Cross-linguistic differences in the production and perception of consonant and vowel intrinsic F0 effects

Connie Ting & Meghan Clayards
McGill University

LabPhon19 – June 27-29, 2024
connie.ting@mail.mcgill.ca

Background

- Vowels vary systematically in f0 based on intrinsic properties of:
  - the vowel → higher f0 for high vs low vowel
  - the preceding obstruent → higher f0 following 'voiceless' vs 'voiced'

- Cross-linguistic differences for effect size in **production** of consonant F0 effects (CF0) and vowel F0 effects (VF0)[1,2]

Effect	English	Mandarin
CF0	~2 st	~0.5 st
VF0	~1.5 st	~0.5 st

- **perceptual** effects of the CF0 effect: higher f0 → more voiceless responses[3,4,5]
- less work has focused on perceptual effects of the VF0 effect
- Previous work shows listeners take VF0 effects into consideration in their perception of vowels (e.g. for relative pitch or height)[6,7]

Gap in previous work:
- ❖ very little cross-linguistic comparisons of CF0 in perception
- ❖ no cross-linguistic comparison of VF0 effects in perception

Research Questions

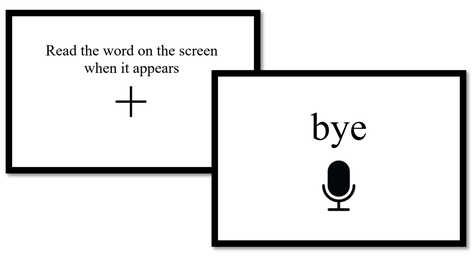
1. Are CF0 and VF0 effects different across English and Mandarin speakers? How?
2. Do English and Mandarin speakers differ in their use (if any) of CF0 and VF0 effects in perception through voicing judgements?

Methods

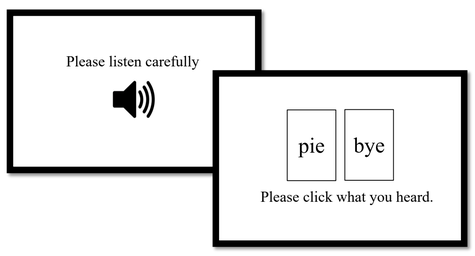
Participants
- 36 native English speakers recruited through Prolific; 4 only did production
- 25 native Mandarin speakers living in China recruited through WeChat

Procedure
- Participants completed the production task followed by the perception task

Production



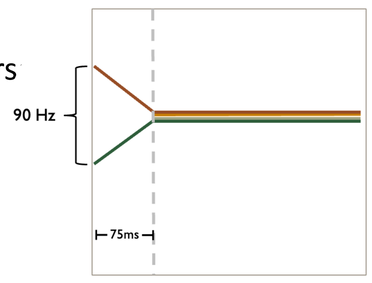
Perception



Stimuli

Production
- English = 10 CV words x 12 reps = 120 trials
- Mandarin = (10 CV words x 4 tones – 2 gaps) x 3 reps = 114 trials

Perception
- 1 male speaker recorded 5 repetitions of syllable pairs
  - /baj, paj/ and /bi, pi/ in all 4 Mandarin tones
- VOT manipulation:
  - 0-48 ms; 7 steps
- F0 manipulation
  - high, low; difference of 90 Hz
- Constant vowel duration = 350 ms



Tone 1 F0 manipulation.

- Perception stimuli were the same across English and Mandarin groups
- Each participant heard 336 syllables (7 VOT x 2 F0 x 2 vowels x 4 tones x 3 reps)

Analysis

Production: A GAMM model for each language group (tone included for Mandarin)

```
mandarin.gam <- bam(f0 ~ voicing +
                    s(f0_time) + s(f0_time, by=voicing) + s(f0_time, by=tone) + s(f0_time, speaker, bs="fs", m=1) +
                    s(f0_time, speaker, by=voicing.ord, bs="fs", m=1) + s(f0_time, consonant, bs="fs", m=1) +
                    s(f0_time, by=vowel) + s(f0_time, speaker, by=vowel, bs="fs", m=1) +
                    s(f0_time, id, bs="fs", m=1) + voicing*tone + vowel*tone,
                    data = mand_gam1_df, method='fREML', discrete=T)
```

