VOT merger in progress and speech rate accommodation in perception: a case study of Daejeon Korean

Yoonjung Kang[1], Suyeon Yun[2], and Na-Young Ryu[3]

[1]University of Toronto, [2]Chungnam National University, [3]Pennsylvania State University

NWAV 51

October 13-15, 2023

Queens College, NY

Speech rate and production

- Compared to slow/normal speech, fast speech is marked by
  - shortening of segments
  - gestural overlap
  - articulatory reduction
  - prosodic reorganization
  - deletion and lenition of segments and syllables
    - Baese-Berk et al. 2018; Cohen Priva and Gleason 2018; Crystal and House 1988; Ernestus 2014; Fougeron and Jun 1998

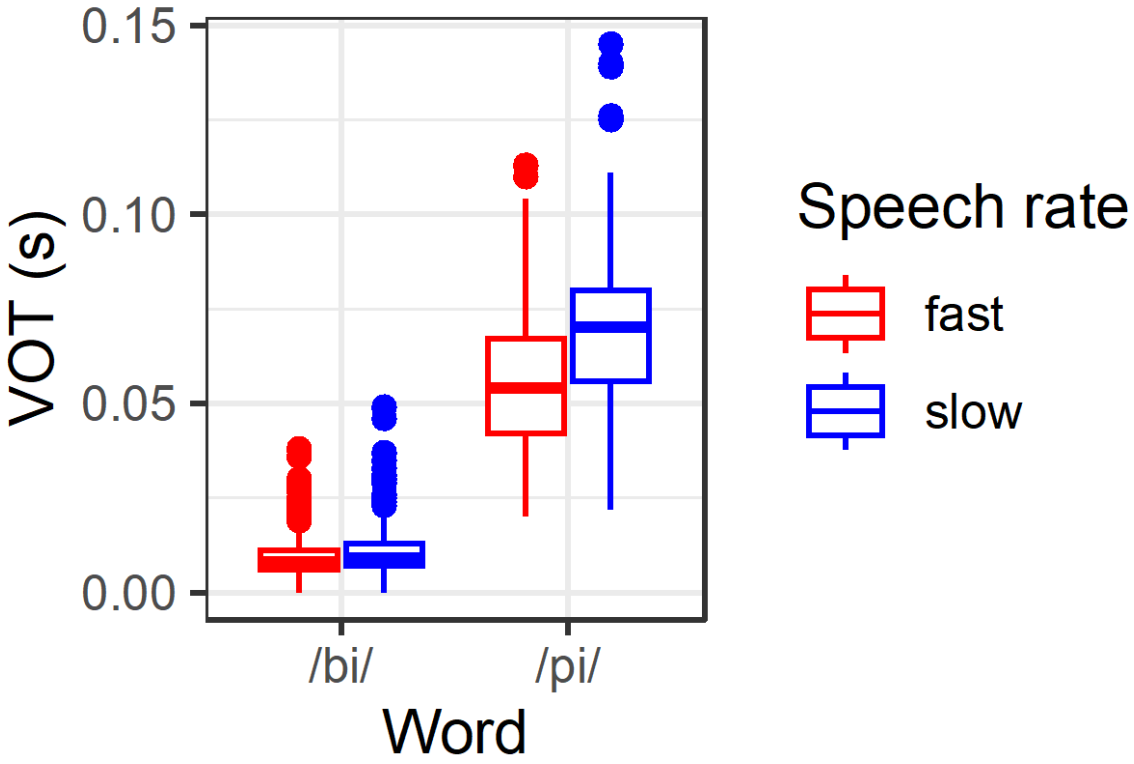
Speech rate and production

- Compared to slow/normal speech, fast speech is marked by
  - shortening of segments
  - gestural overlap
  - articulatory reduction
  - prosodic reorganization
  - deletion and lenition of segments and syllables
    - Baese-Berk et al. 2018; Cohen Priva and Gleason 2018; Crystal and House 1988; Ernestus 2014; Fougeron and Jun 1998
- Such variation could potentially blur the contrast between long and short segments.

