Speech production and articulatory phonetics

Bryan Gick

The complementary themes of physical embodiment and environmental embedding, though frequently touted as bases of human cognition, have not been recognized explicitly in models of phonetics. We explore these concepts more fully as they apply to phonetics, seeking a theory based on richer, higher-dimensional models of both the human body and the human environment. In contrast with previous models, we pursue not an inventory of primitive “sounds”, but rather an inventory of highly specialized body structures, each of which is defined according to its specific function at the interface between the physical organism and its richly multisensory interactive environment. Simulations and experiments will be explored with implications relating to these two themes.

While phonetic theories have traditionally made reference to anatomical terms like “lips” and/or derivative features like [labial], such terms have remained physically and neurophysiologically undefined. Using biomechanically realistic models without predefined articulators in the ArtiSynth platform (www.artisynth.org [e.g., Fels et al. 2003 ICPhS 15. 179-184; Gick et al. 2014 Comp. Meth. Biomech. & Biomed. Eng.: Imag. & Vis. doi: 10.1080/21681163.2013.851627]), simulations and experimental results show that labial speech movements may be viewed as the output of independent neuromuscular “modules” that emerge through use as a learner optimizes the biomechanics of speech production [see Gick & Stavness 2013 Front. Psych.: Cog. Sci. 4, 977]. Embodied modules of this kind constitute a long-sought-after set of discrete and controllable “body parts” our nervous systems can viably define and deploy [e.g., Bizzi & Cheung 2013 Front. Compu. Neurosci. 7: 51]. Such a model offers a real possibility of closing the loop between recently observed speech movement primitives [Ramanarayanan et al. 2013 J. Acoust. Soc. Am. 134. 1378-1394] and newly discovered cortical regions associated with speech production [Bouchard et al. 2013 Nature 495, 327–332]. Additional papers [Li, Honda, Wei & Dang; Chiu & Gick] in this session both describe techniques for 3D FEM-based modeling of the lips, with Chiu & Gick additionally reporting an experiment looking at whether speech modules may be decoupled into lower-level primitives using auditory startle.

Embodied neuromuscular primitives for speech cannot exist without being embedded in an interactive environment; that is, the only structures that are expected to emerge are those that are useful (i.e., that provide consistent sensory feedback through use) in the world. Considering sensory feedback, note that the simple act of closing the lips in speech presents both producer and perceiver with a flood of sensory information: auditory, visual, pressure-tactile, aerotactile, proprioceptive, and so on. Consistent with this view, Ghosh et al. [2010 J. Acoust. Soc. Am. 128, 3079-3087] show that listeners with higher acuity in either auditory or pressure-tactile perceptual modalities are predictably better at producing an audible contrast between the sibilants ‘s’ and ‘sh’. In the context of the long-standing controversy between articulatory and auditory-perceptual "targets" in speech, a reasonable conclusion is that sibilants occupy both articulatory and auditory-perceptual "target space" [see Perkell 2012 J. Neuroling. 25, 382-407]. A still stronger hypothesis could contend that speech draws upon all relevant perceptual modalities, not just articulatory or auditory-acoustic ones, casting speech events in a richly multidimensional space. A follow-up study [Francis et al. 2012 Can. Acoust. 40, 22-23] tests this strong hypothesis by extending the Ghosh et al. [2010] approach to include a test of speakers’ acuity in a third dimension – aerotactile [Gick & Derrick 2009, Nature 462. 502-504] – by measuring how well speakers can distinguish the sensation of air flow across the tongue. An additional paper [Gluth & Hoole] in this session describes a further extension of this paradigm to include the visual modality. A final paper [Niziolek, Nagarajan & Houde] provides an example of the perception-production loop, at the interface between embodiment and embedding.