Acquiring and adapting phonetic categories in a computational model of speech perception

Joe Toscano

Beckman Institute for Advanced Science and Technology
University of Illinois at Urbana-Champaign

Acknowledgements

Cheyenne Munson Toscano
University of Illinois

Dave Kleinschmidt
University of Rochester

Florian Jaeger
University of Rochester

Funding: Beckman Institute

Overview

Two types of learning:

- ***Adaptation*** of phonetic categories by adult listeners
- ***Acquisition*** of phonetic categories by infants during development

Question: Can a single learning mechanism account for both?

Not necessarily the same:

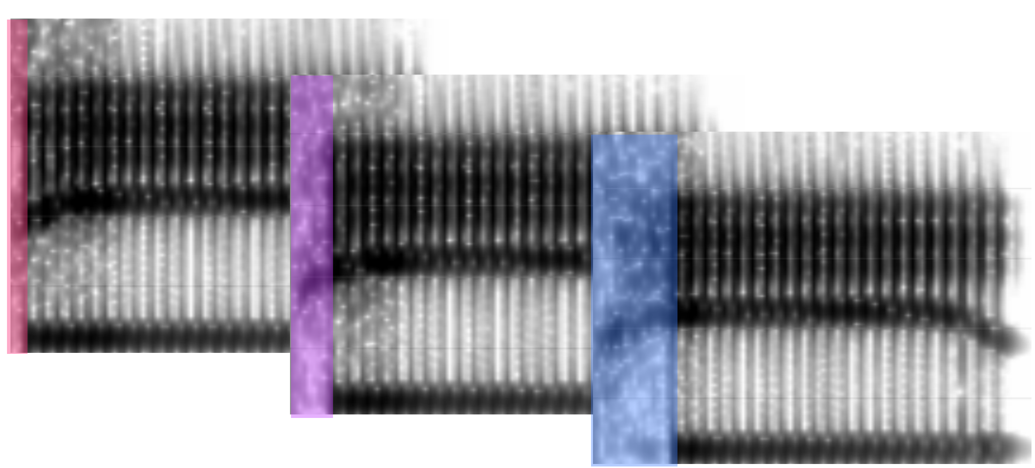
- Typically viewed as distinct processes
- Very different time scales: acquisition is slow; adaptation is rapid
- May require separate representations of phonetic categories

Speech development

Speech perception

Acoustic information

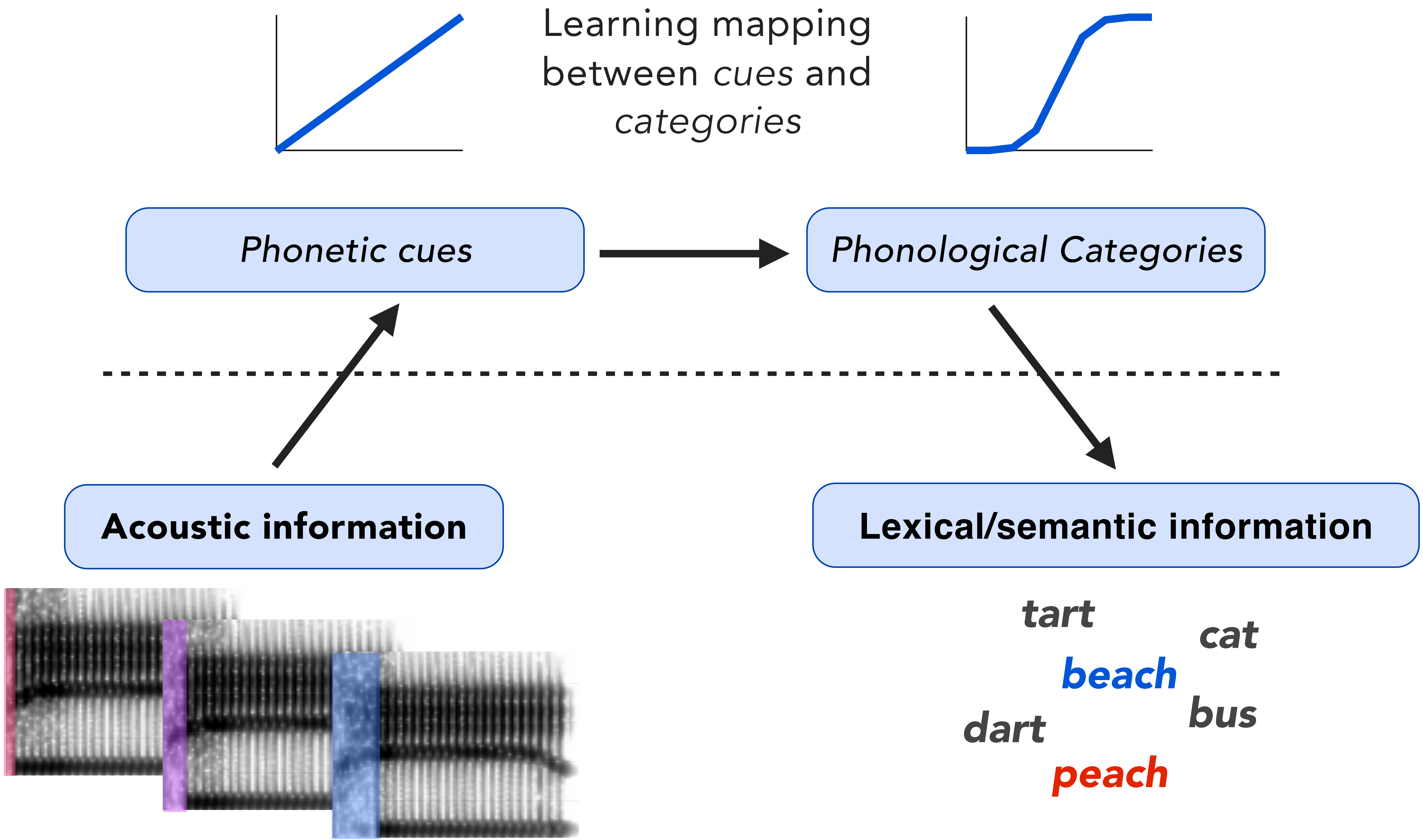
Lexical/semantic information



tart
cat
beach
bus
dart
peach

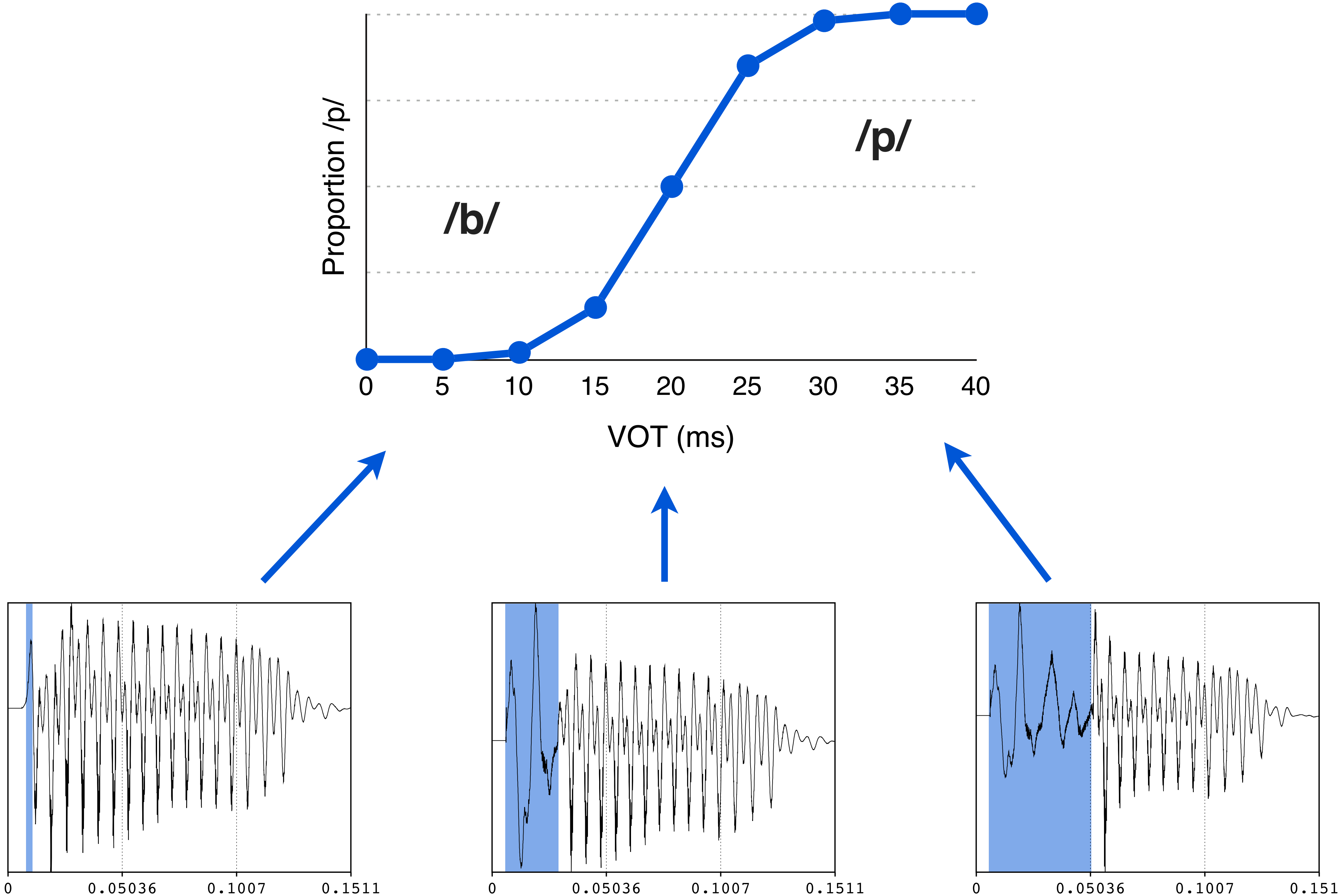
Toscano, McMurray, Dennhardt, & Luck (2010), *Psych Sci*

Speech development



Toscano, McMurray, Dennhardt, & Luck (2010), *Psych Sci*

A model system: *VOT and voicing*

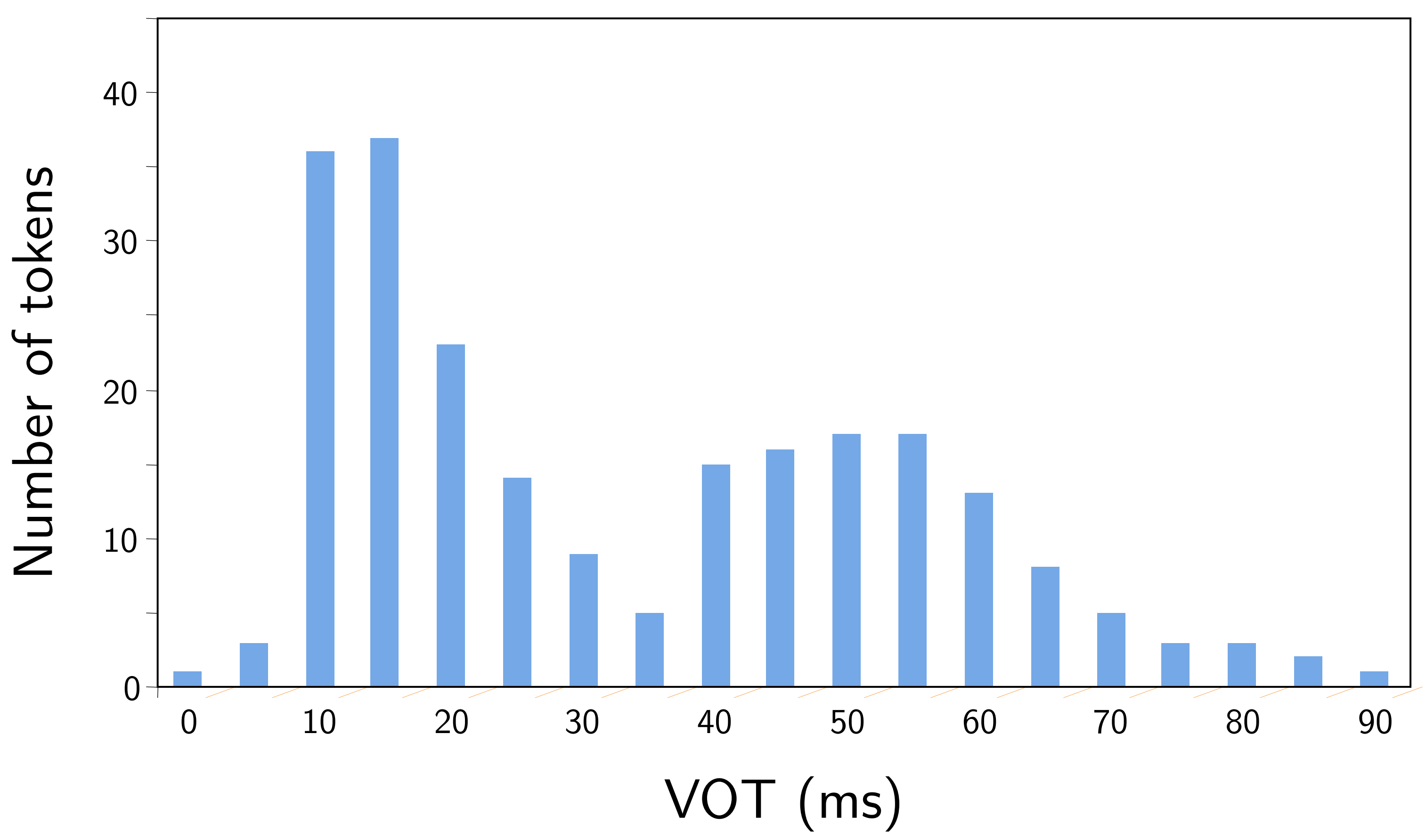


Toscano, McMurray, Dennhardt, & Luck (2010), *Psych Sci*

A model system: *VOT and voicing*

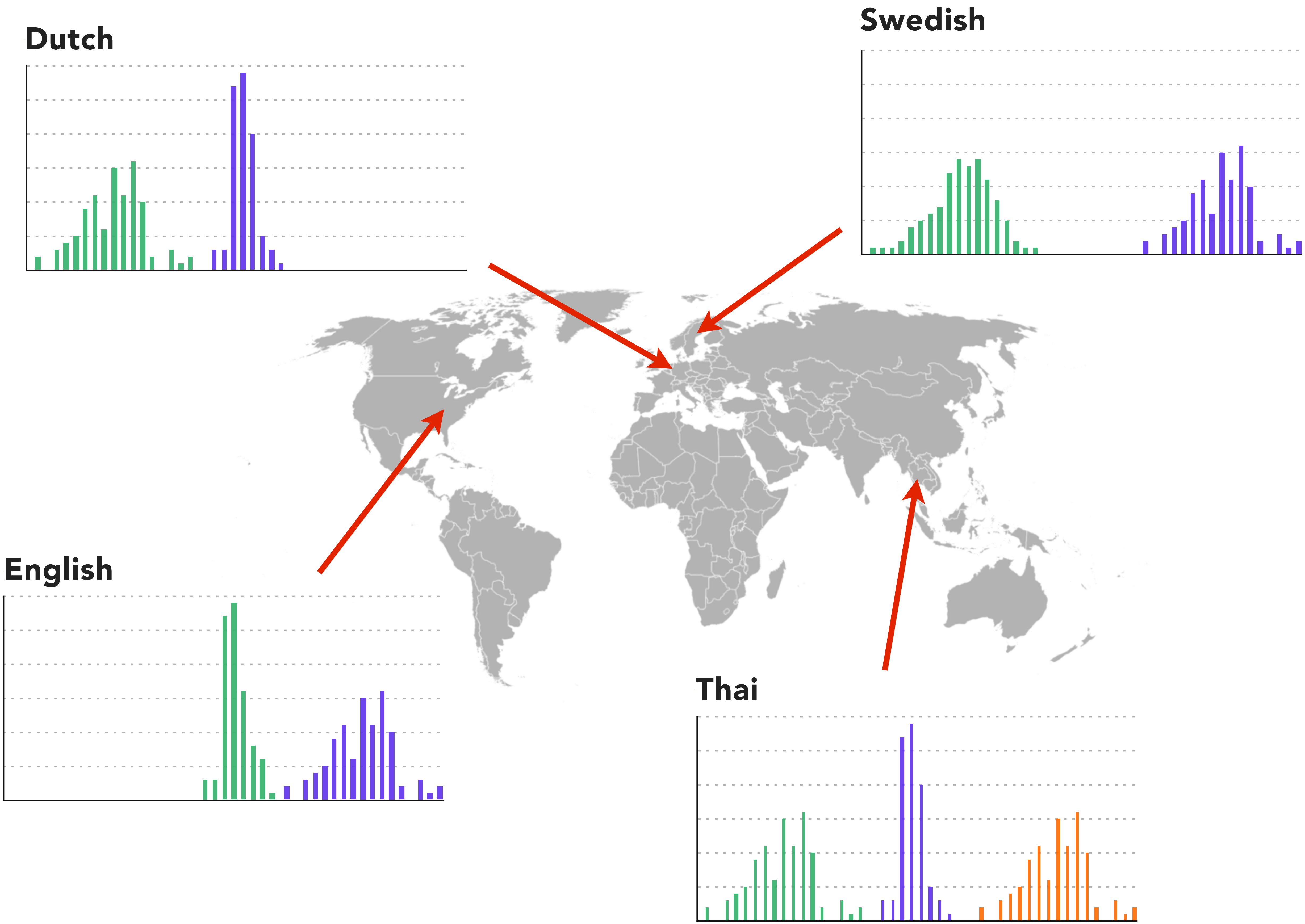
How do listeners learn the mapping between cues and categories?

