Aspects of Modeling the Learning of Vowel Normalization

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

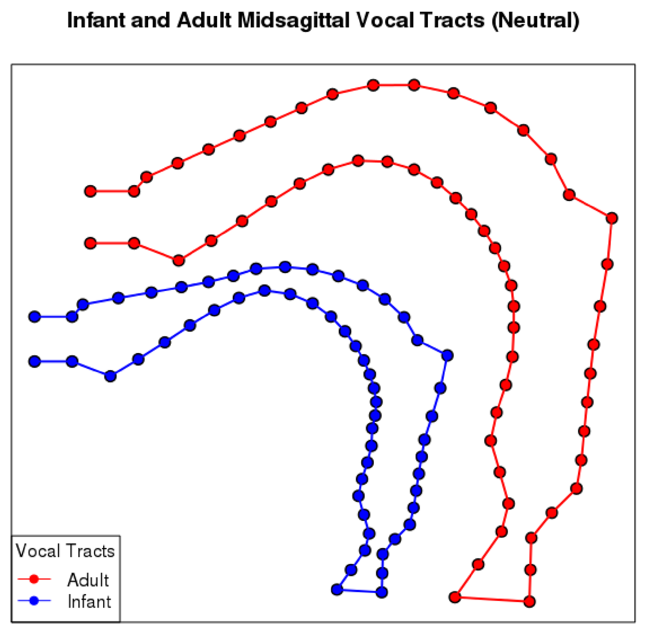
Basic Motivation

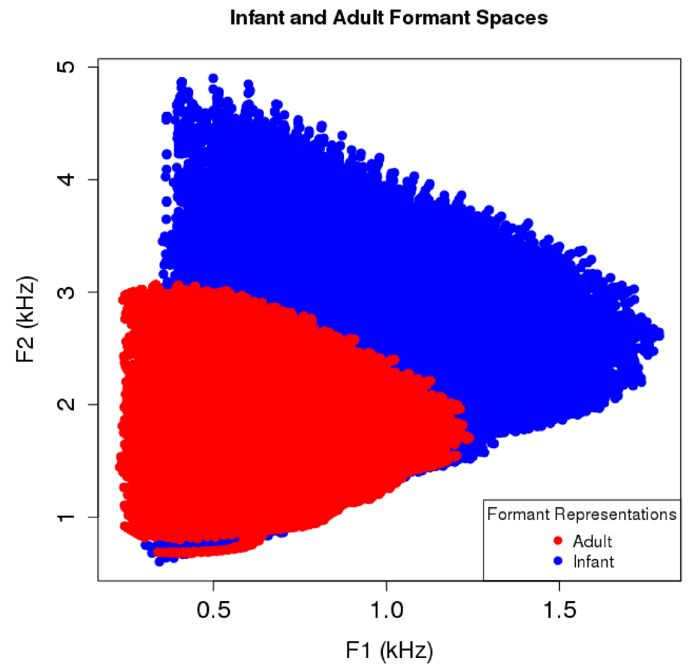
Auditory representations differ

- 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 (Kuhl, 1991; Kuhl & Meltzoff, 1996; Masataka, 2003; Fitch, 2004, 2010).





Objects and Aims

- We limit our inquiry to vowels, and take vowel normalization to be a learned cognitive process which may yield equivalence classes over psychophysical and cognitive representations of vowels.
- We proffer a framework for investigating the learning of vowel normalization based on the idea that infants perform abstractions over psychophysical and cognitive representations of vowels of individual speakers by mapping them to mediating spaces of representations, guided by vocal interaction with their caretakers.

Comparative Considerations

Complex Computational Systems for Dealing with Variation...

- All organisms are faced with the task of sorting out the bewildering amount of variation they sense in their external environments, especially that pertaining to the signals of other organisms, both non- and con-specific.
- Even bacteria possess complex intraspecies communication systems. Quorum-sensing species are equipped with structured "signal detection and relay apparatuses" that separate conspecific signals from an environment teeming with noise, signal mimics, etc. (reviewed in Waters & Bassler, 2005).

