How Does the Brain Go from Sound to Meaning?

Steven Greenberg, Silicon Speech
Santa Venetia, CA USA

Contemporary models of speech recognition by humans and machines are difficult to reconcile with many properties of spoken language. Pronunciation variation, robustness to acoustic interference, categorical perception, and lexical access are among the (many) phenomena the "standard framework" fails to explain. This presentation describes a new theoretical formulation, using hierarchical oscillatory networks (Hi-O Nets), that relates auditory speech processing with other sensory (e.g., vision) and cognitive (e.g., memory) data streams. Within the Hi-O framework, signal-parsing and pattern-matching are crucial stages in going from sound to meaning. They depend on the structured interaction of oscillatory neural activity across a broad range of time constants characteristic of speech (20 – 2000 ms). A multi-time-scale, hierarchical oscillatory framework can account for many phenomena in spoken language, including (1) the ability to understand speech in background noise and other forms of acoustic interference, (2) the effect of sentential and semantic context on speech intelligibility, and (3) the perceptual invariance of highly variable and dynamic acoustic signals. Hi-O Nets will be illustrated and discussed with reference to both classic and more recent perceptual studies.