- One possibility: Track distributional statistics of acoustic cues
- Clusters corresponding to phonological categories
- e.g., English VOT and voicing



Maye, Werker, and Gerken (2002), *Cognition;* Allen & Miller (1999), *JASA*

Cross-linguistic differences



Allen & Miller (1999); Beckman et al. (2012); Lisker & Abramson (1964); Image credit: Roke / Wikimedia Commons

Speech development

*Learning the **distributional statistics** of acoustic cues*

Provides a way of learning the mapping between cues and categories

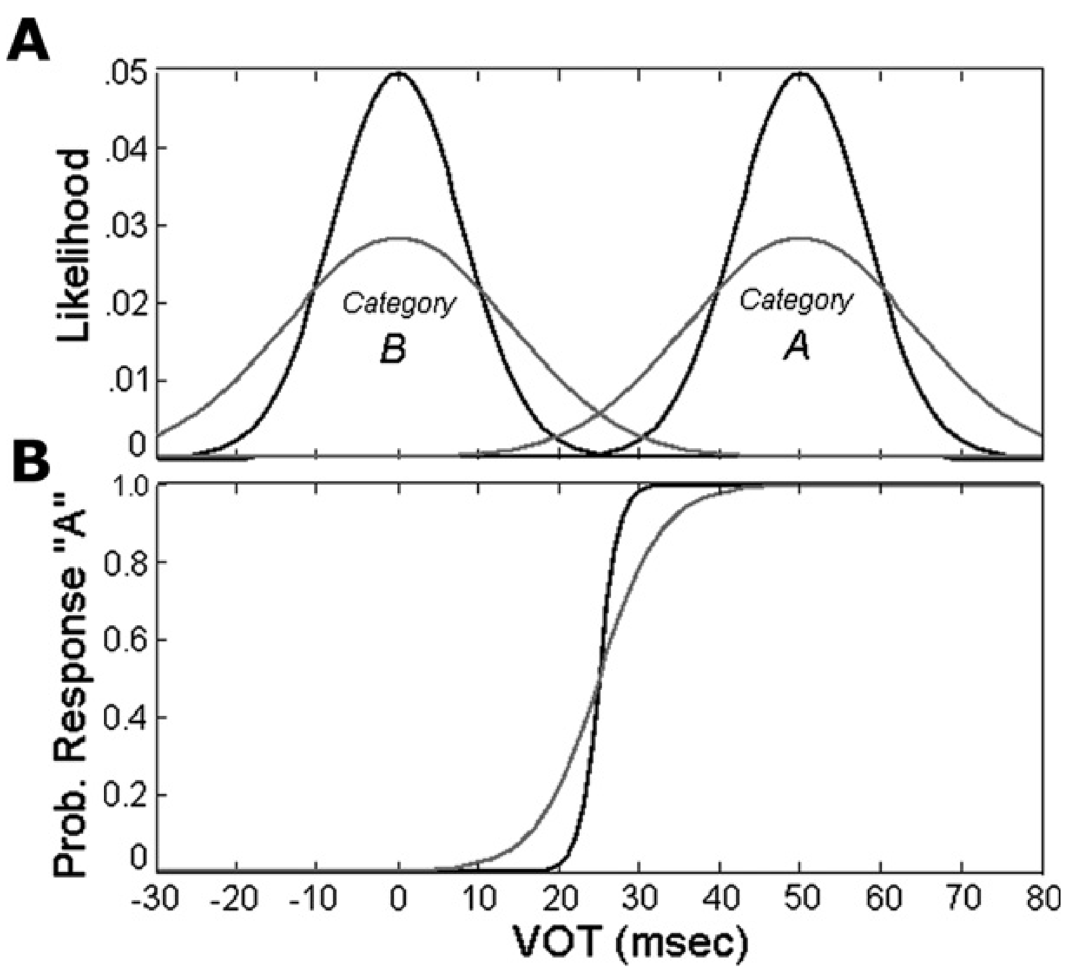
Is this similar to unsupervised perceptual adaptation experiments?

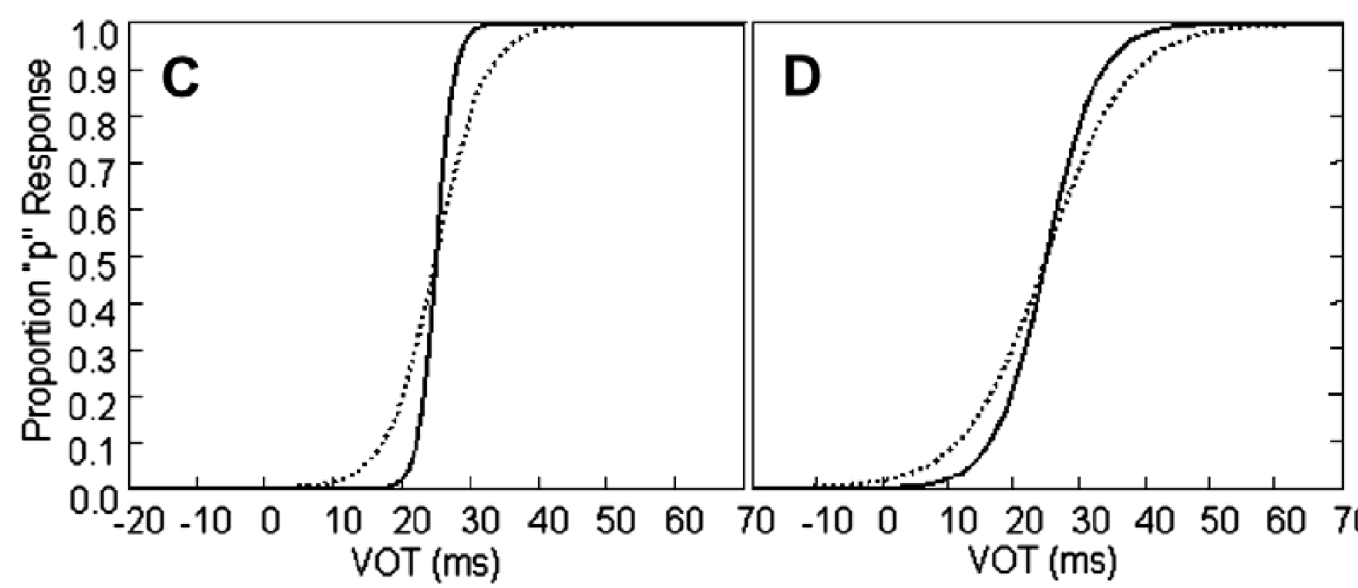
Can adults track changes in the distributional statistics of acoustic cues?

Perceptual adaptation

Listeners rapidly adapt to novel distributions of cues (~1 hr experiments)

- Clayards, Tanenhaus, Aslin, & Jacobs (2008): *Category variance*



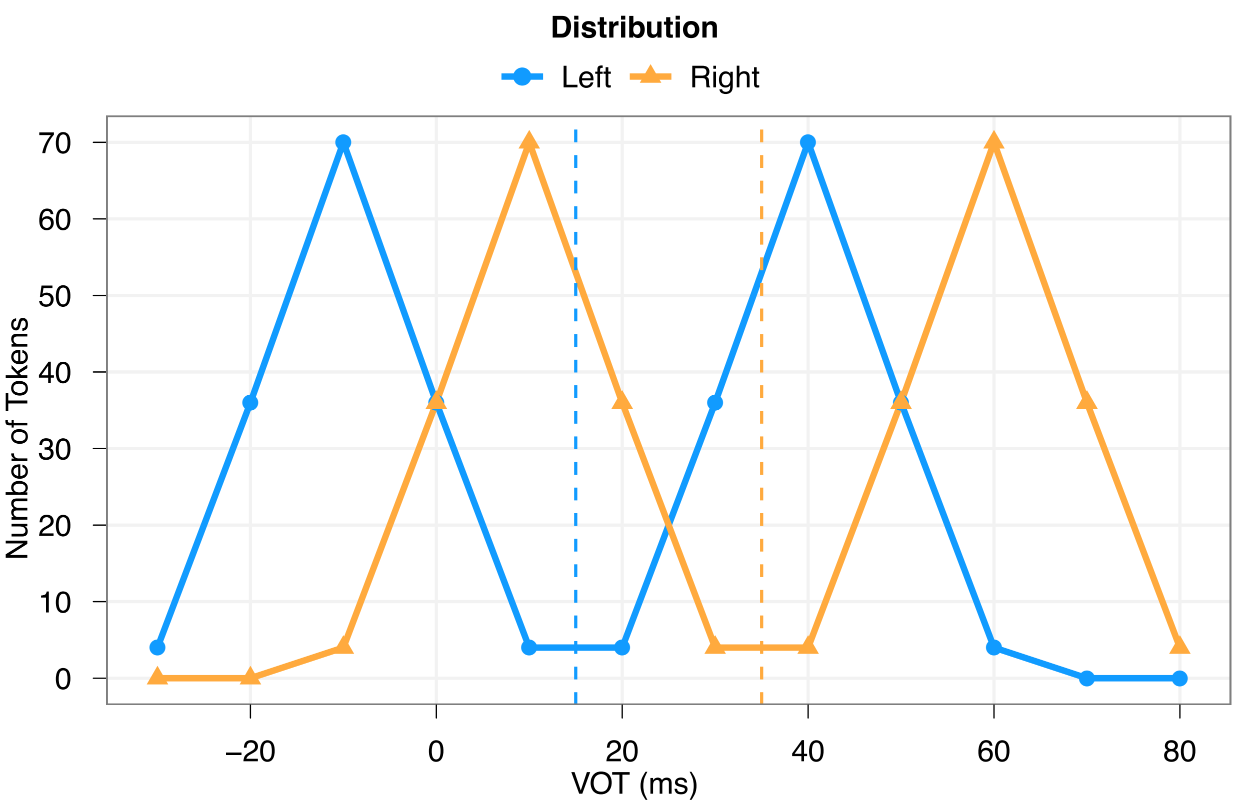


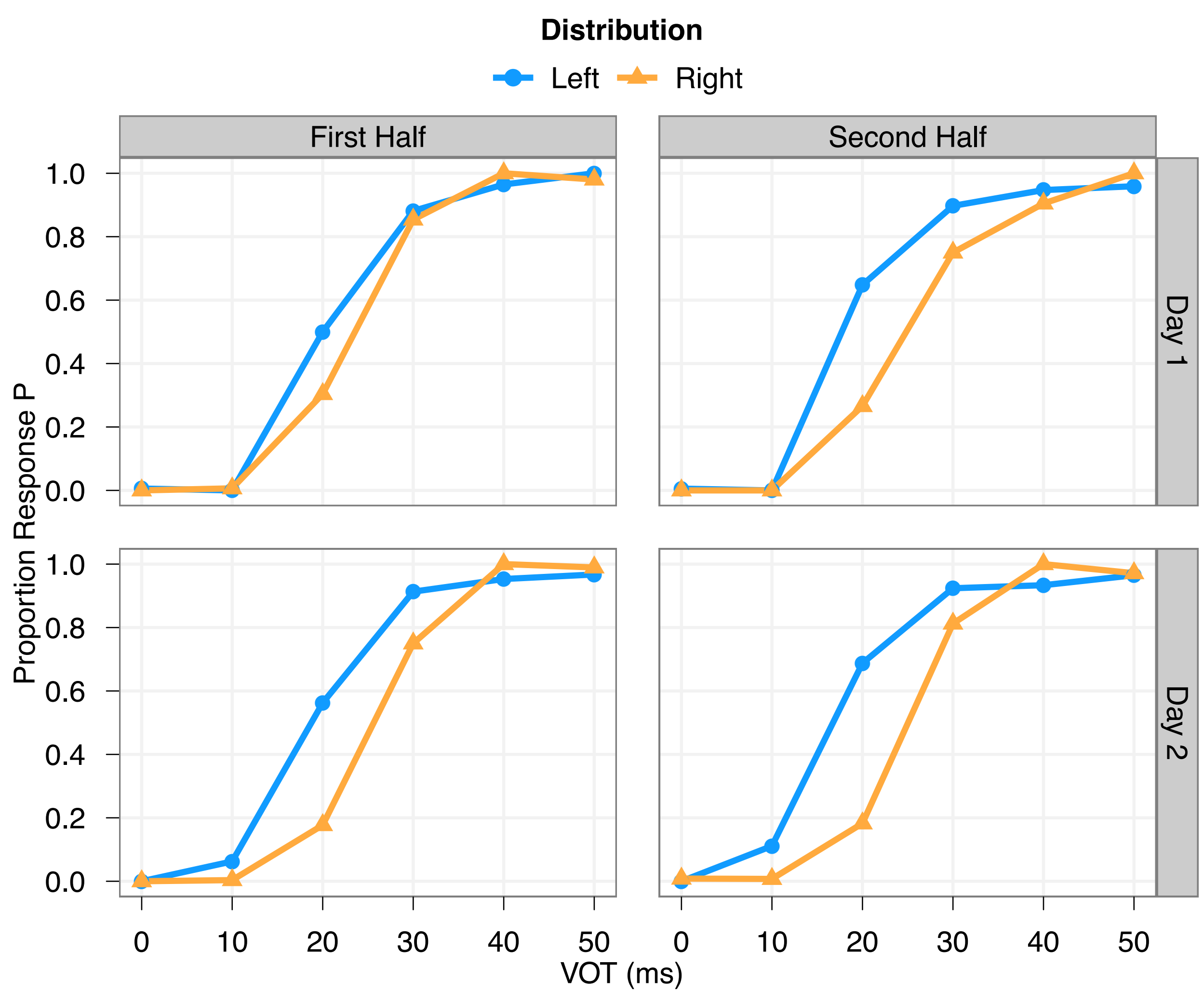
Clayards et al. (2008), *Cognition*

Perceptual adaptation

Listeners rapidly adapt to novel distributions of cues (~1 hr experiments)

- Clayards, Tanenhaus, Aslin, & Jacobs (2008): *Category variance*
- Munson (2011): *Category means*





Munson (2011), dissertation

Language acquisition and perceptual adaptation

Two phenomena

- ***Acquisition*** of speech sounds during development (slow process)
- ***Adaptation*** of speech sounds in adulthood (fast process)

Can a single model account for both?

- Are changes in plasticity needed?
- Are separate representations of long- and short-term categories needed?

