ACOUSTICS2008/1710
Spectral and temporal modulations essential to spoken word, gender and timbre identification

Frédéric Theunissen and Taffeta Elliott
UC Berkeley, Dept. of Psychology, 3210 Tolmant Hall, Berkeley, CA 94720-1650, USA

Human speech and musical sounds contain complex spectral and temporal modulations. Speech intelligibility, perception of melody, and identification of source characteristics (e.g., speaker gender or musical timbre) depend on spectrotemporal modulations but can be surprisingly robust to drastic spectral and temporal degradations. We systematically explored which restricted spectral and temporal modulations are essential to the perception of complex sounds. Degraded sentences and musical sounds were obtained by a novel modulation filtering procedure performed on the sound spectrogram. Temporal modulation filtering smeared the amplitude envelope by removing changes above particular Hz. Spectral modulation filtering smeared the spectral energy across frequency bands by removing changes above particular cyc/kHz. We further complemented this low-pass filtering with more specific notch-filtering. Speech intelligibility, gender recognition and musical instrument identification were assessed in psychophysical experiments. We determined that spectral modulations below ∼3.75 cyc/kHz, and temporal modulations between 1 and 7 Hz are essential for speech comprehension. Gender identification however required the presence of higher spectral modulations. Similarly the timbre and pitch of instruments was affected differentially by notch filters in these two regions of the modulation spectrum. Our research could be used to guide the design of optimal signal processing in hearing aids and cochlear implants.