Formulating the earliest dynamic aspects of phonological acquisition

Andrew R. Plummer

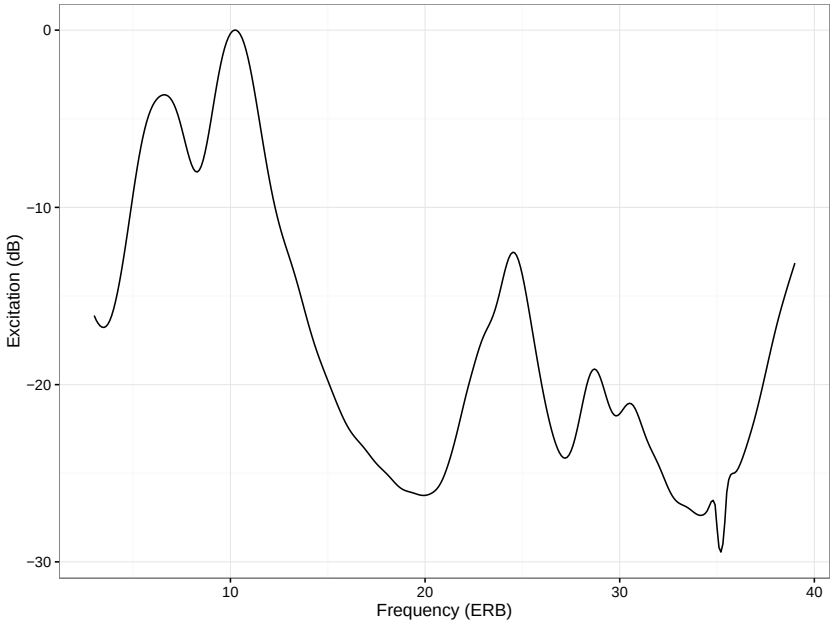
plummer.321@osu.edu

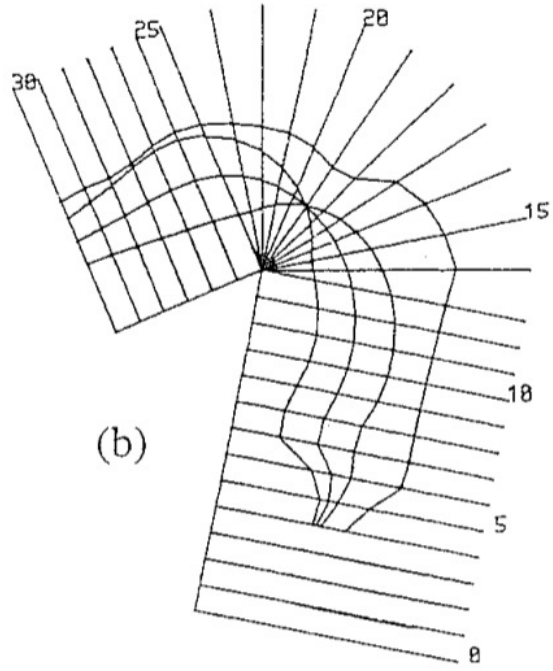
July 13, 2016

Order of the day

- How is speech dynamics typically characterized?
- What is social meaning?
- What are categories in phonology?
- How do these relate to one another?

Dynamic characterizations of speech signals

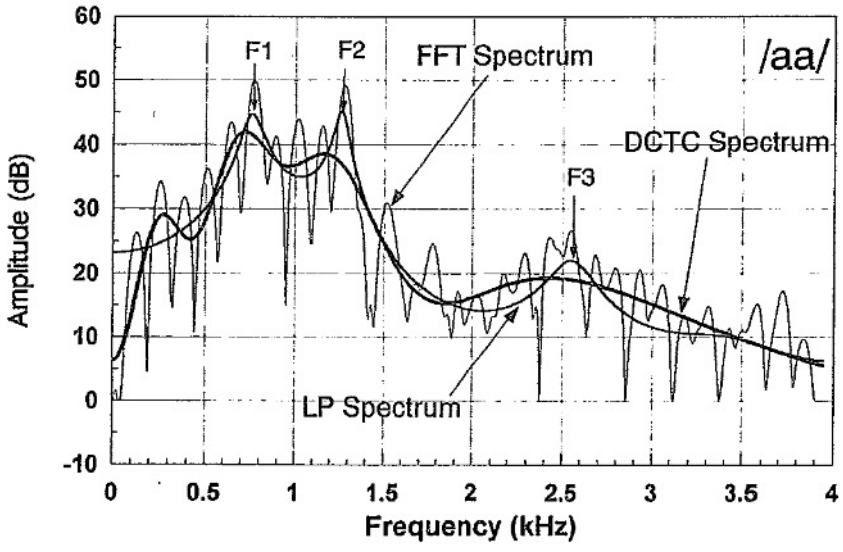


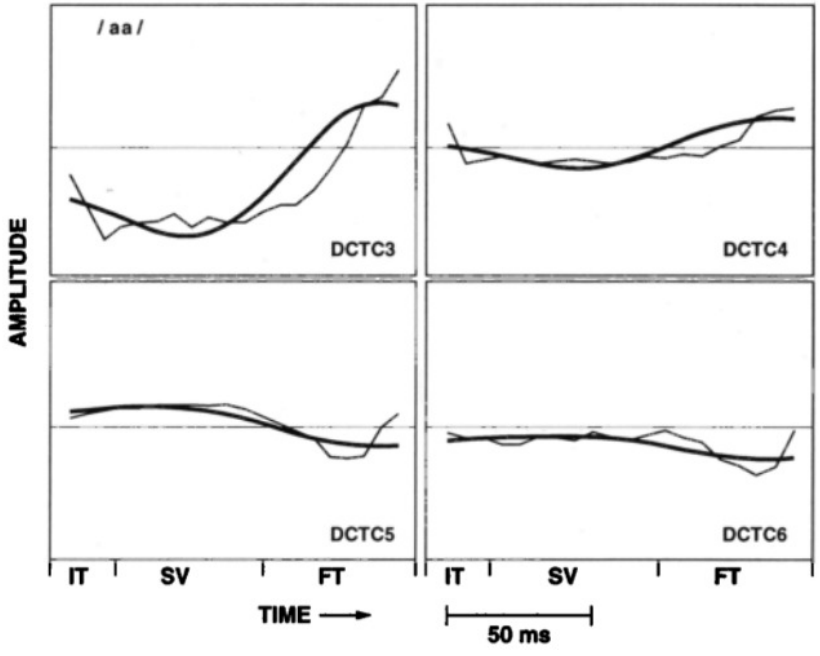


- Auditory space – modeling signal processing capacities of the auditory system and its sensitivity to different types of temporal aspects of processing (see Moore, 2012, for a review).

- 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).

