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).

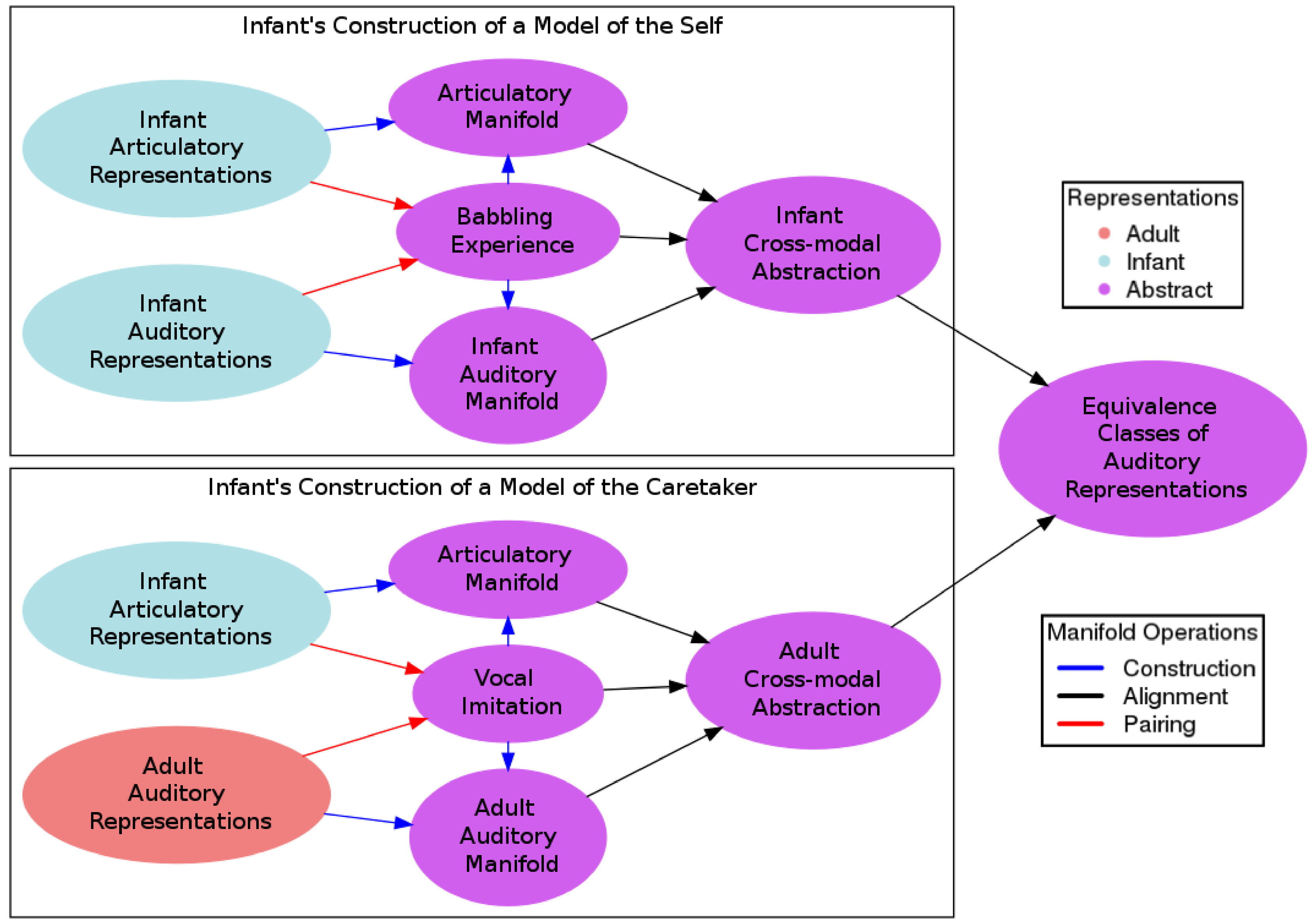
Infant's organization of sensory information using cognitive manifolds

- 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.

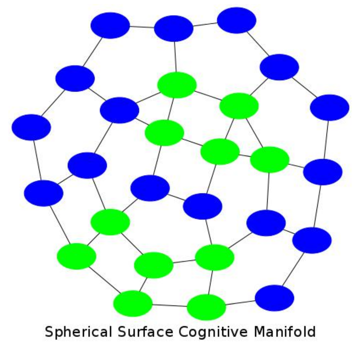
Key Theoretical Claims

- 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



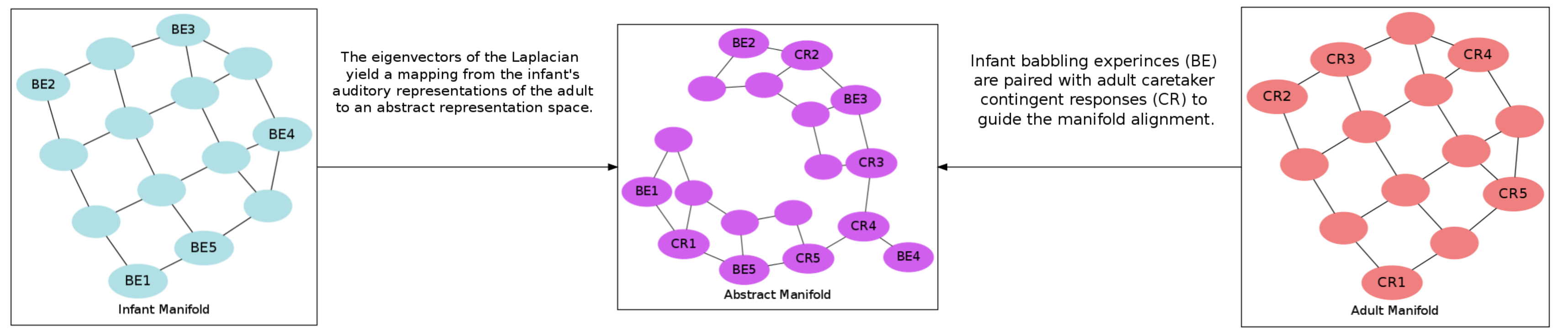
- Cognitive manifolds are weighted graphs whose nodes encode sensory info.
- Organizational aspects of the information is encoded in the weighted edges.
- Mappings on manifolds can be learned using their graph Laplacians, which are operators derived from edge weights.

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).