The time-course and nature of dimension-based statistical learning

Kaori Idemaru, Lori Holt,
Volya Kapatsinski

Adaptive nature of speech processing

- Speech categorization is dynamically tuned by the local environment
  - (e.g., Norris et al., 2003; Holt, 2005; Eisner & McQueen, 2005, 2006; Kraljic & Samuel, 2006, 2007; Clayards et al. 2008; Maye et al. 2008; Idemaru & Holt, 2011)
- A dual nature of speech categories - sensitivity to long-term regularities and short-term deviations

Perceptual statistical learning

- Perceptual adjustment / ‘recalibration’ of phonetic categories
  - E.g., Getting used to an accent or a speaker’s particular way of pronouncing a phone

Perceptual statistical learning

EXPOSURE

TEST

f-training group

polygra[?] hippopotamu[s]
"polygraph" "hippopotamus"
etc

[?] is heard as [f]

[?] is acoustically ambiguous between [f] and [s]
Disambiguated by he lexicon

s-training group

polygra[f] hippopotamu[?]
"polygraph" "hippopotamus"
etc

[?] is heard as [s]

Norris, McQueen, & Culter (2003)

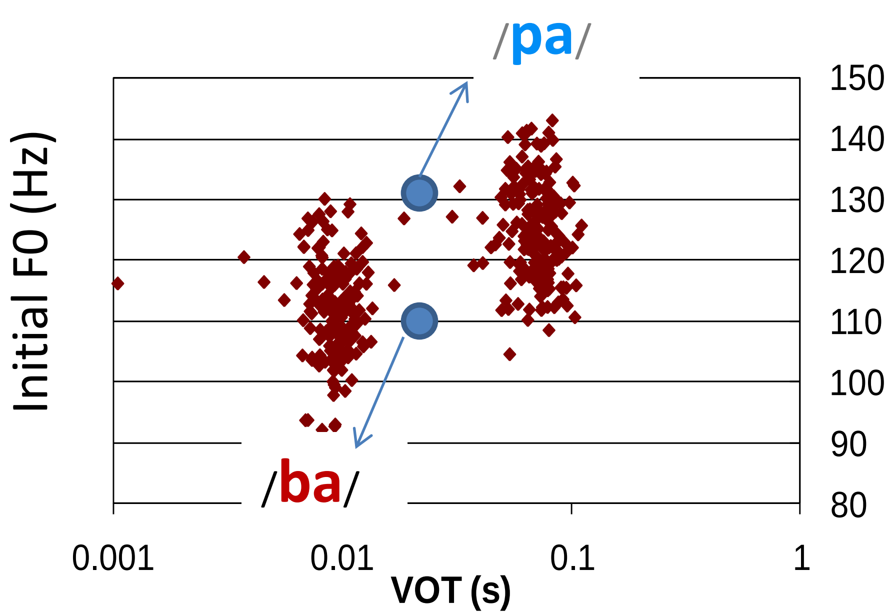
Dimension-based Statistical Learning

- Idemaru & Holt, 2011
  - *Dimension-based*: Perceptual adjustment at the level of fine-grained acoustic dimensions
    - Multiple cues to the same phonetic contrast
  - *Statistical:* implicit detection of correlations between values of the dimensions
    - The meaning of values of one dimension learned based on the values of the other dimension they correlate with

Multi-dimensional speech categories

- 16 covarying acoustic dimensions for English stop voicing (Lisker 1986)