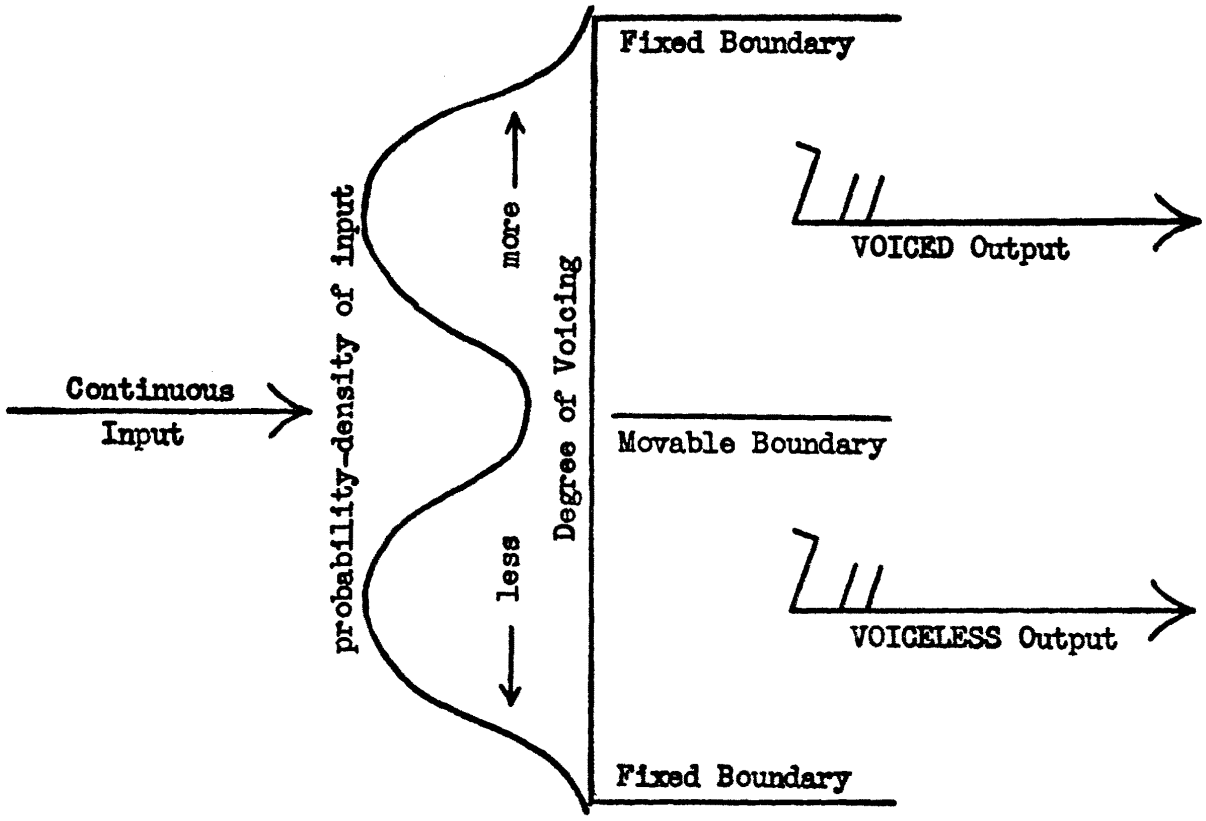
Dynamic characterizations of speech signals





- Acoustic Spaces – 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).

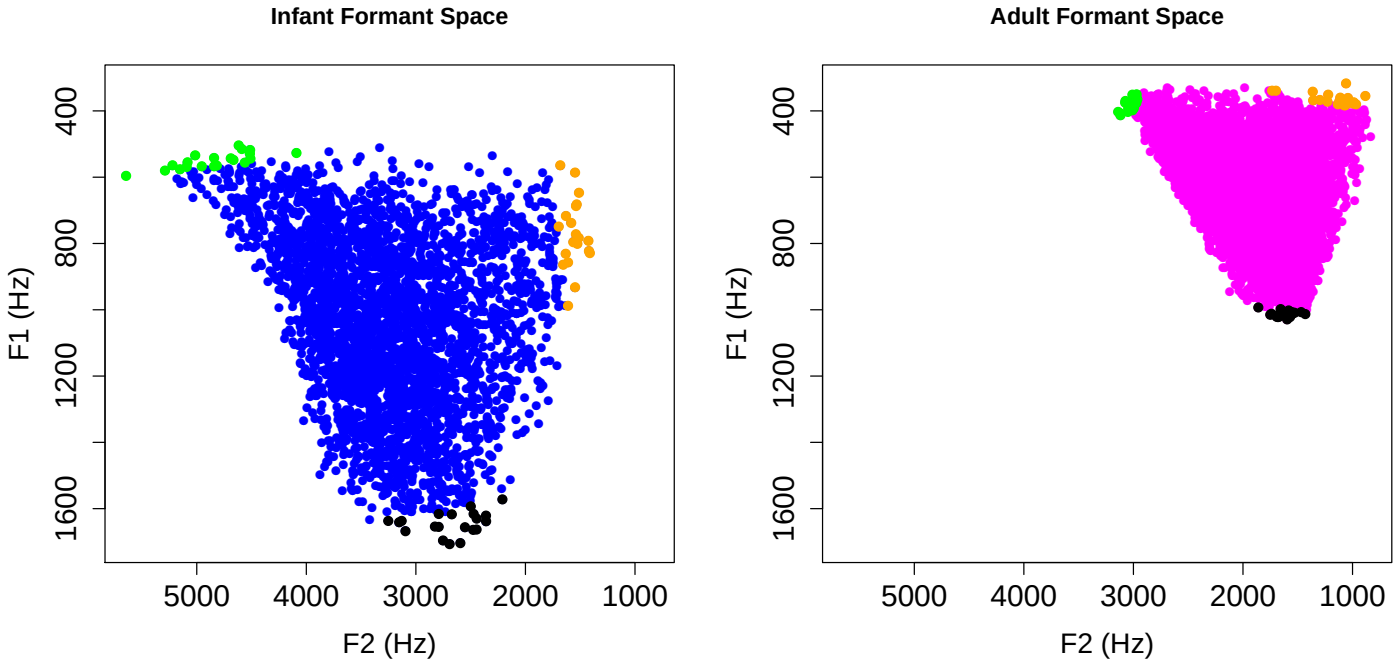
Function spaces for phonological systems



- Use of these kinds of reference frames for representing speech productions permits conceptualizing phonological systems as functions over such a space.

- For example, within Hockett's (1965) sound change model, features are projections of distributions onto axes, and phonological categories are vectors of feature values.