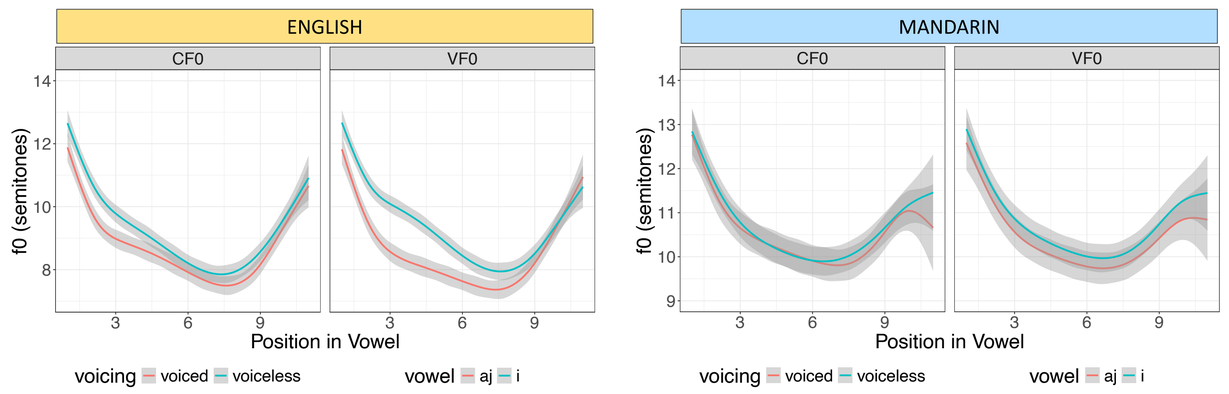
Perception: A Bayesian mixed effects model for English + Mandarin perception

```
fit <-  brm(formula = resp ~ vot + language*vowel + language*f0 + vowel*f0 + tone + tone*f0 + tone*vowel +
        (1+vot+f0+vowel+tone|Participant), family = bernoulli(link = "logit"),
        prior = c(prior(normal(0.5, 0.15), class="b", coef="f0"), prior(normal(0.5, 0.15), class="b", coef="vowel"))
```

Results

PRODUCTION TASK

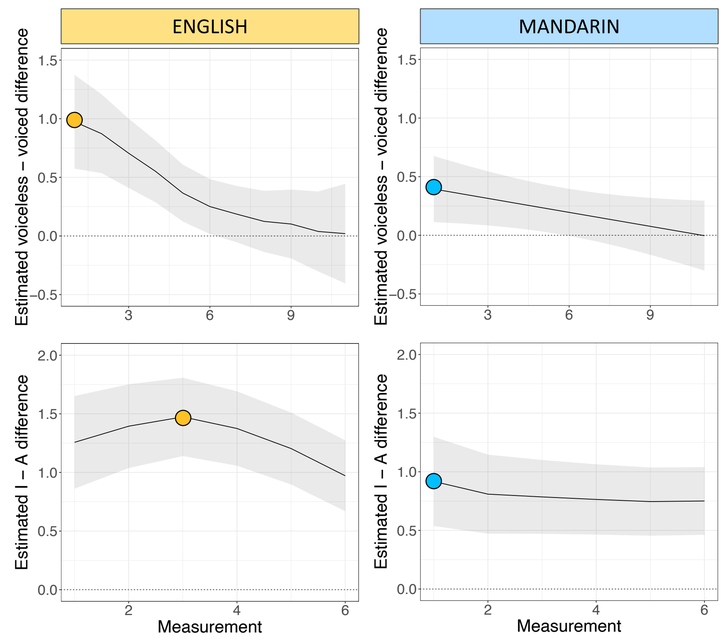
Empirical Plots: f0 measured at 11 points through vowel



