Formulating the earliest dynamic aspects of phonological acquisition

Andrew R. Plummer

Speech dynamics encompasses the production, transmission, and perception of signals among a group of social agents who employ articulatory configuration categories (e.g., combinations of phonational, mandibular, lingual, etc. gestures), and sequential compositions of such categories, in the production of signals for transmission and in the parsing of transmitted signals during perception. Dynamic characterizations of speech signals that map to and from articulatory configuration categories, and sequences of such categories, are formulated with respect to one (or more) types of parameter spaces designed to capture relevant aspects of signal production and perception – e.g., an "articulatory space" where degrees of freedom correspond to possible articulatory postures during production and trajectories through the space correspond to coordinated motions of the articulators (see Saltzman & Munhall, 1989; Browman & Goldstein, 1990, inter alia), an "auditory space" modeling signal processing capacities of the auditory system and its sensitivity to different types of temporal aspects of processing (see Moore, 2012, for a review), or an "acoustic space" where parameters encode acoustic properties of signals and trajectories are used to model the transitional aspects of the properties over signal duration (see Zahorian & Jagharghi, 1993; Fox & Jacewicz, 2009, inter alia).

While much insight into speech dynamics has been achieved using these types of formulations, there are still enormous gaps in understanding how these largely incommensurate parameter spaces are organized and coordinated within an individual agent as that agent develops from an infant into a member of a speech community with adult-like perception and production capabilities. Moreover, conceptualization of how mappings from (sequences of) articulatory configuration categories to (coordinated) parameter spaces emerge within social agents, both at the community level and during ontogeny, in the face of massive disparities in parameter space representations of vocalizations of different social agents has barely begun to take shape. These lacunae are especially prominent in the case of an infant agent interacting with caretaker agents. A number of models have been put forward that attempt to shed light on these issues (see Guenther, 1995; Howard & Messum, 2011; Rasilo et al., 2013; Warlaumont et al., 2013, inter alia). However, they model the coordination of parameter spaces in terms of direct mappings between them, making it difficult to incorporate the influence of the dynamics of both the infant's physiological growth and their social interactions with caretakers on category emergence over the course of development.

In this presentation, we focus on the emergence of vowel categories during ontogeny, and take as our point of departure the conceptual framework put forward in Plummer & Beckman (2015). Within this framework vowel categories emerge as a set of culture-specific mappings that infants use to relate parameter space representations of vocalizations from different talkers with affective and affiliative information. Within the corresponding computational modeling architecture (Plummer et al., 2010; Plummer, 2014), infants generate structures, called "manifolds," over parameter space representations of their productions and those of their caretakers, and "align" the structures based on affiliative information in

vocal interaction with caretakers, in order to generate the culture-specific mappings that relate the differing parameter space representations. We formulate a dynamic extension to the modeling architecture as a first step toward building a model capable of accounting for the the following phenomena: the intra-modal and inter-modal dynamics of vowel posture target gestures that infants begin to establish in the first two to five months of life, the intra-modal and inter-modal dynamics of sequences of two vowel postures that infants begin to establish in the canonical babble and variegated babble periods, and the inter-talker dynamics that infants and caretakers establish in coordinating their vowel postures, or their CV gestural configurations, at both stages.

References

Browman, C., & Goldstein, L. (1990). Gestural specification using dynamically-defined articulatory structures. *Journal of Phonetics*, *18*(3), 299–320.

Fox, R., & Jacewicz, J. (2009). Cross-dialectal variation in formant dynamics of American English vowels. *Journal of the Acoustical Society of America*, *126*, 2603–2618.

Guenther, F. H. (1995). Speech sound acquisition, coarticulation, and rate effects in a neural network model of speech production. *Psychological Review*, *102*, 594–621.

Howard, I. S., & Messum, P. (2011). Modeling the development of pronunciation in infant speech acquisition. *Motor Control*, *15*, 85–117.

Moore, B. C. J. (2012). *An Introduction to the Psychology of Hearing*. Emerald, sixth ed.

Plummer, A. R. (2014). *The Acquisition of Vowel Normalization: Theory and Computational Framework*. Ph.D. thesis, The Ohio State University.

Plummer, A. R., & Beckman, M. E. (2015). Framing a socio-indexical basis for the emergence and cultural transmission of phonological systems. *Journal of Phonetics*, *53*, 66–78.

Plummer, A. R., Beckman, M. E., Belkin, M., Fosler-Lussier, E., & Munson, B. (2010). Learning speaker normalization using semisupervised manifold alignment. In *Proceedings of INTERSPEECH 2010*. Tokyo.

Rasilo, H., Räsänen, O., & Laine, U. K. (2013). Feedback and imitation by a caregiver guides a virtual infant to learn native phonemes and the skill of speech inversion. *Speech Communication*, *55*(9), 909–931.

Saltzman, E., & Munhall, K. G. (1989). A dynamical approach to gestural patterning in speech production. *Ecological Psychology*, *1*, 333–382.

Warlaumont, A. S., Westermann, G., Buder, E. H., & Oller, D. K. (2013). Prespeech motor learning in a neural network using reinforcement. *Neural Networks*, *38*, 64–75.

Zahorian, S. A., & Jagharghi, A. J. (1993). Spectral-shape features versus formants as acoustic correlates for vowels. *Journal of the Acoustical Society of America*, *94*, 1966–1982.