

Cocktail Party Processing

DeLiang Wang

The Ohio State University

Speech segregation, or the cocktail party problem, has proven to be extremely challenging. This presentation describes a computational auditory scene analysis (CASA) approach to the cocktail party problem. This approach performs auditory segmentation and grouping in a two-dimensional time-frequency representation that encodes proximity in frequency and time, periodicity, amplitude modulation, and onset/offset. In segmentation, our model decomposes the input mixture into contiguous time-frequency segments. Grouping is first performed for voiced speech where detected pitch contours are used to group voiced segments into a target stream and the background. In grouping voiced speech, resolved and unresolved harmonics are dealt with differently. Grouping of unvoiced segments is based on supervised classification of acoustic-phonetic features. This CASA approach has led to major advances towards solving the cocktail party problem.