ACOUSTICS2008/723
Perception of speech properties from extremely brief segments

Sue Harding and Martin Cooke
Sheffield University, Computer Science Department, Regent Court, 211 Portobello St., S1 4DP Sheffield, UK

A glance at a visual scene enables observers to become rapidly aware of its most important characteristics. Here, we describe experiments using very brief segments of natural speech which demonstrate that a surprising amount of information can be determined from only a few milliseconds of the auditory signal. Segments with durations ranging from 2.5 to 80 ms were extracted from six vowels and six fricatives spoken by males and females. Listeners identified the phoneme and/or gender, or whether a vowel or consonant had been presented. While listeners' performance dropped close to chance for the 2.5 ms stimuli for most tasks, for the vowel/fricative distinction listeners obtained scores above 70% even for such short segments. Listeners performed well above chance for the 10 ms stimuli for three out of four tasks. Combining results within tasks showed that listeners also distinguished voiced from unvoiced phonemes in less than 10 ms. Threshold values from logistic fits indicate the order in which information becomes available: vowel/fricative distinction (3.0 ms), voicing distinction (6.7 ms), phoneme identification (11.9 ms) and gender identification (15.3 ms). By exploiting the "gist" of an auditory scene, listeners may be able to deploy prior knowledge rapidly to constrain further interpretation.