Approach:

- Simulations with a computational model of speech categorization
- Examine parameter space of model to see if there are common learning rates for both acquisition and adaptation

Overview

Modeling approach

- Gaussian mixture model
- Statistical learning and competition

Acquisition during development

- Simulation 1: Determining the number of categories and their properties

Adaptation in the same model

- Simulation 2: Perceptual learning of shifted VOT distributions

Other aspects of perceptual learning in the model

- Simulation 3: Speaking rate adaptation
- Simulation 4: Learning new phonetic categories
- Simulation 5: Learning the categories of a second language

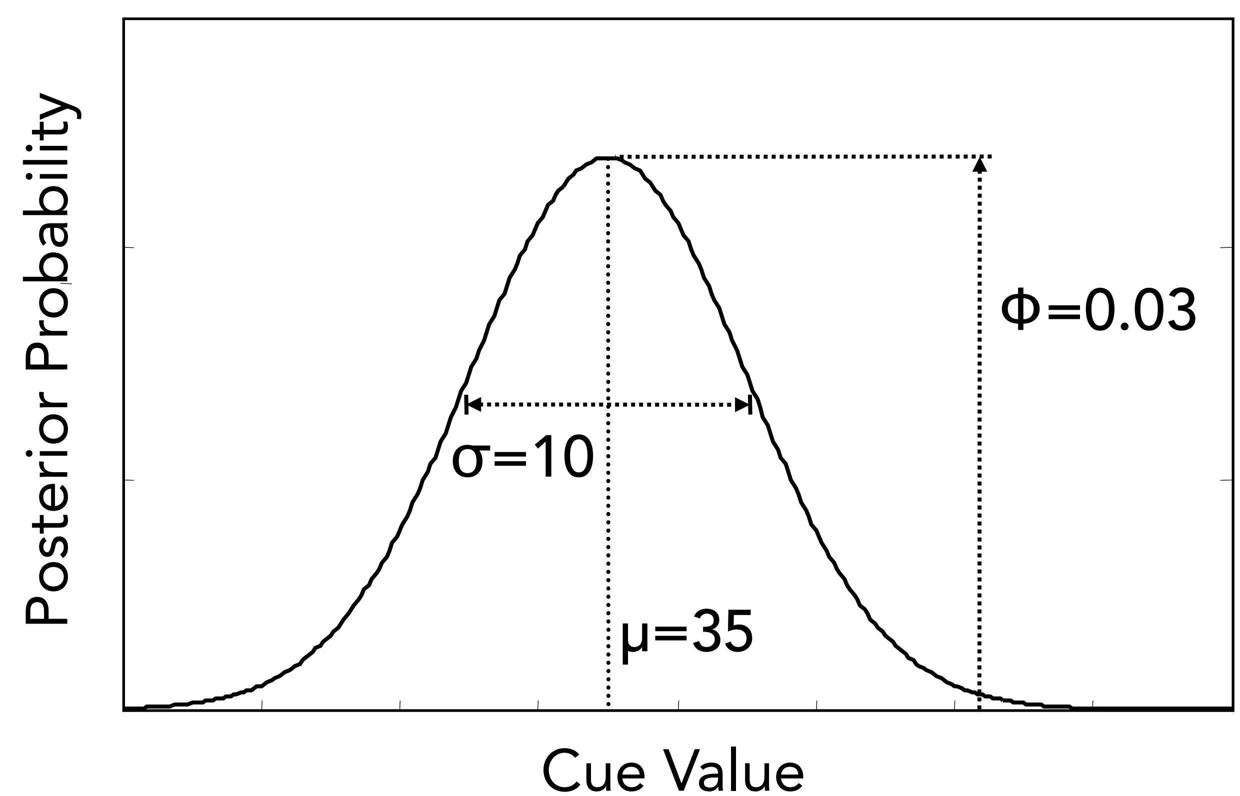
Model of speech perception

VOT example

- Clusters corresponding to phonological categories
- Different patterns across languages (Lisker & Abramson, 1964)

Gaussian mixture model (GMM)

- Categories defined by Gaussian distributions
- Mean (μ)
- Standard deviation (σ)
- Likelihood (Φ)



McMurray, Aslin, & Toscano (2009); Toscano & McMurray (2010)

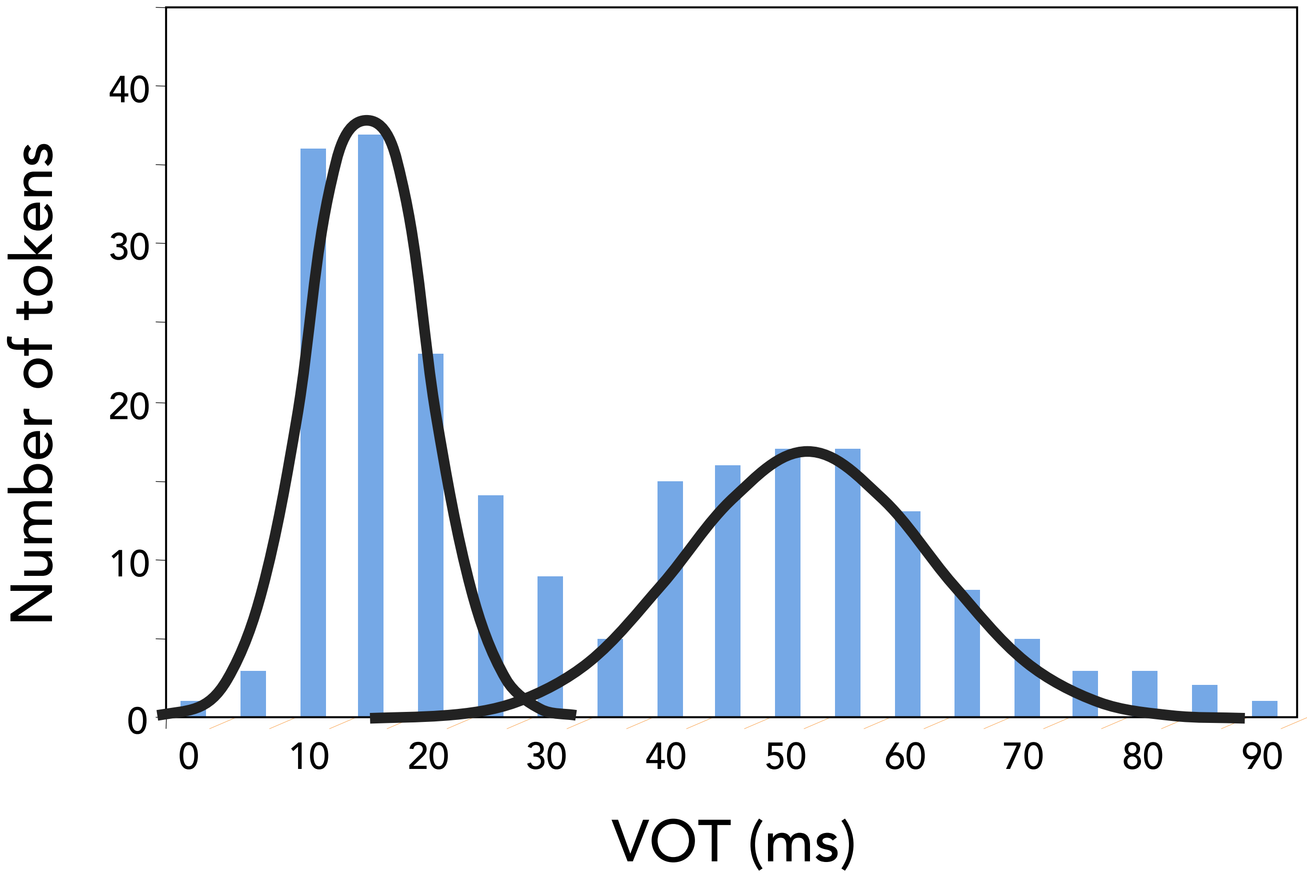
Model of speech perception

VOT example

- Clusters corresponding to phonological categories
- Different patterns across languages (Lisker & Abramson, 1964)

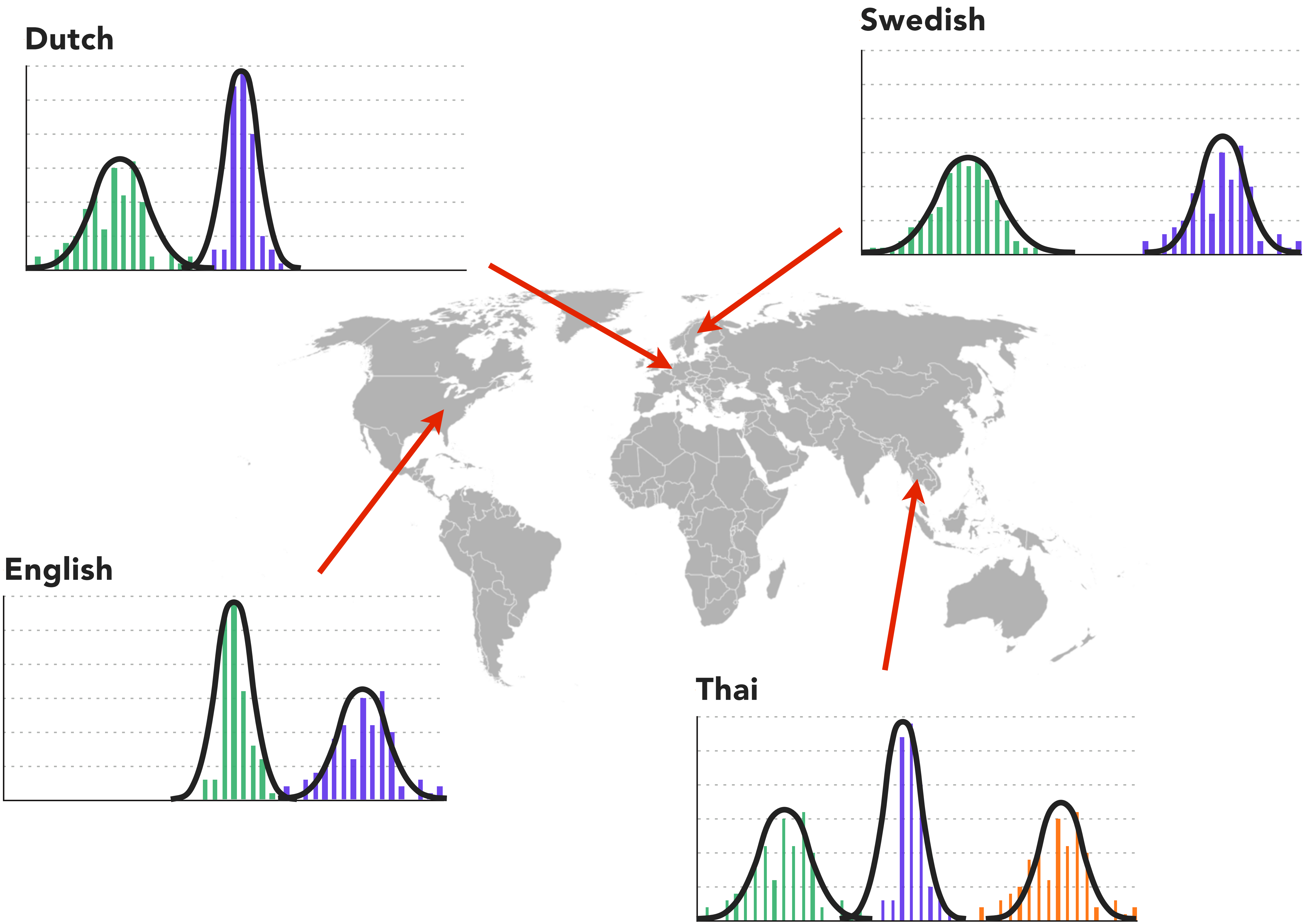
Gaussian mixture model (GMM)

- Categories defined by Gaussian distributions
- Model consists of a mixture of Gaussians along a cue dimension



McMurray, Aslin, & Toscano (2009); Toscano & McMurray (2010)

Speech sounds across the world's languages



Allen & Miller (1999); Beckman et al. (2012); Lisker & Abramson (1964); Image credit: Roke / Wikimedia Commons

Overview

Modeling approach

- Gaussian mixture model
- Statistical learning and competition

Acquisition during development

- Simulation 1: Determining the number of categories and their properties

Adaptation in the same model

- Simulation 2: Perceptual learning of shifted VOT distributions

Other aspects of perceptual learning in the model

- Simulation 3: Speaking rate adaptation
- Simulation 4: Learning new phonetic categories
- Simulation 5: Learning the categories of a second language

Acquiring phonetic categories

Learning the distributional statistics of acoustic cues

Why is this a hard problem?

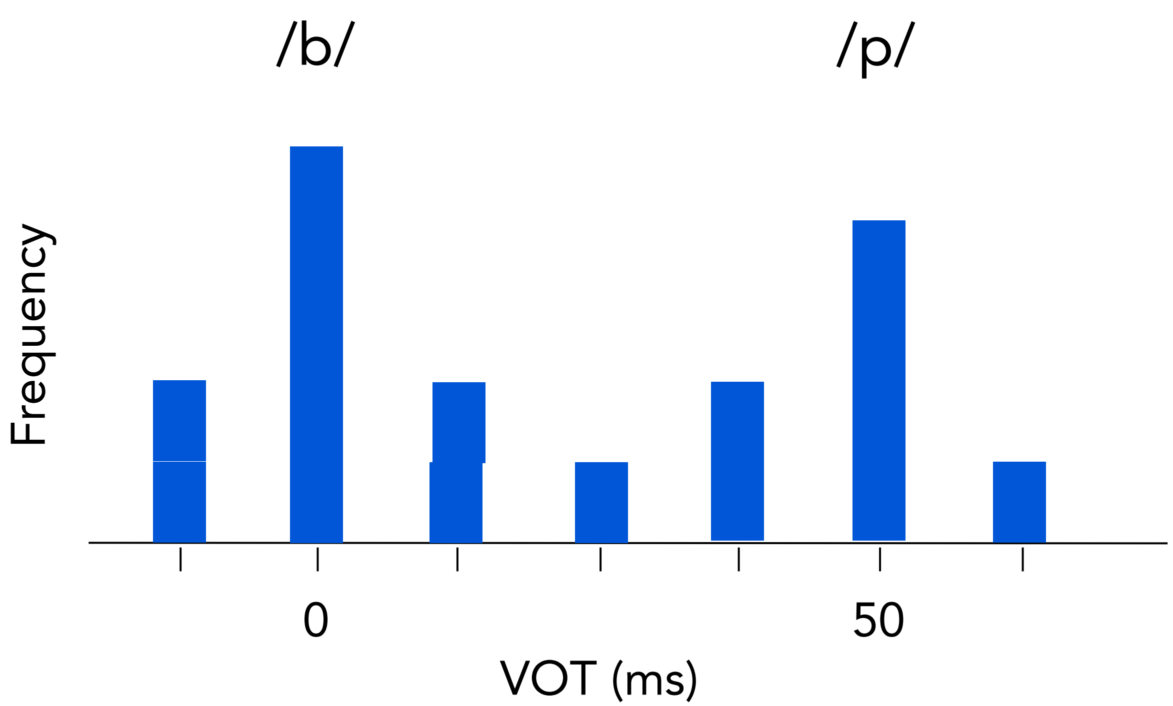
- Can't specify number of categories *a priori*
- Speech sounds are unlabeled
- Learning is incremental

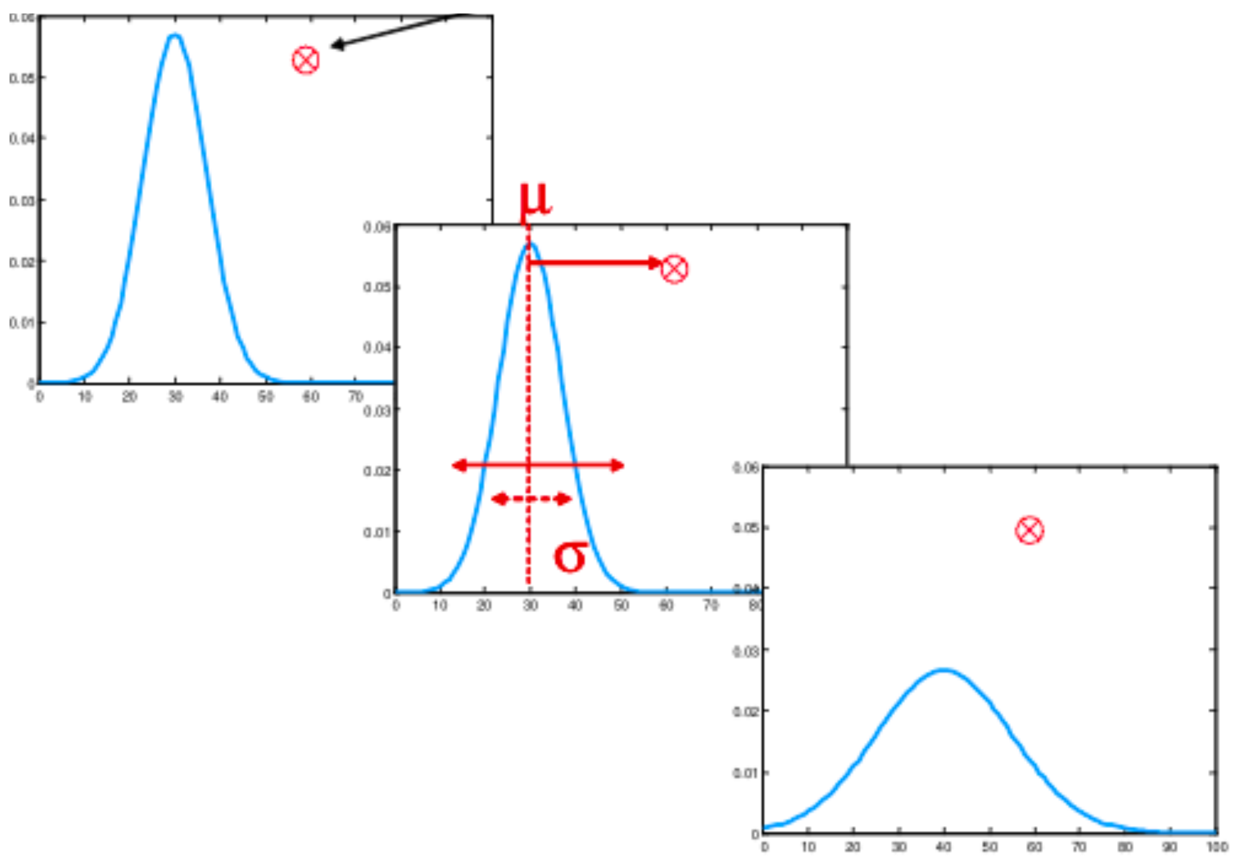
McMurray, Aslin, & Toscano (2009); Toscano & McMurray (2010)

Acquiring phonetic categories

Learning in the model

- Statistical learning (Saffran, Aslin, & Newport, 1996; Maye, Werker, & Gerken, 2002)
- Track the distributional statistics of acoustic cues





McMurray, Aslin, & Toscano (2009); Toscano & McMurray (2010)

Acquiring phonetic categories

Learning in the model

- Statistical learning (Saffran, Aslin, & Newport, 1996; Maye, Werker, & Gerken, 2002)
- Track the distributional statistics of acoustic cues

