Speech Perception Across the Lifespan: Using a Gaussian Mixture Model to Understand Changes in Cue Weighting Between Younger and Older Adults

Sarah Vrabic
svrabic@villanova.edu

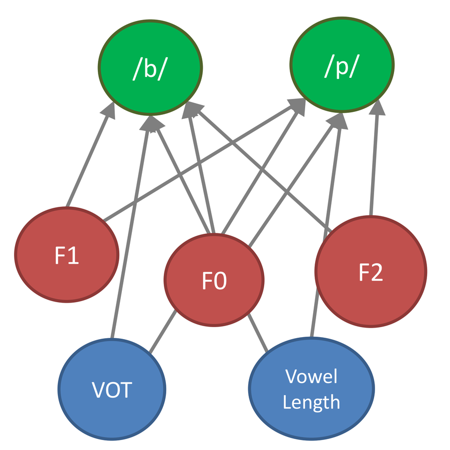
Elke Nordeen
enordeen@villanova.edu

Joe Toscano
joseph.toscano@villanova.edu

VILLANOVA UNIVERSITY

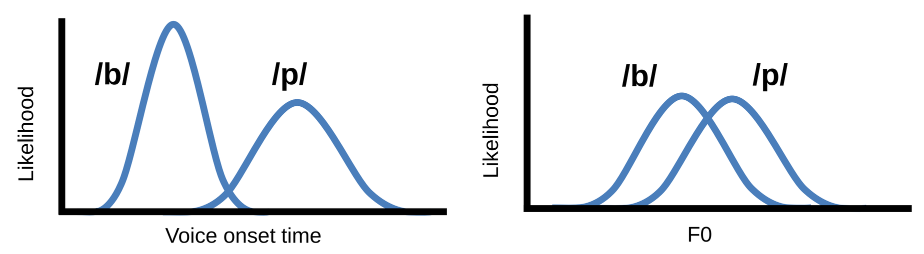
WRAP Lab
http://wraplab.co

INTRODUCTION

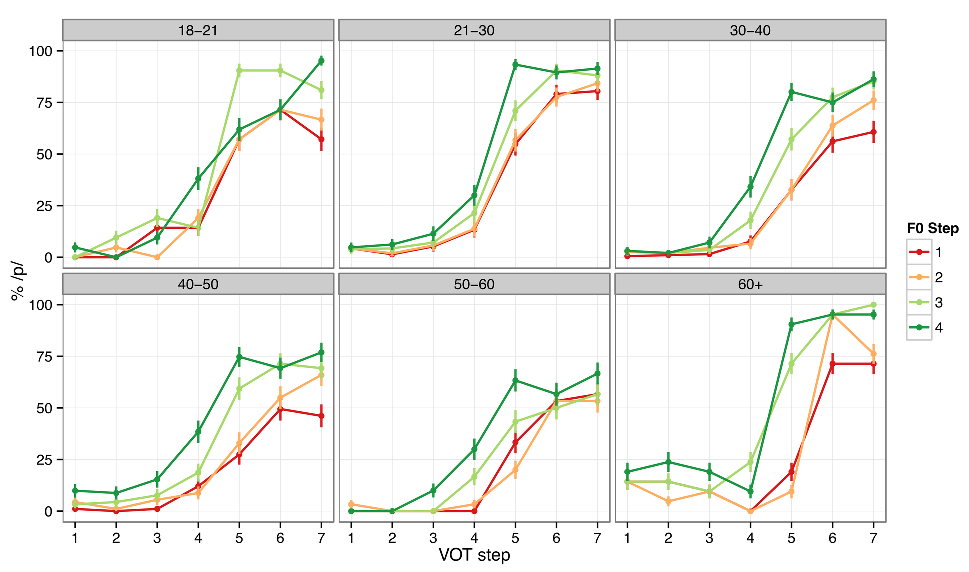


- Listeners weight and combine multiple acoustic cues to understand speech
- Both temporal and spectral cues must be integrated (e.g., voicing)

