A correlation metric in the envelope power spectrum domain for speech intelligibility prediction

Relaño-Iborra, H.; May, Tobias; Zaar, Johannes; Scheidiger, Christoph; Dau, Torsten

Publication date: 2016

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

Link back to DTU Orbit

Citation (APA):
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) [1], based on the auditory processing of the multi-resolution speech-based Envelope Power Spectrum Model (mr-sEPSM) [2], with the correlation back-end of the Short-Time Objective Intelligibility measure (STOI) [3]. The sEPSMcorr can accurately predict NH 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 losses can be incorporated based on the audiogram, but other types of impairment (e.g., loss of compression, reduced frequency selectivity) 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].

**Evaluation**

- Speech mixed with stationary or non-stationary interferences:
  - Speech shaped noise (SSN), which was also used to fit the model
  - Amplitude modulated SSN (SAM) with f_{mod} = 8 Hz and modulation depth of 1.
  - The speech like, but non-semantically international speech test signal (ISTS)
- Speech in the presence of reverberation, \( t_{reverb} = 0, 0.4, 0.7, 1.3, 2.3 \) s
- Ideal Binary Mask processing (IBM) with four interferers.
- Speech subjected to Phase jitter distortion:
  \[ r(t) = \text{Re}(s(t)e^{j\omega t}) + \sigma(t)\cos(\omega t)) \]

**Towards realistic cochlear processing**

sEPSMcorr includes auditory thresholds, such that sensitivity losses can be incorporated based on the audiogram. However, other types of impairment (e.g., loss of compression, reduced frequency selectivity) cannot be simulated using this model.

Therefore, the predictive power of the sEPSMcorr back end is further investigated in combination with a more realistic auditory pre-processing. The CASP model [4] is considered, as its front end can be tuned on the audiogram, but other types of impairment (e.g., spectral masking, amplitude-modulation detection, and forward masking [4] as well as individual HI results from simultaneous and forward masking and notch-noise experiments [6]).

**References**