Aspects of modeling the learning of vowel normalization

We put forward a framework for the investigation of the learning of vowel normalization based on the idea that infants perform abstractions over their psychophysical representations of the vowels of individual speakers by mapping them to mediating spaces of representations, guided by vocal imitative interaction with their caretakers, as a first step in the phonological acquisition process. The framework is accompanied by a computational methodology for modeling the abstraction, which involves the "alignment" of cognitive structures, called "manifolds," that an infant builds from the psychophysical representations of the vowels of individual speakers. We conclude with a simple demonstration of the main algorithm involved in implementation of the methodology.

We take *vowel normalization* to be a cognitive process "in which interspeaker vowel variability is reduced in order that perceptual vowel identification may then be performed by reference to relative vowel quality rather than absolute [psychophysical] parameters of vowels" (Johnson, 1990, p. 230). In this connection, we take the following to be a minimal collection of aspects essential to the modeling of the learning of normalization. The first is a "reference frame," which "can be thought of as a coordinate frame that best captures the form of information represented in a particular part of the nervous system" (Guenther, 2003, p. 209), though construed more broadly to include physical and cognitive domains (à la Saltzman, 1995). The second essential aspect is that of a "manifold," a structure embedded within a reference frame and used to organize representations. Specifically, we make use of "vowel manifolds" (Jansen & Niyogi, 2006, 2007), physical structures over vowel signals hypothesized to motivate an infant's formation of "perceptual manifolds" (Seung & Lee, 2000; Niyogi, 2004), and "cognitive manifolds" (Plummer, 2012), cognitive structures used by an infant in the normalization process. The third essential aspect is a computation over manifolds, called "manifold alignment" (Wang, 2010), which maps representations on two (or more) manifolds to a mediating "latent space" (Ham et al., 2005; Ma & Fu, 2012), where vowel identification may take place, or further cognitive computations. Alignment computations are guided by social interaction between an infant and adult caretakers characterized by specific types of vocal exchanges (Howard & Messum, 2011; Masataka, 2003; Fitch, 2010; Gros-Louis et al., 2006; Goldstein & Schwade, 2008). These exchanges broadly involve: (i) structured turn-taking between an infant and caretakers (Masataka, 2003), and (ii) caretaker responses differentiated according to the nature of infant vocalizations (Gros-Louis et al., 2006; Goldstein & Schwade, 2008).

We briefly exemplify the approach with the following simplified dramaturgical dyadic exchange. Let V_I be a set of representations of vowels derived from an infant, and V_A a set derived from an adult caretaker, both within a single acoustic reference frame (Figure 1). The adult may impart their systematic knowledge of the vowel categories [i], [u], and [a] to the infant by responding in a positive manner to infant productions in V_I judged by the adult to be good examples of [i], [u], and [a], respectively, with their own productions taken to be good examples of [i], [u], and [a]. The yellow, orange, and green points, respectively, in Figure 1 approximate a series of vocal exchanges involving good examples of infant (left) and adult (right) [i], [u], and [a], as judged by the adult. The infant may then represent the positive interaction, pairing representations of their good productions with the corresponding representations of the positive adult responses. These pairs of positive representations guide the alignment of the manifolds the infant constructs over V_I and V_A, yielding the aligned structures in Figure 2, which may then be used for vowel identification, as well as for further cognitive computation.

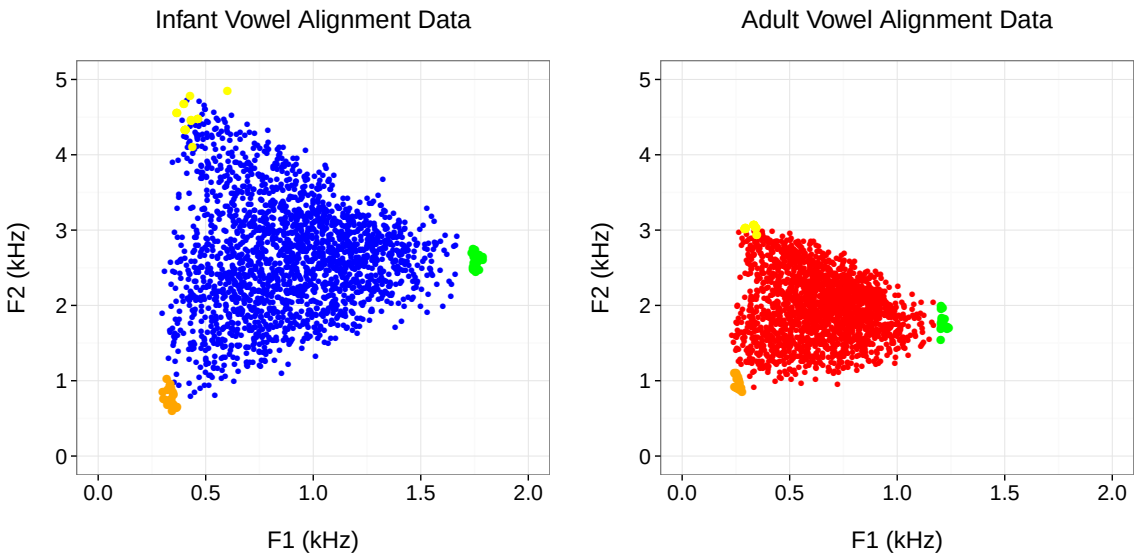

Figure 1: Infant vowel representations V_I (left) and adult vowel representations V_A (right), in a formant-based acoustic reference frame, together with "good" examples of infant and adult [i] (yellow), [u] (orange), and [a] (green) as judged by the adult.