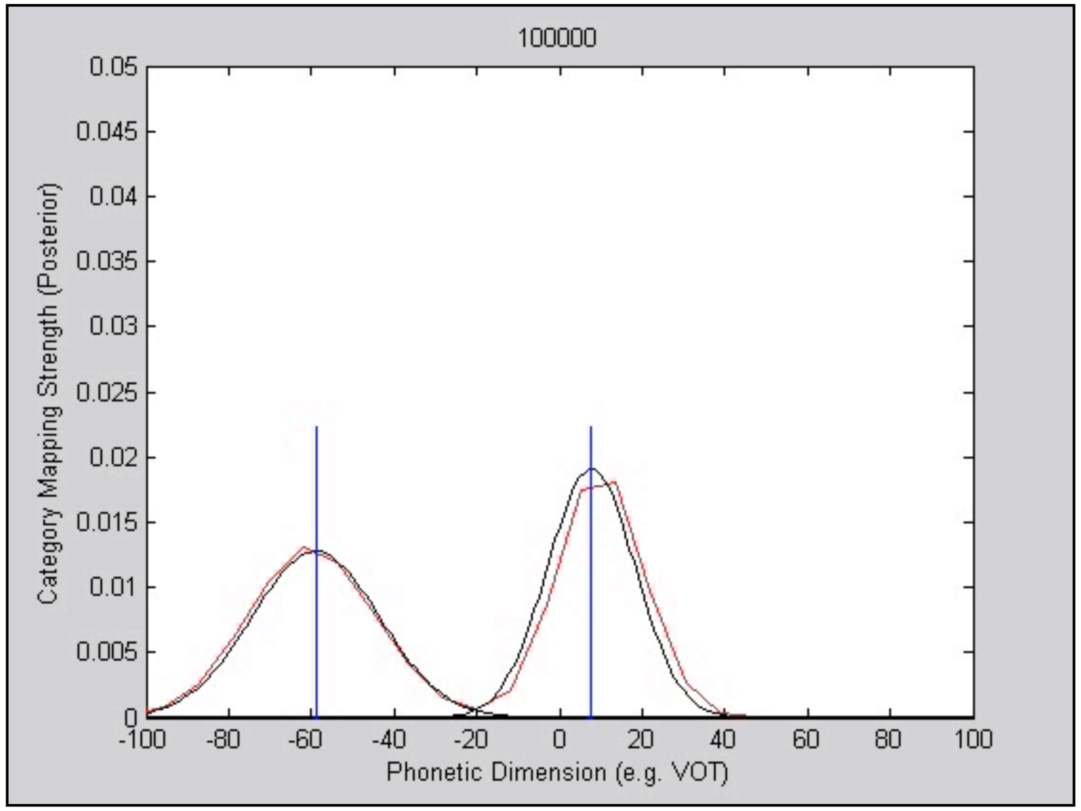
Competition

- Allows the model to determine the correct number of categories

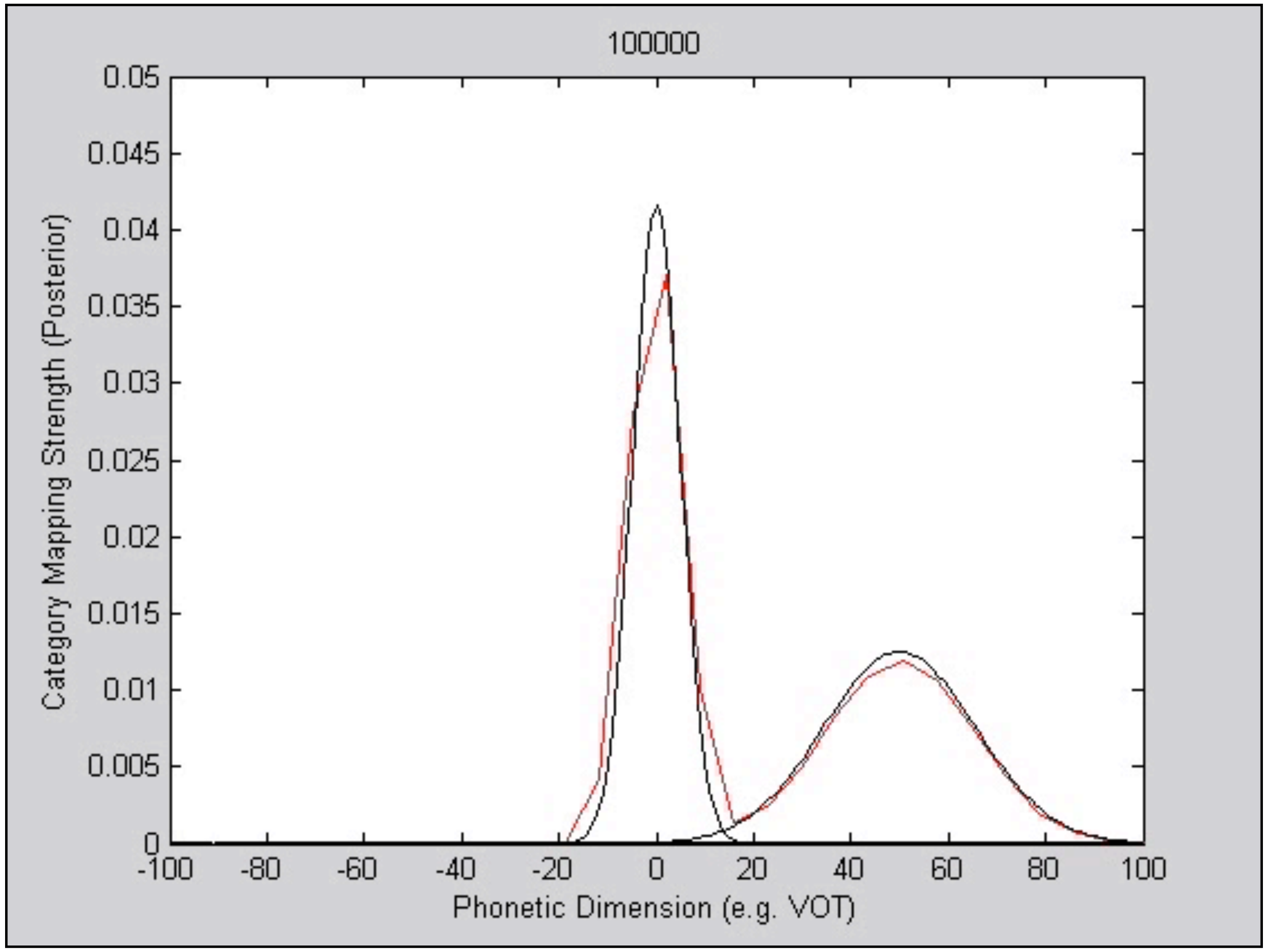
McMurray, Aslin, & Toscano (2009); Toscano & McMurray (2010)

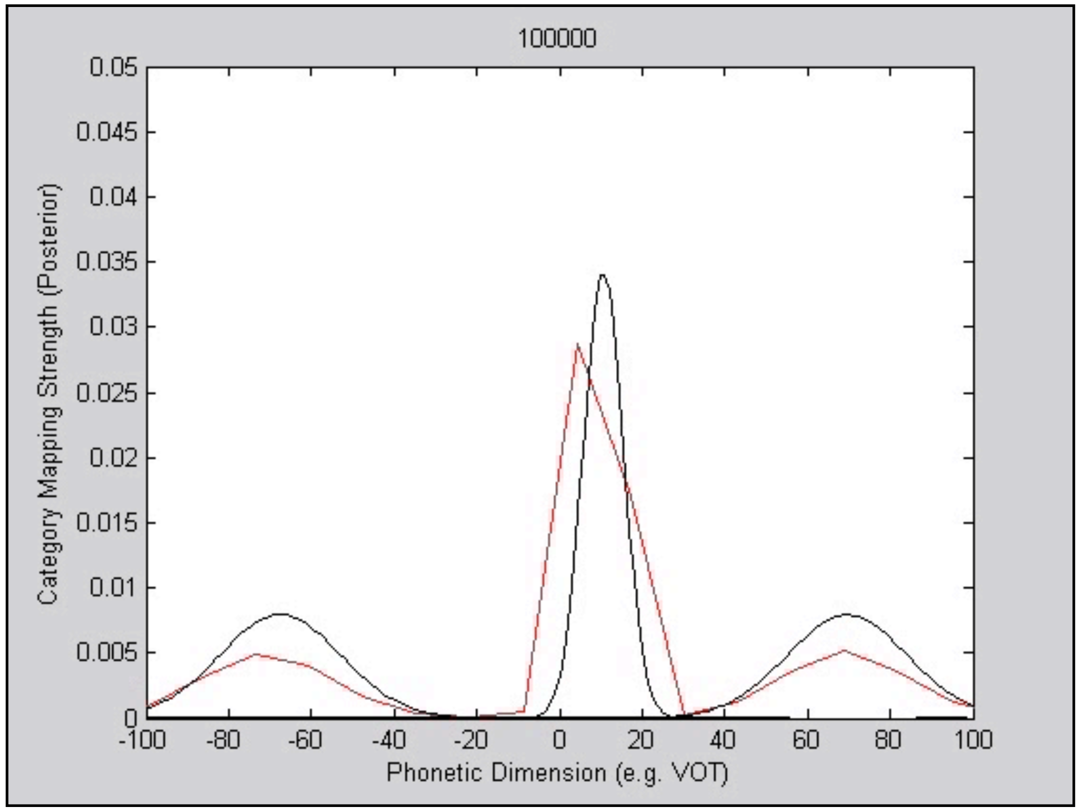
Acquiring phonetic categories

Spanish VOTs



English VOTs





Thai VOTs

McMurray, Aslin, & Toscano (2009); Toscano & McMurray (2010)

Acquiring phonetic categories

The model can learn the correct categories for a variety of acoustic cues and phonological distinctions across different languages

Makes few assumptions:

- Unsupervised, incremental learning
- Competition between categories
- Small number of parameters (3) used to describe each category

McMurray, Aslin, & Toscano (2009); Toscano & McMurray (2010)

Overview

Modeling approach

- Gaussian mixture model
- Statistical learning and competition

Acquisition during development

- Simulation 1: Determining the number of categories and their properties

Adaptation in the same model

- Simulation 2: Perceptual learning of shifted VOT distributions

Other aspects of perceptual learning in the model

- Simulation 3: Speaking rate adaptation
- Simulation 4: Learning new phonetic categories
- Simulation 5: Learning the categories of a second language

Learning and adapting categories in a single model

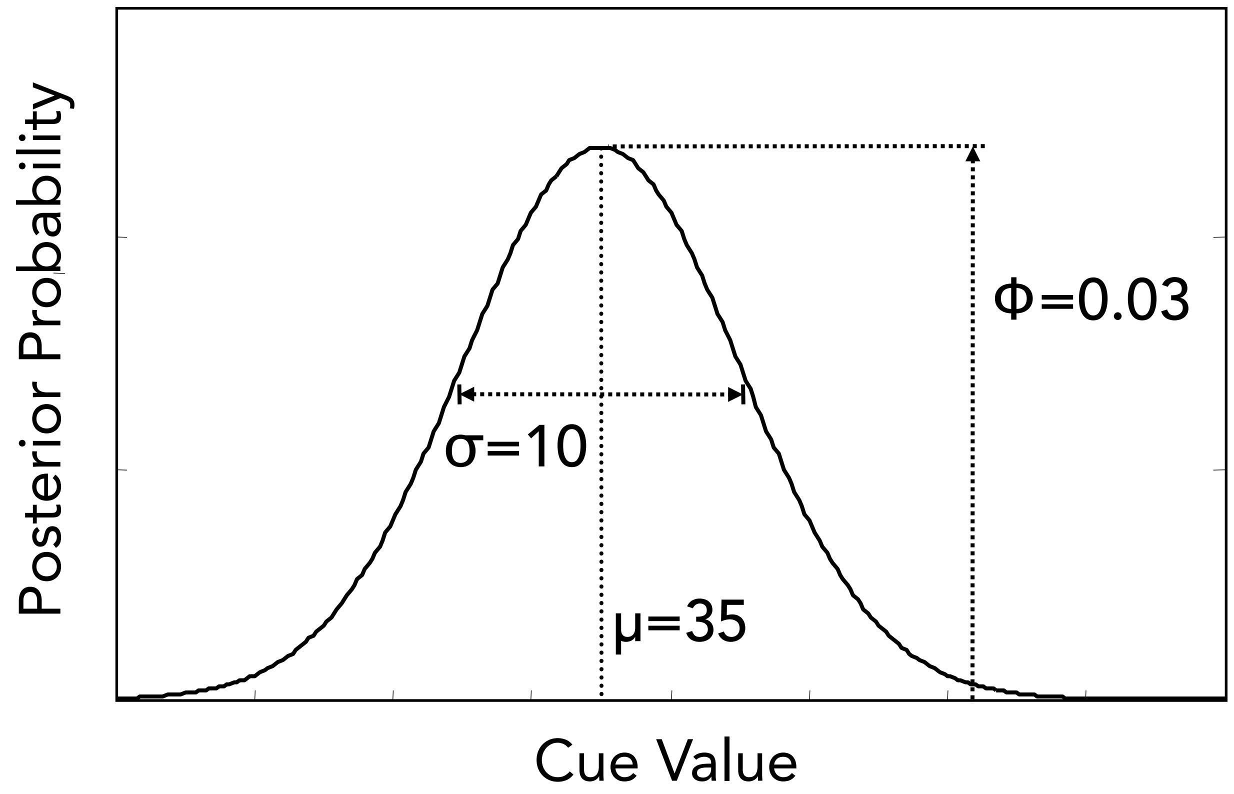
Can the same model adjust its categories in an adaptation experiment?

- Without changes in learning rates?
- Without separate long- and short-term representations of categories?

Examined this by exploring model parameter space

Compared model's responses with listeners from Munson (2011)

Learning and adapting categories in a single model



Gaussian mixture model (GMM)

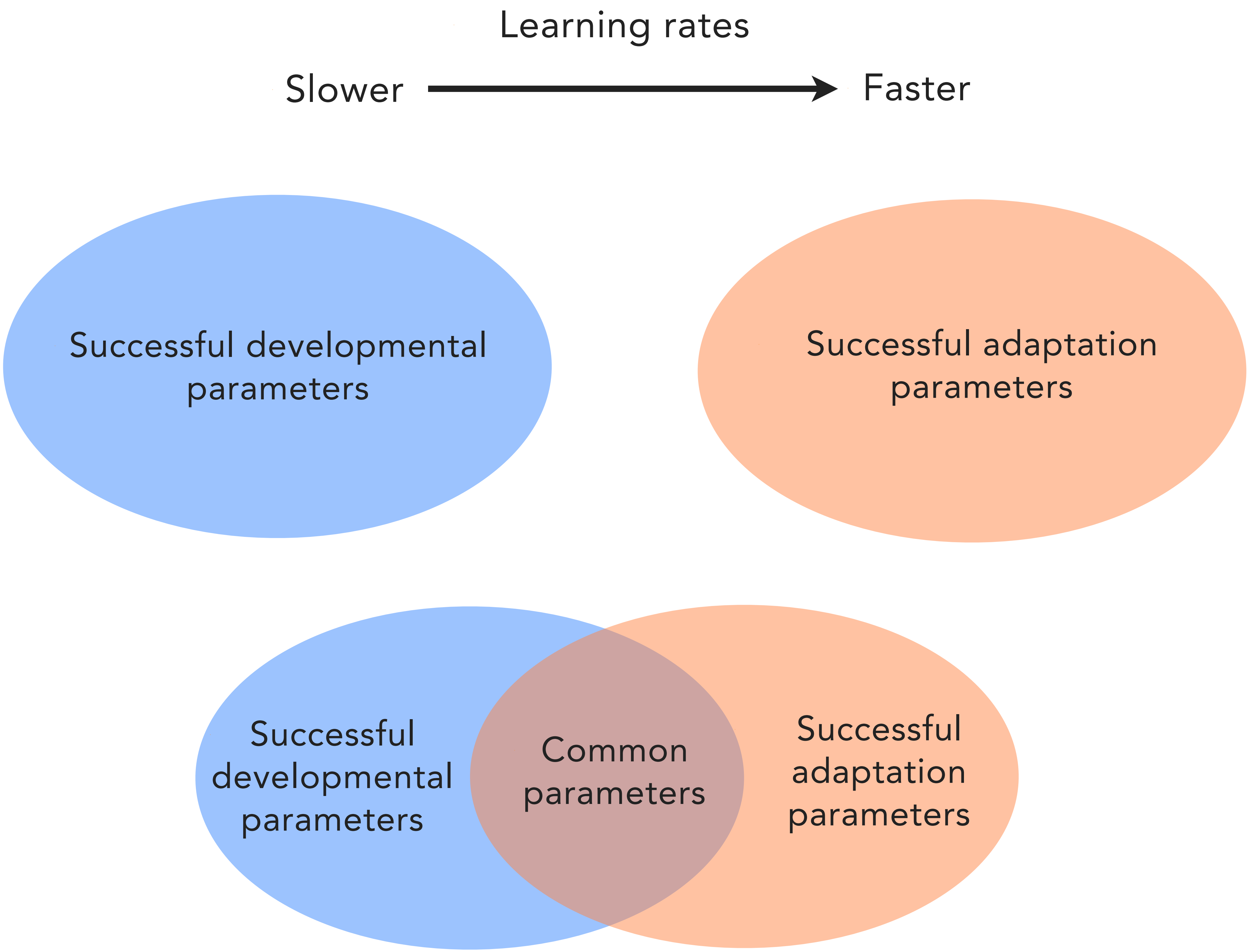
- Categories defined by Gaussian distributions
- Mean (μ)
- Standard deviation (σ)
- Likelihood (Φ)

Each parameter has a learning rate associated with it

μ	0.5	1	2	4	8	...
σ	0.1	0.2	0.4	0.8	1.6	...
Φ	0.01	0.02	0.04	0.08	0.16	...

