Manifold Alignment, Vocal Imitation, and the Perceptual Magnet Effect

Andrew R. Plummer
The Ohio State University, Columbus, OH, USA

Introduction

Auditory representations differ

- People who have different vocal tracts have different vocalizations.
- Vocalizations of different talkers are represented differently even after applying auditory models (e.g., Moore et al., 1997) to a spectrum.

Computation of equivalence classes of representations

- Humans are able to impose equivalence classes over these differing representations, providing a basis for high-fidelity communication.
- This ability is apparent very early in infancy, demonstrated by the nature of the vocal exchanges between infants and their caretakers by four months of age (Masataka, 2003; Fitch, 2004, 2010).

Equivalence classes are language-specific, and interactively constructed

- The perceptual magnet effect (Kuhl, 1991; Guenther & Gjaja, 1996) suggests that computation of equivalence classes is sensitive to the ambient language.
- The influence of vocal imitation suggests that the equivalence classes are constructed via social interaction (Kuhl & Meltzoff, 1996; Masataka, 2003).

Objects and Aims

- We limit our inquiry to vowels, and take vowel normalization to be a cognitive process which yields equivalence classes of auditory representations of vowels.
- We briefly lay out a general approach to the theory of the acquisition of vowel normalization during infancy, along with a computational modeling framework.

Aspects of the Theory

Infant's construction of a model "self"

- An infant constructs/refines an internal model (Wolpert, Ghahramani, & Jordan, 1995) over articulatory and auditory representations.
- Construction of the self involves cross-modal abstraction over representations yielded by the distinct modalities (Davenport, 1976; Kuhl & Meltzoff, 1982).

Infant's construction of a model of "others" and alignment with the self

- Humans "trace patterns upon" their internal representations of the external world (Lippmann, 1922; on James, p. 16), which includes "other" humans.
- Relating the patterns imposed on other conspecifics to those imposed on the self enables "social learning" (as in Meltzoff's (2007) "like-me" framework).

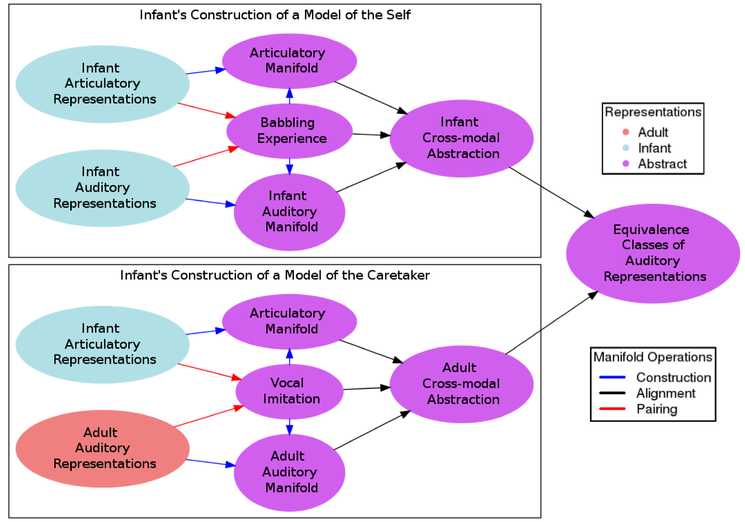
Infant's organization of sensory information using cognitive manifolds

- A cognitive manifold describes what our brains might know about something that is very complex and multi-dimensional by building a much lower-dimensional "map" of it.
- For example, a map of the world is a two-dimensional manifold built to describe what we need to know to navigate the three-dimensional surface of our planet.

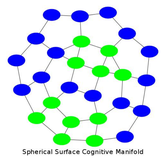
Key Theoretical Claims

- Infants construct cognitive manifolds – During the earliest stage of spoken language acquisition, infants construct cognitive manifold representations over their own vowel productions, and those of their caretakers.
- Vowel normalization is manifold alignment – The computation of equivalence classes of auditory representations of different talkers, including those of an infant learner, is the alignment of cognitive manifolds constructed by the infant.
- Vocal imitation guides alignment – The cognitive manifold alignment is guided by vocal imitative exchanges between infants and caretakers, with the perceptual magnet effect as a consequence of the exchanges.

System Architecture



Cognitive Manifolds



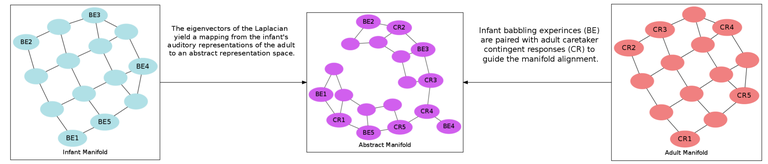
- Cognitive manifolds are weighted graphs whose nodes encode sensory info.
- Organizational aspects of the information is encoded in the weighted edges.
- Mappings on manifolds can be learned using their graph Laplacians, which are operators derived from edge weights.