- Two prominent cues

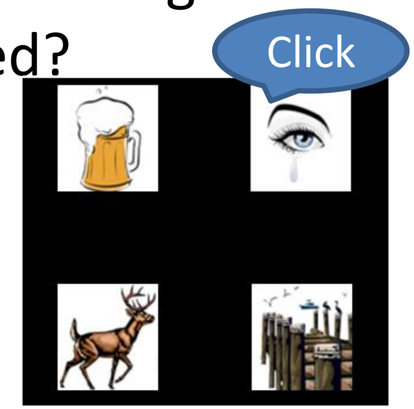


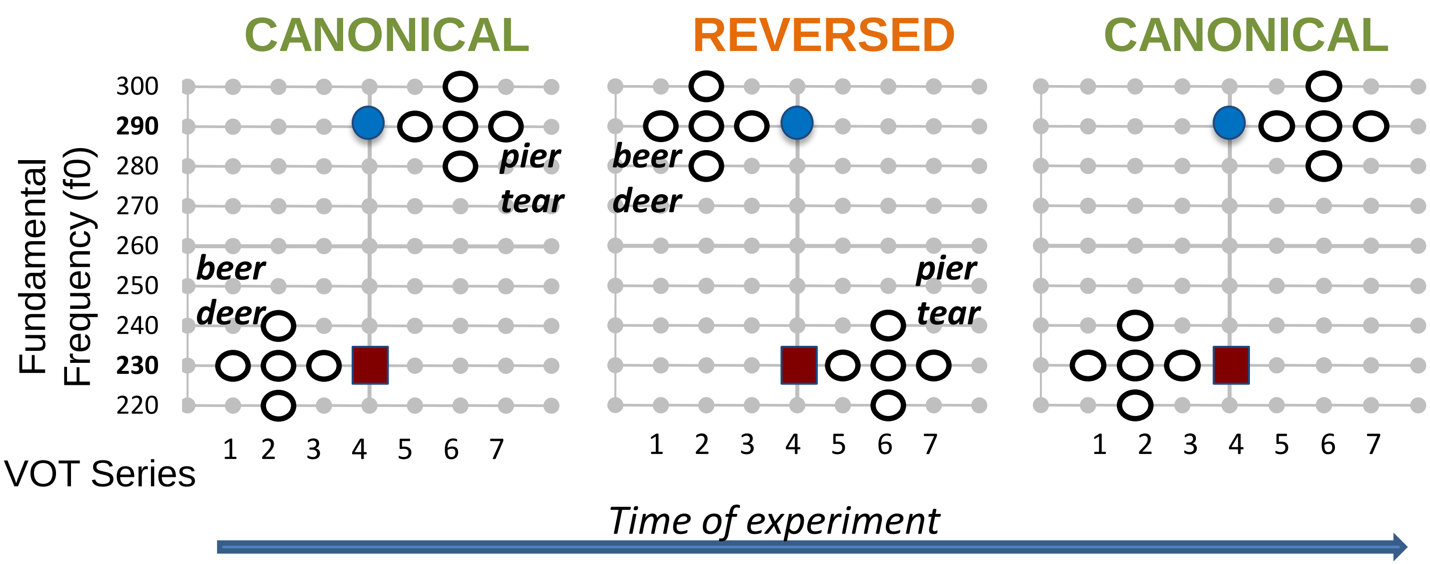
  – F0 – VOT
    - Correlation in production
    - Perceptual sensitivity to this correlation

Holt & Wade (2004)