McMurray, Aslin, & Toscano (2009)

Learning and adapting categories in a single model



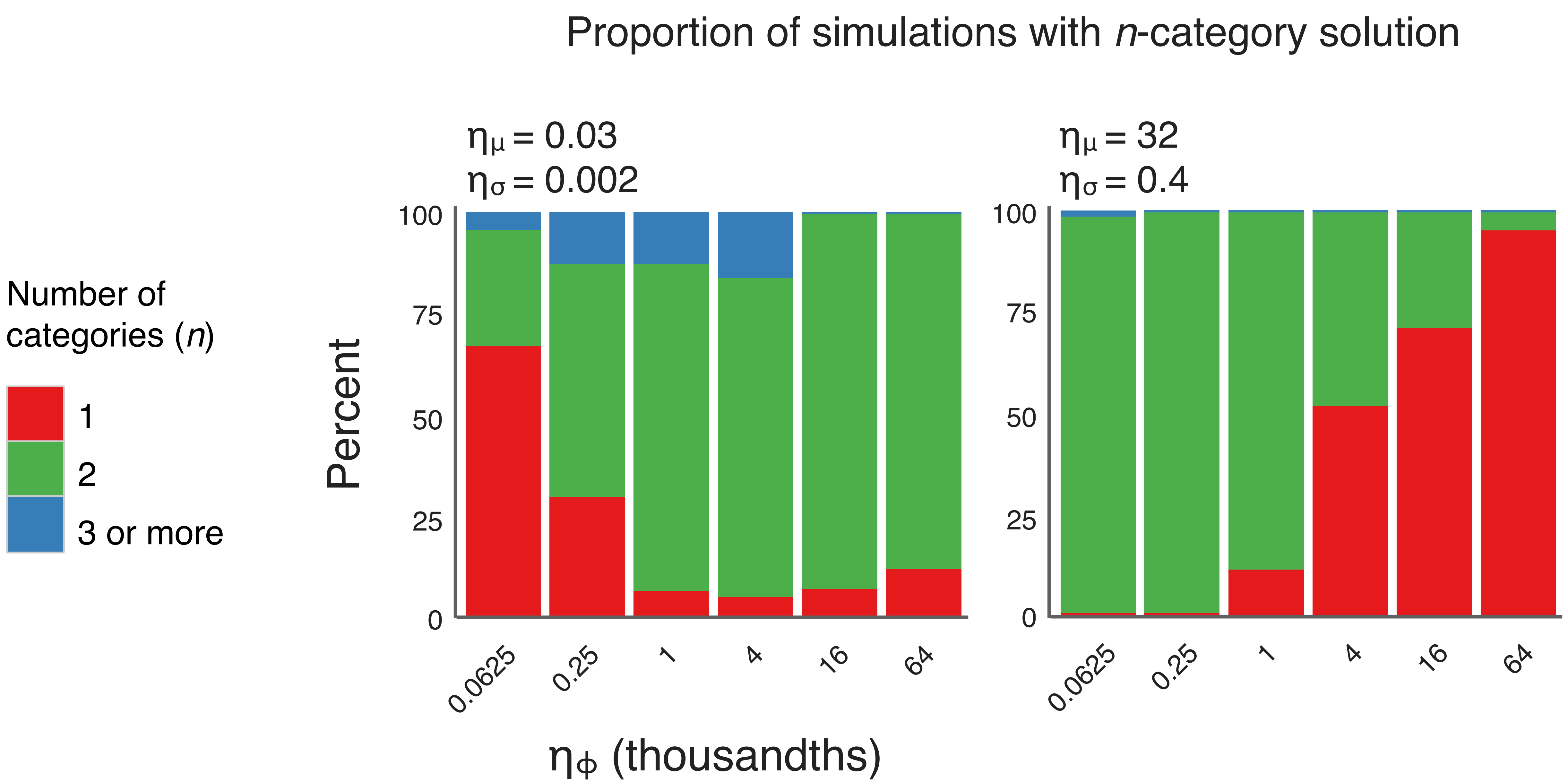
Learning and adapting categories in a single model

Ran simulations exploring the parameter space of the model

- Which learning rates yield successful development (generally slower?)
- Which yield successful perceptual learning (generally faster?)
- Are there learning rates that are common to both?

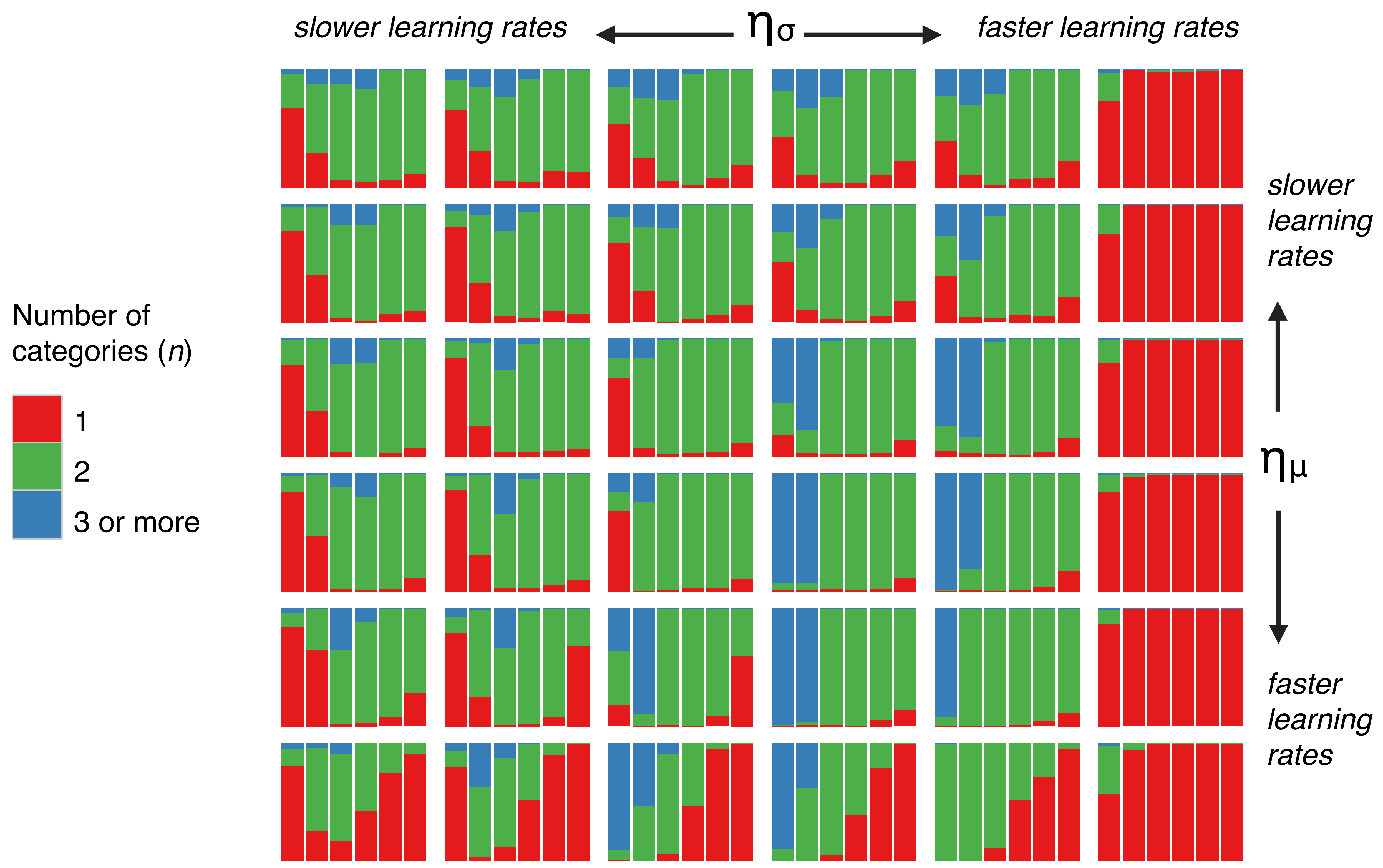
Learning and adapting categories in a single model

Which learning rates yield successful development?



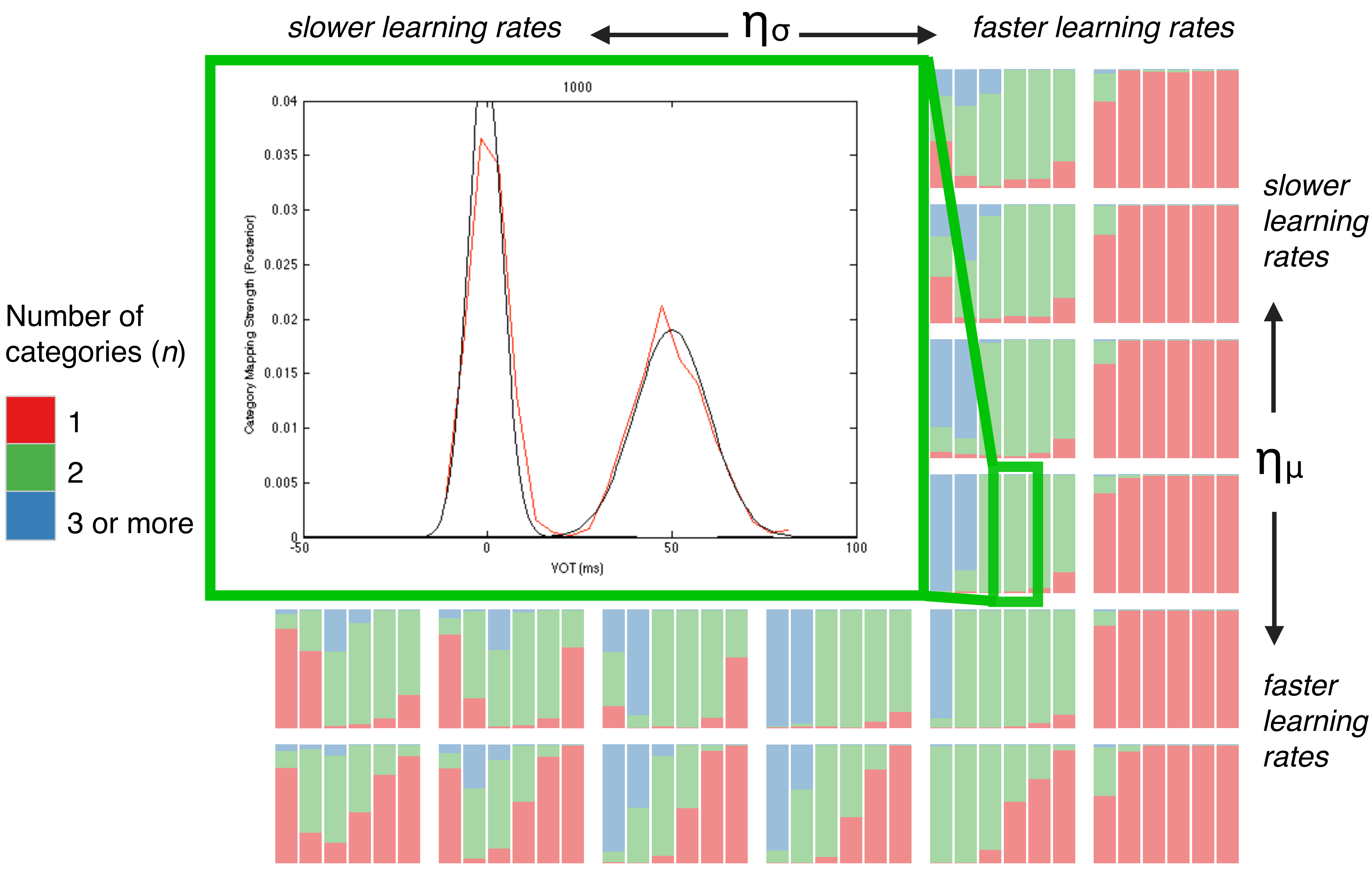
Learning and adapting categories in a single model

Which learning rates yield successful development?



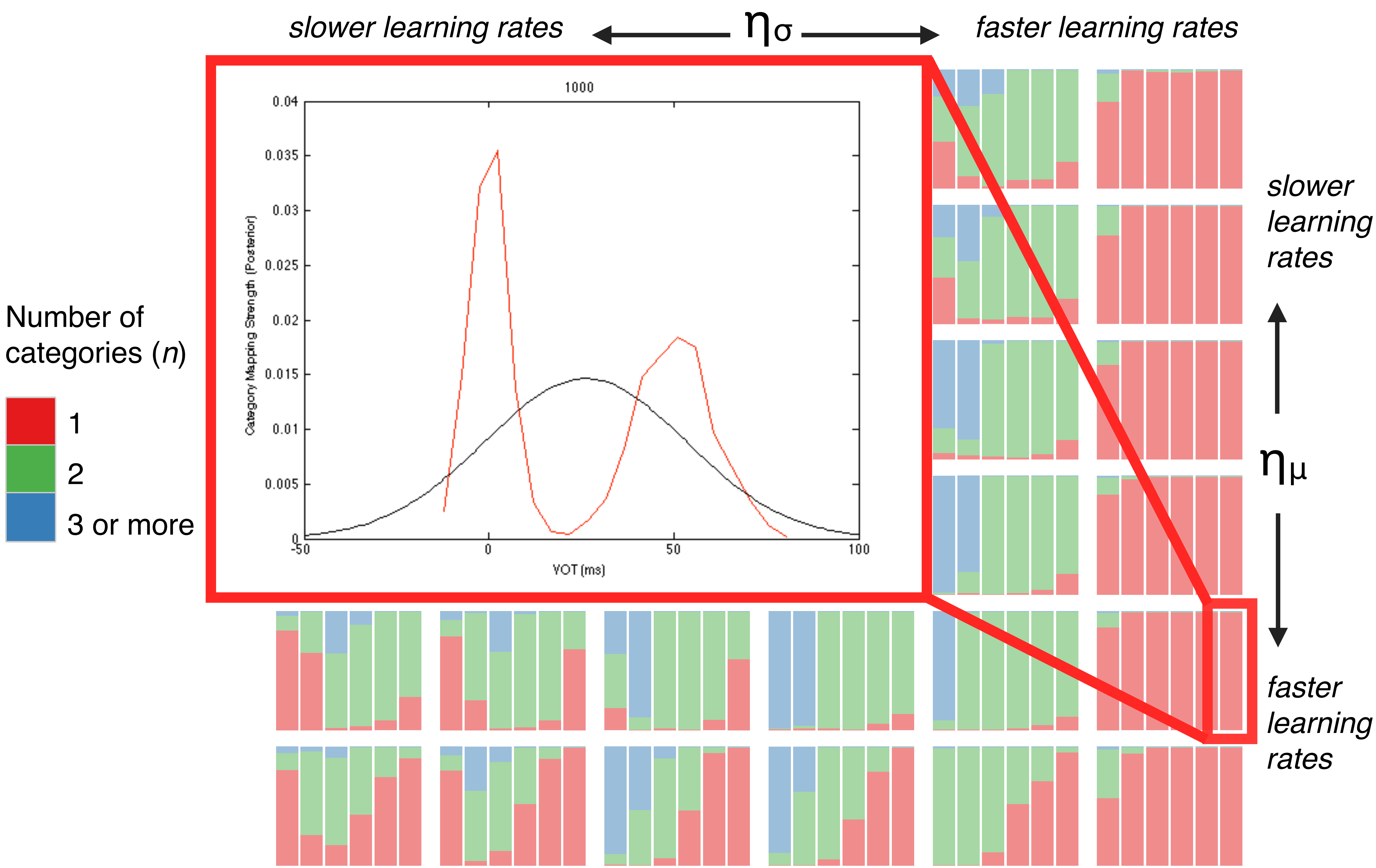
Learning and adapting categories in a single model

Which learning rates yield successful development?