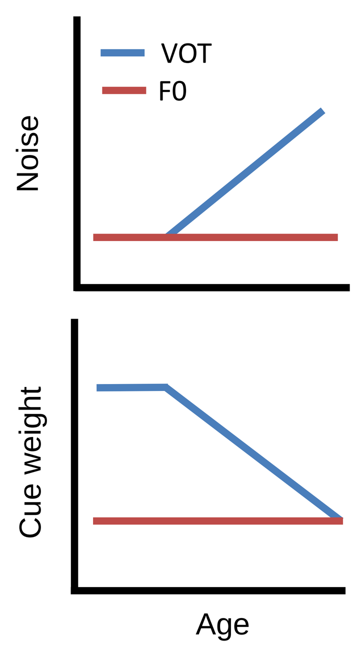
Voicing Categories
Temporal Cues
Spectral Cues



- Cues also vary in reliability
- Word-initial voicing cues:
  - Voice onset time (VOT) — Highly reliable
  - F0 onset — Less reliable
- Generally, listeners weight cues by reliability (Toscano & McMurray, 2010)
- For example, listeners generally weight VOT much more highly than F0



- Cue weights also change over time (Nittrouer, 2002)
- Normal-hearing older adults rely less on VOT than young adults, applying more weight to F0, even though F0 is less reliable (Toscano & Lansing, submitted)



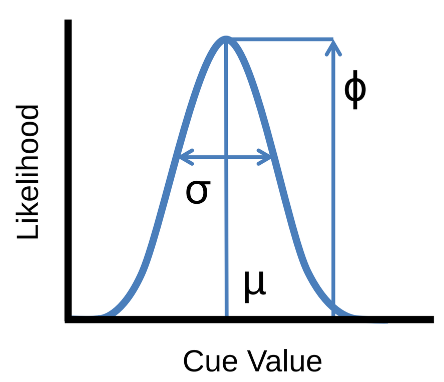
Possible mechanism:

- Older listeners show decreased sensitivity to fine-grained *temporal* differences in sounds (i.e., deficits in temporal acuity)
- This could render temporal cues (VOT) less reliable for older adults
- Could produce greater reliance on spectral cues (F0 onset)

- **Approach:** Simulate cue weighting changes using a Gaussian mixture model (GMM) (Toscano & McMurray, 2010)
- Model uses unsupervised learning mechanism (maximum likelihood estimation) to identify number of phoneme categories and their statistical properties
- Cue weights determined based on reliability
- Tested multiple mechanisms that could lead to changes observed with older listeners

MODEL ARCHITECTURE

Language structure: Training data based on statistical distributions from Lisker and Abramson (1964) for VOT, and Dmitrieva et al. (2015) for F0 (in semitones)

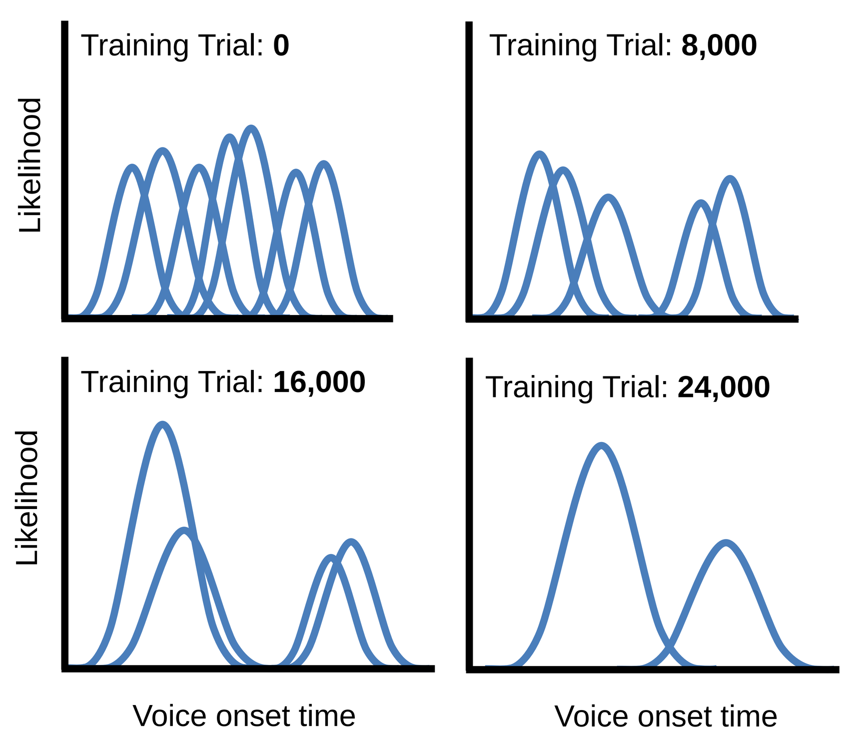


Model structure: Each phoneme category modeled by a Gaussian distribution with three parameters:

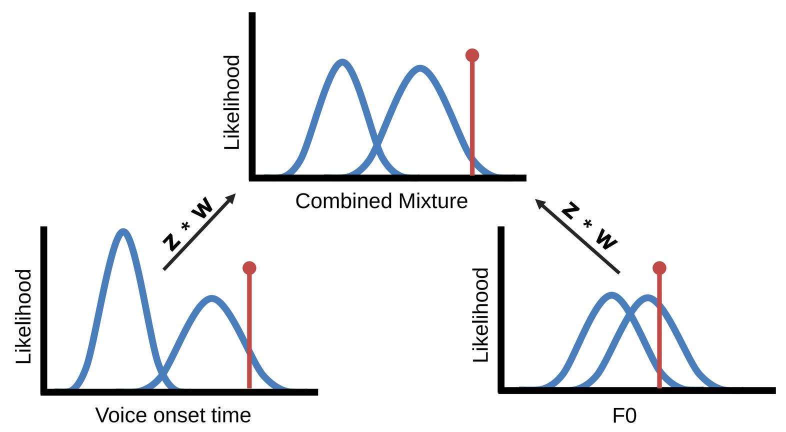
- μ: category mean
- σ: standard deviation
- ϕ: Likelihood

Training procedure

- Model presented with VOT and F0 values, drawn from cue distributions above
- Cue values used to update cue-specific mixtures



- Inputs converted to z-scores and weighted by reliability to update combined (i.e., voicing) mixture
- Categorization functions derived from posterior probabilities of combined mixture