Perceptual compensation

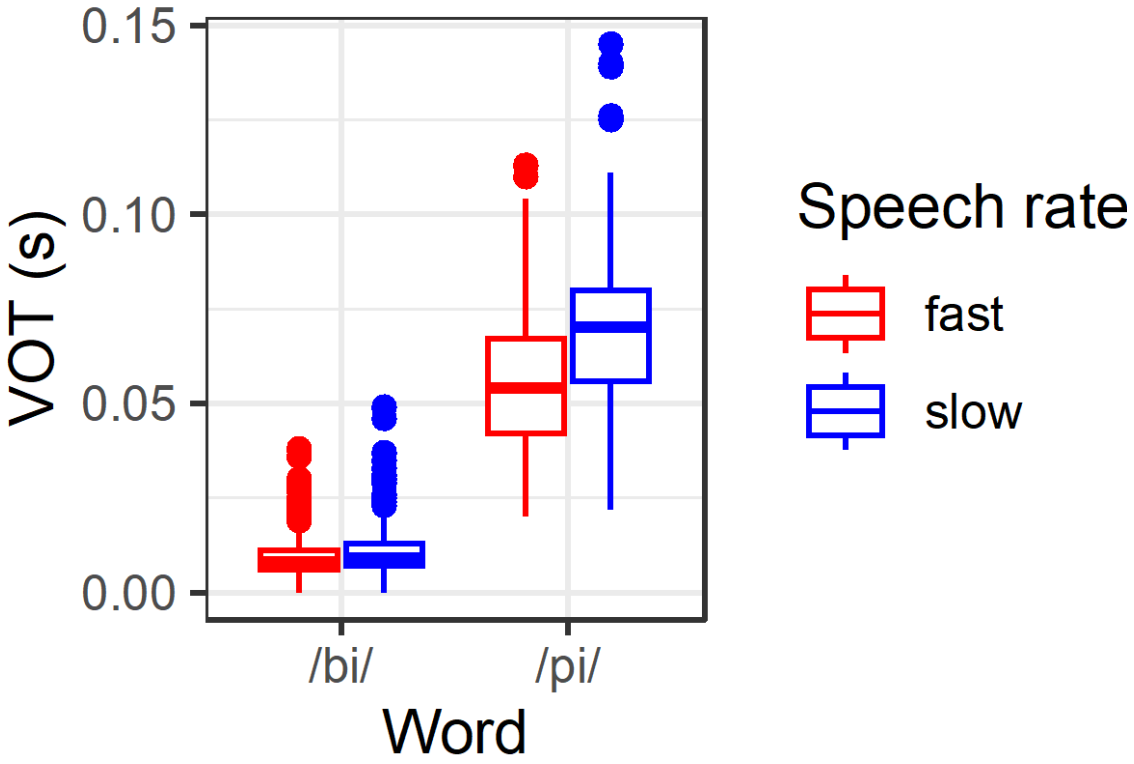
- Listeners take into account the context of speech and adjust their perception in a process known as **perceptual compensation**.
    - Drager 2011; Hay et al. 2006a; Hay et al. 2006b; Johnson et al.1999; Koops 2008; Mann 1980; Mitterer 2006; Niedzielski 1999; Schertz et al. 2019; Strand 1999; Strand and Johnson 1996; Yu 2010
- Speech rate and perceptual compensation
    - Listeners are more likely to perceive a given duration as “long” when embedded in fast speech than in slow speech, other things being equal.

Speech rate & VOT **production** (English)

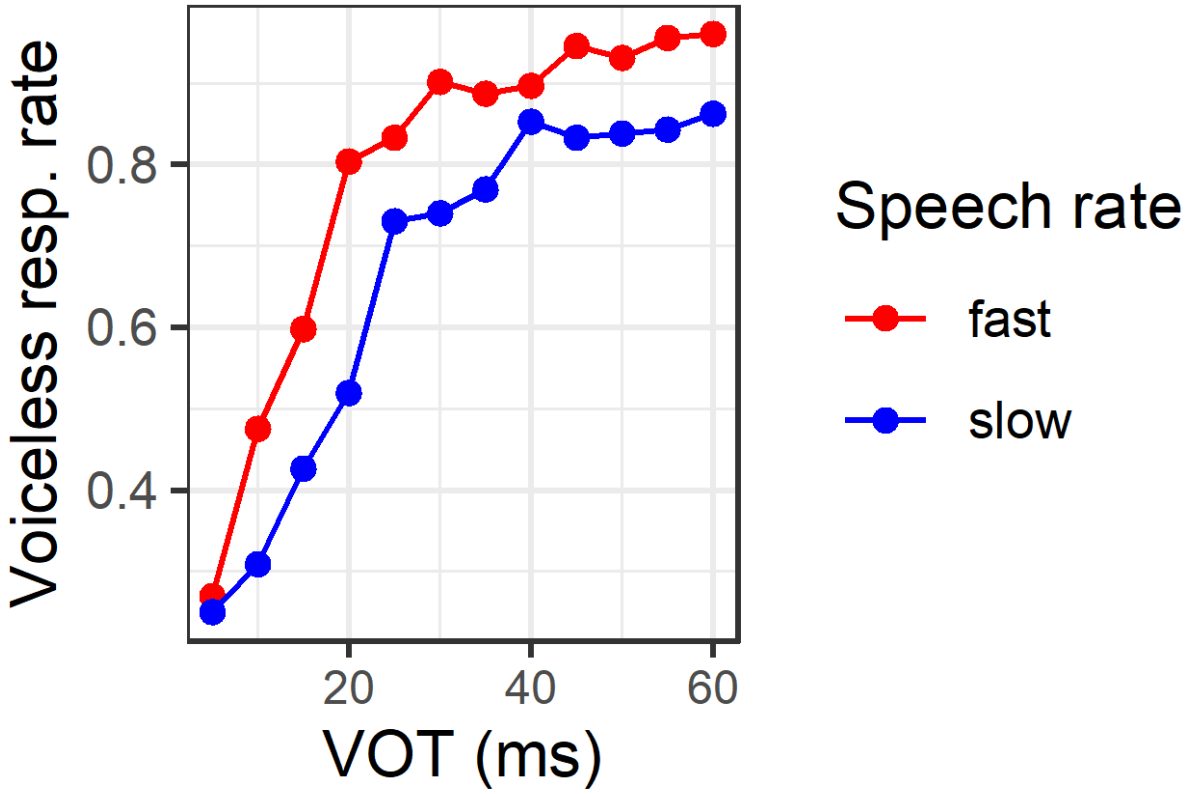


Kang, et al. (2018)

