ACOUSTICS2008/571
Estimation of Target-to-Interferer Ratio using the Auditory Image Model

Michael Carlin
Air Force Research Laboratory, 525 Brooks Road, Rome, NY 13441, USA

Performance in speech processing applications such as speaker recognition becomes severely degraded when the short-time audio being analyzed contains more than one speaker. However, it has been shown that if the audio is only minimally-corrupted by the interfering speech, i.e., the Target-to-Interferer Ratio (TIR) is sufficiently large, then accurate recognition results can still be achieved. During phonation, estimation of TIR is especially critical since uncorrupted vowel sounds contain important speaker-discriminating information. This research investigates a method to estimate the relative intensity of interfering speech using the Auditory Image Model (AIM) of Patterson et al. (J. Acoust. Soc. Am., Vol. 98, pp. 1890-1894, 1995). The proposed TIR estimator attempts to exploit both the apparent high resolution in the simulated Neural Activity Pattern and variation in cross-channel Strobe Point correlation when observing overlapping vowel sounds. Experiments were conducted for five canonical male vowels which were perceptually-scaled using the STRAIGHT algorithm (Chapter in Speech Separation by Humans and Machines, P. Divenyi, ed., Kluwer Academic Publishers, 2005) and superimposed at varying levels of TIR. Results suggest that the proposed approach is a promising step towards both detecting the presence and relative intensity of an interfering speaker.