Adding noise to inputs

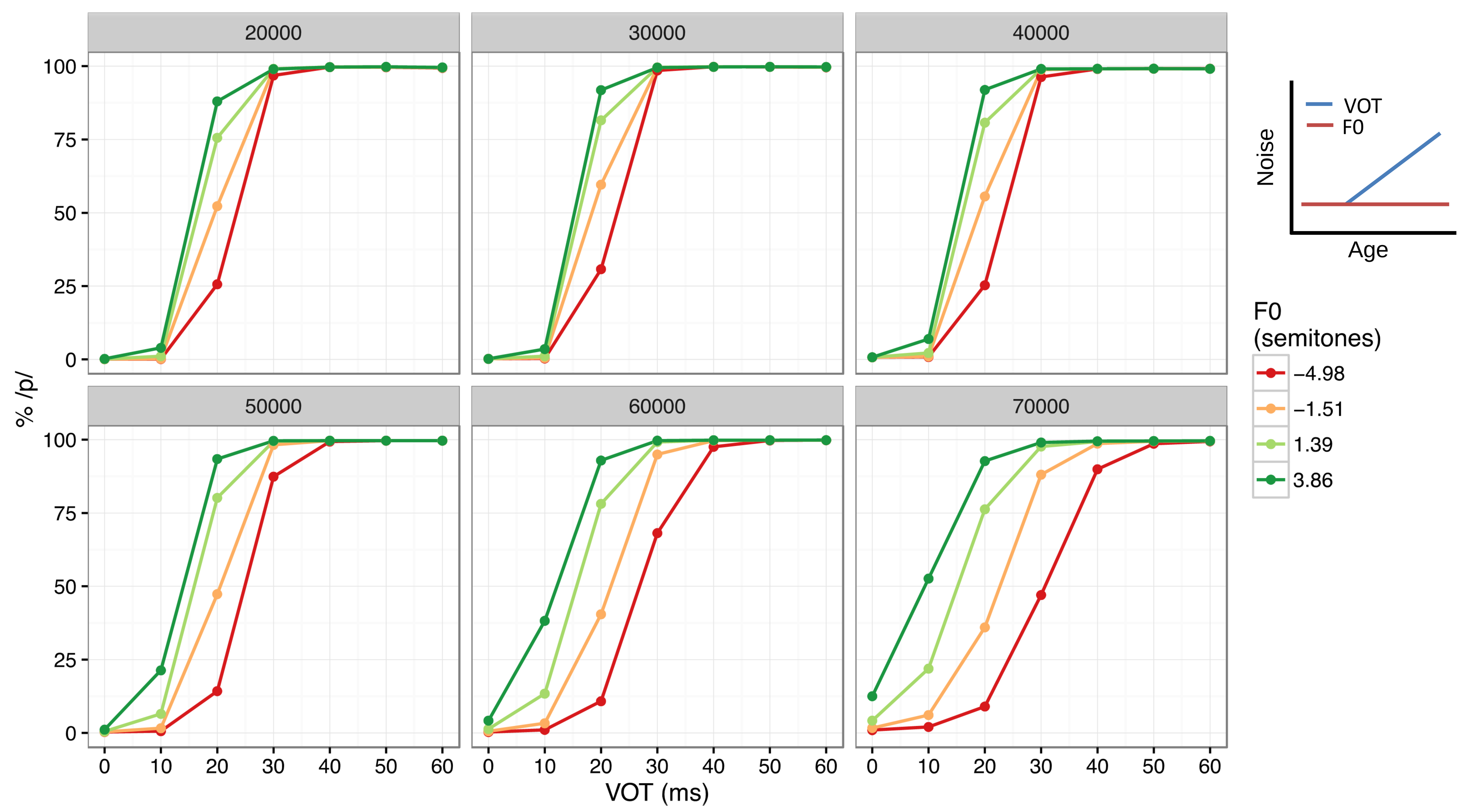
- After the initial developmental stage (t_0=35,000), random values added to cue inputs, which adds noise to the cue distribution
- Magnitude of noise was proportional to the training trial number, such that at t=70,000, the noise added was twice the amount at t_0=35,000

$$x_i = \left(\frac{t}{t_0}\right) r s_{ij} + m_{ij}$$

where,

- x_i : input for a given cue
- i : cue dimension
- j : phoneme category that input is derived from
- s_{ij} : standard deviation of category
- m_{ij} : mean of category
- r : random number from a normal distribution
- t : current trial number
- t_0 : trial number at which noise is initially added