Speech rate & VOT **production** (English)



Speech rate & VOT **perception** (English)



Kang, et al. (2018)

Korean stops

	Fortis (unaspirated)	Lenis (slightly aspirated)	Aspirated (heavily aspirated)
Plosives	/p’ t’ k’/	/p t k/	/p^h t^h k^h/

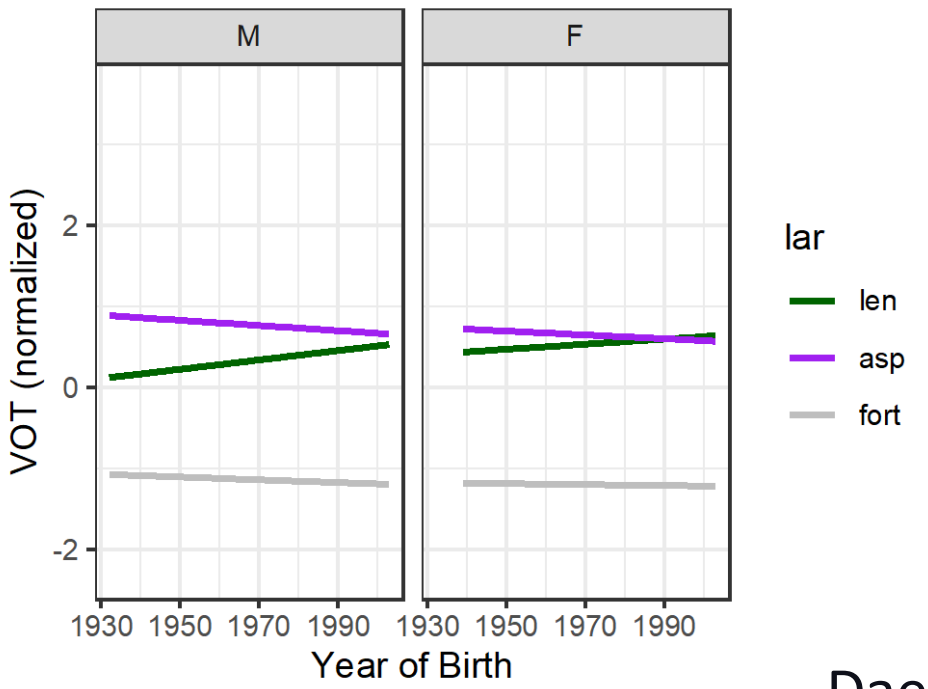
Short VOT Long VOT

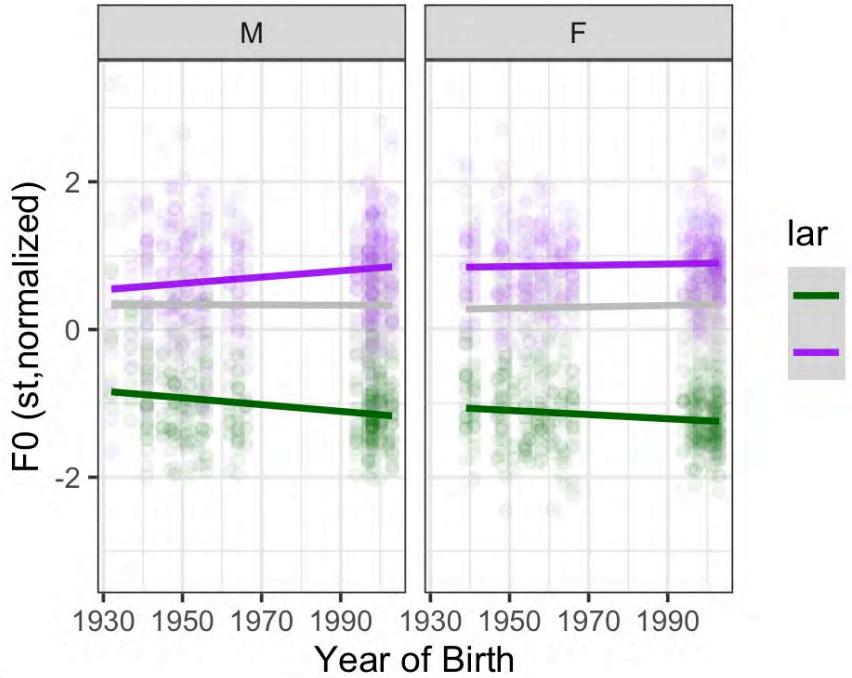
- Three-way contrast in stops

/t’al/ ‘daughter’ /tal/ ‘moon’ /t^hal/ ‘mask’

VOT merger in Korean stops

- **Aspirated (long VOT)** and **lenis (intermediate VOT)** stops are merging in VOT, while the F0 of following vowel is taking over as the primary cue.