Previous study

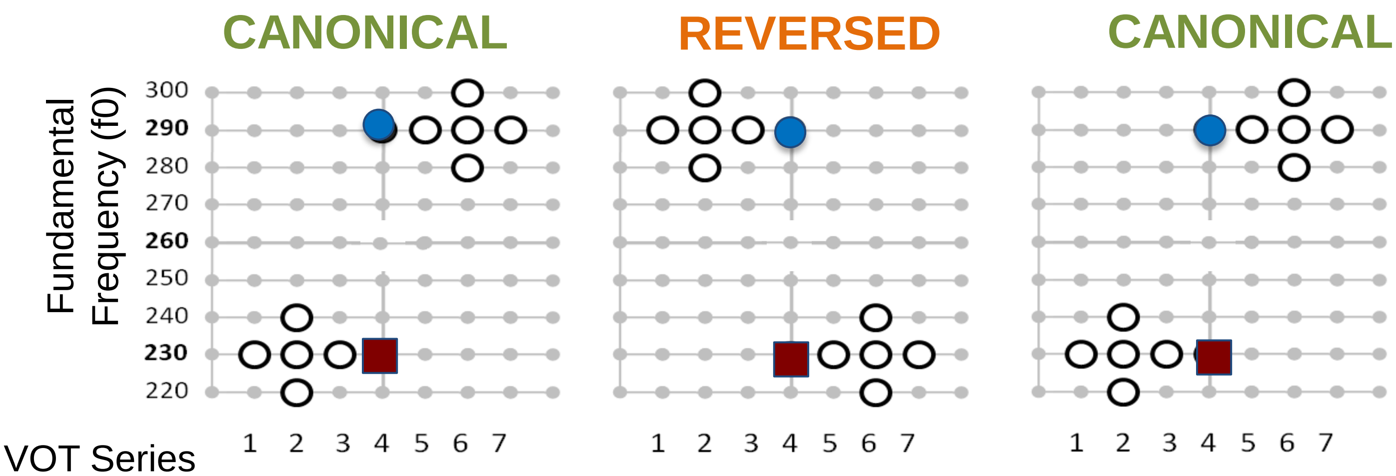
- Does listeners' pattern of perception change, if this long-term regularity of F0/VOT correlation is perturbed?
  - Word recognition task - beer, pier, deer, tear
  - F0-VOT correlation manipulated *implicitly*

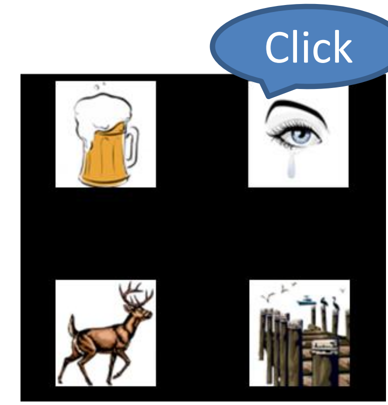


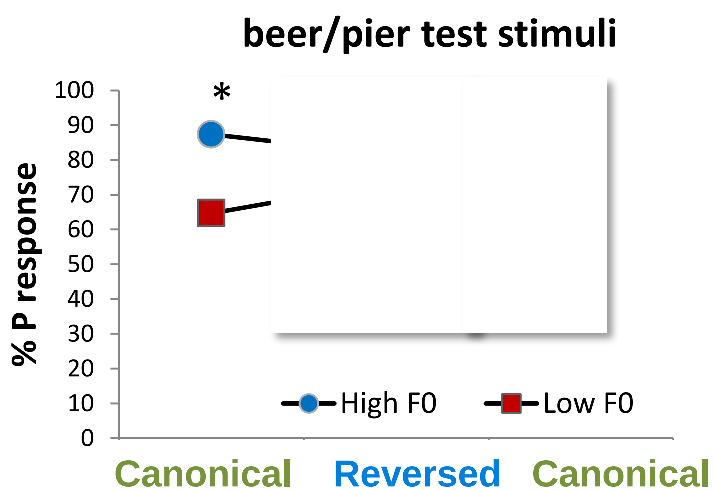


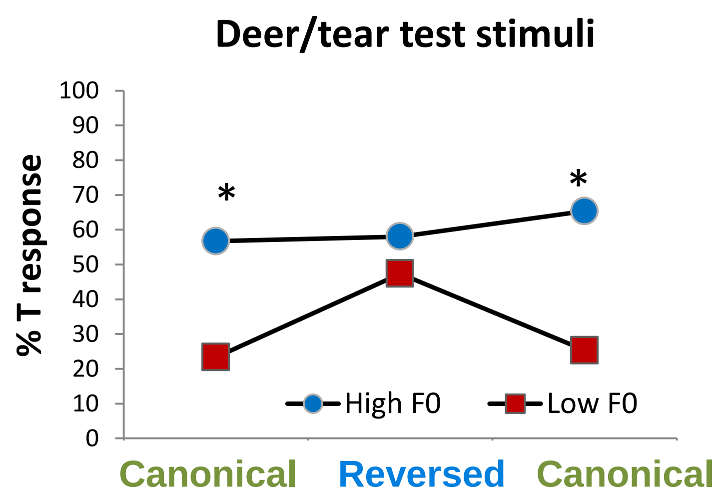
Idemaru, K. & Holt, L. (2011). Word-recognition reflects dimension-based statistical learning. Journal of Experimental Psychology: HPP