Manifold Alignment: Perceptual Magnet Example

- Let E_I and E_A denote sets of infant and adult caretaker auditory representations and denote imitation pairing as $\chi_{im} : E_I \times E_A \to \{0, 1\}$. In the depiction below,

$$\chi_{im} = \{\langle BE1, CR1\rangle, \langle BE2, CR2\rangle, \langle BE3, CR3\rangle, \langle BE4, CR4\rangle, \langle BE5, CR5\rangle\}.$$

- The infant manifold $M(E_I)$ and adult manifold $M(E_A)$ are aligned (Ham et al., 2005) by combining their graph Laplacians, using the pairs χ_{im}.
- The alignment process yields mappings from $M(E_I)$ and $M(E_A)$ to a space of abstract representations of those in E_I and E_A such that the abstract representations of the points in each pair in χ_{im} are close to each other.



Expansion to Cross-modal Modeling

- Let A_I denote the set of infant articulatory representations corresponding to the infant auditory representations in E_I, and $M(A_I)$ an articulatory manifold.
- Denote cross-model pairing as $\chi_{cm} : E_I \times A_I \to \{0, 1\}$, and let

$$\chi_{cm} = \{\langle BE1, AR1\rangle, \langle BE2, AR2\rangle, \langle BE3, AR3\rangle, \langle BE4, AR4\rangle, \langle BE5, AR5\rangle\}.$$

- The infant constructs a model self by aligning $M(E_I)$ and $M(A_I)$, using the pairs χ_{cm}, mapping points in E_I and A_I to an abstract cross-modal space C_I.
- The infant constructs a model of the adult caretaker by aligning $M(E_A)$ and $M(A_I)$, using the pairs χ_{cm}, mapping points in E_A and A_I to an abstract cross-modal space C_A.
- Let $C_I(\mathbf{e}_I)$ denote the abstract representation of $\mathbf{e}_I \in E_I$, and $C_A(\mathbf{e}_A)$ that of $\mathbf{e}_A \in E_A$. Finally, $M(C_I)$ and $M(C_A)$ are aligned using the pairs

$$\chi_{abs} = \{\langle C_I(BEi), C_A(CRi)\rangle \mid i = 1, \ldots, 5\}.$$

Conclusions and Acknowledgments

- We have put forward a potentially useful conceptualization and computational framework for the investigation of vowel normalization in infants that incorporates a wide array of related phenomena.

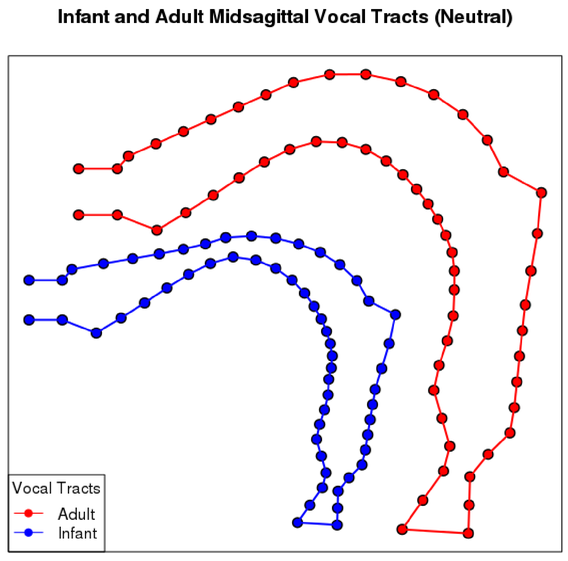
- The author wishes to thank Mary Beckman, Eric Fosler-Lussier, Misha Belkin, William Schuler, and Pat Reidy for their contributions this project.
- Work supported by NSF grants BCS 0729306 (to Mary Beckman) and BCS 0729277 (to Benjamin Munson), and by an OSU Center for Cognitive Science seed grant (to Mary Beckman, Mikhail Belkin, & Eric Fosler-Lussier).

