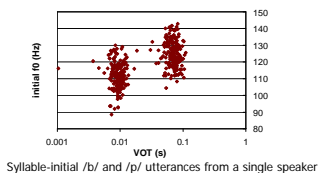


Re-tuning Speech Sound Categories: An Eyetracking Study

Kaori Idemaru, Dept of East Asian Languages and Literatures, University of Oregon
 Lori L. Holt, Dept of Psychology and the Center for the Neural Basis of Cognition, Carnegie Mellon University
 The 156th meeting of the Acoustical Society of America, Nov 10-14, 2008

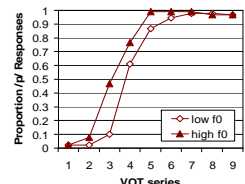
Introduction

- Adaptive nature of auditory system:** The acoustics of speech are highly variable across speakers and even within speakers. However, human listeners show an impressive ability to recognize words despite the variation in signal. This seems to suggest the highly detailed and adaptive nature of the auditory representation.
- Evidence of talker adaptation:** There has been accumulating evidence that
 - Listeners do encode fine-grained details of speaker's voice (e.g., Bradlow & Pisoni 1999, Maye et al 2008), and
 - Listeners are sensitive to statistical characteristics of relevant acoustic parameters in the local input (Holt & Lotto, 2006; Clayards et al 2008, Dahan et al 2008).
- Given these findings, we examine the hypothesis that listeners update their speech categories dynamically to reflect the distributional characteristics of the current input.
 - Here we examine listeners' sensitivity to correlations between two acoustic attributes, F0 and VOT.
- F0/VOT correlation in English stops:** English stop voicing is defined by multiple acoustic parameters. Among them, VOT is a primary cue, but F0 shows robust correlation with voicing. It is used as a secondary cue.



In English and many languages, F0 covaries with VOT in natural productions...

such that F0 is lower near voiced consonants than near voiceless consonants



The correlation of F0 and VOT in English productions is echoed in perception of these consonants...

sounds ambiguous for VOT are more often heard as voiceless when following F0 is high (e.g. [2])

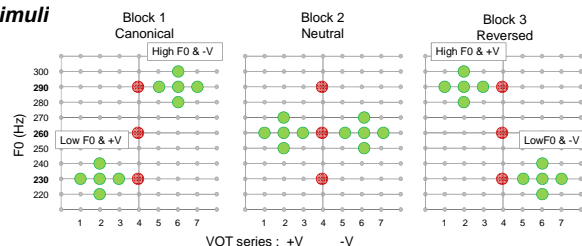
This F0/VOT correlation in perception may be a learned pattern influenced by the highly regular co-occurrence in the signal (Kingston & Diehl 1994, Holt et al 2001)

- This study:** Although F0 and VOT are correlated in English, there is considerable variation across speakers. What counts as "low F0" for one speaker may be a "high" F0 for another.
- Here we exploit the natural F0/VOT covariation to examine
 - To what extent listeners track cue correlation in online speech perception, and
 - To what extent experience with shifts in cue correlation influences speech categories

Method

- Experiment overview:** 14 native-English listeners heard minimal-pair words, *beer-pier* and *deer-tear*, varying in VOT. The vowel F0 was gradually shifted over the course of the experiment from the canonical English correlation (higher F0 for voiceless stops) to the reversed pattern (lower F0 for voiceless stops). Listeners' use of F0 as a cue to voicing was measured throughout the experiment using eyetracking and overt word identification.

Stimuli



Exposure stimuli

Beer, pier, deer, & tear with shifting F0-voicing correlation pattern

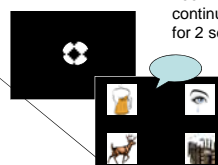
Critical stimuli

Beer/pier, deer/tear with ambiguous VOT and high, mid, and low F0 values

Procedure



Participants heard 780 words in an identification task. They were asked to look at and click on the object on a monitor corresponding to the word they heard.