Our previous study









Time course

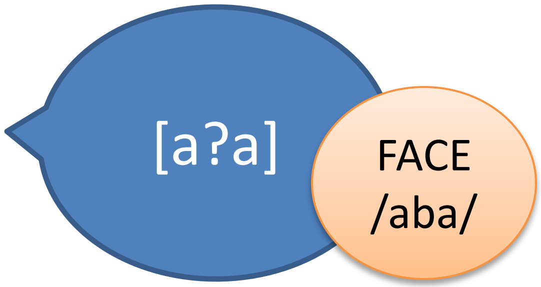
- How quickly does the learning emerge?
  - Competition: Listener’s long-term generic representation vs. Short-term deviation in the current input, specific to some aspect of the situation (e.g., speaker)

  - Expect rapid learning as long as it is specific

Time course of perceptual learning

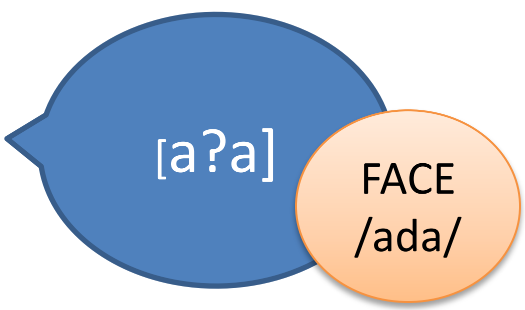
EXPOSURE

TEST



[?] is heard as [b]

[?] is ambiguous between [b] and [d]



[?] is heard as [d]

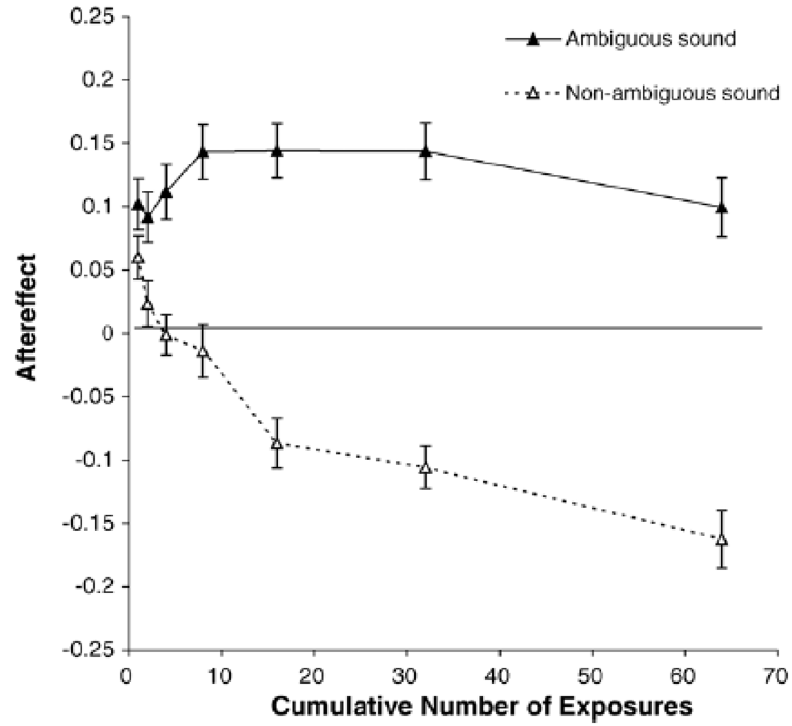


Fig. 2. Mean aftereffects as functions of cumulative number of exposures (1–64), for the pooled data of Groups 64 and 256, in the ambiguous sound condition (adapters A?Vb and A?Vd) and the non-ambiguous sound condition (adapters AbVb and AdVd).

