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 Iborra1), 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 (sEPSMcorr) [2], based on the auditory processing of the multi-resolution speech-based Envelope Power Spectrum Model (mr-sEPSM) [2], combined with the computation back-end of the Short-Time Objective Intelligibility measure (STOI) [3]. The sEPSMcorr can accurately predict HI data for a broad range of listening conditions, e.g., additive noise, phase jitter and ideal/binary mask processing.

The sEPSMcorr model includes audibility thresholds, such that sensitivity loss can be conditions, e.g., additive noise, phase jitter and ideal/binary mask processing.

The sEPSMcorr model includes audibility thresholds, such that sensitivity loss can be incorporated based on the audiogram, but other types of hearing impairments (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 sEPSMcorr back-end was further investigated in combination with a more realistic auditory pre-processing front-end adopted from the computational auditory signal processing and perception model (CASP) [4]. Here, the speech-based CASP (sCASP) was evaluated in HI conditions and compared to the sEPSMcorr.

The sEPSMcorr model
The sEPSMcorr model includes audibility thresholds, such that sensitivity loss can be incorporated based on the audiogram, but other types of hearing impairments (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 sEPSMcorr back-end was further investigated in combination with a more realistic auditory pre-processing front-end adopted from the computational auditory signal processing and perception model (CASP) [4]. Here, the speech-based CASP (sCASP) was evaluated in HI conditions and compared to the sEPSMcorr.

sCASP model
The model is presented 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 (AMSSN) with \(f_{\text{mod}} > 10 \text{ Hz} \) and modulation depth of 1. and the speech like, but non-semantic international speech test signal
- Noisy speech in the presence of reverberation: T60 = 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 interferers.
  - Phase Jitter distortion
    \( \gamma = \gamma(t) \) if \( \text{SNR(t)} \) > LC otherwise

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 sEPSMcorr model provides similar (and in some conditions better) results than the sEPSMcorr.

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.

<table>
<thead>
<tr>
<th>Model</th>
<th>Human data</th>
<th>mr-sEPSM</th>
<th>STOI</th>
<th>sEPSMcorr</th>
<th>sCASP</th>
</tr>
</thead>
<tbody>
<tr>
<td>additive noise</td>
<td>0.94</td>
<td>0.90</td>
<td>0.94</td>
<td>0.89</td>
<td>0.90</td>
</tr>
<tr>
<td>Reverberant speech</td>
<td>0.93</td>
<td>0.91</td>
<td>0.93</td>
<td>0.92</td>
<td>0.91</td>
</tr>
<tr>
<td>Jittered speech</td>
<td>0.91</td>
<td>0.90</td>
<td>0.91</td>
<td>0.90</td>
<td>0.90</td>
</tr>
</tbody>
</table>

Summary of results
The sEPSMcorr model provides similar (and in some conditions better) results than the sEPSMcorr.

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 [8]; López-Poveda and Barrios, 2013 [9]), 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.


helaila@elektro.dtu.dk