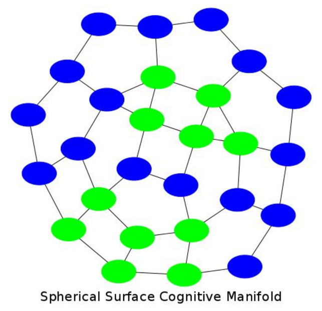
...Operating over Highly Differentiated Input

- Signals used by organisms for communication are highly differentiated, both component-wise, and by function, with the latter rarely considered in modeling.
- Young passerine songbirds acquiring song must have access to (i) the vocalizations of other conspecifics, and (ii) their own self-produced vocalizations, for acquisition to successfully occur (reviewed in Doupe & Kuhl, 1999).
- Moreover, vocalizations further differentiated by social or emotional contact affect the acquisition of song to the point where cross-fostered birds learn the songs of their foster parents, even when given (audio-visual) access to conspecific song.
- "Mental signaling" of organisms appears to substantially influence learning, e.g., through the creation of sensorimotor "internal models" (see Wolpert, et al., 1995), or the broader creation of "meaningful internal representations" not necessarily dependent on external signaling (see Harms, 2004).

Main Aspects of the Framework

- Reference Frames – Infants organize vowel phenomena within reference frames assumed to be metric spaces over psychophysical representations (inter alia).
- Infants construct cognitive manifolds – During the earliest stage of spoken language acquisition, infants construct cognitive manifolds over psychophysical representations of their own vowels, 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, involves the alignment of cognitive manifolds constructed by the infant.
- Vocal exchanges guides alignment – Cognitive manifold alignment is guided by vocal exchanges between infants and caretakers.

Reference Frames and Cognitive Manifolds



- A reference frame "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), or other physical/cognitive domain.
- Reference frames are modeled simply as metric spaces.
- A cognitive manifold describes what our minds might know about something that is very complex by multi-dimensional by building a lower-dimensional "map" of it.
- Cognitive manifolds are modeled as weighted graphs whose vertices correspond to psychophysical and cognitive representations within reference frames, and whose edges correspond to relations between representations.