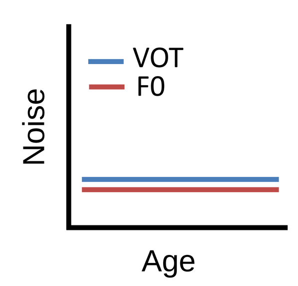
RESULTS

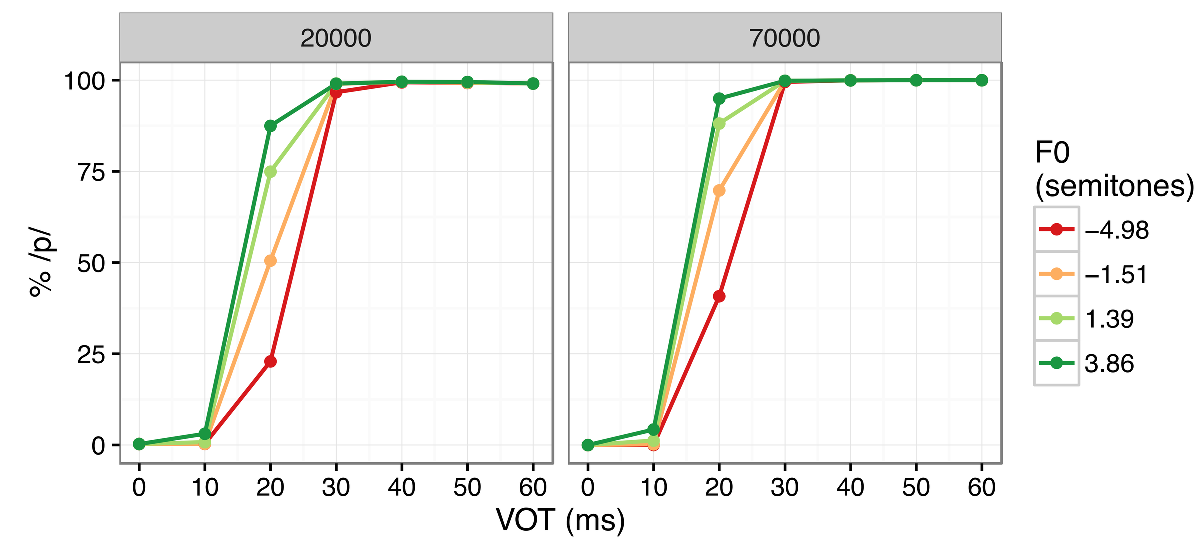


- Trial number in model corresponds to approximate age in human listeners:
  - 20,000-30,000: adolescence through young adulthood
  - 40,000-50,000: middle-age
  - 60,000-70,000: older adults
- At all stages, VOT weighted higher than F0 onset, consistent with responses from listeners and predictions based on overall reliability of VOT and F0 as voicing cues
- Relative weight of F0 onset increases over time, leading to a larger effect of F0 on categorization responses
- Mirrors results seen with human listeners—greater reliance on spectral information (F0) with increasing age

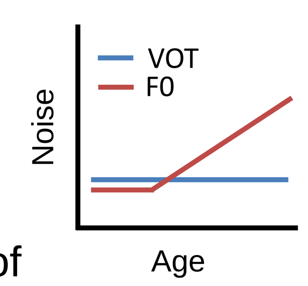
Alternative Hypotheses:

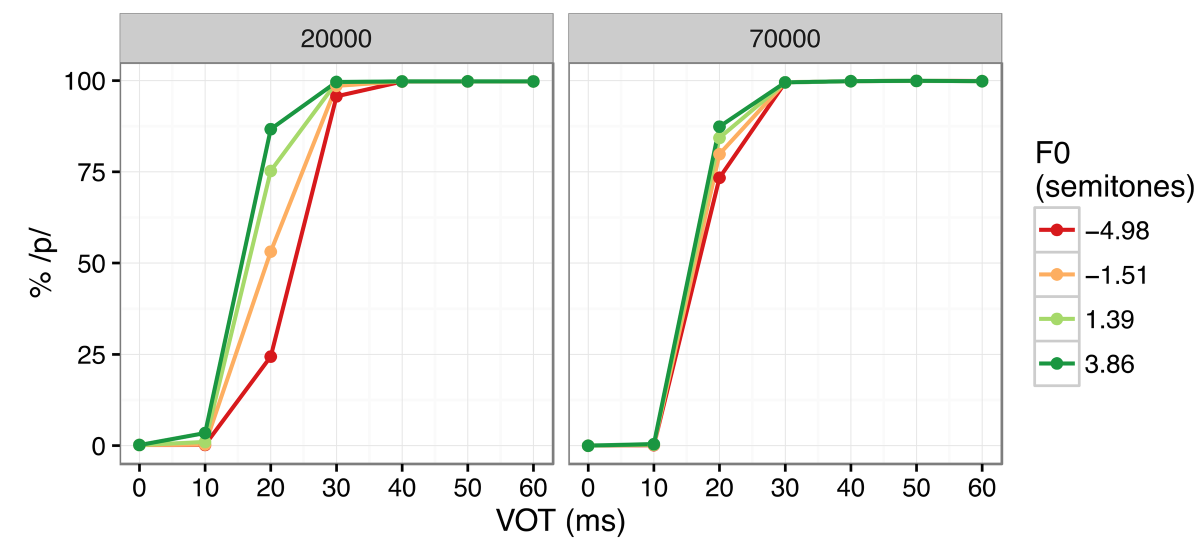
- No noise added to either cue dimension
- **Result:** no change in cue weights over time; inconsistent with data from human listeners