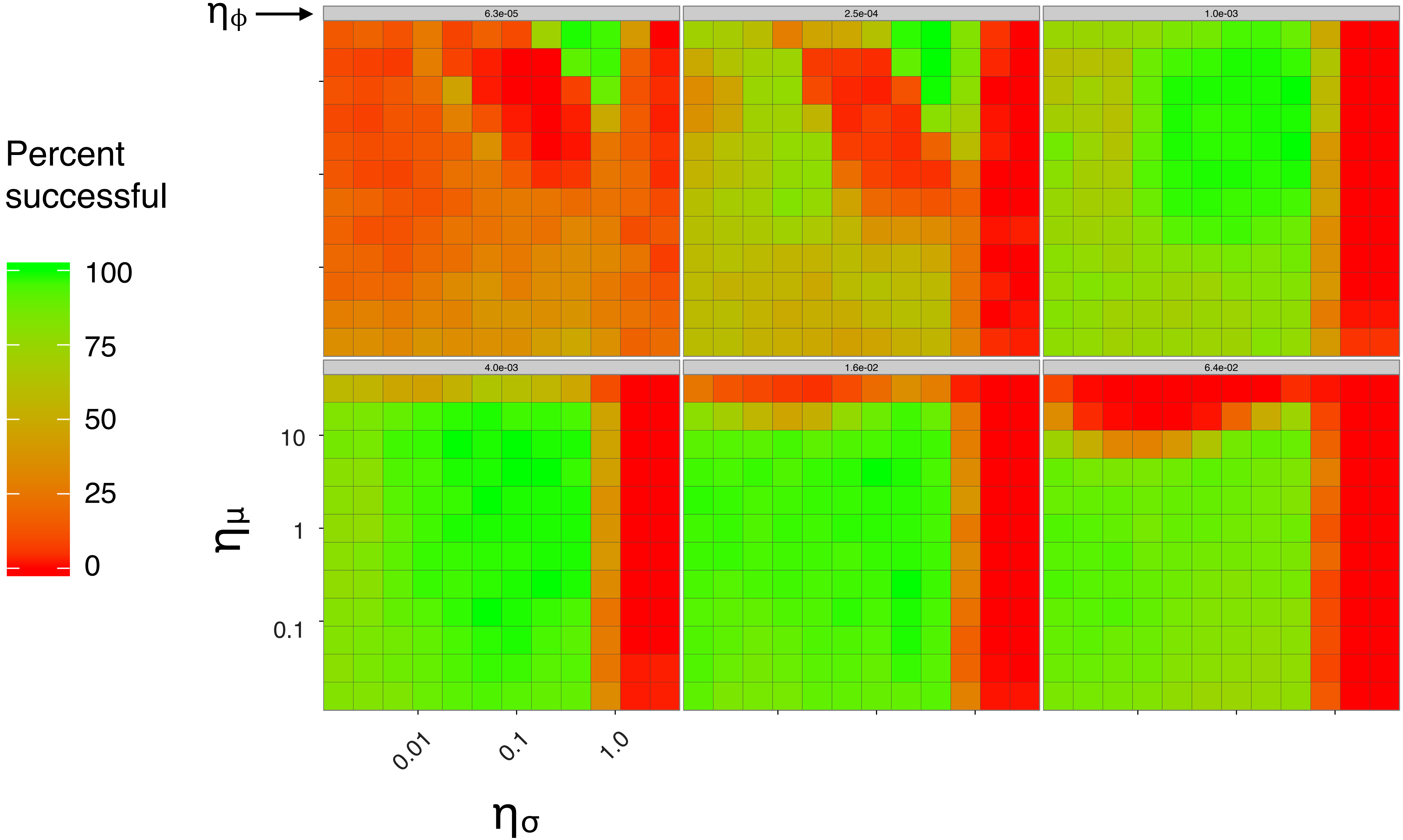
Learning and adapting categories in a single model

Which learning rates yield successful development?



Learning and adapting categories in a single model

Which learning rates yield successful development?



Learning and adapting categories in a single model

Results of developmental simulation

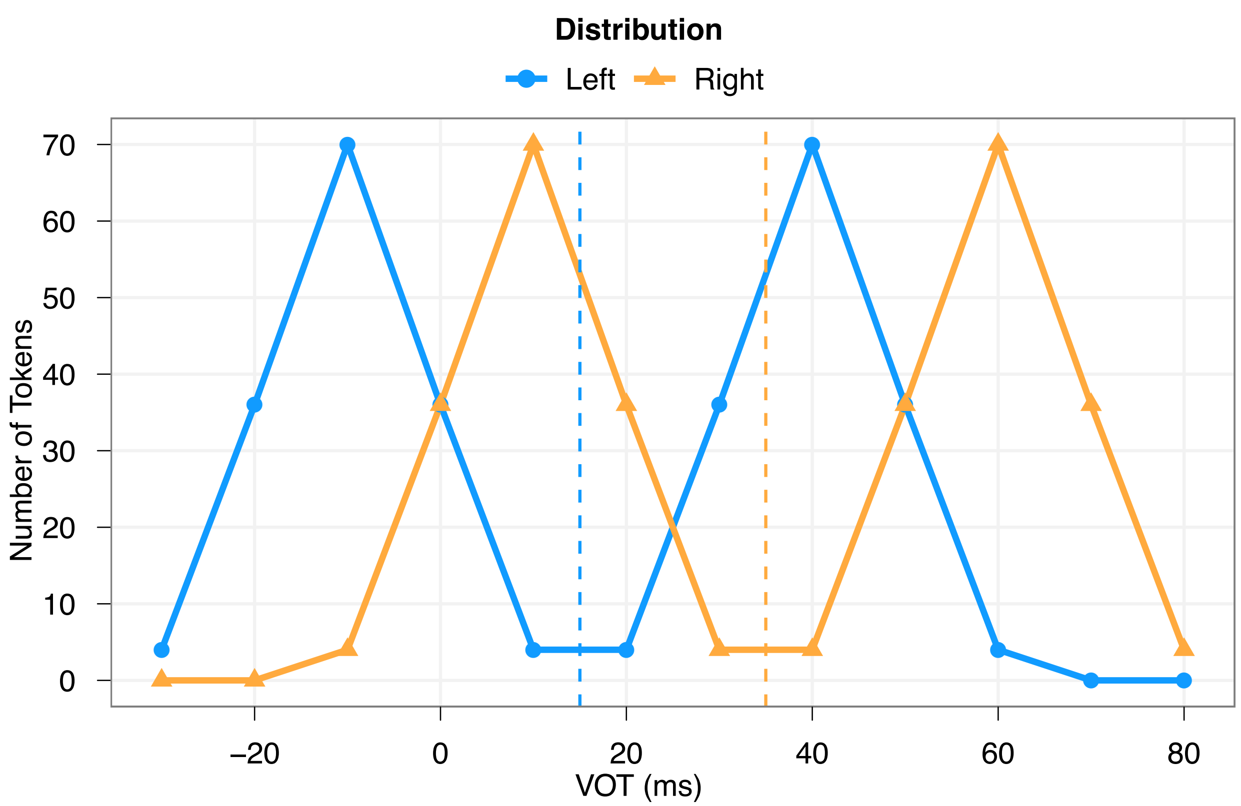
- A range of learning rates leads to successful category acquisition
- Demonstrates that the model is relatively flexible in its ability to discover the category structure over development

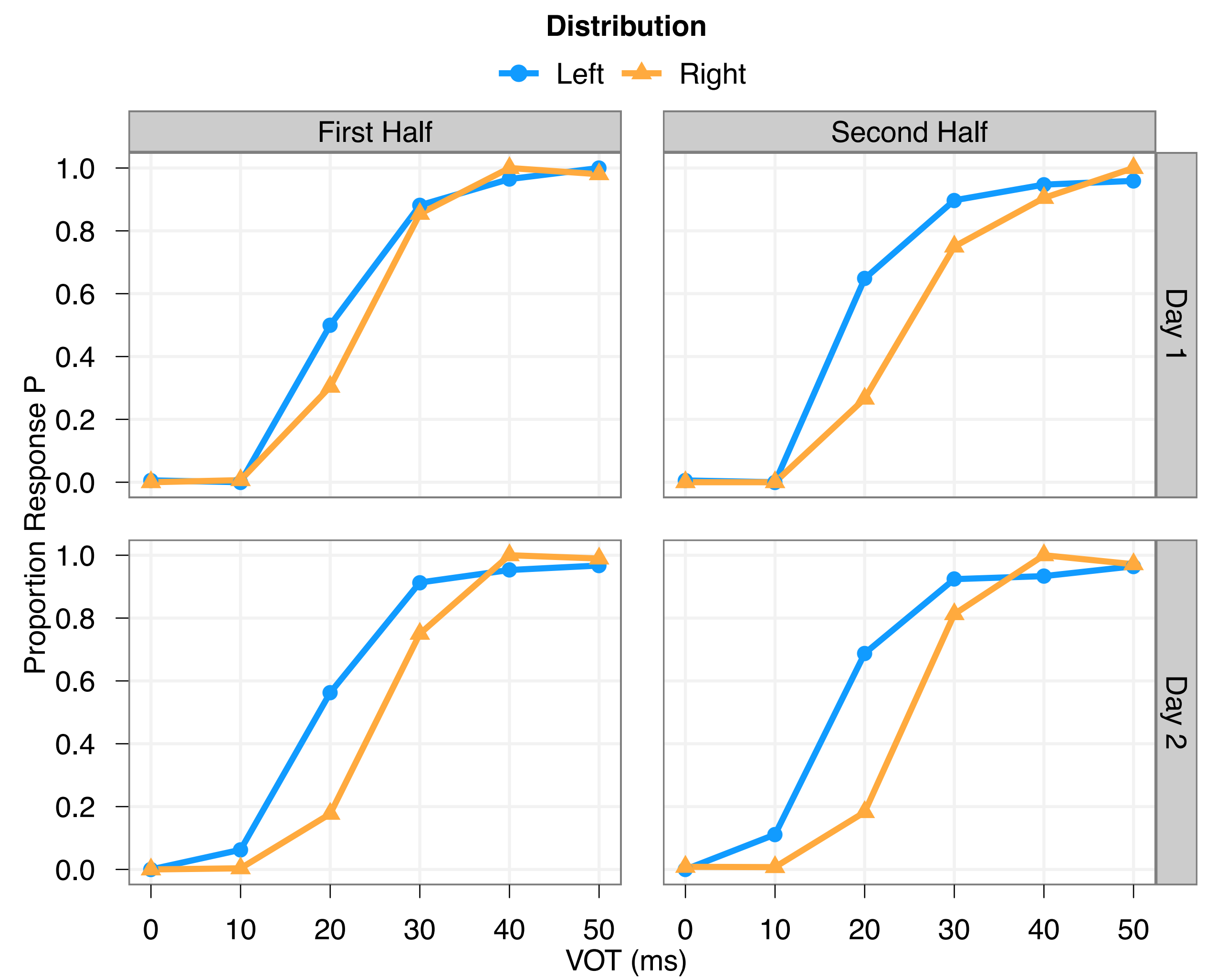
Next question: do some of these learning rates also lead to successful adaptation?

Learning and adapting categories in a single model

Can the model capture learning effect seen for listeners in Munson (2011)?

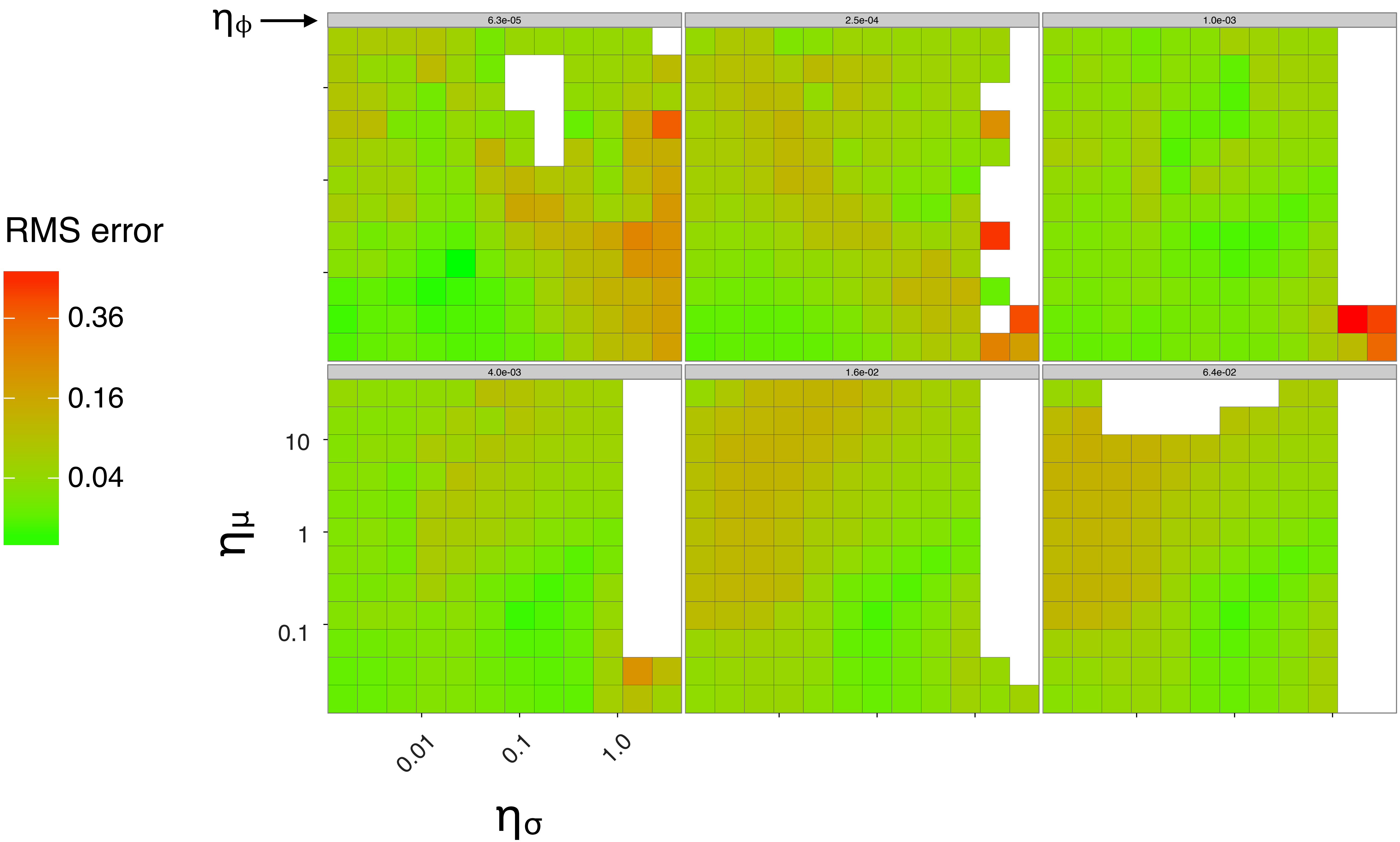
- Tested model in same adaptation experiment
- Compared model and listener responses across sets of learning rates





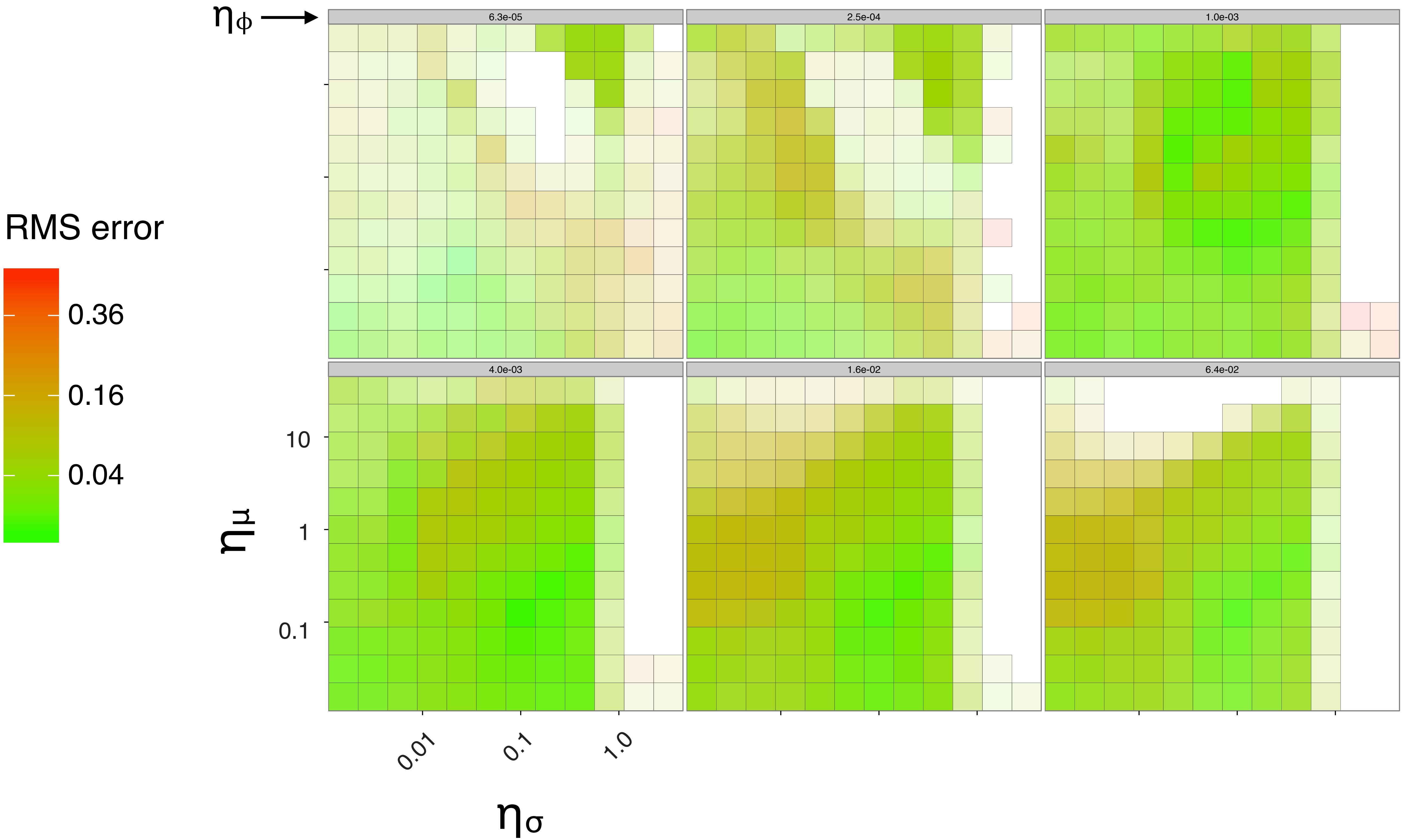
Learning and adapting categories in a single model

Can the model capture learning effect seen for listeners in Munson (2011)?