Daejeon Korean

Kang, Gao, Yun & Ryu (this conference)

VOT merger in Korean stops

- **Aspirated (long VOT)** and **lenis (intermediate VOT)** stops are merging in VOT, while the F0 of following vowel is taking over as the primary cue.
- Pan-Korean change: A similar trend is reported for almost all dialects of Korean examined.

  Silva (2022, 2006), Wright (2007), Kang & Guion (2008), Jin (2008), Kang & Han (2012, 2013), Beckman et al. (2014), Kang (2014), Kim & Byun (2014), Shin (2015), Kang & Nagy (2016), Byun (2016), Jin & Silva (2017), Ahn (2017), Bang et al. (2018), H. Lee & Jongman (2018), Kang, Han, Ryu, Schertz, & Yun (to appear), Kang, Schertz & Han (2022)

VOT merger in Korean stops

- **Aspirated (long VOT)** and **lenis (intermediate VOT)** stops are merging in VOT, while the F0 of following vowel is taking over as the primary cue.
- The change is led by younger and female speakers, resulting in **age- and gender-based variation**.

Goals of the study

- Do listeners compensate for speech rate when the target duration contrast (VOT) is undergoing merger and VOT is no longer a reliable cue?
- If they do, do they make adjustment differently depending on
  - **the listeners' own** speech pattern?
    - Ladefoged & Broadbent 1957, Janson 1983, Miller & Grosjean 1997, Hay et al. 2006, Drager 2011, Fridland & Kendall 2012, Kendall & Fridland 2012
  - **the talker's** (expected) speech pattern?
    - Johnson et al. 1999, Strand 1999, Niedzielski 1999, Hay et al. 2006a,2006b, Koops et al. 2008, Drager 2010, Schertz et al. 2017

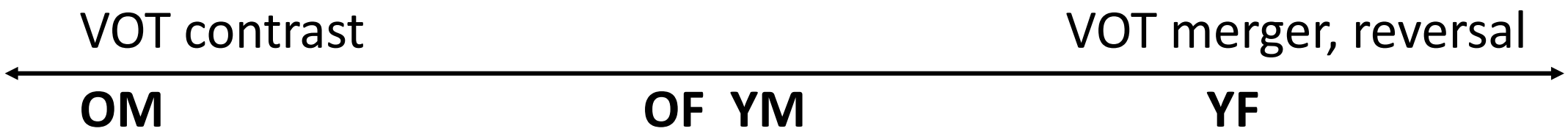
Language and participants

- Daejeon Korean
  - Spoken in the city of Daejeon in the central region of South Korea.
  - 5th largest city in Korea (pop. 1.5 million)
- Participants
  - 81 speakers of Daejeon Korean

