ACOUSTICS2008/950
Do we perceive articulatory gestures when we listen to speech?

Pierre Divenyi and Adam Lammert

VA Northern Calif. Health Care Syst. and East Bay Inst. for Research and Education, Bldg. R4, 150 Muir Rd., Martinez, CA 94553, USA

Articulatory synthesis methods, classic and contemporary, have demonstrated that it is possible to generate speech from an ensemble of functions derived from articulatory gestures. Such gesture-to-waveform transforms suggest that, inversely, the speech signal should be also decomposable into the same set of gesture, or gesture-like, functions. These functions vary slowly in time and their association with the speech waveform (words as well as sentences) can be established by machine learning algorithms. In a recent study at our laboratory, listeners were asked to type the word or the sentence they heard, with speech (degraded in diverse ways) as the stimulus. The subjects' responses were synthesized, time-aligned with the stimulus, and decomposed into a set of eight gestures, as specified by the Haskins Laboratories TADA system (http://www.haskins.yale.edu/tada_download/index.html). When the running distance between input and response gesture functions is calculated, results indicate a significant degree of gesture information transmitted even during severely degraded speech segments, suggesting that the perceptual system may track speech via underlying functions similar to gestures. Epochs at which this running distance estimate fails, i.e., exceeds a certain threshold, may be considered to signal periods during which insufficient bottom-up information had to be supplemented using higher-order linguistic knowledge.