A computational model of a contextual effect on consonant perception in nasal vowel-nasal syllables

Susan L Denham and Georg F Meyer

1: Plymouth Institute of Neuroscience, University of Plymouth, UK
2: MacKay Institute of Communication and Neuroscience, Keele University, UK

An important challenge in auditory perception is to discover how the various components in a complex acoustic mixture are grouped and segregated in order to create a meaningful representation of sound sources in the environment. Grouping of simultaneous components is strongly promoted by common onsets and common periodicity (Darwin and Carlyon, 1995). However, it has also been shown that the temporal context of the sound mixture exerts an important influence on the grouping and segregation process (Hukin and Darwin, 1995; Bregman, 1990). Not surprisingly, temporal context is also important in speech perception. For example, when a synthetic nasal /m/, with formant frequencies at 250, 1000 and 2000 Hz is presented in isolation it is correctly identified in about 86% of presentations. However, when it is preceded by vowels with formant frequencies at 375 and 2700 Hz, and varying second formant frequencies, the percept changes from an /m/ to an /n/ when the second formant exceeds 1400 Hz (Meyer and Barry, 1999). The paradigm commonly used for investigating the effect of temporal context on auditory perception is that of two-tone streaming (Bregman, 1990). In a recent study of population responses in primary auditory cortex of awake monkeys to such alternating tone sequences, the differential suppression of non-best frequency tones by alternating best frequency tones was found (Fishman et al., 2001). Furthermore, consistent with human streaming perception, this suppression increased with increasing frequency difference and presentation rate. The timing and frequency distribution of suppression was also consistent with forward masking (Brosch and Schreiner, 1997) and it was suggested that there may be a common mechanism underlying both phenomena. A computational modelling study showed that that mechanism could be depression of thalamocortical synapses, and also helped to explain the paradoxical observation that differential suppression is greater at larger frequency differences while forward masking lasts longer at the best frequency of the cell. The same computational model was presented with the vowel nasal stimuli used in the human perceptual experiment (Meyer and Barry, 1999) and the patterns of activity in response to the final consonant were compared. It was found that while the activity in the peripheral auditory model was not influenced by the preceding vowel, activity in the ‘cortical’ layer was, and the relationship between the perception of a final /m/ with the frequency of the second formant closely resembled human perceptual judgments.