	Younger (20s)	**Older (50s +)**
Female	20	21
Male	20	20

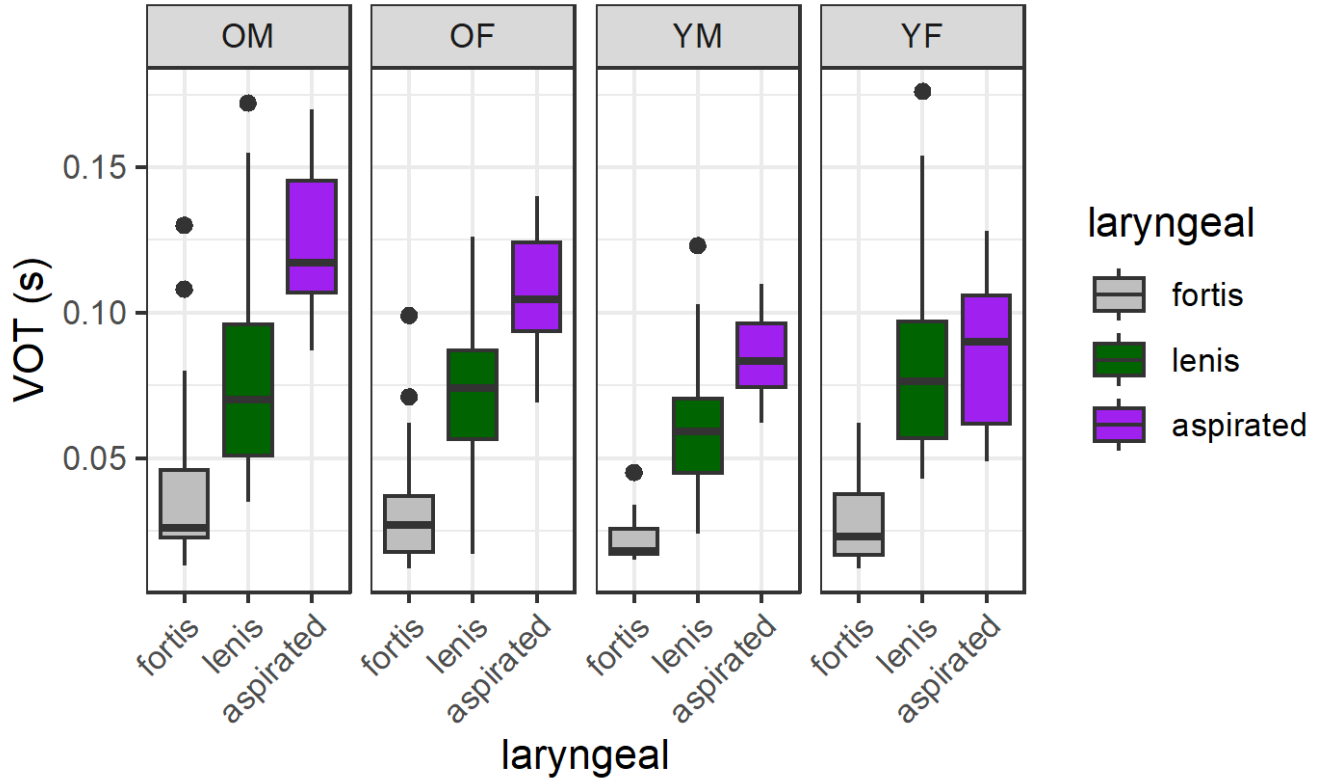
Perception stimuli talkers

- Self-identified Daejeon natives
- Representing four age/gender groups expected to be at different stages of sound change.