Model Results: Estimated difference smooths

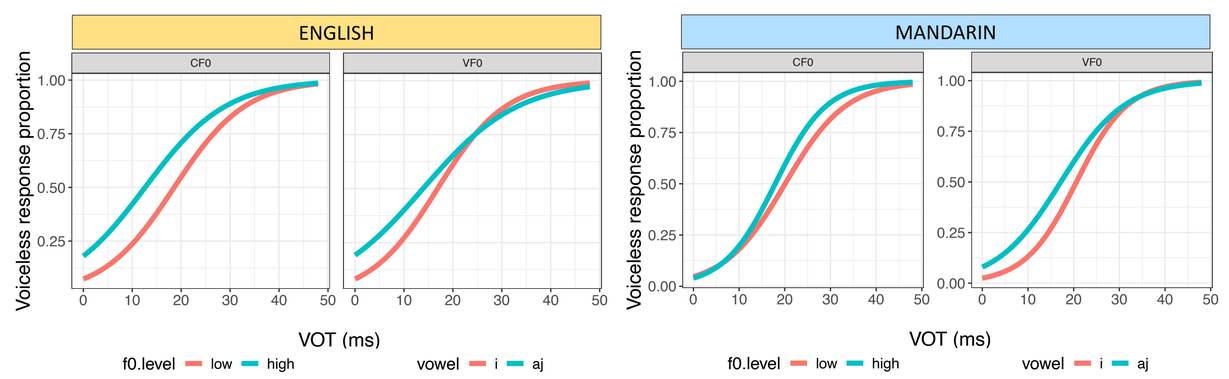
CF0 effect size (st):
- English: ~1
- Mandarin: ~0.4

VF0 effect size (st):
- English: ~1.5
- Mandarin: ~0.9

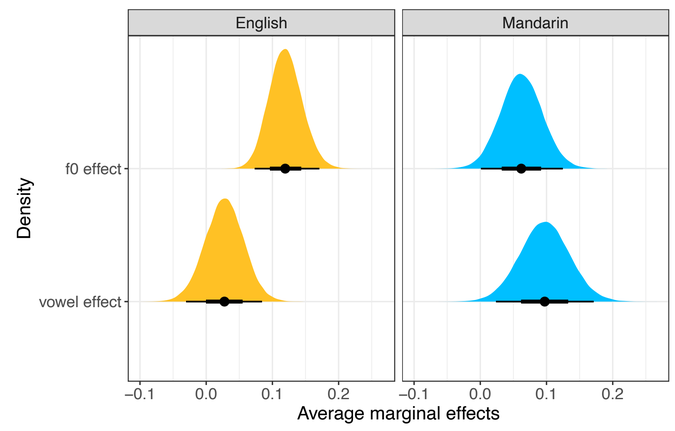


PERCEPTION TASK

Empirical Plots: voiced vs voiceless response by VOT for different levels of f0 + vowel



Model Results:



Posterior draws:
- English: CF0 > VF0 = 99.2%
- Mandarin: CF0 > VF0 = 24.1%

- English CF0 > Mandarin CF0 = 97.4%
- Mandarin VF0 > English VF0 = 95.1%

Conclusions

- **Production**:
  - VF0 > CF0 across languages, not in line with previous cross-linguistic study[2]
  - IF0 effects overall larger in English than in Mandarin, replicating previous work[1,2]
- **Perception**:
  - English CF0 use > Mandarin CF0 use (matching production)
  - Mandarin VF0 use > English VF0 use

❖ Languages vary in their use of CF0 and VF0 effects in both production & perception
❖ CF0 *and* VF0 effects can be examined through voicing judgements for cross-linguistic comparisons

References

[1] Whalen & Levitt. 1995. The universality of intrinsic f0 of vowels. J. Phon., 23, 349-366. **[2]** Ting et al. 2024. The cross-linguistic distribution of vowel and consonant intrinsic f0 effects. Language (in press). **[3]** Haggard et al. 1970. Pitch as a voicing cue. J. Acoust. Soc. Am., 47, 613-617. **[4]** Shultz et al. 2012. Differential cue weighting in perception and production of consonant voicing. JASA Express Lett., 132, EL95-EL101. **[5]** Schertz & Khan. 2020. Acoustic cues in production and perception of the four-way stop laryngeal contrast in Hindi and Urdu. J. Phon., 81, 100979. **[6]** Hombert. 1977. Development of tones from vowel height? J. Phon., 5, 9-16. **[7]** Reinholt-Peterson. 1986. Perceptual compensation for segmentally conditioned fundamental-frequency perturbations. Phonetica, 43, 31-42.

This research was supported by the Social Sciences and Humanities Research Council of Canada (#435-2020- 0209) and the Centre for Research on Brain, Language and Music.

McGill UNIVERSITY
SSHRC CRSH
CRBLM Centre for Research on Brain, Language and Music