Speech Segregation by Oscillatory Correlation

DeLiang Wang

Department of Computer Science and Engineering, and Centre for Cognitive Science
The Ohio State University
Columbus, Ohio 43210
E-mail: dwang@cse.ohio-state.edu

What neural mechanisms underlie auditory scene analysis? Both theoretical and empirical investigations of the brain point to the mechanism of oscillatory correlation as a plausible paradigm for scene representation. In this presentation I describe an oscillatory correlation approach to the problem of speech segregation, or cocktail-party processing. In the oscillatory correlation approach, a perceptual stream corresponds to a synchronized assembly of neural oscillators and different streams correspond to desynchronized oscillator assemblies. This approach has been employed for double-vowel separation and segregation of voiced speech. An oscillator model for double-vowel separation synchronizes auditory channels that define the spectral components of each vowel on the basis of periodicity analysis. This model is able to replicate the perceptual observation that listeners’ ability to identify concurrent vowels improves with increasing difference in fundamental frequency between the vowels. For voiced speech segregation, a two-layer network of relaxation oscillators is used. The first layer performs the task of auditory segmentation whereby an auditory scene is broken into a collection of auditory segments, each of which corresponds to a contiguous time-frequency region. The second layer performs the task of grouping, in which segments are organized into distinct streams. Lateral connections between oscillators encode harmonicity and proximity in frequency and time. Prior to the oscillator network are a model of the auditory periphery and a stage in which mid-level auditory representations, such as correlogram, are formed. This model of speech segregation has been evaluated using a corpus of voiced speech mixed with a variety of interfering sounds. Further developments of this model are discussed. Finally, I will speculate on possible roles of oscillatory correlation in the broad framework of computational audition.