The normalization issue

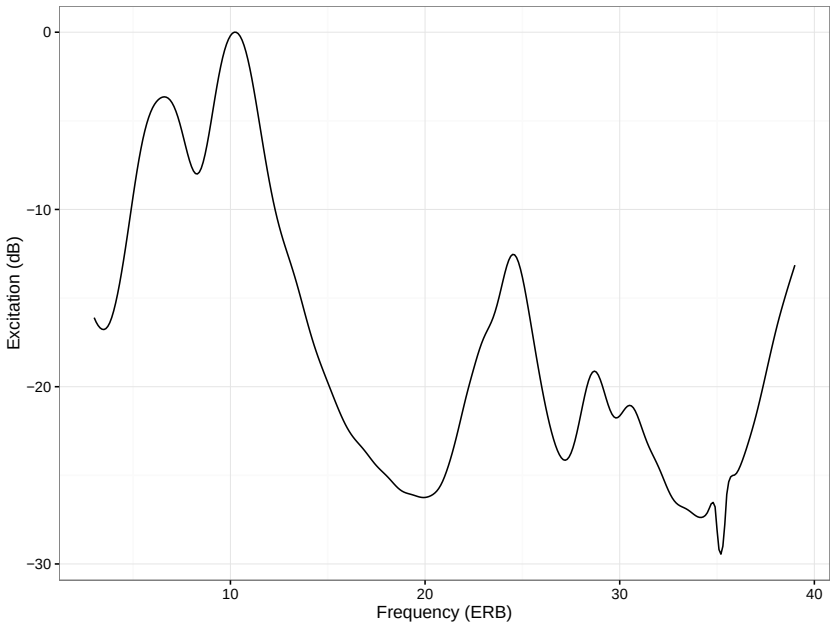


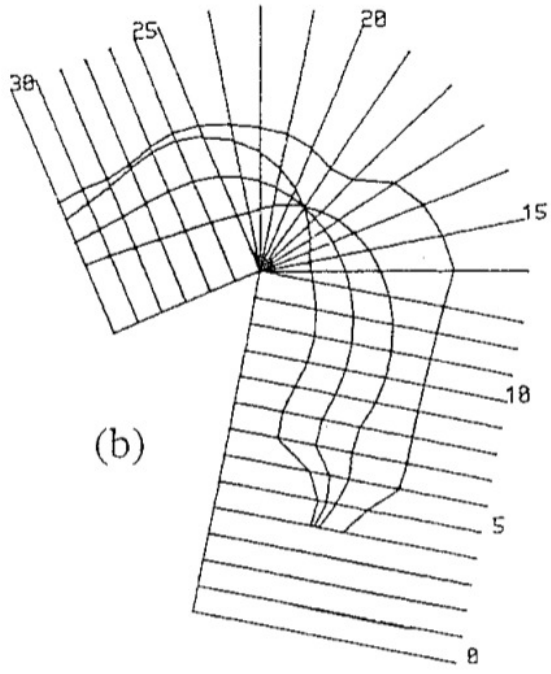
- The example illustrates a key issue – representations of vowels differ substantially across talkers.
- This issue is typically addressed using a normalization computation that attempts to eliminate differences in representations due to vocal tract length differences across talkers.

Technical solutions vs. psychological aspects

- Hindle (1978) distinguished between “technical” solutions to normalization and its “psychological” aspects.
- The type of solution that researchers adopt in modeling impacts the logical structure of their underlying conceptual framework.
- The choice impacts not only “dynamic” characterizations of speech, but also models of how phonological systems emerge during ontogeny, how they are transmitted and mutated within and across cultures, and how they emerged within the species.

Toward the psychological aspects?

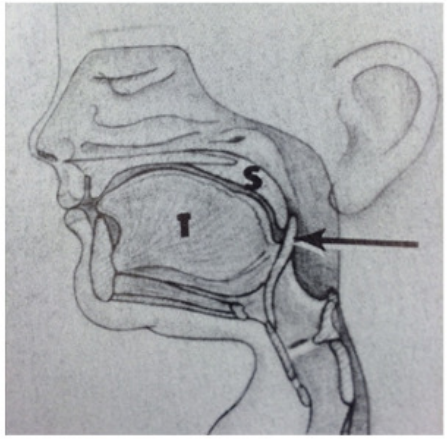


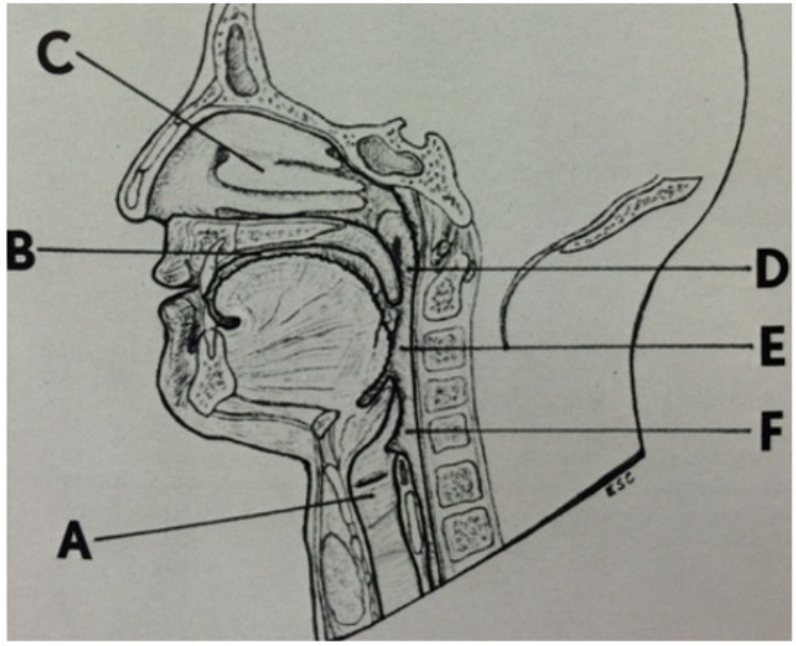


- Infants must organize and coordinate largely incommensurate reference frames as they develop to possess adult-like phonological systems (see Kuhl & Meltzoff, 1982, 1996; inter alia).

- 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), but ...

Dynamics during ontogeny

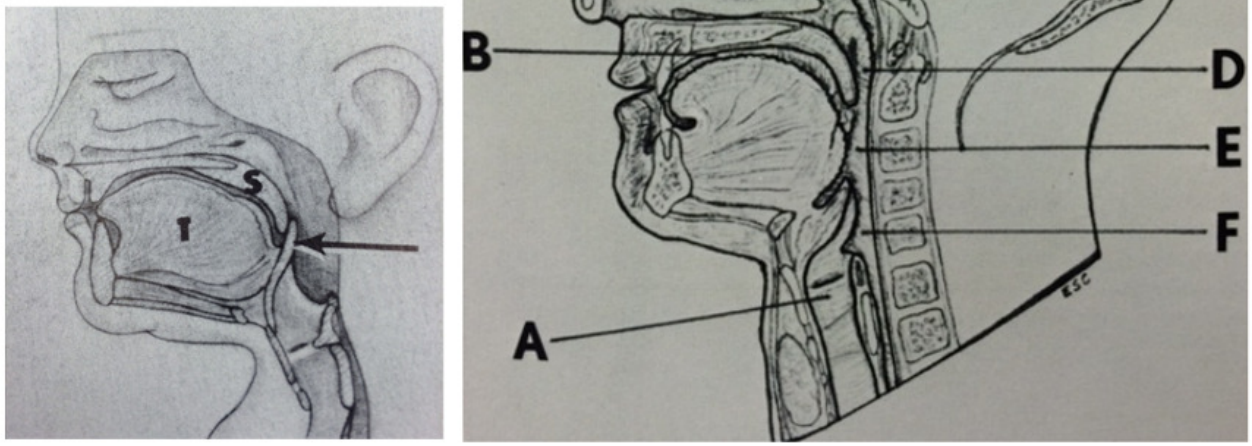




- An infant's physiological growth involves a massive morphological restructuring of the vocal tract which continues well into the first year of life and beyond (see Crelin, 1987; Voperian et al. 2011; inter alia).