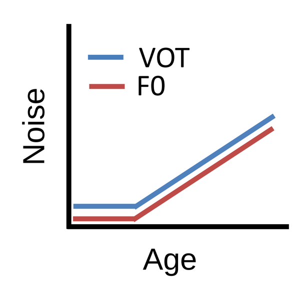


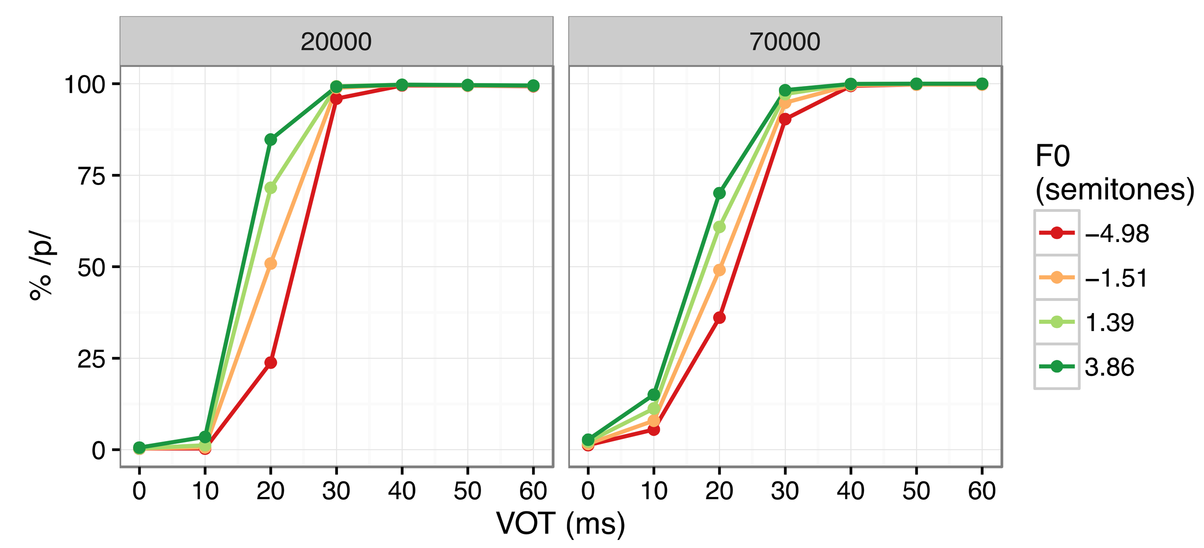
- Noise added to F0 dimension
- **Result:** Decrease in use of F0; opposite of pattern observed with human listeners





- Noise added to both dimension
- **Result:** Overall decrease in cue weights, with no relative change over time between VOT and F0





DISCUSSION

- Using a weighted Gaussian mixture model, we simulated and found evidence of decreased reliance on VOT over time, caused by increased cue variability
- This leads to the same pattern of differences observed between younger and older listeners
- Supports a mechanism whereby decreased temporal acuity with age leads to decreased reliance on temporal speech cues
- Alternative hypotheses did not yield the same pattern of results
- Implications: Cue weights continue to change over the lifespan; no critical period/change in the model's plasticity needed to account for these changes
- Suggests that cue weighting may be approx. optimal, yet both the model and listeners deviate from the optimal strategy (given the language properties) over time

ACKNOWLEDGEMENTS & REFERENCES

Dmitrieva, O., Llanos, F., Shultz, A. A., & Francis, A. L. (2015). Phonological status, not voice onset time, determines the acoustic realization of onset *f*0 as a secondary voicing cue in Spanish and English. *Journal of Phonetics, 49,* 77-95. doi: 10.1016/j.wocn.2014.12.005

Lisker, L., & Abramson, A.S (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word, 20,* 384-422.

Nittrouer, S. (2002). Learning to perceive speech: How fricative perception changes, and how it stays the same. *Journal of the Acoustical Society of America, 112,* 711-9

Toscano, J. C., & Lansing, C. R. (submitted). Effects of age on temporal and spectral cue weights in speech.

Toscano, J. C., & McMurray, B. (2010). Cue integration with categories: Weighting acoustic cues in speech using unsupervised learning and distributional statistics*. Cognitive Science, 34,* 434-464. doi: 10.1111/j.1551-6709.2009.01077.x