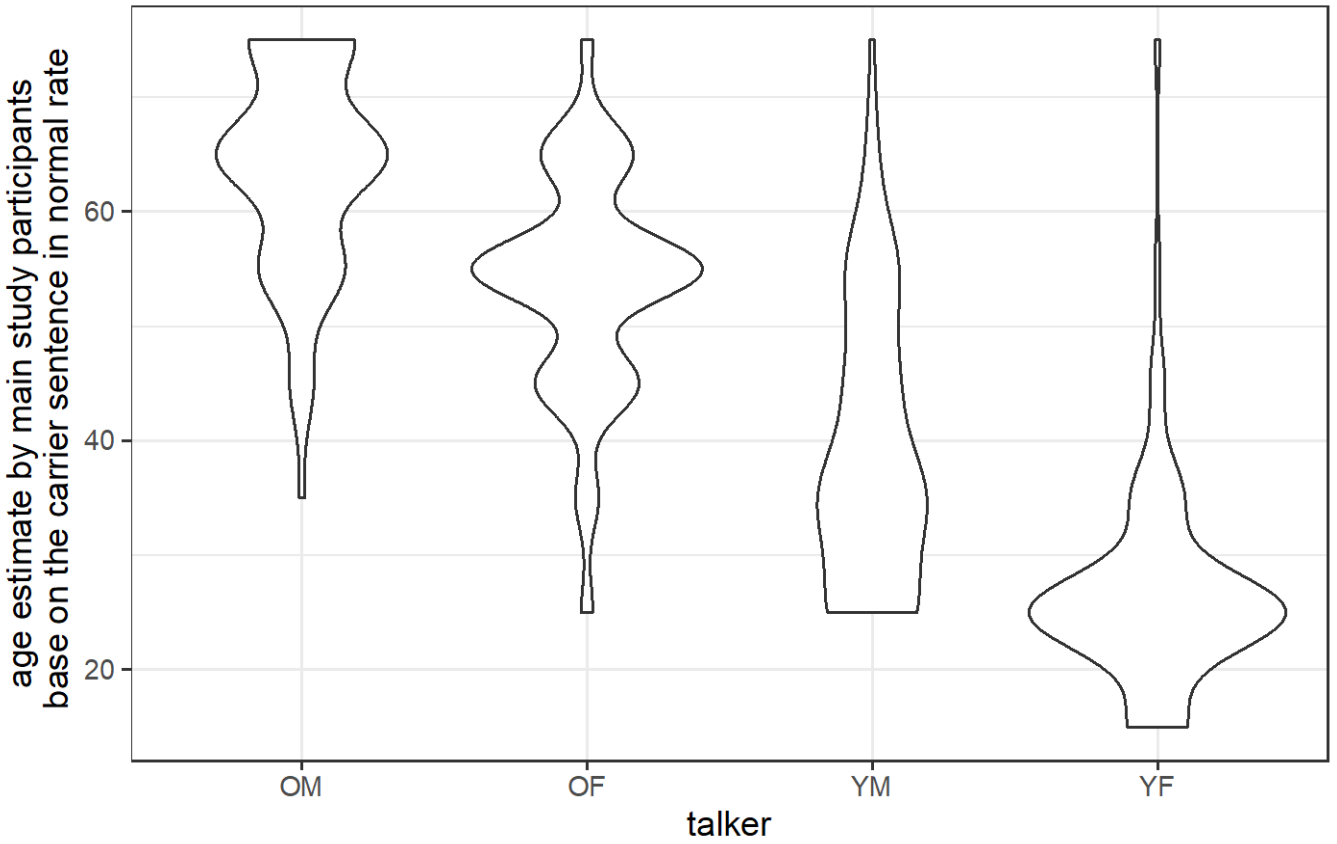
Perception stimuli talkers

- Production matches the expected pattern based on their age/gender



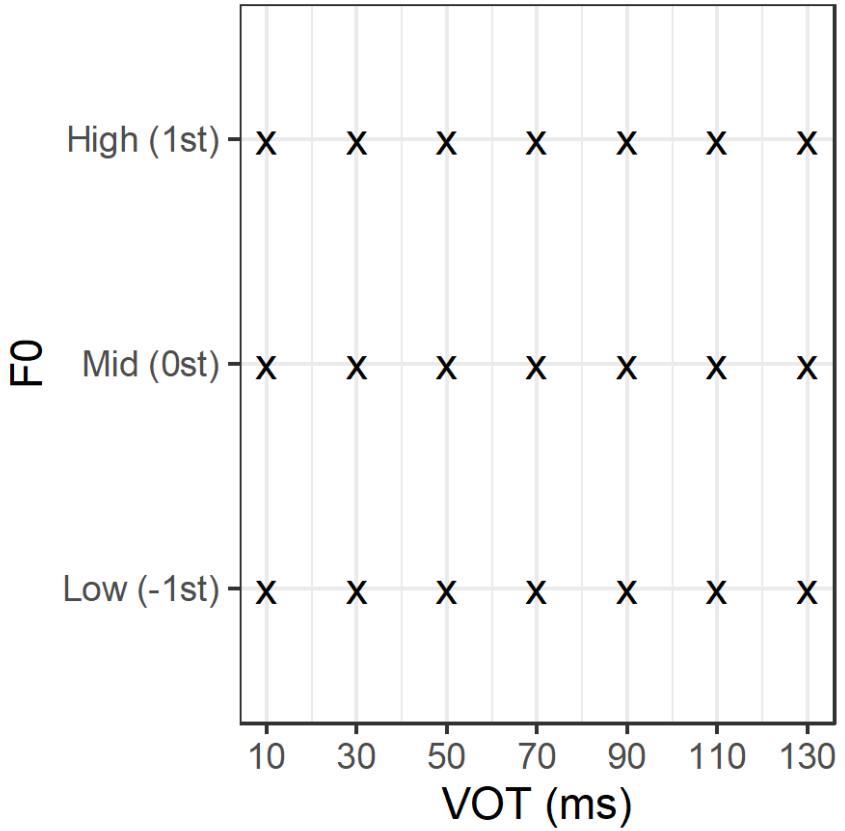
Perception stimuli talkers

- Judged to be younger vs. older by our main study participants as well as pilot participants



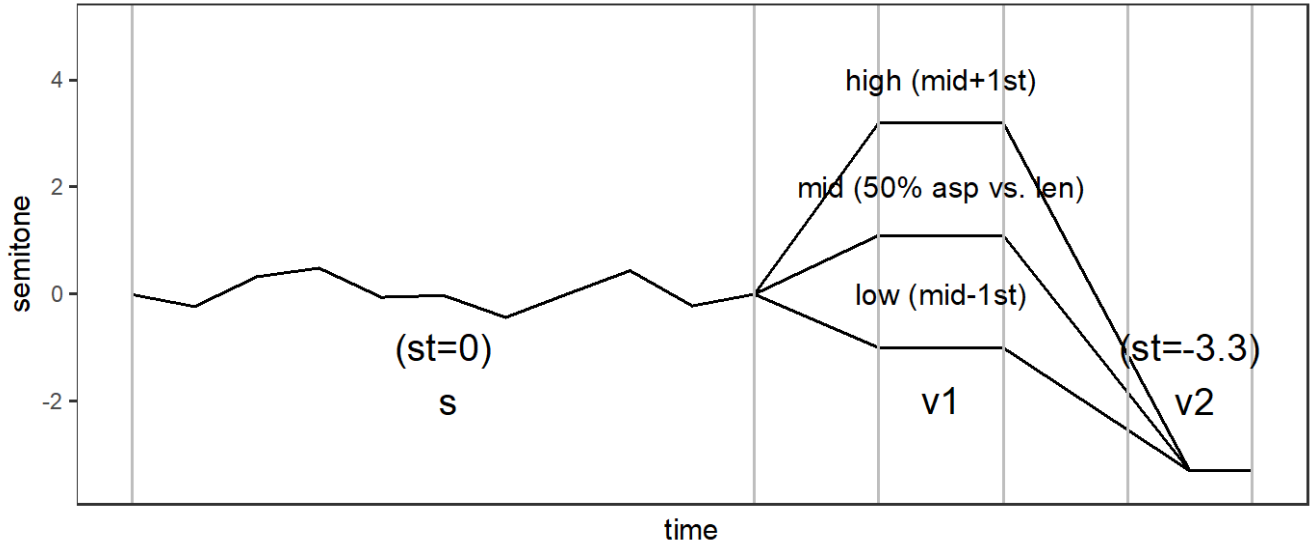
Perception stimuli

- Comparable acoustic space across the participants
- A monosyllable /Pa/ ([p^ha] baseline)
- Manipulated to vary in:
  - VOT (7 steps, 10-130 ms)
  - F0 (3 steps, Low-Mid-High, 1 semitone intervals)



F0 range and manipulation

- Mid: 50% lenis-aspirated response level for each talker, determined by a 29 speaker pilot
- Low and High are 1 semitone away from Mid



Perception stimuli

- Carrier sentence:
  - 문장 맨 마지막 말은 __다. “The last word of the sentence is __.”
  - Manipulated to vary between **slow** (120 % of mean duration of all natural productions by the four talkers) and **fast** rates (80 %).
- 168 stimuli (2 rates * 7 VOT levels * 3 F0 levels * 4 talkers)