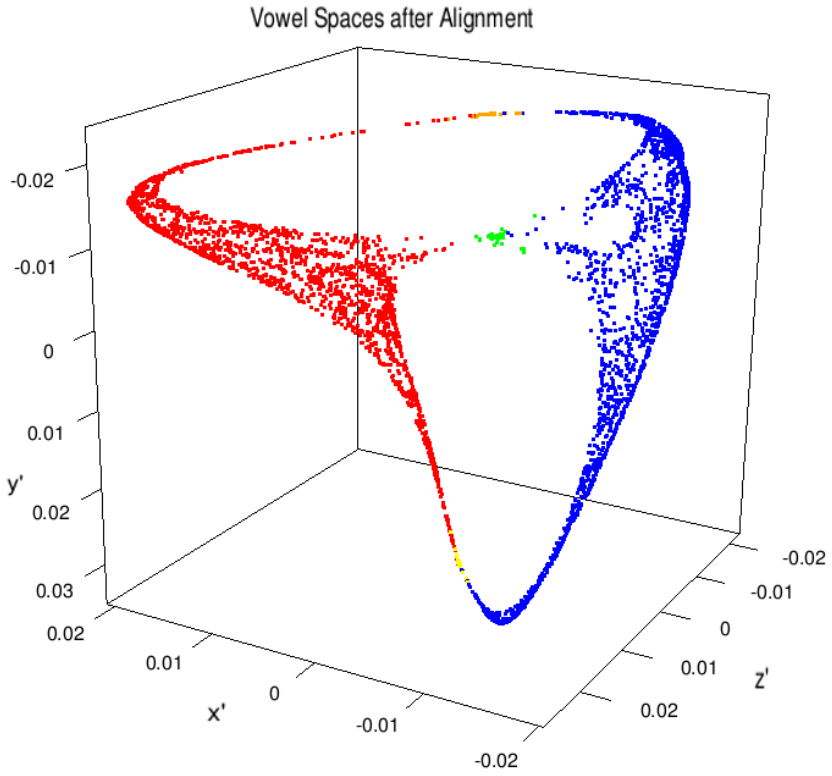

Figure 2: Representations in a reference frame where the alignment of manifolds over adult and infant formant representations has been achieved using the "good" infant and adult productions.

References

Fitch, W. T. (2010). *The Evolution of Language*. Cambridge University Press.

Goldstein, M. H., & Schwade, J. A. (2008). Social feedback to infants' babbling facilitates rapid phonological learning. *Psychological Science*, *19*(5), 515–523.

Gros-Louis, J., West, M. J., Goldstein, M. H., & King, A. P. (2006). Mothers provide differential feedback to infants' prelinguistic sounds. *International Journal of Behavioral Development*, *30*(5), 112–119.

Guenther, F. H. (2003). Neural control of speech movements. In N. O. Schiller, & A. Meyer (Eds.) *Phonetics and Phonology in Language Comprehension and Production: Differences and Similarites*, (pp. 209–239). Walter de Gruyter.

Ham, J., Lee, D. D., & Saul, L. K. (2005). Semisupervised alignment of manifolds. In Z. Ghahramani, & R. Cowell (Eds.) *Proc. of the Ann. Conf. on Uncertainty in AI*, vol. 10, (pp. 120–127).

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

Jansen, A., & Niyogi, P. (2006). Intrinsic fourier analysis on the manifold of speech sounds. In *in IEEE Proceedings of International Conference on Acoustics, Speech, and Signal Processing*, (pp. 241–244).

Jansen, A., & Niyogi, P. (2007). Semi-supervised learning of speech sounds. In *Proceedings of INTERSPEECH 2007*.

Johnson, K. (1990). Contrast and normalization in vowel perception. *Journal of Phonetics*, *18*, 229–254.

Ma, Y., & Fu, Y. (2012). *Manifold Learning Theory and Applications*. CRC Press.

Masataka, N. (2003). *The Onset of Language*. Cambridge, UK: Cambridge University Press.

Niyogi, P. (2004). Towards a computational model of human speech perception. In *Proceedings of the Conference on Sound to Sense, MIT (In Honor of Ken Stevens' 80th birthday)*.

Plummer, A. R. (2012). Aligning manifolds to model the earliest phonological abstraction in infant caretaker vocal imitation. In *13th Annual Conference of the International Speech Communication Association (INTERSPEECH 2012)*. Portland, OR.

Saltzman, E. (1995). Dynamics and coordinate systems in skilled sensorimotor activity. *Mind as motion: Explorations in the dynamics of cognition*, (pp. 149–173).

Seung, H. S., & Lee, D. D. (2000). The manifold ways of perception. *Science*, *290*(5500), 2268–2269.

Wang, C. (2010). *A Geometric Framework For Transfer Learning Using Manifold Alignment*. Ph.D. thesis, University of Mass. Amherst.