Learning and adapting categories in a single model

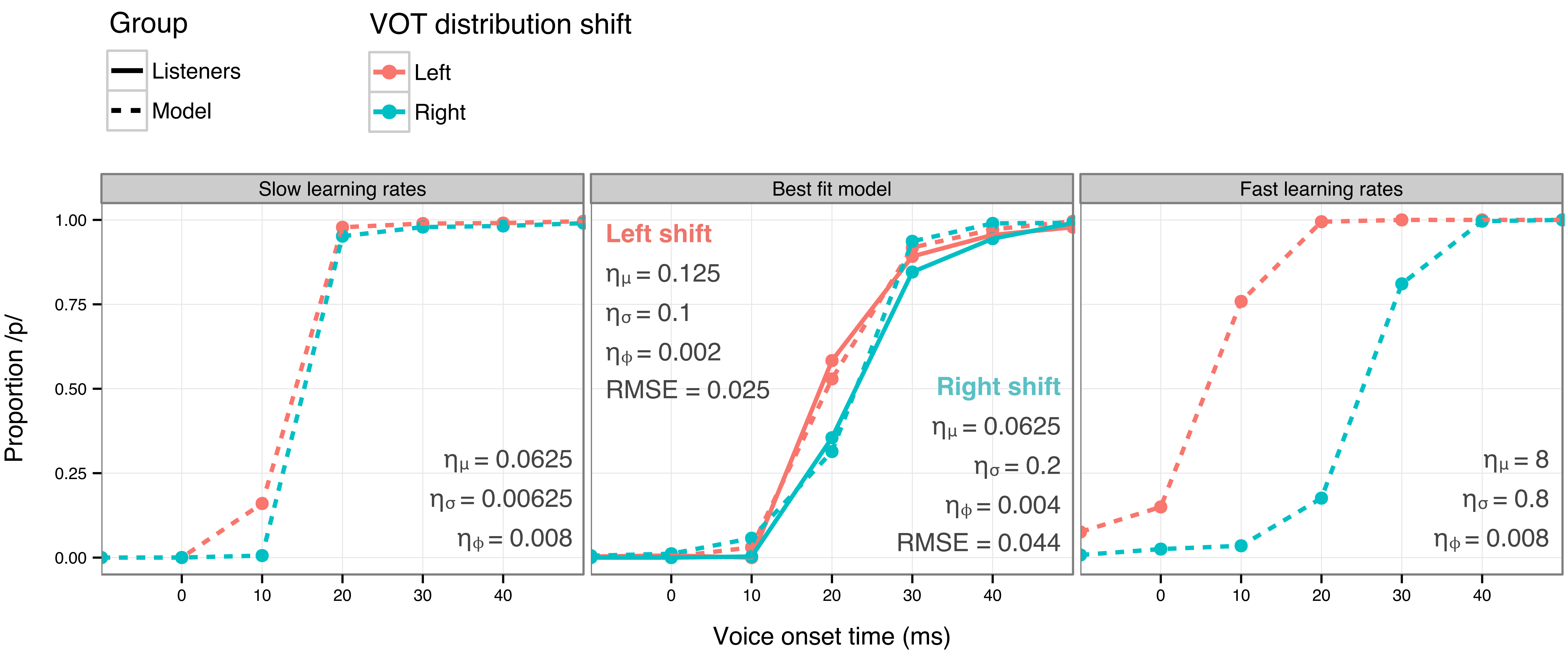
Can the model capture learning effect seen for listeners in Munson (2011)?



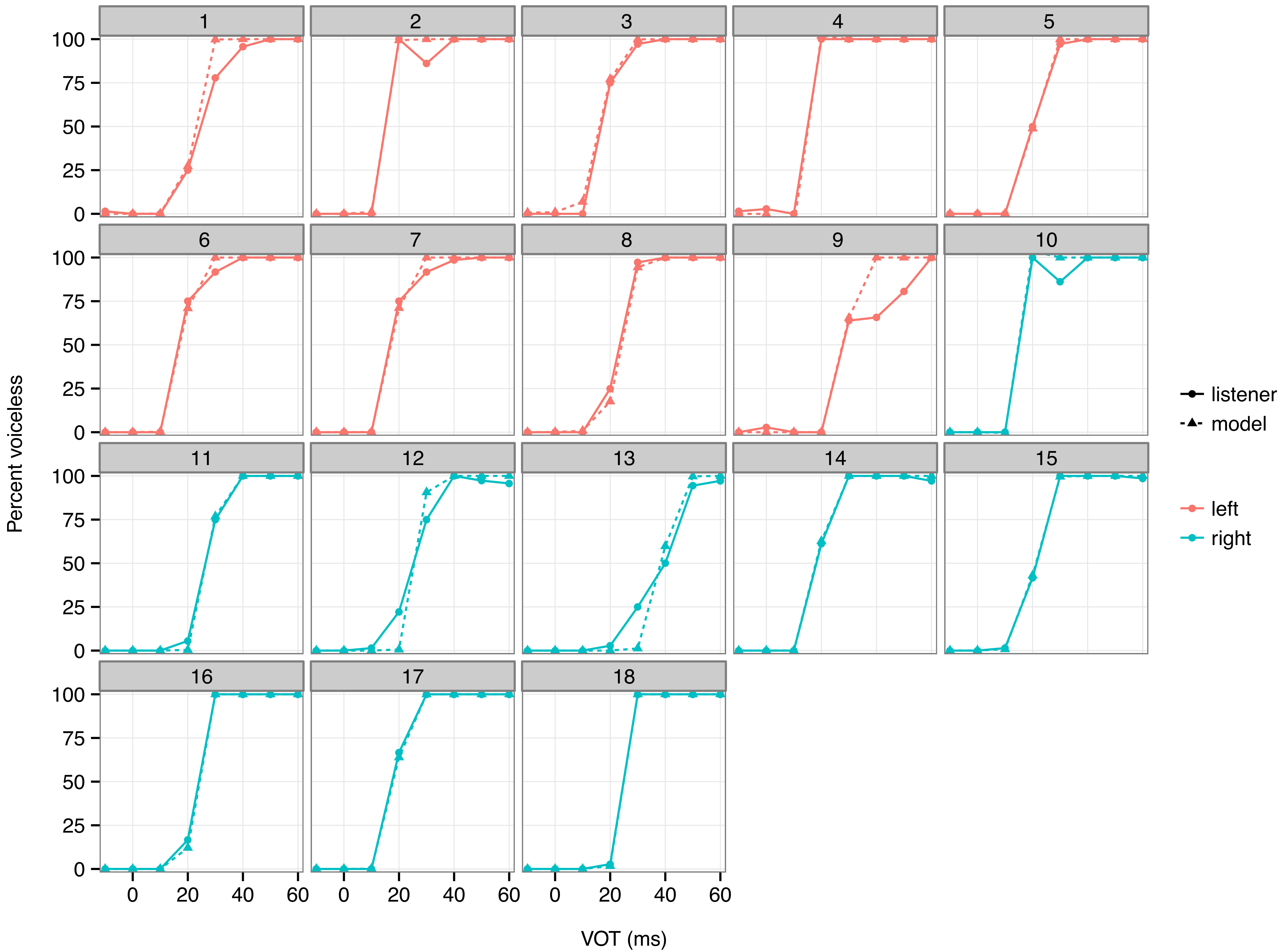
Learning and adapting categories in a single model

Can the model capture learning effect seen for listeners in Munson (2011)?

- Model accurately captures responses to left- and rightward shifted distributions
- Can also model individual differences



Learning and adapting categories in a single model



Learning and adapting categories in a single model

A single model can capture both ***acquisition*** of speech sound categories during development and ***adaptation*** in adulthood

- Simple unsupervised learning procedure
- No changes in model plasticity over development
- Represents a “minimal description” of the process

Overview

Modeling approach

- Gaussian mixture model
- Statistical learning and competition

Acquisition during development

- Simulation 1: Determining the number of categories and their properties

Adaptation in the same model

- Simulation 2: Perceptual learning of shifted VOT distributions

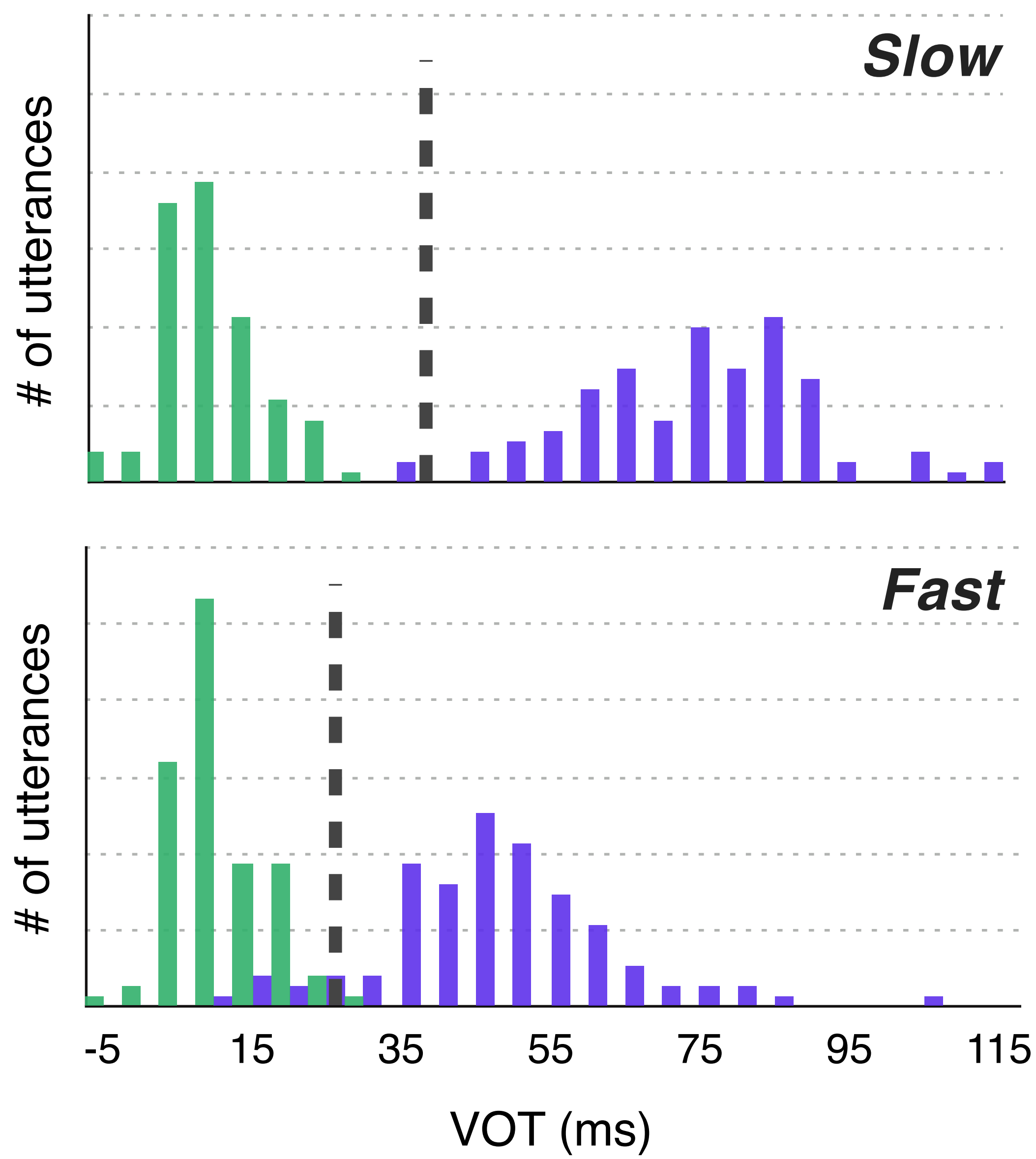
Other aspects of perceptual learning in the model

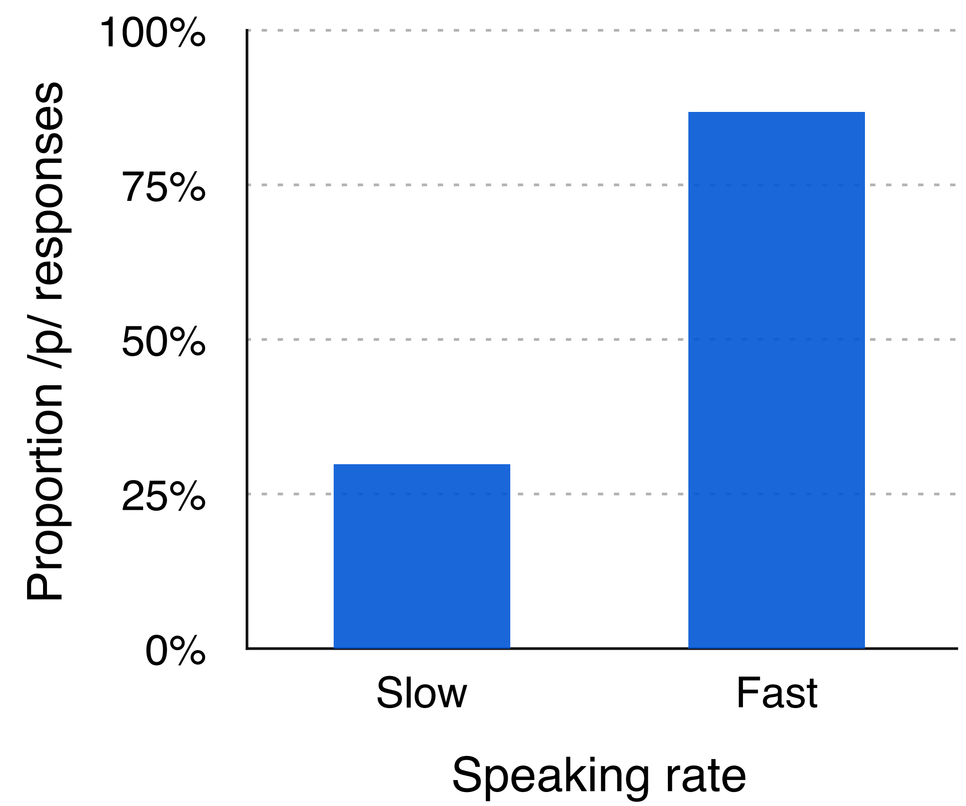
- Simulation 3: Speaking rate adaptation
- Simulation 4: Learning new phonetic categories
- Simulation 5: Learning the categories of a second language

Adapting phonetic categories

Simulation 2: *Speaking rate adaptation*

- Can the model update its VOT representations in the context of variable speaking rates?