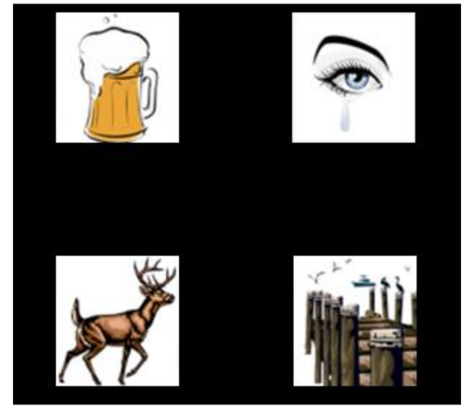
Vroomen, Linden, Gelder, Bertelson (2007)

Present study

- Uses eyegaze data
  - Information about the strength of the competitor ([b] vs. [p])
    - More looks
    - Faster looks
  - Can be used to look at learning during training, not just test stages
    - Both F0 and VOT reweighting can be detected
    - Can look at parts of VOT continua that the learners are trained on

Method

- 40 Listeners
- Task: word recognition



- Experiment structure
  - Training: exposure to accented *beer, pier, deer, tear*

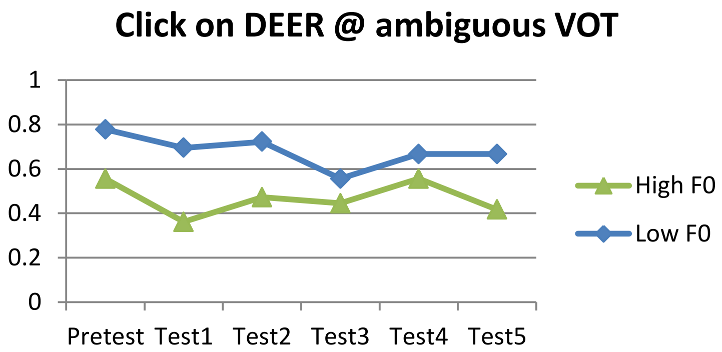
Method

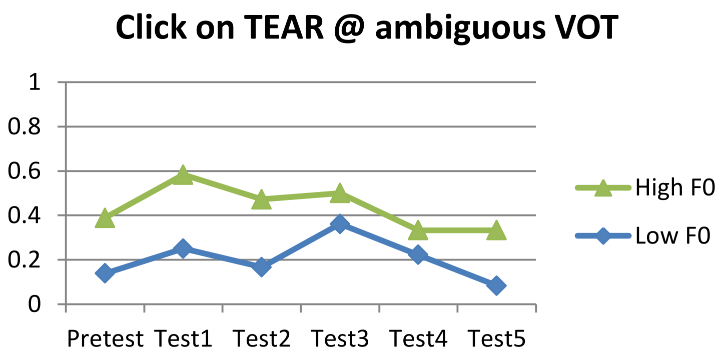
- Stimuli

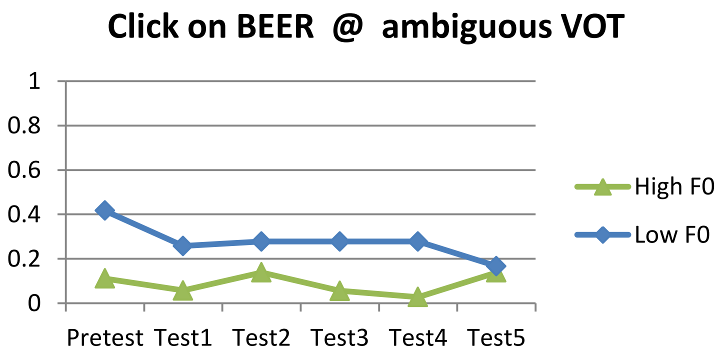
- Pretest
  - Beer/Pier: VOT (5, 10, 15 ms) x F0 (230, 260, 290 Hz)
  - Deer/Tear: VOT (15, 20, 25 ms) x F0 (230, 260, 290 Hz)
- Accented words
  - Beer: Short VOT (-20, -10, 0 ms) x High F0 (280, 290, 300 Hz)
  - Pier: Long VOT (20, 30, 40 ms) x Low F0 (220, 230, 240 Hz)
  - Same patterns for Deer/Tear
- Test
  - B/P: Mid VOT (10 ms) x High, Mid, Low F0 (230, 260, 290 Hz)
  - D/T: Mid VOT (20 ms) x High, Mid, Low F0 (230, 260, 290 Hz)