- After the larynx begins to descend at about two months of age, infants can begin engaging in the simultaneous composition of an exhalation-phase phonation gesture with different oral gestures that reflect the language-specific voice qualities and vowel timbres of the ambient language.

Dynamics during ontogeny



- As control of the respiratory and articulatory systems develops, infants become capable of rhythmically coordinating constriction gestures and oral postures with a single exhalation-phase phonation gesture.

- This “canonical babble” (Oller & Eilers, 1988) enables the creation of novel sequences to productively expand the repertoire of language-specific phonological forms.

Dynamics during ontogeny

- Vocal interactions between infants and caretakers encompass a rich set of dynamic phenomena that is scarcely even cataloged, and yet they are typically modeled as a nearly trivial form of face-to-face “vocal imitation” (see Rasilo et al., 2013; Messum & Howard, 2015; inter alia).

- General models of the emergence of phonological systems within individual agents and communities (see Moulin-Frier et al., 2015) tend to endow vocal exchanges between agents with adult-like properties, e.g., joint attention to a “reference” object, which are counter-factual with respect to early vocal interactions between infants and caretakers.

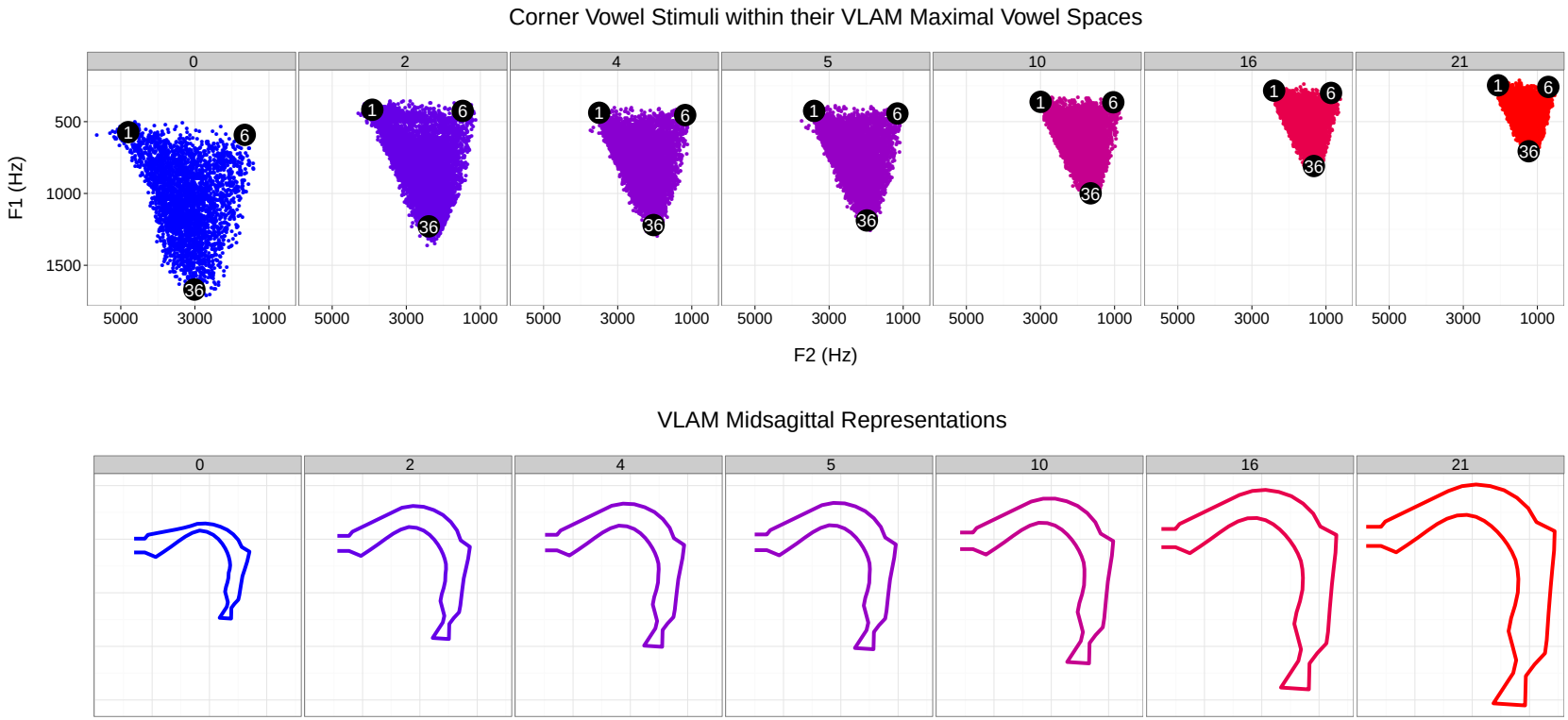
- Rather, early vocal interactions between infants and caretakers seem to exhibit dyadic mutual attention in “imitation games” that exhibit complex development trajectories (see Masataka, 2003) and potentially dyad-specific idiosyncrasies.

Basic modeling requirements

- Models of how talkers across language communities parse the vocalizations of others in terms of their language-specific phonetic and social categories.

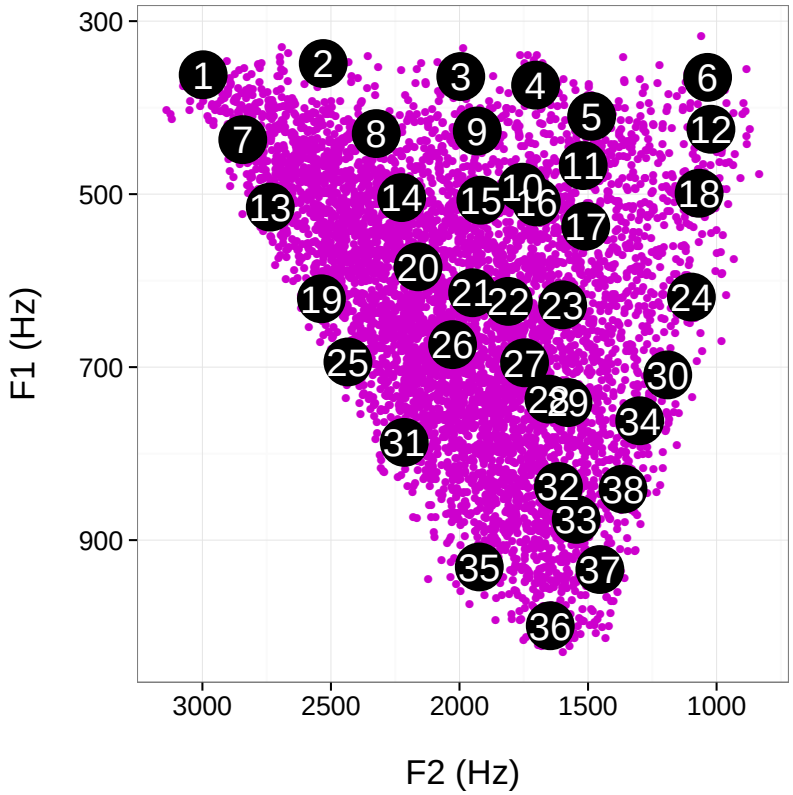
- Models of how infants coordinate and generate references frames for mapping sensory representations to categories over the course of ontogeny that reflect the affective and affiliative information they intake via early vocal exchanges with their caretakers, i.e., the dynamics of social meaning.

