O-01

Brain Mechanisms for Speech Segmentation

Maria Chait*, Steven Greenberg**, Takayuki Arai***,Jonathan Z. Simon**** and David Poeppel****

*Neuroscience and Cognitive Science Program, Cognitive Neuroscience of Language Lab, Department of Linguistics, UMCP.
**The Speech institute Oakland, CA
***Department of Electrical and Electronic Engineering, Sophia University, Tokyo
****Departments of Biology and Electrical and Computer Engineering, UMCP
*****Neuroscience and Cognitive Science Program, Cognitive Neuroscience of Language Lab, Departments of Linguistics and Biology, UMCP

The basic units into which the acoustic speech signal is segmented is an issue of central importance for speech research. A growing body of work is pointing to the perceptual reality of both the phonetic segment (~30 ms) and the syllable (~300 ms) during the course of speech processing.

In this study we propose a new method of systematically examining the extraction and subsequent combination of these informational constituents of speech. The original wide band speech signal is split into 14 frequency bands with an FIR filter bank spanning the range 0-6kHz spaced in 1/3 octave steps along the cochlear frequency map. The amplitude envelope from each band is computed by means of a Hilbert transform and then either low (0-3Hz) or high (22-40Hz) band passed before being combined again with the original carrier signal. The result for each original signal (S) is S_low and S_high, containing only low or high modulation frequencies.

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. This data demonstrates that intelligibility crucially depends on both the slowly varying as well as the rapidly varying components of speech and suggests a binding process, in which a conjunction of low and high frequency information creates an emergent representation that forms the basis for successful speech processing.

In order to investigate the temporal properties of this binding mechanism we introduce a time shift in the onset of S_low relative to S_high. Our data shows that Asynchronies <30ms have no effect on intelligibility, performance declines sharply between 30-150 ms, remaining constant beyond that interval. The significance of these findings to models of speech processing is further discussed.