Older Male		Younger Female	
Slow	**Fast**	**Slow**	**Fast**
			
VOT = 70ms		VOT = 70ms	
F0 = Mid (234 Hz)		F0 = Mid (103 Hz)	

Perception stimuli

- Carrier sentence:
  - 문장 맨 마지막 말은 __다. “The last word of the sentence is __.”
  - Manipulated to vary between **slow** (120 % of mean duration of all natural productions by the four talkers) and **fast** rates (80 %).
- 168 stimuli (2 rates * 7 VOT levels * 3 F0 levels * 4 talkers)

Older Male		Younger Female	
Slow	**Fast**	**Slow**	**Fast**
			
VOT = 70ms		VOT = 70ms	
F0 = Mid (234 Hz)		F0 = Mid (103 Hz)	

Perception stimuli

- Carrier sentence:
  - 문장 맨 마지막 말은 __다. “The last word of the sentence is __.”
  - Manipulated to vary between **slow** (120 % of mean duration of all natural productions by the four talkers) and **fast** rates (80 %).
- 168 stimuli (2 rates * 7 VOT levels * 3 F0 levels * 4 talkers)

Older Male		Younger Female	
Slow	**Fast**	**Slow**	**Fast**
			
VOT = 70ms		VOT = 70ms	
F0 = Mid (234 Hz)		F0 = Mid (103 Hz)	

Perception stimuli

- Carrier sentence:
  - 문장 맨 마지막 말은 __다. “The last word of the sentence is __.”
  - Manipulated to vary between **slow** (120 % of mean duration of all natural productions by the four talkers) and **fast** rates (80 %).
- 168 stimuli (2 rates * 7 VOT levels * 3 F0 levels * 4 talkers)

Older Male		Younger Female	
Slow	**Fast**	**Slow**	**Fast**
VOT = 70ms		VOT = 70ms	
F0 = Mid (234 Hz)		F0 = Mid (103 Hz)	

Identification (3 Alternative Forced Choice)

파다	바다	빠다

Statistical analysis

- A mixed-effects logistic regression model
  - Dependent variables: laryngeal type (**Lenis = -0.5**, **Aspirated. = 0.5**)
  - Fixed effects:
    - VOT (10, 30, 50, 70, 90, 110, 130) → z-scored
    - F0 (-1,0,1) → z-scored
    - Speech rate (**Fast = - 0.5**, **Slow = 0.5**)
    - Listener Gender (Male = -0.5, Female = 0.5)
    - Listener Age (Old = -0.5, Young = 0.5)
    - Talker Gender (Male = -0.5, Female = 0.5)
    - Talker Age (Old = -0.5, Young = 0.5)
    - Interactions: (VOT + F0 + Rate) * (ListenerGender*ListenerAge + TalkerGender*TalkerAge)
  - Random effects: Speaker
  - Stepwise regression, backward elimination (Buildmer)