Cognitive Manifold Alignment

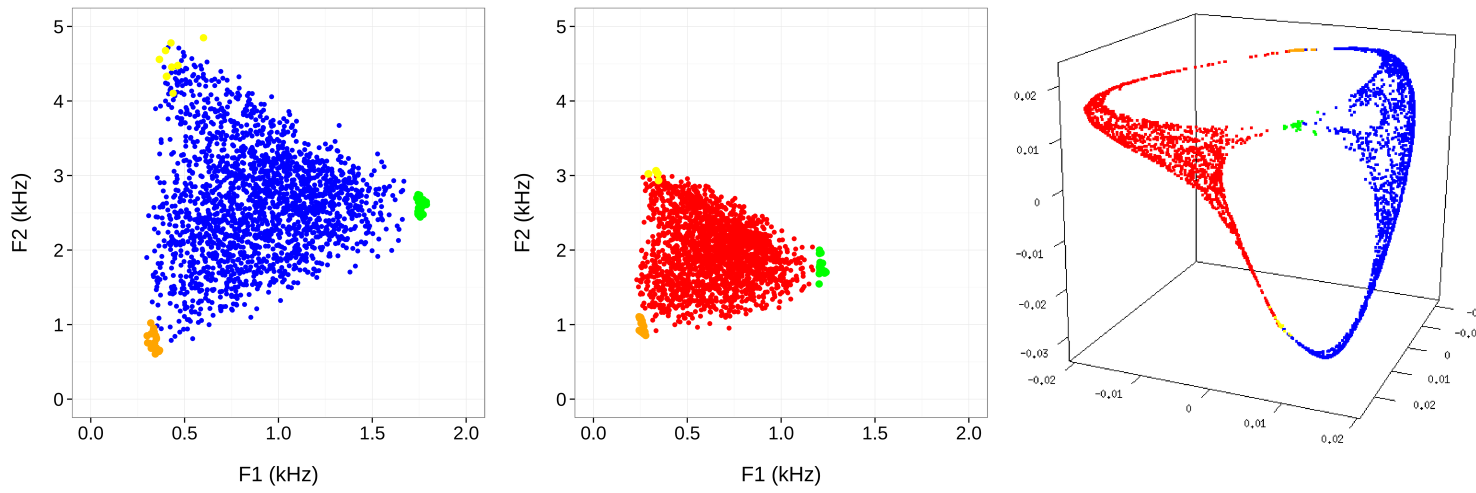


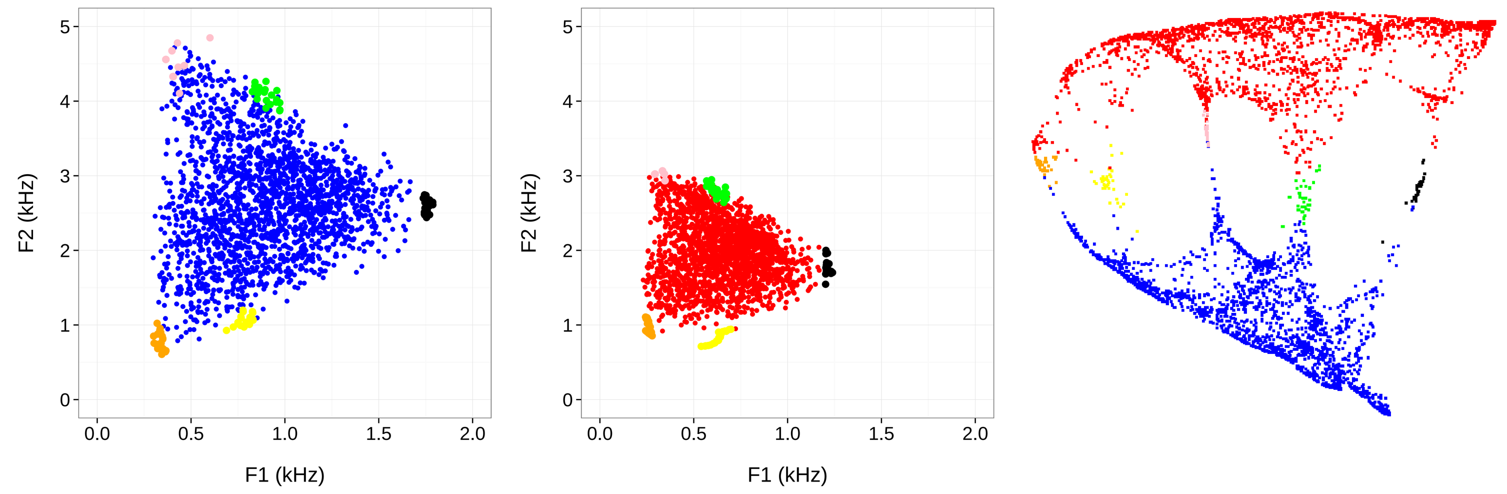
Figure: Infant vowel representations V_I (left) and adult vowel representations V_A (mid), in a formant-based acoustic reference frame, together with "good" examples of infant and adult [i] (yellow), [u] (orange), and [a] (green) as judged by the adult. (Left) Aligned representations of the representations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

- A manifold alignment can be viewed as a function that takes two (or more) disjoint manifolds and essentially links the two structures, creating a single, connected structure which facilitates the transfer of information from one to another.
- We exemplify the modeling of manifold alignment using V_I and V_A, as follows:
  - Suppose we are given two cognitive manifolds, say M_I and M_A, over V_I and V_A, respectively. Let $V_I \times V_A$ be the cartesian product of V_I and V_A, and let

$$\chi_{voc} : V_I \times V_A \rightarrow \{0, 1\}.$$

  - We construct a weighted graph M_Z whose vertices are those of M_I and M_A, and whose edges are those of M_I and M_A together with edges corresponding to the nonzero values of the vocal exchanges represented by χ_{voc}.
  - The eigenvectors of the graph Laplacian of M_Z yield the aligned representations of V_I and V_A in a new reference frame.