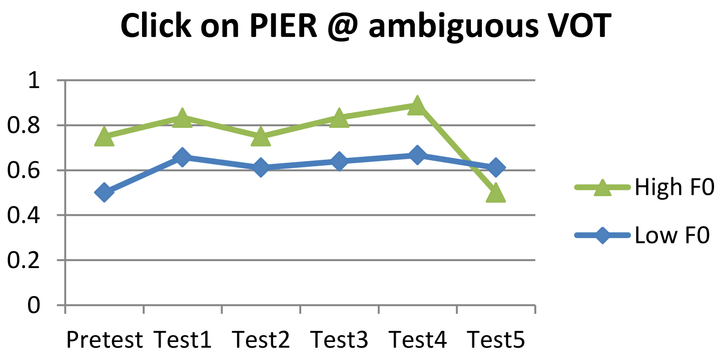
Click data at Tests

- See unlearning of F0 use for p/b late

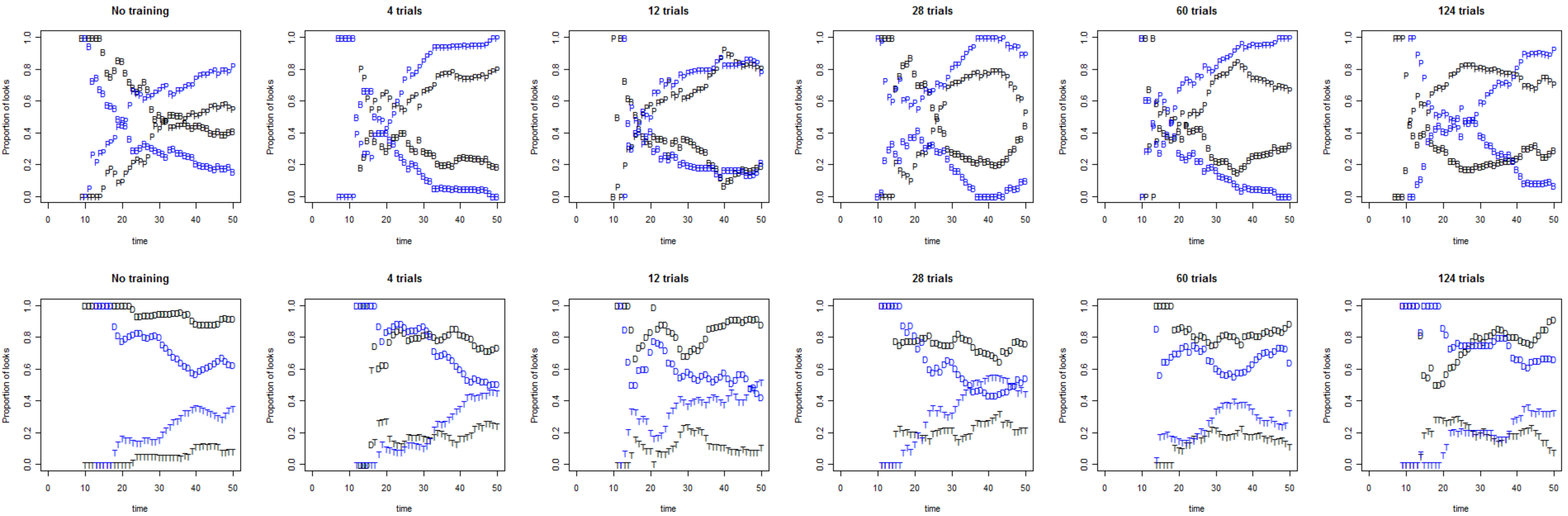








Gaze data



- Blue = High F0
- Black = Low F0

Interpreting eyegaze data in training

- Stimuli with High F0 & Short VOT

	prior knowledge	learning
– Higher F0 → more looks to	P/T	B/D
– Shorter VOT → less looks to	P/T	B/D

