Two time scales in speech segmentation

Maria Chait University of Maryland
Steven Greenberg ICSI, Berkeley
Takayuki Arai Sophia University
Jonathan Simon University of Maryland
David Poeppel University of Maryland

The basic units into which the speech-signal is segmented is an issue of central importance for speech research. Evidence points to the perceptual reality of both the phonetic segment (~30ms) and the syllable (~300ms) during the course of speech processing. We propose a new method of systematically examining the extraction and combination of these informational constituents of speech. We build on accumulating evidence regarding the importance of temporal envelope for speech processing to create a specially crafted stimulus. The original signal is split into 1/3 octave-wide bands, the amplitude envelope from each band is extracted and low or high-passed before being combined again with the carrier signal. The result for each signal (S) is S_low and S_high, containing only low or high modulation information. We demonstrate that although each of these, when presented separately in intelligibility judgment tasks, has low (c.a. 40% and 15%) intelligibility, the dichotic presentation of S_low with S_high results in significantly better (c.a. 70%) performance, suggesting a binding mechanism between the syllabic and segmental information. To investigate the properties of this mechanism, we introduce a time-shift in the onset of S_low relative to S_high. Asynchronies <45ms have no effect on intelligibility, performance declines sharply between 45-150 ms, remaining constant thereafter. This evidence suggests a process of multi-resolution analysis (MRA) of speech: segmental and supra-segmental information are extracted simultaneously but separately from the input stream from 'short' (~30ms) and 'long' (~300ms) windows of integration. These streams are then bound together to create a stable representation which is the perceptual unit that is used for subsequent higher order perceptual computations. Crucially, according to this model, supra-segmental units as well as phoneme-sized units are equally fundamental. We discuss these findings in light of other experimental evidence and suggest that the envelope carries information that is critical for our ability to segment speech and that the precise information extracted from these temporal-integration windows depends on phonological and prosodic constraints related to the listeners' native language. These segmentation mechanisms, or 'temporal windows' are plastic and 'stretch'/'shrink' up to a certain degree, but once the 'segmentation cues' begin to come too fast (time-compressed speech experiments) or too slow (temporally-segmented speech experiments), they fail. The MRA model is an attempt to bridge the gap between speech psychophysics and auditory neuroscience, by providing common terminology to describe, understand and integrate the experimental results in these often independent fields.