, the predictive power of the sEPSM\textsuperscript{corr} back-end was further investigated in combination with a more realistic processing and perception model (CASP)\cite{Jepsen2008}. Here, the speech-based CASP (sCASP) was incorporated based on the audiogram, but other types of hearing impairment (HI) prediction models. Recently, a model was presented, termed correlation-based speech-based envelope power spectrum model (sEPSM\textsuperscript{corr})\cite{Jepsen2008}.

The sEPSM\textsuperscript{corr} model

The CASP model offers more flexibility to model hearing impairments, beyond the audiogram, due to the Dual Resonance Non-linear filterbank (DRNL),\cite{Relaño-Iborra2016}. The model has been shown to account for psychoacoustic data from individual HI subjects.

sCASP model

Towards prediction of HI data

The CASP model offers more flexibility to model hearing impairments, beyond the audiogram, due to the Dual Resonance Non-linear filterbank (DRNL),\cite{Relaño-Iborra2016}. The model has been shown to account for psychoacoustic data from individual HI subjects.

Test conditions

The models were evaluated in conditions with:
- Speech mixed with stationary or non-stationary interferers: Speech shaped noise (SSN), which was also used to fit the model. Amplitude modulated SSN (AMS) with $\rho = 0.8$ for and modulation depth of 1, and the speechlike, but non-stereoscopic international speech test signal (STIs).
- Noisy speech in the presence of reverberation: $T_{60} = 0, 0.4, 0.7, 1.3$ and 2.3 s
- Noisy speech subjected to different types of non-linear processing
  - Ideal Binary Mask processing (IBM) with four interferences
  - Phase Jitter distortion

Fitting of the models

The models are fitted per speech material to the condition of clean speech with SSN by fitting a sigmoid function between the model outputs and the human scores.

Results

The sCASP model provides similar (and in some conditions better) results than the sEPSM\textsuperscript{corr} model. The model can now serve as foundation for the development of a HI model, since the DRNL-based framework allows for fitting to individual impairments.

Outlook

- Investigate the model’s ability to account for individual hearing impairments using the parameters available in the CASP framework.
- Consider additional processing stages that could account for inner hair-cell loss and auditory nerve deafferentation (Sumner et al., 2002; López-Poveda and Barrias, 2013)\cite{Sumner2002}, as they are likely to be determinant in speech-in-noise related tasks.
- Determine the conditions on which the HI model will be tested with special focus on supra-threshold distortions that might be challenging for HI subjects.

Summary of results

The sCASP model provides similar (and in some conditions better) results than the sEPSM\textsuperscript{corr} model.

The model can now serve as foundation for the development of a HI model, since the DRNL-based framework allows for fitting to individual impairments.