Training data

Model Architecture:
- Acoustic-phonetic categories represented as a mixture of Gaussians
  (McMurray et al., 2009; Toscano & McMurray, 2010; Valabha et al., 2007)
- Vowel categories represented by bivariate (2D) Gaussians (F1xF2 space); gender categories as univariate (1D) Gaussians (F0)
- Each 1D Gaussian has 3 parameters: mean (μ), standard deviation (σ), and likelihood (p)
- 2D Gaussians have μ- and σ-values for each dimension as well as a parameter for the correlation between F1 & F2 (ρ)

Training data:
- Model trained on distributions of English vowel sounds based on measurements from from Hilbrandt et al. (1995)

METHOD

INTRODUCTION

The current models address this by specifying context a priori, a type of supervised learning (Cole, Lindeburg, Monroe, & McMurray, 2008; McMurray & Jongman, 2011).

• But this is not developmentally plausible: infants use unsupervised statistical learning to acquire speech sound categories (Saffran, Aslin, & Newport, 1996; Maye, Werker, & Garner, 2003).

• Goal: Develop a model that compensates for contextual differences using developmentally-realistic learning processes

RESULTS

Simulation 1: Female talkers only
- Baseline performance for female talkers
- Learned correct number of categories

Simulation 2: Male talkers only
- Baseline performance for male talkers
- Some vowel mergers (e.g., /a/ and /A/)

Simulation 3: Both groups of talkers; no context compensation
- Representative model runs shown
- Numerous mergers; low classification accuracy
- Same vowel often mapped onto separate categories (one male, one female)

Simulation 4: Both groups of talkers; context compensation via f0 mixture
- Representative model runs shown
- Generally, model was successful in separating vowels based on talker gender (estimated via f0)
- Categories in each F1xF2 space mapped onto correct vowels for each group of talkers

DISCUSSION

Testing procedure
- Success evaluated by measuring (1) number of above-threshold categories after training and (2) classification accuracy for a set of 500 test tokens
- Euclidean distance between model component means and test token calculated; shortest distance corresponds to model classification

ACCOUNTING FOR TALKER VARIABILITY

• Accounting for talker variability allowed the model to successfully learn vowel categories
• When categories are mapped to separate vowel spaces according to gender, classification accuracy improves 16.5% percent (Sim 4) relative to performance without compensation (Sim 3)
• Provides a developmentally-plausible learning mechanism that makes minimal assumptions about pre-existing knowledge to account for contextual variability in speech

ACKNOWLEDGEMENTS & REFERENCES

This work was supported by a Psychonomic Society Graduate Student Travel Award to TC.