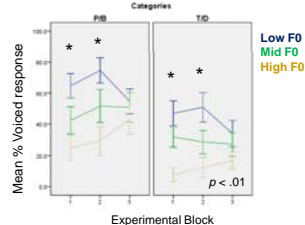
Looming checkerboard continues until Ss fixate for 2 sec

The block structure was implicit to the participants. Sampling rate = 60Hz

Mouse click on an object triggers next trial

Results

Identification (mouse-click) data

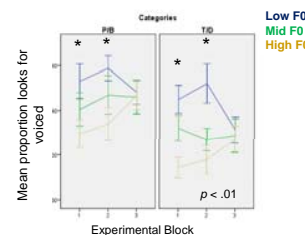


Listeners' responses to ambiguous VOT showed clear effect of F0 in Block 1 and 2 ($p < .01$), but the F0 effect disappeared in Block 3.

In Block 3, listeners ceased to use F0 as a cue to categorize voicing

Results

Eyegaze data

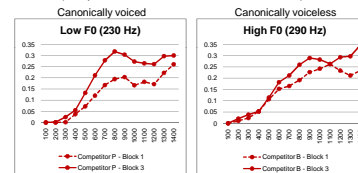


These data show the overall proportion of looks to the voiced (beer or deer) choice.

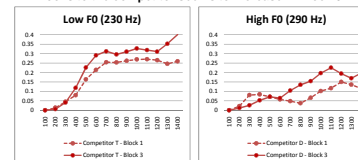
Listeners' responses to ambiguous VOT showed clear effect of F0 in Block 1 and 2 ($p < .01$), but the F0 effect disappeared in Block 3.

In Block 3, listeners ceased to use F0 as a cue to categorize voicing

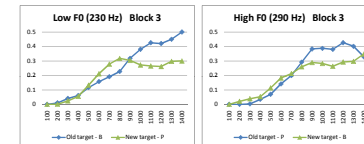
Strength of Competitor Sounds (competitor defined in the canonical correlate)



Looks to the competitor seems to increase in Block 3.



Response to OLD and NEW targets during the Reverse Exposure (canonical vs. reverse target)



Up to around 700-800 ms since stimulus onset, there seem to be more looks to the NEW target.

Due to voiceless bias to D/T stimuli, responses here look different from above. Low F0 attracted more looks to the new target (voiceless).

Conclusion & Future Directions

- Listeners show tuning of their voicing categories in response to changes in probabilistic distribution of key acoustic parameters in the input
 - Canonical F0 perception effect disappeared during reversed exposure. Listeners discard use of F0 rather quickly.
- F0 effects in voicing appear to be amenable to change if given appropriate input.
- Cue reversal (high F0-voiced, low F0-voiceless) may require longer exposure. Examination of long-term exposure is needed.

Reference

Bradlow, A. R., & Pisoni, D. B. (1999). Recognition of spoken words by native and non-native listeners: Talker-, listener-, and item-related factors. *The Journal of the Acoustical Society of America*, 106, 2074.

Clayards, M., Tanenhaus, M. K., Aslin, R. N., & Jacobs, R. A. (2008). Perception of speech reflects optimal use of probabilistic speech cues. *Cognition*, 108(3), 804-9. doi: S0010-0277(08)00101-7.

Dahan, D., Drucker, S. J., & Scarborough, R. A. (2008). Talker adaptation in speech perception: Adjusting the signal or the representations? *Cognition*, 108(3), 710-718.

Holt, L. L., Lotto, A. J., & Kluender, K. R. (2001). Influence of fundamental frequency on stop-consonant voicing perception: A case of learned covariation or auditory enhancement? *Journal of the Acoustical Society of America*, 109, 764-774.

Kingston, J., & Diehl, R. L. (1994). Phonetic Knowledge. *Language*, 70(3), 419-454.

Maye, J. (2008). The Weckud Wetch of the West: Lexical Adaptation to a Novel Accent. *Cognitive Science: A Multidisciplinary Journal*, 32(3), 543-562.

Acknowledgement

This work was supported by National Institute of Health (5R01DC004674) and National Organization for Hearing Research.