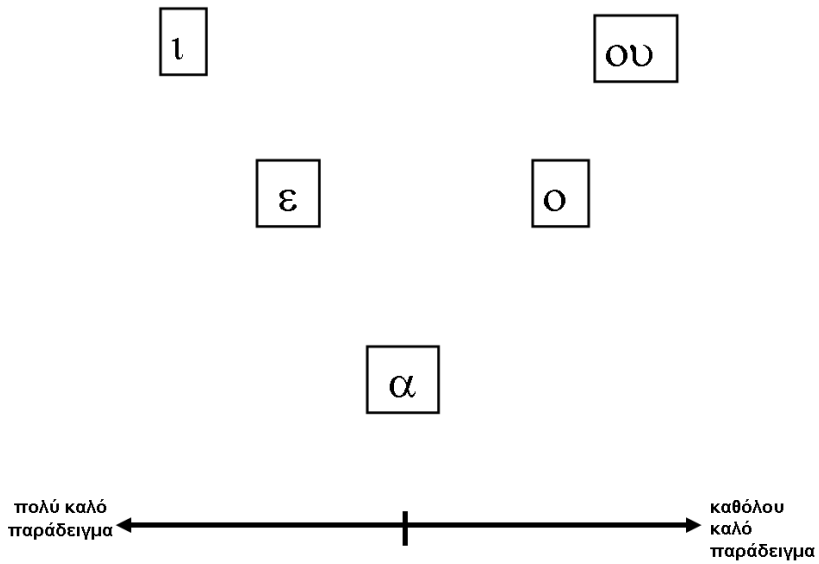
Articulatory synthesizer



- VLAM: Boë & Maeda (1998)