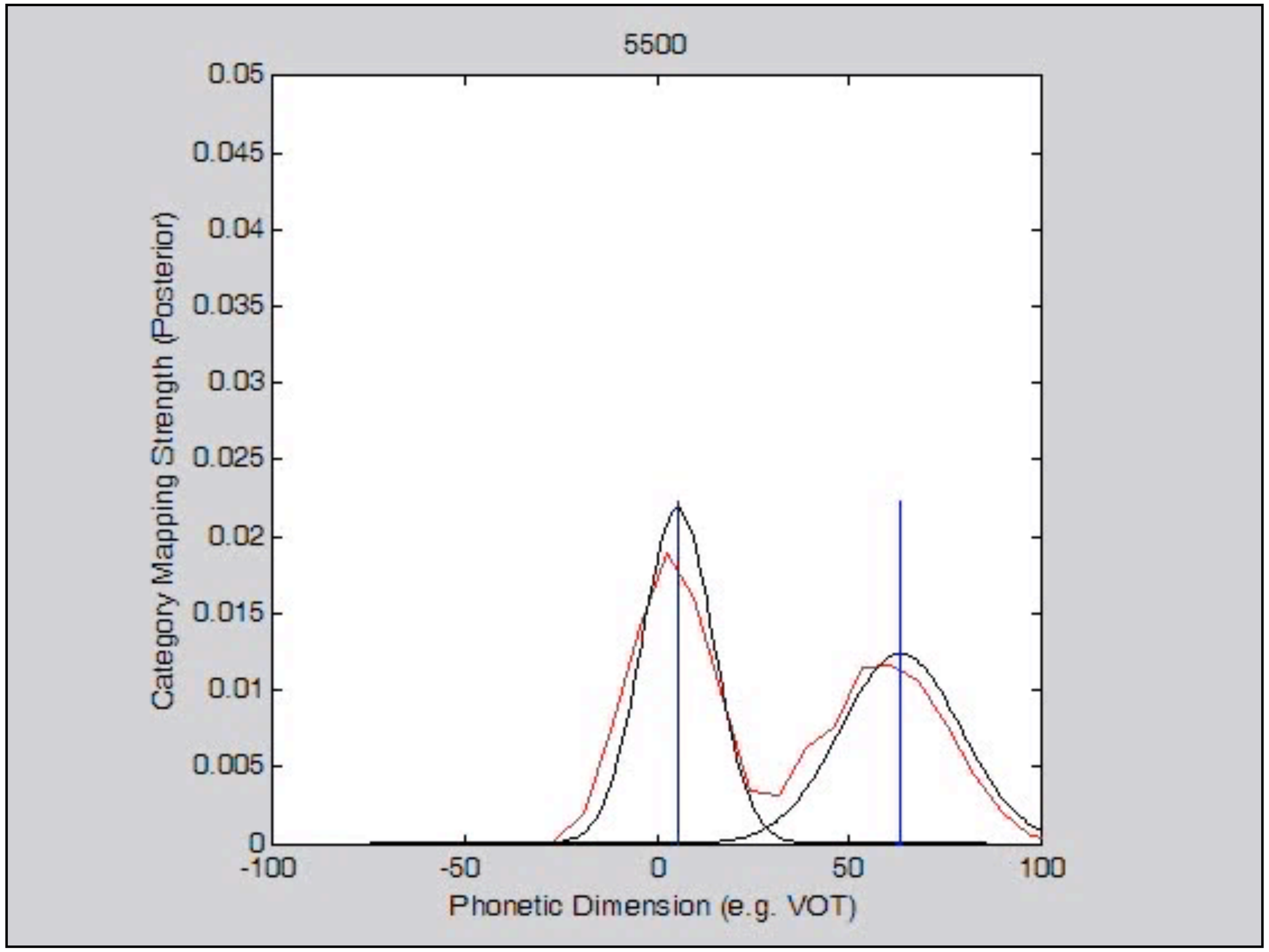


Toscano & McMurray (2012), *Attn Percep & Psychophys*; Toscano & McMurray (submitted)

Adapting phonetic categories

Simulation 2: *Speaking rate adaptation*

- Can the model update its VOT representations in the context of variable speaking rates?

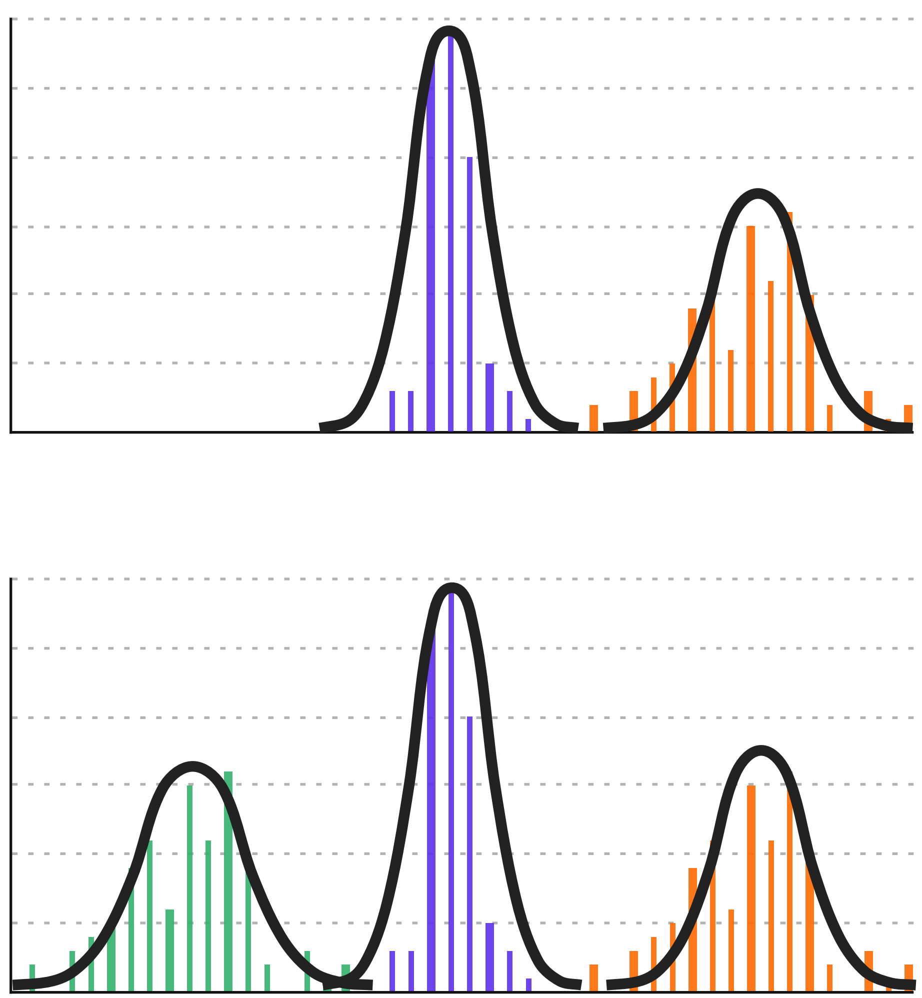


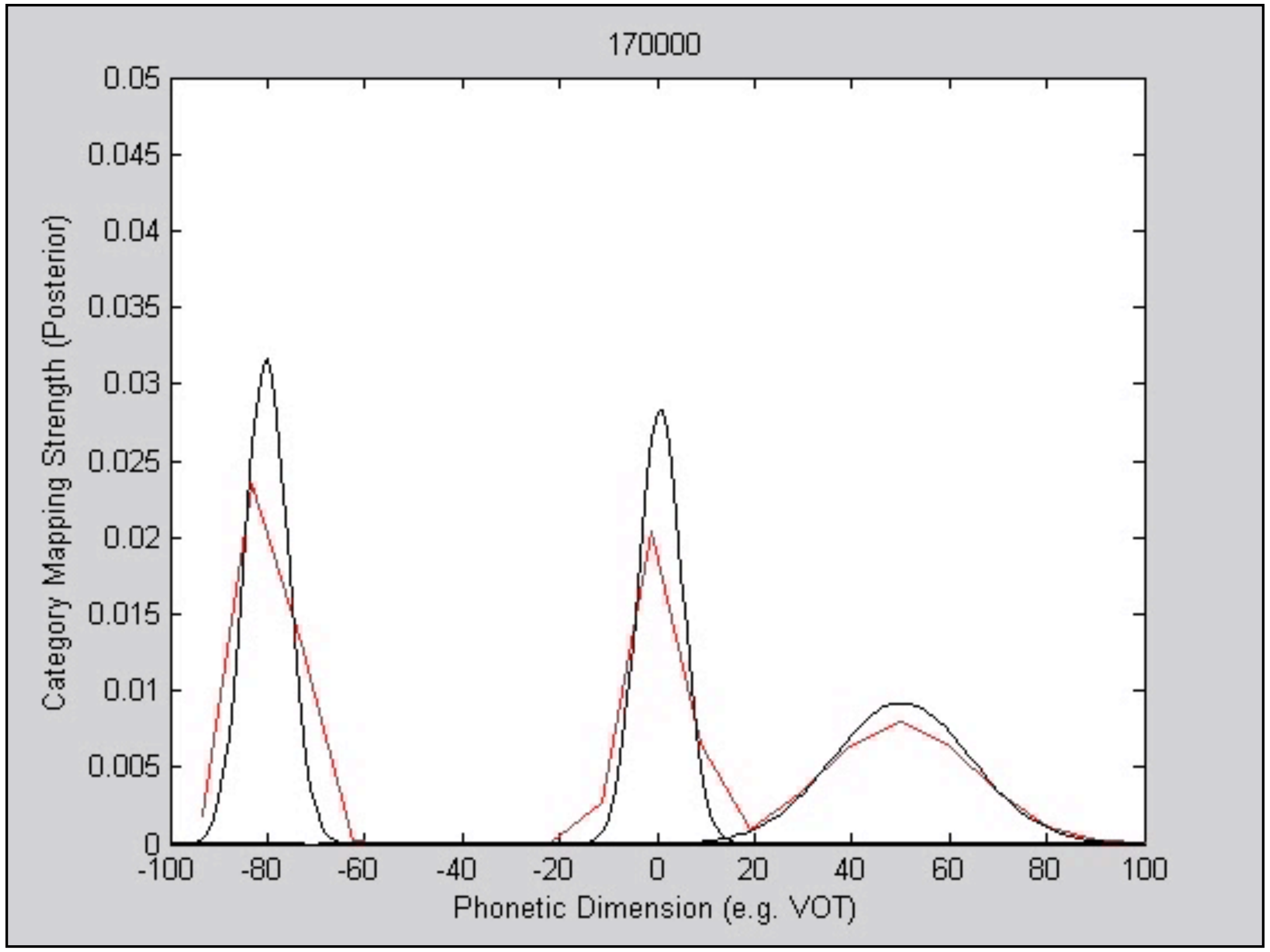
McMurray, Horst, Toscano, & Samuelson (2009)

Adapting phonetic categories

Simulation 3: *Learning a new category*

- Pisoni, Alsin, Perry, & Hennessy (1982)
- 3-way voicing distinction based on VOT

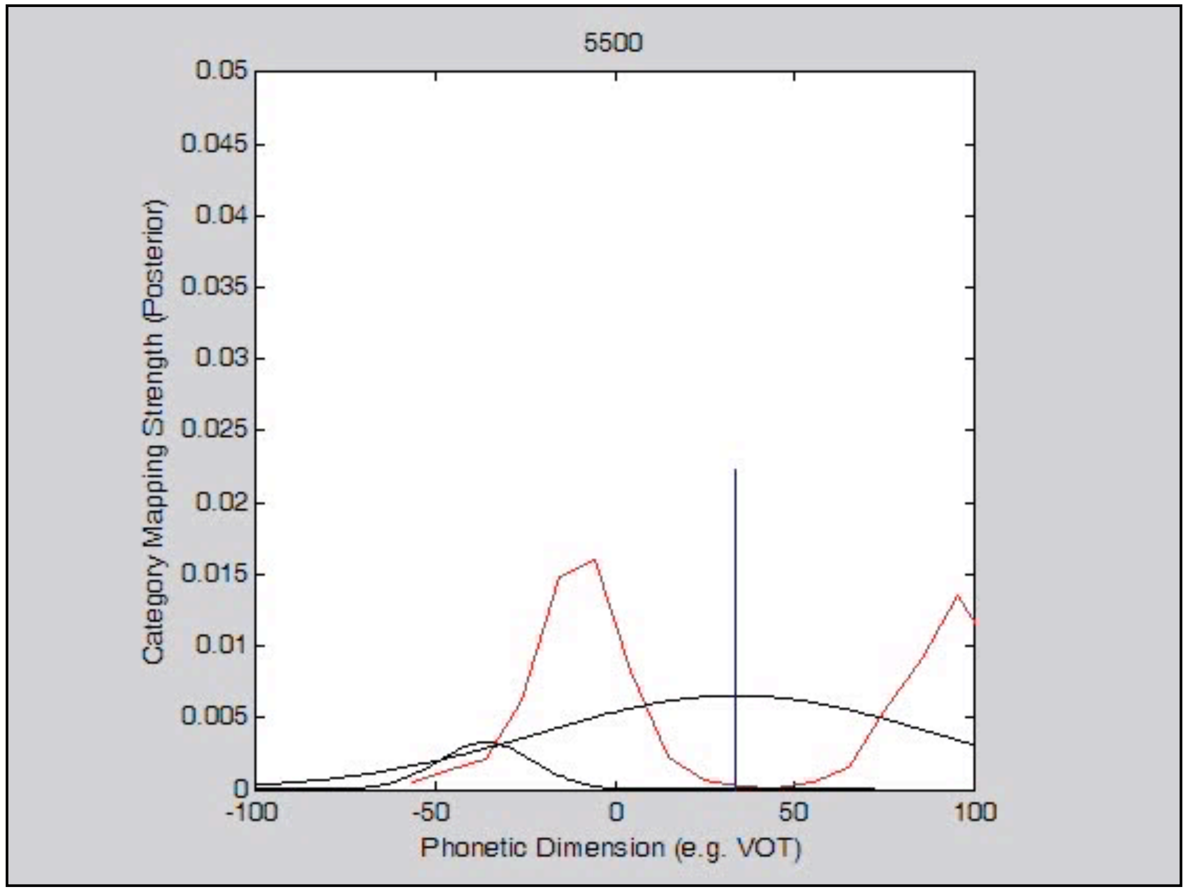




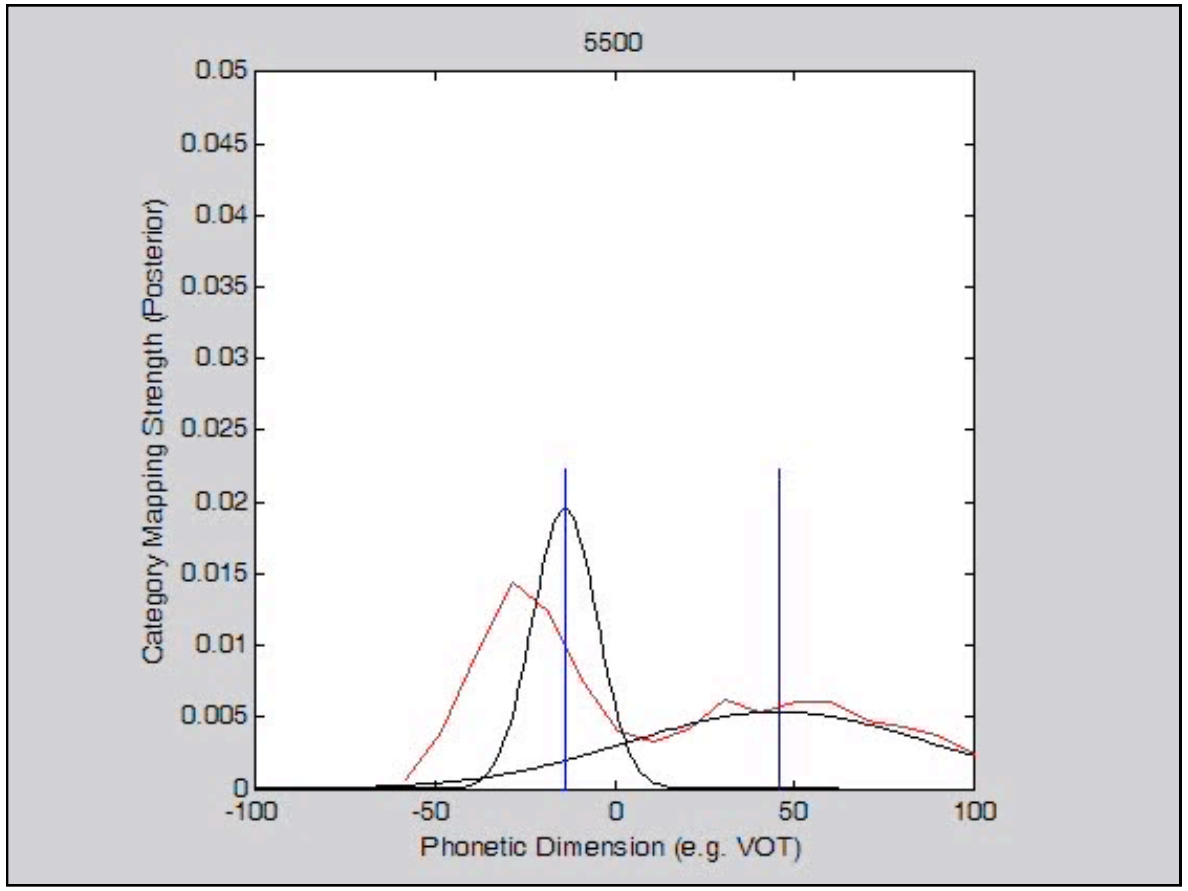
Potential implications for second language learning

Gradual vs. discontinuous changes in language environment

Discontinuous shift



Gradual shift



Summary and conclusions

A single model can capture both ***acquisition*** of phonetic categories during development and ***adaptation*** in adulthood

- Simple unsupervised learning procedure
- No changes in model plasticity over development
- Represents a "minimal description" of the process
- No need to have separate representations for acquisition and adaptation

This suggests that

- aspects of perceptual adaptation can be explained by changes to long-term representation of phonetic categories
- the same learning mechanism can operate over vastly different time-scales

Thanks!