A multi-resolution envelope-power based model for speech intelligibility

A multi-resolution envelope-power based model for speech intelligibility

The speech-based envelope power spectrum model (sEPSM) presented by Jørgensen and Dau [(2011). J. Acoust. Soc. Am. 130, 1475-1487] estimates the envelope power signal-to-noise ratio (SNR_{env}) after modulation-frequency selective processing. Changes in this metric were shown to account well for changes of speech intelligibility for normal-hearing listeners in conditions with additive stationary noise, reverberation, and nonlinear processing with spectral subtraction. In the latter condition, the standardized speech transmission index [(2003). IEC 60268-16] fails. However, the sEPSM is limited to conditions with stationary interferers, due to the long-term integration of the envelope power, and cannot account for increased intelligibility typically obtained with fluctuating maskers. Here, a multi-resolution version of the sEPSM is presented where the SNR_{env} is estimated in temporal segments with a modulation-filter dependent duration. The multi-resolution sEPSM is demonstrated to account for intelligibility obtained in conditions with stationary and fluctuating interferers, and noisy speech distorted by reverberation or spectral subtraction. The results support the hypothesis that the SNR_{env} is a powerful objective metric for speech intelligibility prediction. © 2013 Acoustical Society of America.