

Measuring Speech Perception with Recovered Envelope Cues using the Peripheral Auditory Model

*Nursadul Mamun^{*1}, Khadija Akter², Hussnain Ali¹, John H. L. Hansen¹*

¹EE, University of Texas, Dallas, Dallas, Texas, United States; ²ETE, Chittagong University of Engineering and Technology, Chittagong, Bangladesh

Abstract— Speech perception refers to how understandable speech produced by a speaker would be by a listener. The human auditory system usually interprets this information using both envelope (ENV) and temporal fine structure (TFS) cues. While ENV is sufficient for understanding speech in quiet, TFS cues are necessary for speech segregation in noisy conditions. In general, ENV can be recovered from the TFS (known as recovered ENV); however, the degree of ENV recovery and its significance on speech perception are not clearly known/understood. In order to systematically assess the relative contribution of the recovered ENV for speech perception, this study proposes a new speech perception metric. The proposed metric employs a phenomenological model of the auditory periphery developed by Zilany and colleagues (J. Acoust. Soc. Am. 126, 283–286, 2014) to simulate the responses of the auditory nerve fibers to both original and recovered ENV cues. The performance of the proposed metric was evaluated under different types of noise (both steady-state and fluctuating noise), as well as several classes of distortion (e.g., peak-clipping, center-clipping, and phase jitter). Finally, to validate the proposed metric, the predicted scores were compared with subjective evaluation scores from behavioral studies. The proposed metric indicates a statistically significant correlation for all cases and accounts for a wider dynamic range compared to the existing metrics.