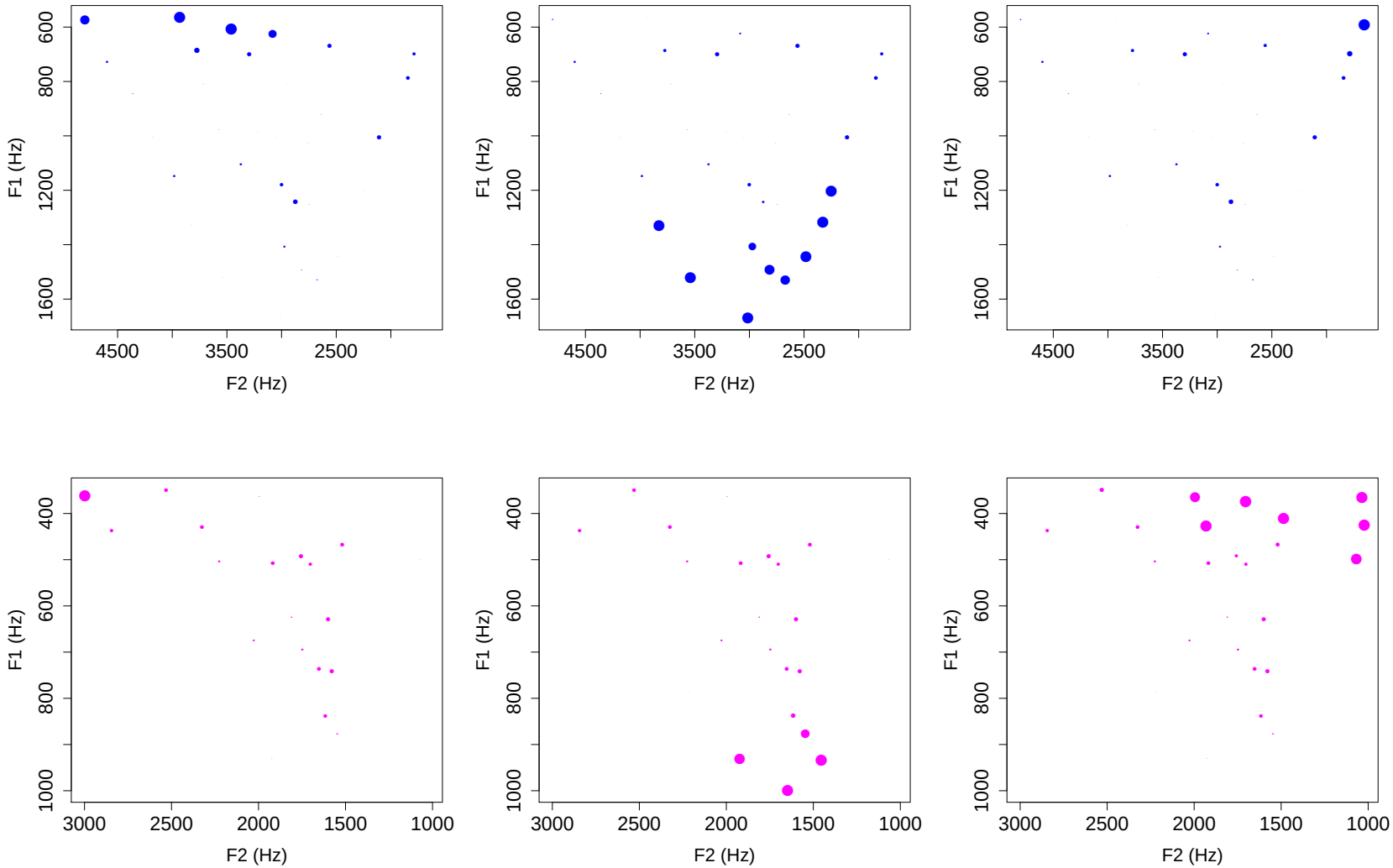
Perceptual categorization experiments





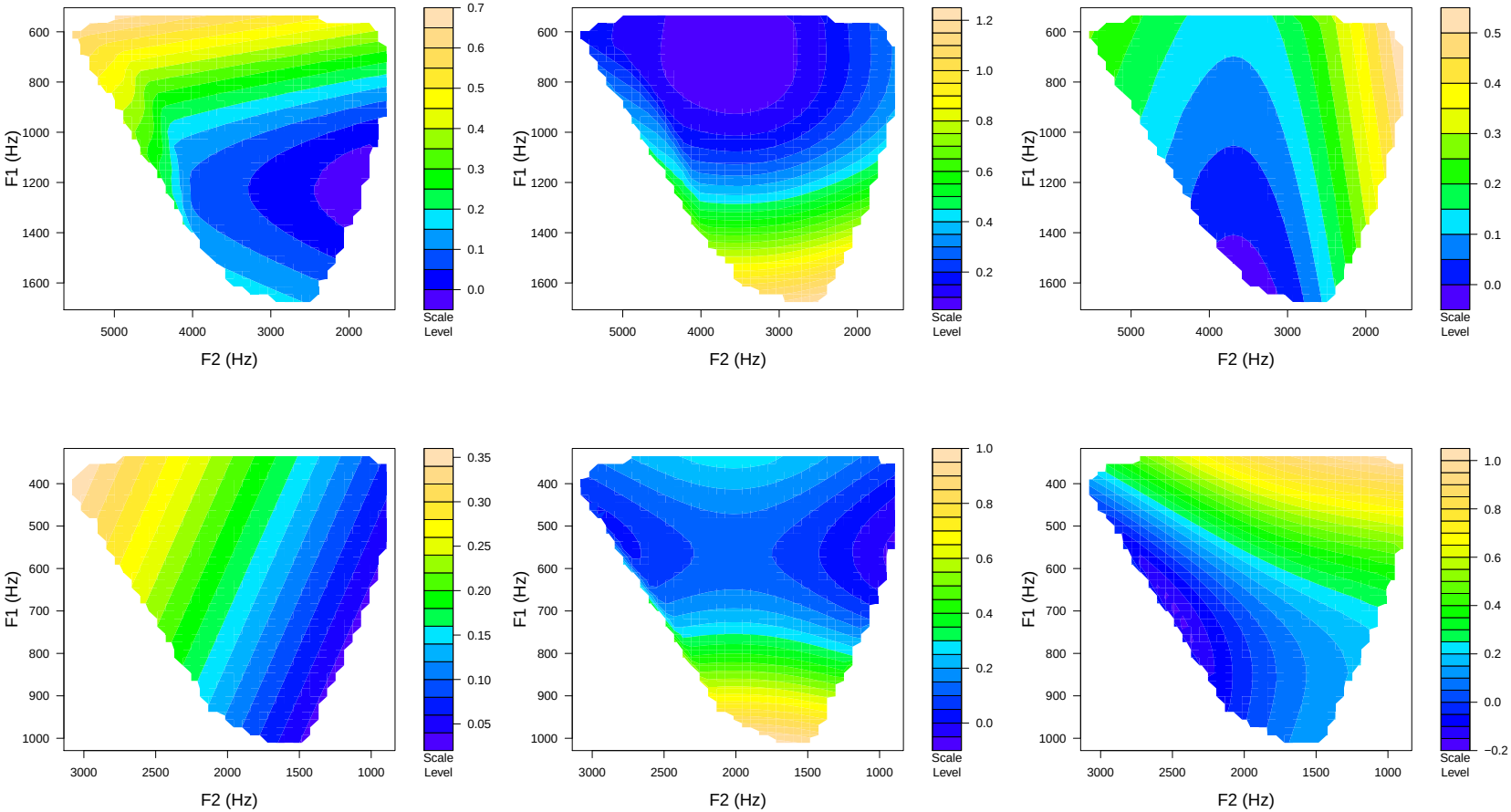
- Munson et al. (2010), Plummer et al. (2013)

Perceptual categorization modeling



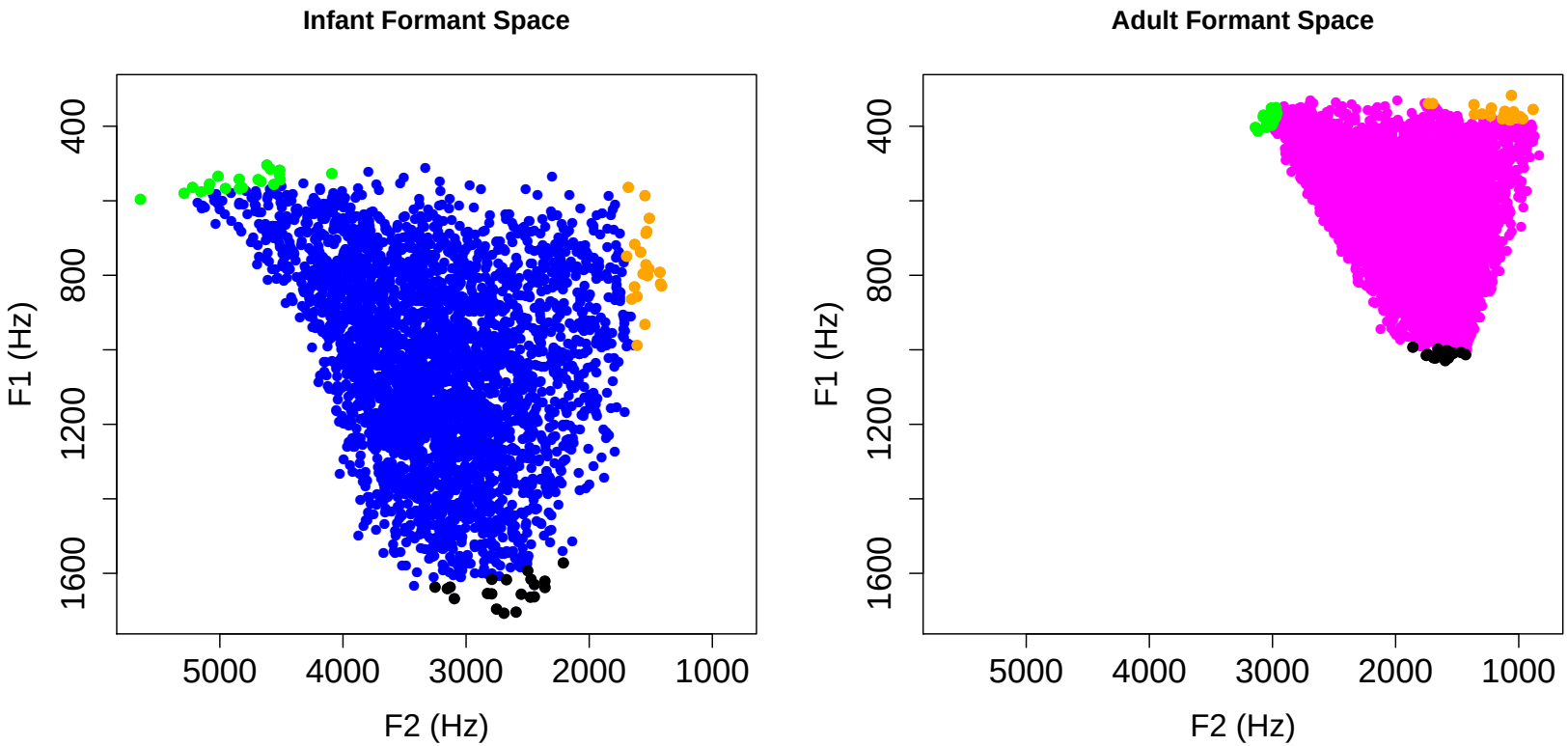
- Plummer et al. (2013)

Perceptual categorization modeling



- Plummer et al. (2013)

Vocal exchanges

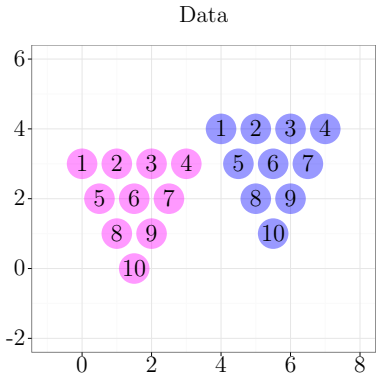


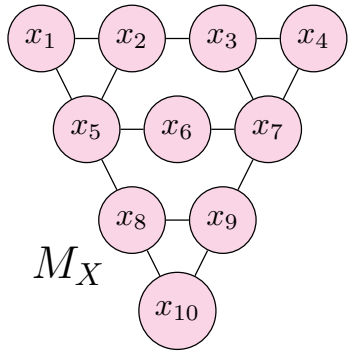
- Plummer et al. (2013)