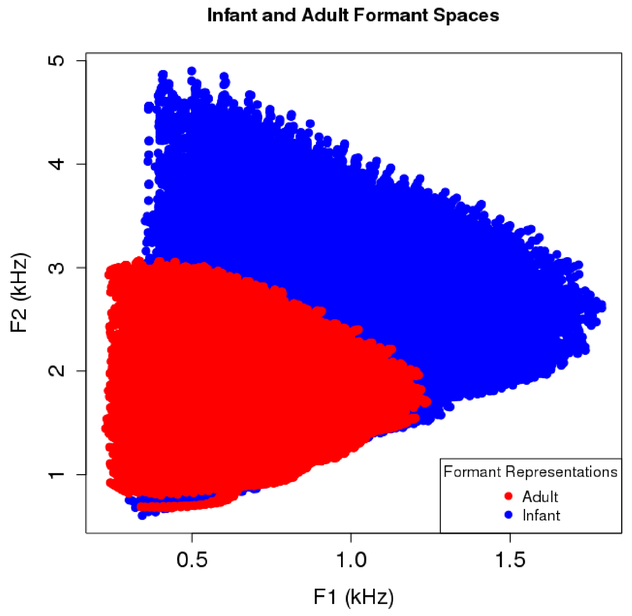
Mail: plummer@ling.ohio-state.edu

Data and Figures

Articulatory and Acoustic Data

- We use Maeda's and Boe's (1997) Variable Linear Articulatory Model (VLAM) to model the vowel productions of an infant and an adult caretaker.
- We use the 6 month-old setting of the VLAM as our model infant, and the 10 year-old setting for the adult caretaker (as it was perceived to be most similar to a young female adult in a cross-language perception study (Munson et al., 2010)).





Representations

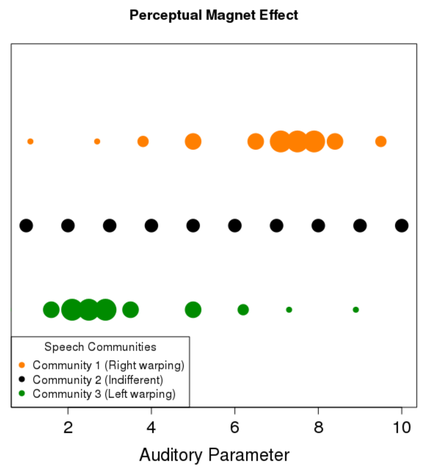
- The articulatory representations (1, left) are principal component vector arguments for the VLAM, each yielding a vowel signal with a formant representation (1, middle).

$$\mathbf{a} = \langle a_1, \ldots, a_7 \rangle, \quad \mathbf{f} = \langle f_1, f_2, f_3 \rangle, \quad \mathbf{e} = \langle e_1, \ldots, e_{361} \rangle \qquad (1)$$

- The auditory representations (1, right) are "excitation patterns" derived from the vowel signals using the transformations described in Moore et al. (1997)

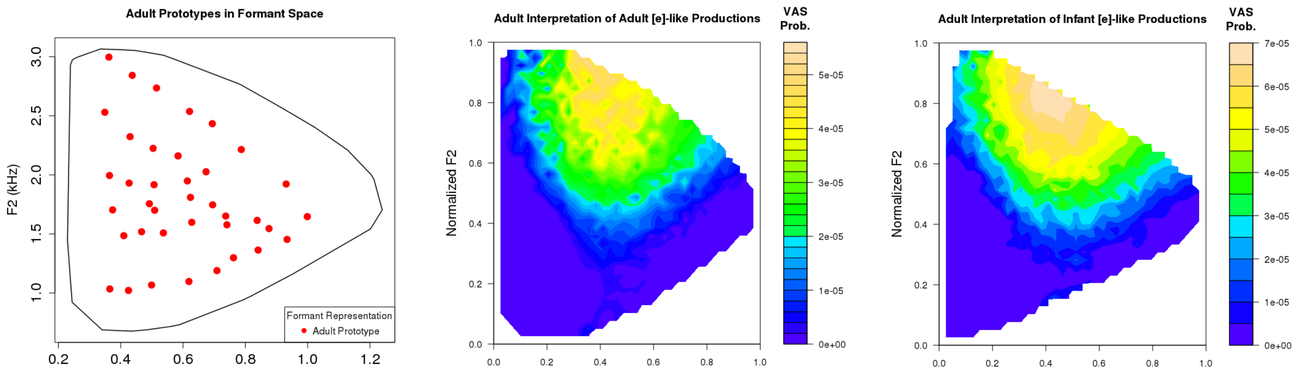
Perceptual Magnet Effect

- The perceptual magnet effect (Kuhl, 1991) names the phenomenon wherein the perception of a vowel in a given language is influenced by "perceptual magnets" located in a perceptual metric space over representations of vowels in that language.

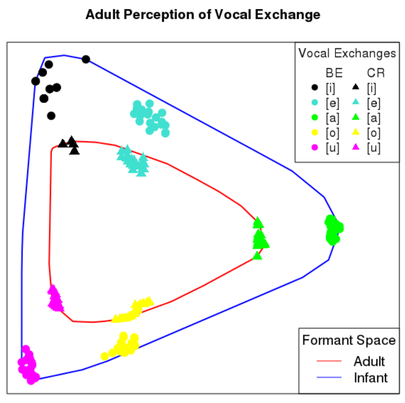


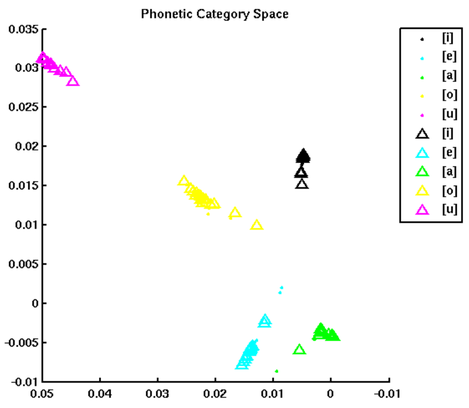
Adult Perception of Infant Vocalizations, and Contingent Responses