- Stimuli with Low F0 & Long VOT

– Lower F0 → more looks to	B/D	P/T
– Longer VOT → less looks to	B/D	P/T

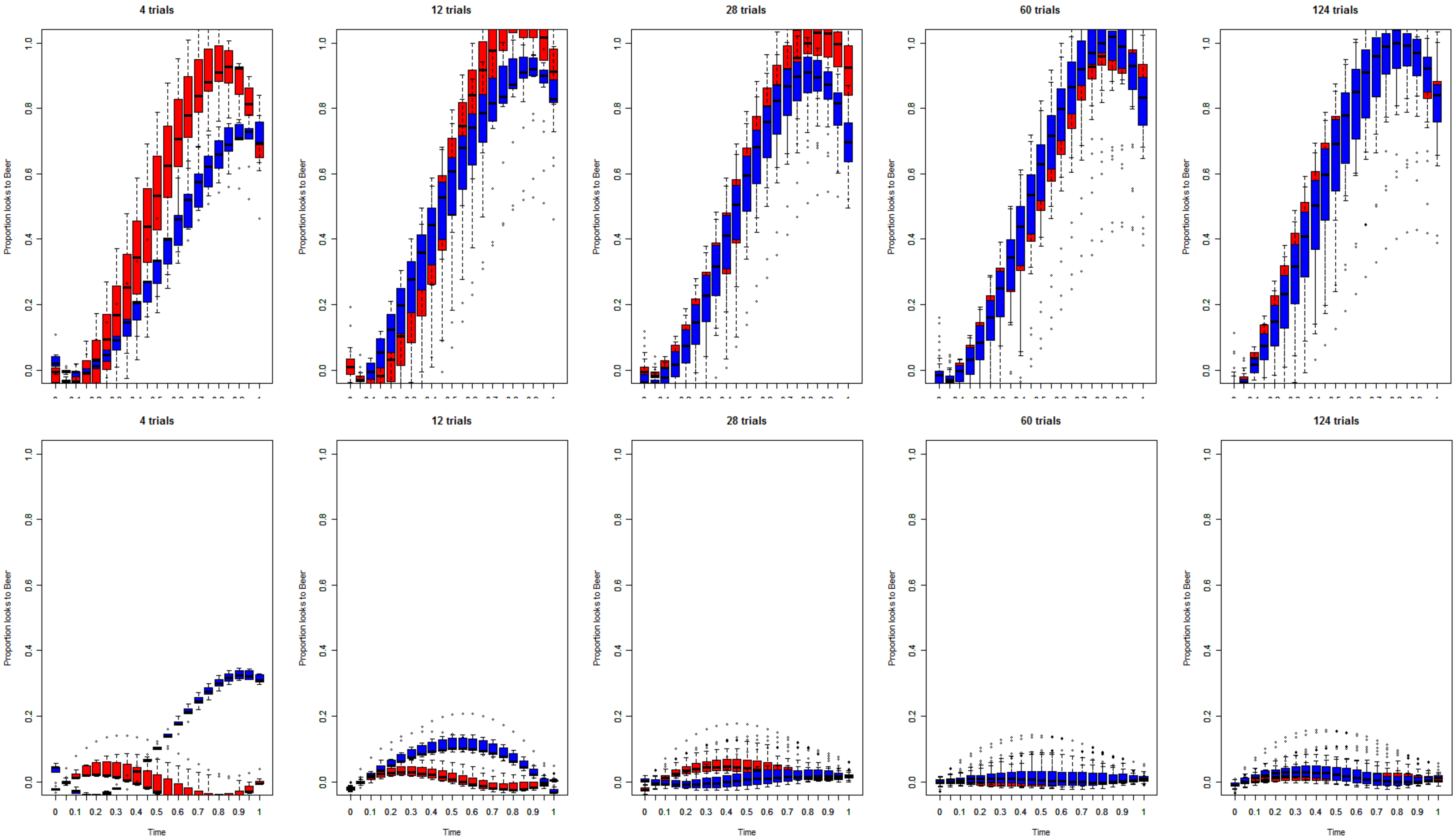
Modeling and Analyses

- lmer(Deer ~
  (1|subject)
  +poly(time,3)
  +(poly(time,3)|subject)
  +stage*CUE}
  +poly(time,3)*stage*CUE)

CUE = VOT or F0

Mirman et al. (2008)

F0=290; VOT=-10 (red) vs. 10 (blue) on 'Deer' (top) and 'Tier" (bottom) responses



We look at looks to dominant response within a trial

Summary of significant effects within stages

Test 1				Beer-Pier ↓
Tr.2	Deer ↑	Deer ↑	Pier ↑	Beer ↑
Test 2		Deer-Tear ↓		
Tr.3		Deer ↓		
Test 3		Deer-Tear ↓		Beer-Pier ↓
Tr. 4	Tear ↑	Tear ↑	Pier ↓	Pier ↓ Beer ↑
Test 4		Deer-Tear ↓		Beer-Pier ↓
Tr. 5	Tear ↑ Deer ↑	Tear ↑ Deer ↓	Pier ↑ Beer ↓	Pier ↑ Beer ↑
Test 5				

Yellow = Learning

Blue = Prior Knowledge

Conclusions

- VOT (primary cue to voicing) can be reweighted based on F0 (secondary cue)
- This reweighting can happen fast, within 8 trials, and tends to recede (also Vroomen et al. 2007)
- Primary and secondary cue are not strictly ranked
- Reweighting can be specific to a trained interval on an acoustic continuum
- Perhaps, learning is not *dimension-based* after all
  - The effect of a small VOT difference on voicing perception can be reweighted for a certain range of VOT and/or F0 (Can be modeled in General Recognition Theory, Ashby & Townsend 1986, Silbert et al. 2009)