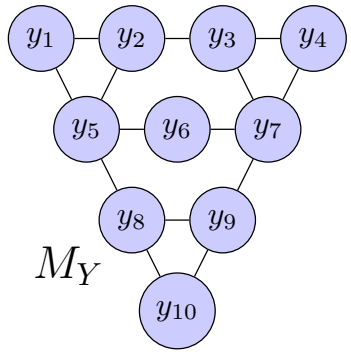
Infant sensory spaces

- Articulatory space – Infant's articulatory representations are sequences of VLAM vectors for 0.5 year old age setting.
- Auditory space – auditory representations are sequences of erb vectors.
- Auditory space contains auditory representations of both infant and caretaker vocalizations.

Organizing sensory spaces

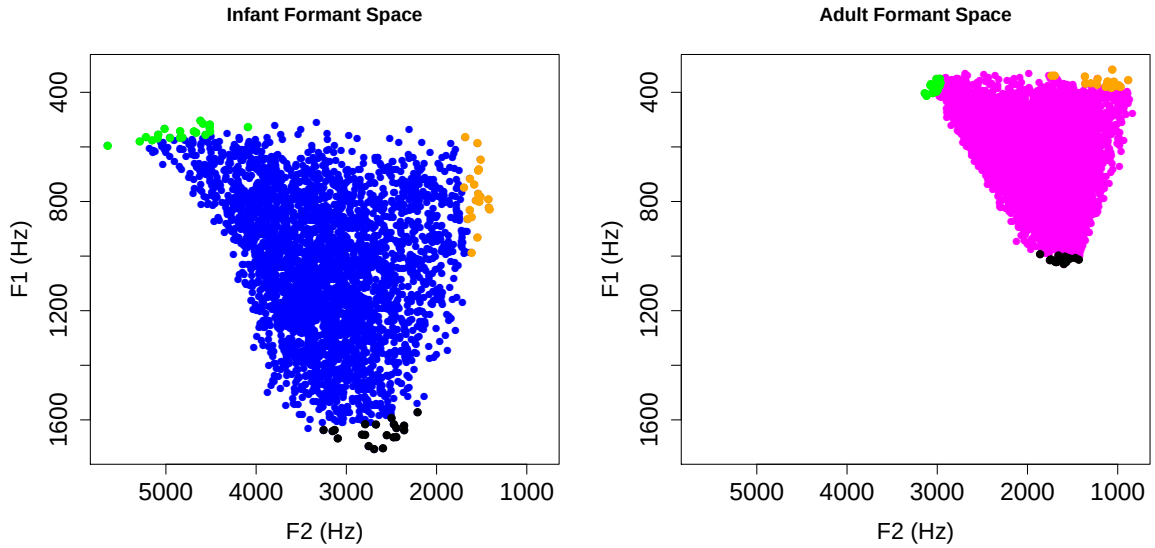






- Manifolds – a weighted graph structure is imposed on reference frame representations.
- Infant articulatory manifold – weighted graph over articulatory representations.
- Infant auditory manifold – weighted graph over auditory representations of infant vocalizations.
- Caretaker auditory manifold – weighted graph over auditory representations of caretaker vocalizations.

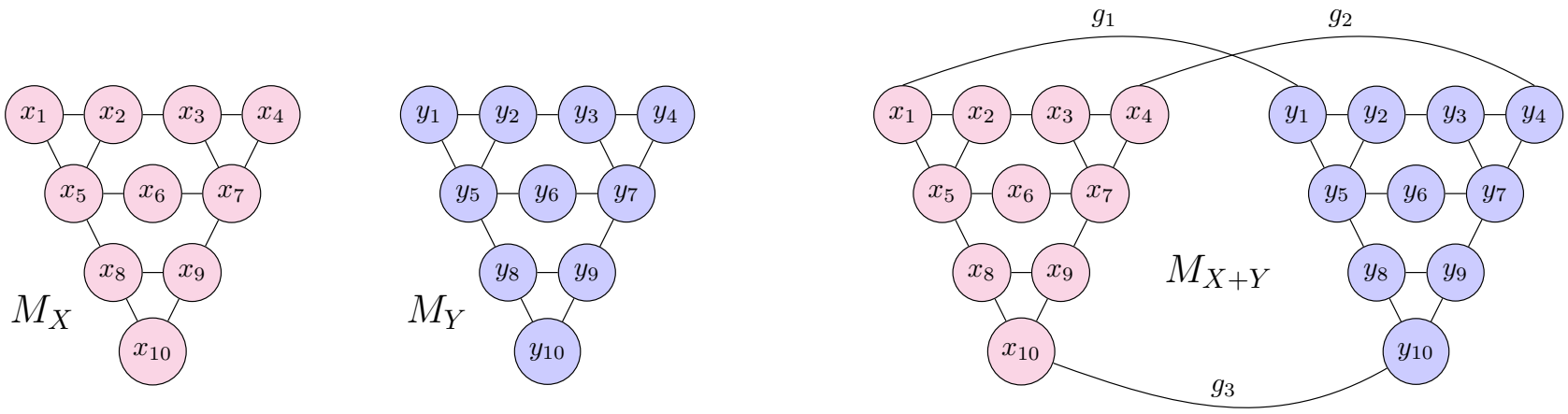
Alignment



- Infant's internalization of vocal exchanges yields coupled representations of infant articulatory and infant auditory representations with an affiliative weight that permits alignment of articulatory and auditory manifolds.

- It also yields coupled representations of infant articulatory and caretaker auditory representations with an affiliative weight that permits alignment of infant articulatory and caretaker auditory manifolds.