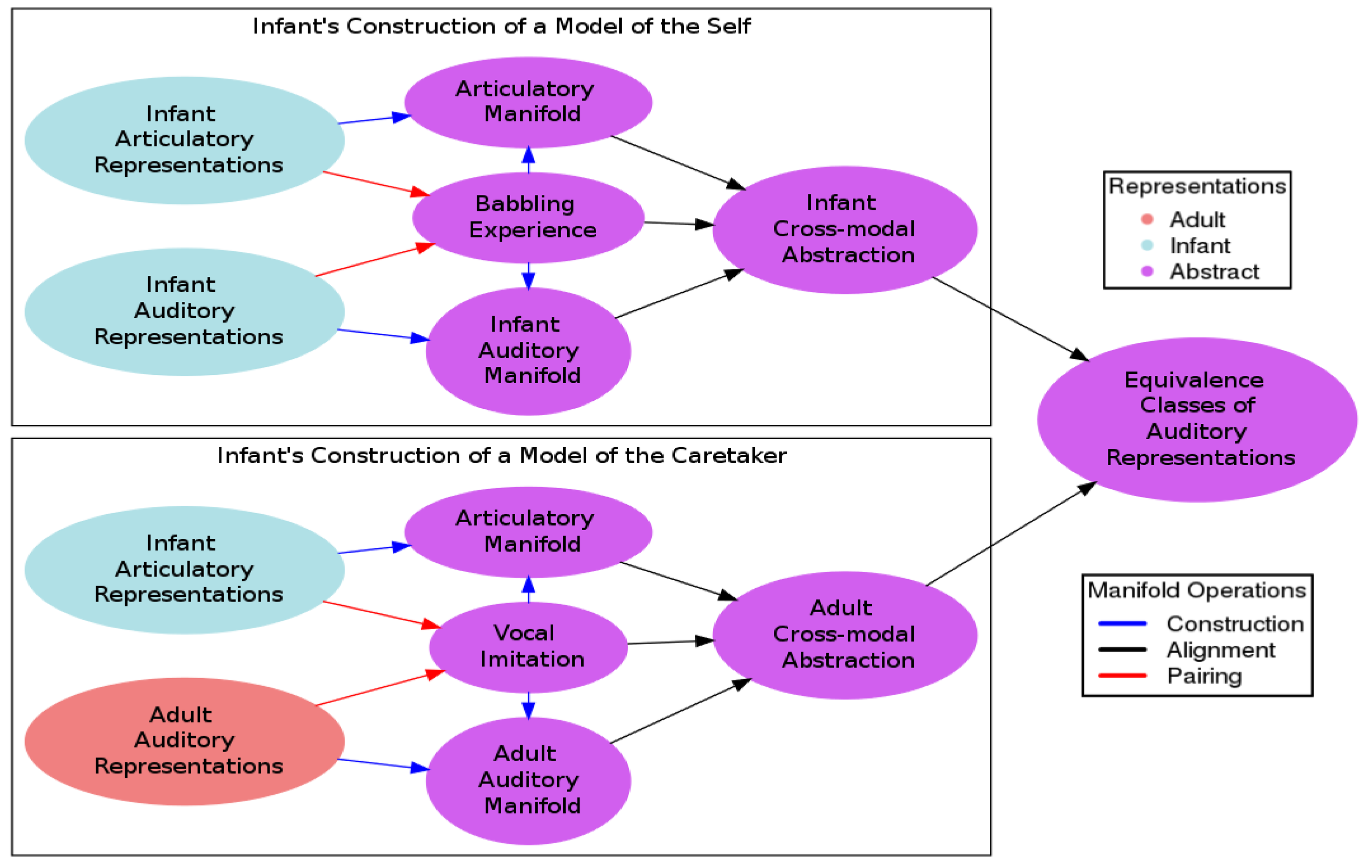
Greek Vowel System Transfer using Auditory Representations



The framework provides for investigation of different representations on the acquisition process, as well as different patterns in vocal exchange:

- The "good" examples of infant and adult [i] (pink), [e] (green), [u] (orange), [o] (yellow), and [a] (black) were provided by an adult Greek listener.
- High-dimensional representations corresponding to auditory "excitation patterns" (Moore et al., 1997) were used to achieve the alignment.

Cross-modal Abstraction and Alignment of Internal Models



The framework can accommodate more complex computations, e.g.:

- cross-modal abstractions over representations yielded by distinct modalities (Davenport, 1976; Kuhl & Meltzoff, 1982; Masataka, 2003),
- relating models imposed on other conspecifics to those imposed on the self, enabling "social learning" (as in Meltzoff's (2007) "like-me" framework).

General Summary

- Variation in signals seems to engender complex computational systems for imposing equivalence classes on collections of signals, especially those produced by conspecifics.
- In the case of vocal learning, the computational systems of certain organisms, including passerine songbirds and humans, operate over highly differentiated input, both in composition and function, some of which is characterized socially.
- Our framework makes room for the complexities of the computational system involved in vowel category learning, including variation in differentiated system input, and the ways in which it influences the learning process.

Acknowledgements

- 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).