Direct realism in speech perception: Perceiving vocal tract actions and postures

Catherine Best, MARCS Auditory Laboratories, UWS

(presentation for HCSNet Perception-Action Workshop, UWS/Bankstown, Friday July 11, 2008)

Communication by speech is so obviously wrapped around acoustic signals that a corollary assumption has gone largely unquestioned, specifically that speech perception is auditory. Combine this auditory assumption with the principles of information processing, computational modelling, statistical tracking, exemplar-trace and/or neuro-cognitive theoretical perspectives, and the conclusion seems inescapable: Speech perception must be accomplished by the registration, processing, representation and retrieval of acoustic cues and patterns.

But if we step back, is the story just so? A fundamental limitation of auditory views is that they generally leave the crucial relationship between speech production and perception "missing in action." Most have steadfastly ignored production. In rare mentions of it, associationist learning of acoustic-to-production correlations have been invoked (FLMP, WRAPSA). Yet, the inadequacy of associationist accounts is underscored by decades of computational failure at solving the inverse mapping problem between acoustics and articulation. That, plus the rapidity with which infants learn to speak the ambient language, offer a vital challenge to auditory views.

Alternative views take account for the fact that speech not only involves an acoustic signal, but is generated by the actions and postures of the human vocal tract, yielding not only acoustic but visible, kinesthetic/proprioceptive, and even haptic information. Thus, speech is coherently multi-modal (or amodal), an observation consistent with the fact that normal infants learn speech in face-to-face interactions, and cannot do so from purely-auditory inputs.

I will discuss the two main models of speech perception that consider speech production: 1) the Motor Theory (Liberman and colleagues); and 2) the ecological, direct realist approach (Best, C. Fowler), founded on Gibson's view of perception. Similarities and differences of the two models will be considered in the context of evidence on speech perception by deaf and blind perceivers, on visual influences in speech perception (speech in noise, McGurk effects) even in infants, on the influence of rarely-experienced haptic information on speech perception, and on mirror neurons.