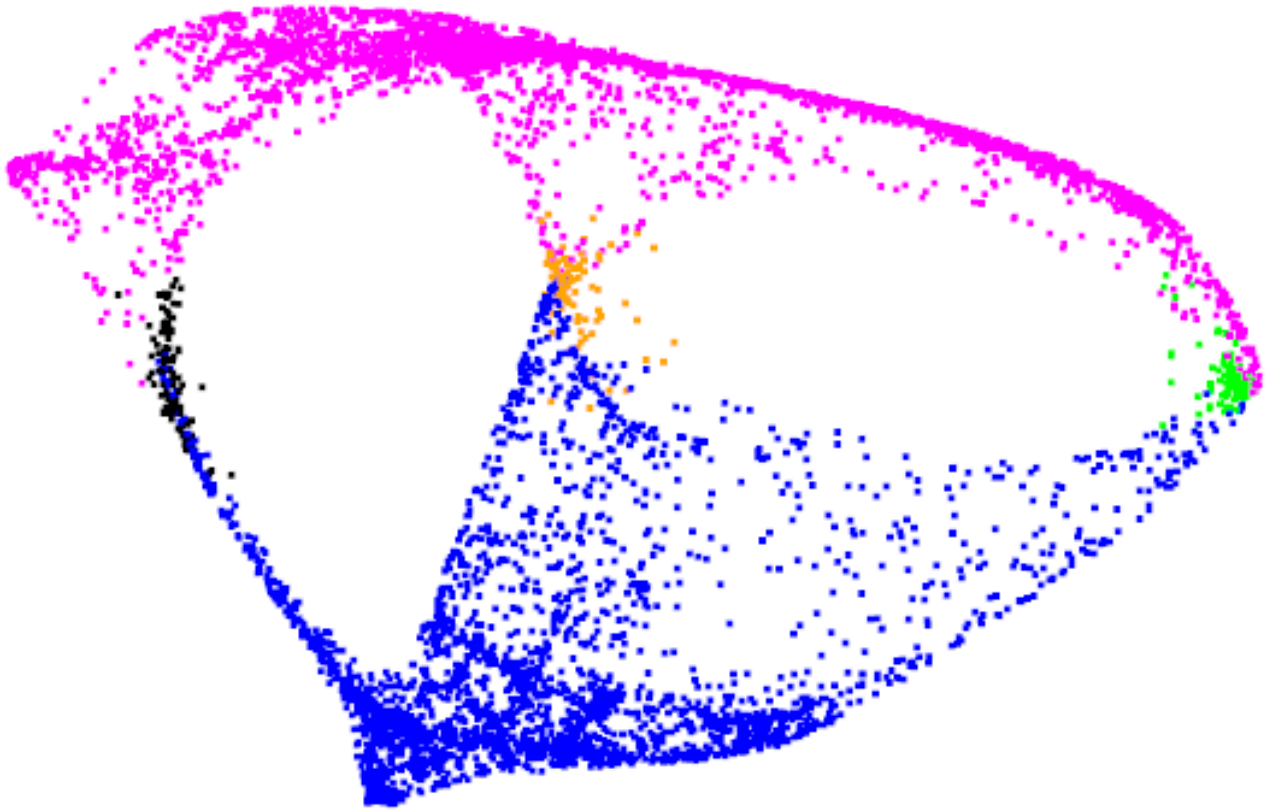
Alignment



- Infant creates a model of self from coupled representations of infant articulatory and infant auditory representations with an affiliative weight that permits alignment of articulatory and auditory manifolds.

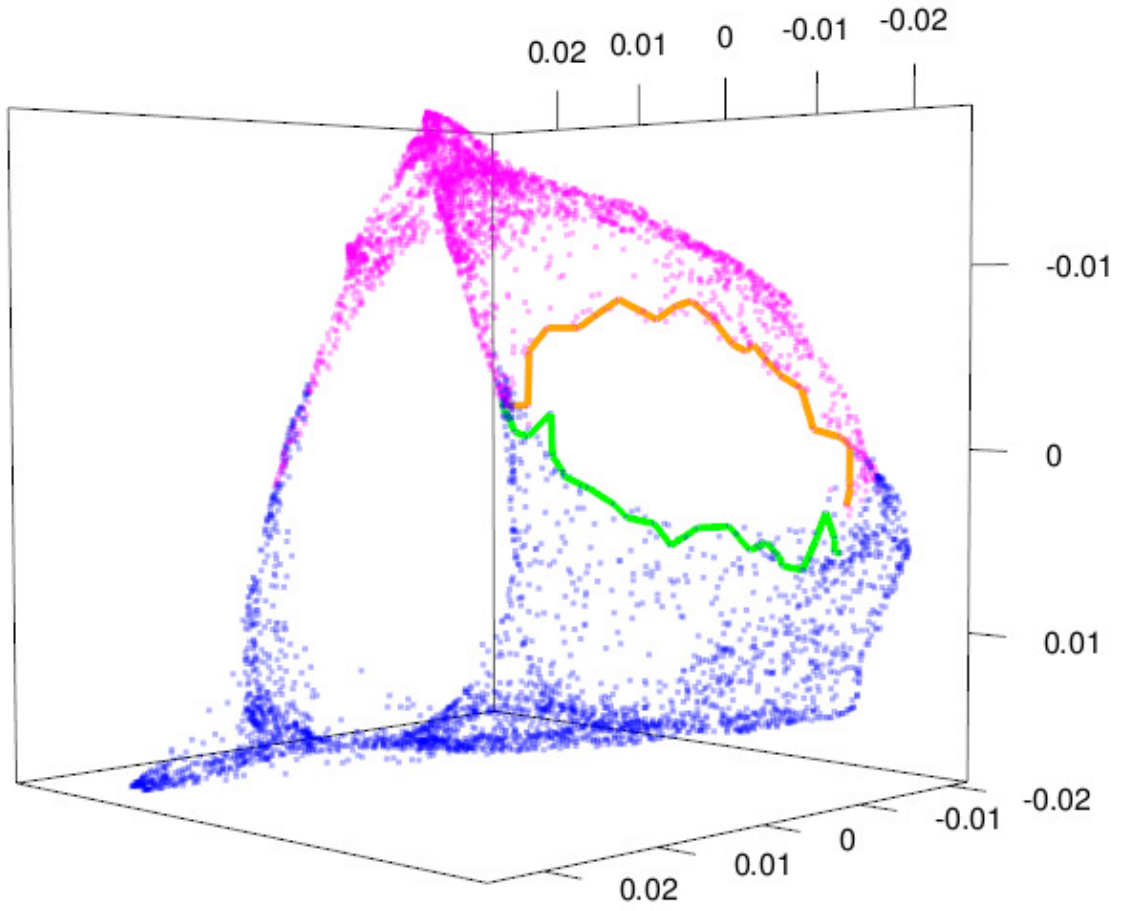
- Infant creates a model of caretaker from coupled representations of infant articulatory and caretaker auditory representations with an affiliative weight that permits alignment of infant articulatory and caretaker auditory manifolds.

Mapping across models



- Laplacian eigenmaps (Belkin & Niyogi, 2003).

Paths and reference frames



- paths in graphs correspond to “trajectories” in reference frames.

Take Home Points

- Technical/physicalist approaches to speech dynamics do not account for cognitive organization of speech dynamics within an individual across the lifespan, or a community across cultural lifespans.
- Modeling isn't done solely to make sense of experiments or data. Sometimes you need it as a platform for creative leaps to move things forward.
- Compositionality in language is extremely deep and complex. Right now, we need imagination more than quantification to characterize it.

Thank you

- Asimina Syrika and Kiyoko Yoneyama for recruiting and testing the Greek- and Japanese-speaking participants,
- Hyunju Chung and Chanelle Mays for recruiting and testing the Korean- and the English-speaking participants,
- Catherine McBride-Chang for providing the resources to recruit the Cantonese-speaking subjects,
- Mary Beckman and Ben Munson for numerous important contributions to various aspects of this presentation.
- Lucie Ménard, Pat Riedy, and Jeff Holliday for various supporting technical components of the computational modeling framework, and
- Eric Fosler-Lussier and Misha Belkin for providing initial direction concerning the computational modeling framework.
- Work supported by NSF grants BCS-0729306 and BCS-0729277.