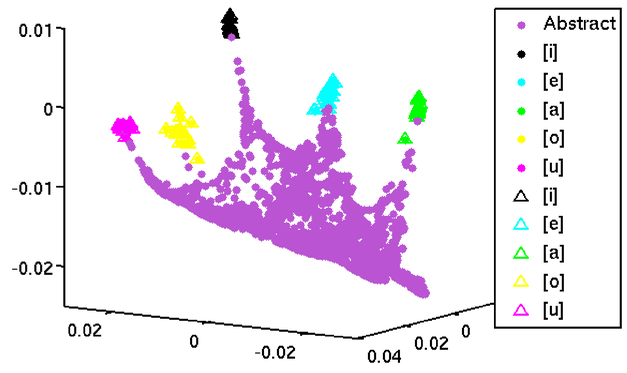
- We use 38 perceptually categorized vowel stimuli (Munson et al. 2010) generated by the VLAM set at 10 years of age to model the vowel categories of an "average" Greek-speaking caretaker.
- We additively interpolate the categorizations over the formant space of the caretaker and "project" the categories to the infant's acoustic space, modeling the caretaker's interpretations of infant vocalizations.



Simulation Results







- In response to an infant's vocalization, the caretaker interprets and responds, yielding vocal imitation pairs (left) that guide manifold alignment.
- Cross-modal manifold alignment yields a basis for vowel categorization (middle), while the intra-modal alignment (right) yields perceptual warping.

Mail: plummer@ling.ohio-state.edu WWW: http://www.learningtotalk.org

Manifold alignment, vocal imitation, and the perceptual magnet effect

Andrew R. Plummer

The Ohio State University

plummer@ling.ohio-state.edu

Poster Presented at the International Child Phonology Conference, June 5, 2012

References

Boë, L.-J. and Maeda, S. (1997). Modélization de la croissance du conduit vocal. Éspace vocalique des nouveaux-nés et des adultes. Conséquences pour l'ontegenèse et la phylogenèse. In *Journée d'Études Linguistiques: "La Voyelle dans Tous ces États"*, pages 98–105. Nantes, France.

Davenport, R. K. (1976). Cross modal perception in apes. *Annals of the New York Academy of Sciences*, 280:143–149.

Fitch, W. T. (2004). Evolving honest communication systems: Kin selection and "mother tongues". In Oller, D. K. and Griebel, U., editors, *Evolution of Communication Systems: A Comparative Approach*, pages 275–296. MIT Press., Cambridge, Massachusetts.

Fitch, W. T. (2010). Musical protolanguage: Darwin's theory of language evolution revisited. Language Log.

Guenther, F. H. and Gjaja, M. N. (1996). The perceptual magnet effect as an emergent property of neural map formation. *J. of the Acoustical Society of America*, 100:1111–1121.

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

James, W. (1890). *The Principles of Psychology*. Macmillan, London.

Kuhl, P. K. (1991). Human adults and human infants show a "perceptual magnet effect" for the protoypes of speech categories, monkeys do not. *Perception & Psychophysics*, 50:93–107.

Kuhl, P. K. and Meltzoff, A. N. (1982). The bimodal perception of speech in infancy. *Science*, 218(4577):1138–1141.

Kuhl, P. K. and Meltzoff, A. N. (1996). Infant vocalizations in response to speech: Vocal imitation and developmental change. *Journal of the Acoustical Society of America*, 100(4):2425–2438.

Lippmann, W. (1922). *Public Opinion*. The Macmillan Company, New York, NY.

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

Meltzoff, A. (2007). The 'like me' framework for recognizing and becoming an intentional agent. *Acta Psychologica*, 124:26–43.

Moore, B. C. J., Glasberg, B. G., and Baer, T. (1997). A model for the prediction of thresholds, loudness, and partial loudness. *Journal of the Audio Engineering Society*, 45(4):224–240.

Munson, B., Ménard, L., Beckman, M. E., Edwards, J., and Chung, H. (2010). Sensorimotor maps and vowel development in English, Greek, and Korean: A cross-linguistic perceptual categorizaton study (A). *Journal of the Acoustical Society of America*, 127:2018.

Wolpert, D. M., Ghahramani, Z., and Jordan, M. I. (1995). An internal model for sensorimotor integration. *Science*, 269(5232):1880–1882.