References

Ashby, F. G., & Townsend, J. T. (1986). Varieties of perceptual independence. *Psychological Review*, 93, 154-179.

Idemaru, K. & Holt, L. L. (2011). Word recognition reflects dimension-based statistical learning. *Journal of Experimental Psychology: Human Perception & Performance,* 37, 1939-1956.

Mirman, D. Dixon, J.A., & Magnuson, J.S. (2008). Statistical and computational models of the visual world paradigm: Growth curves and individual differences. *Journal of Memory and Language*, 59, 475-94.

Norris, D.G., McQueen, J.M., and Cutler, E.A. (2003) Perceptual learning in speech. *Cognitive Psychology, 47*, 204-238.

Poellmann, K., McQueen, J. M., & Mitterer, H. (2011). The time course of perceptual learning. In W.-S. Lee, & E. Zee (Eds.), *Proceedings of the 17th International Congress of Phonetic Sciences 2011 [ICPhS XVII]* (pp. 1618-1621).

Silbert, N. H., J. T. Townsend, & J. J. Lentz. (2009). Independence and separability in the perception of complex non-speech sounds. *Attention, Perception, & Psychophysics*, 71, 1900-15.

Vroomen, J., van Linden, S., de Gelder, B., & Bertelson, P. (2007). Visual recalibration and selective adaptation in auditory-visual speech perception: Contrasting build-up courses. *Neuropsychologia, 45*,572-577.