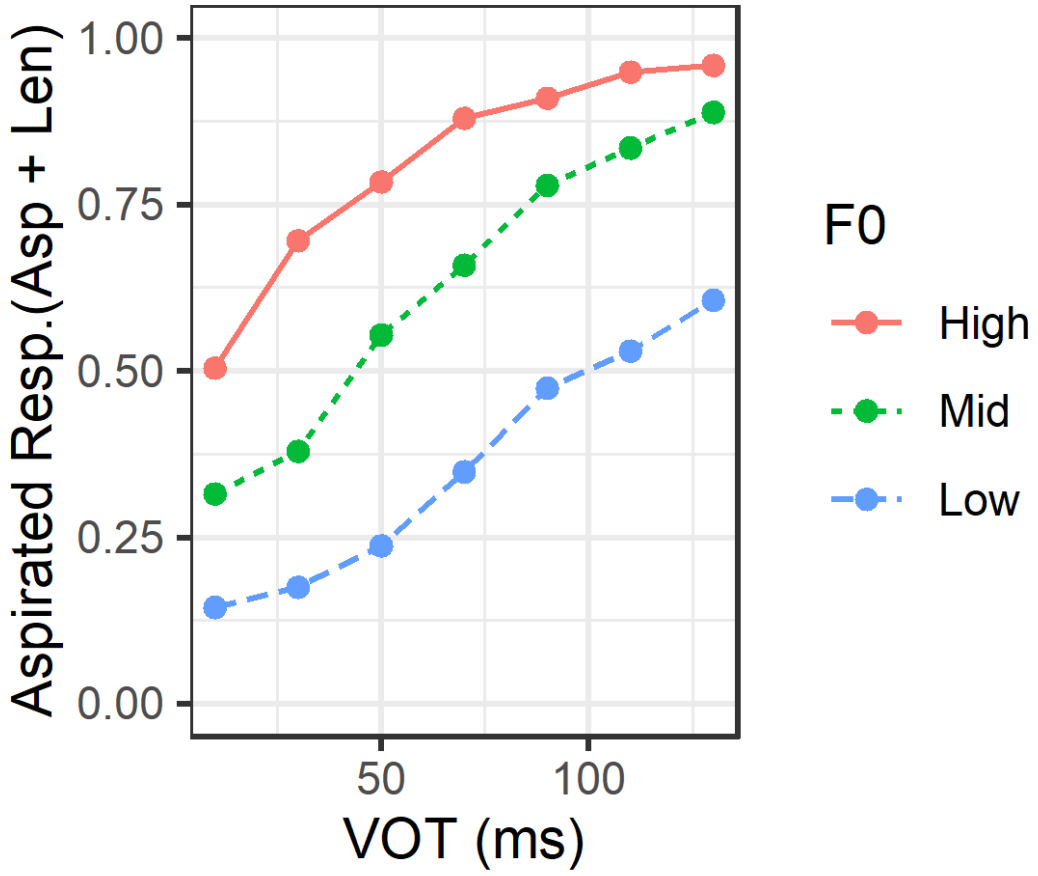
Statistical analysis

- A mixed-effects logistic regression model
  - Dependent variables: laryngeal type (**Lenis = -0.5**, **Aspirated. = 0.5**)
  - Fixed effects:
    - VOT (10, 30, 50, 70, 90, 110, 130) → z-scored
    - F0 (-1,0,1) → z-scored
    - Speech rate (**Fast = - 0.5**, **Slow = 0.5**)
    - Listener Gender (Male = -0.5, Female = 0.5)
    - Listener Age (Old = -0.5, Young = 0.5)
    - Talker Gender (Male = -0.5, Female = 0.5)
    - Talker Age (Old = -0.5, Young = 0.5)
    - Interactions: (VOT + F0 + Rate) * (ListenerGender*ListenerAge + TalkerGender*TalkerAge)
  - Random effects: Speaker
  - Stepwise regression, backward elimination (Buildmer)

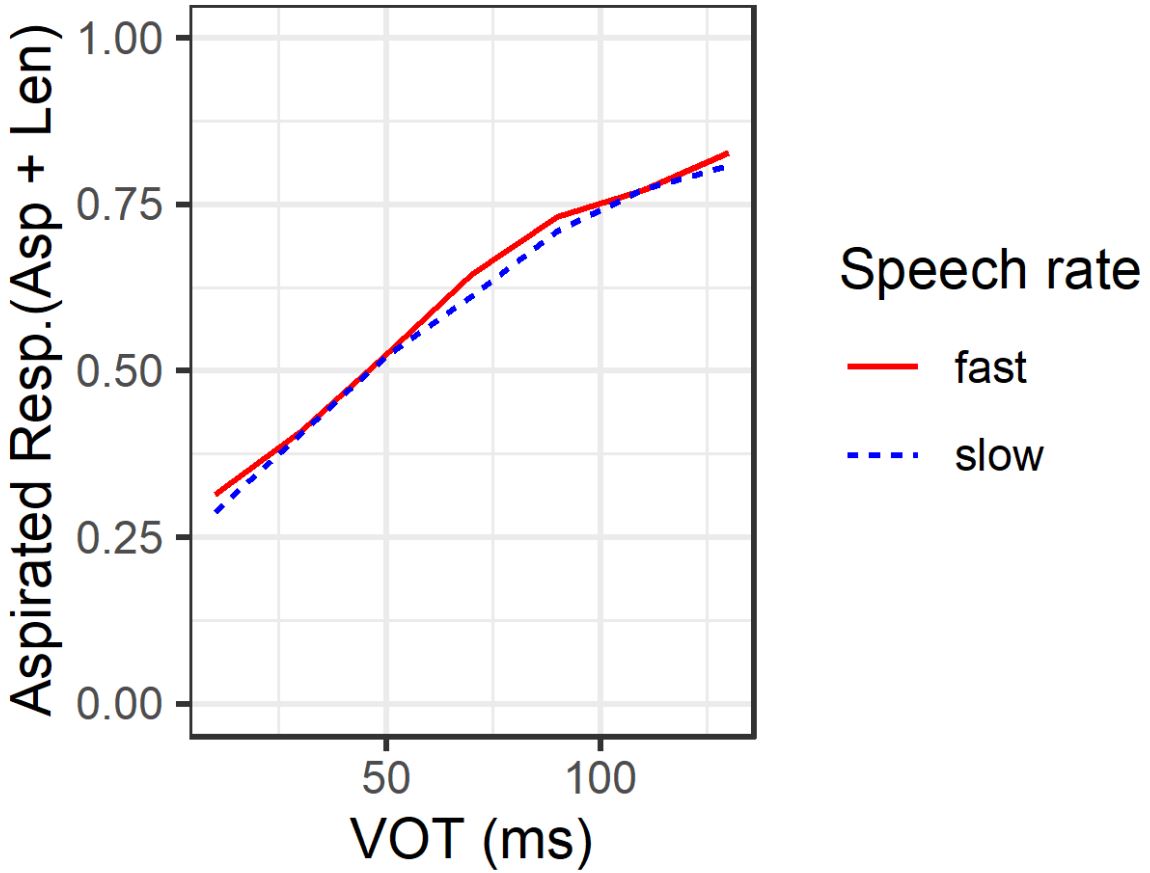
Are more aspirated stops chosen in fast speech?

Is the rate effect modulated by the age and gender of the listener or the talker?

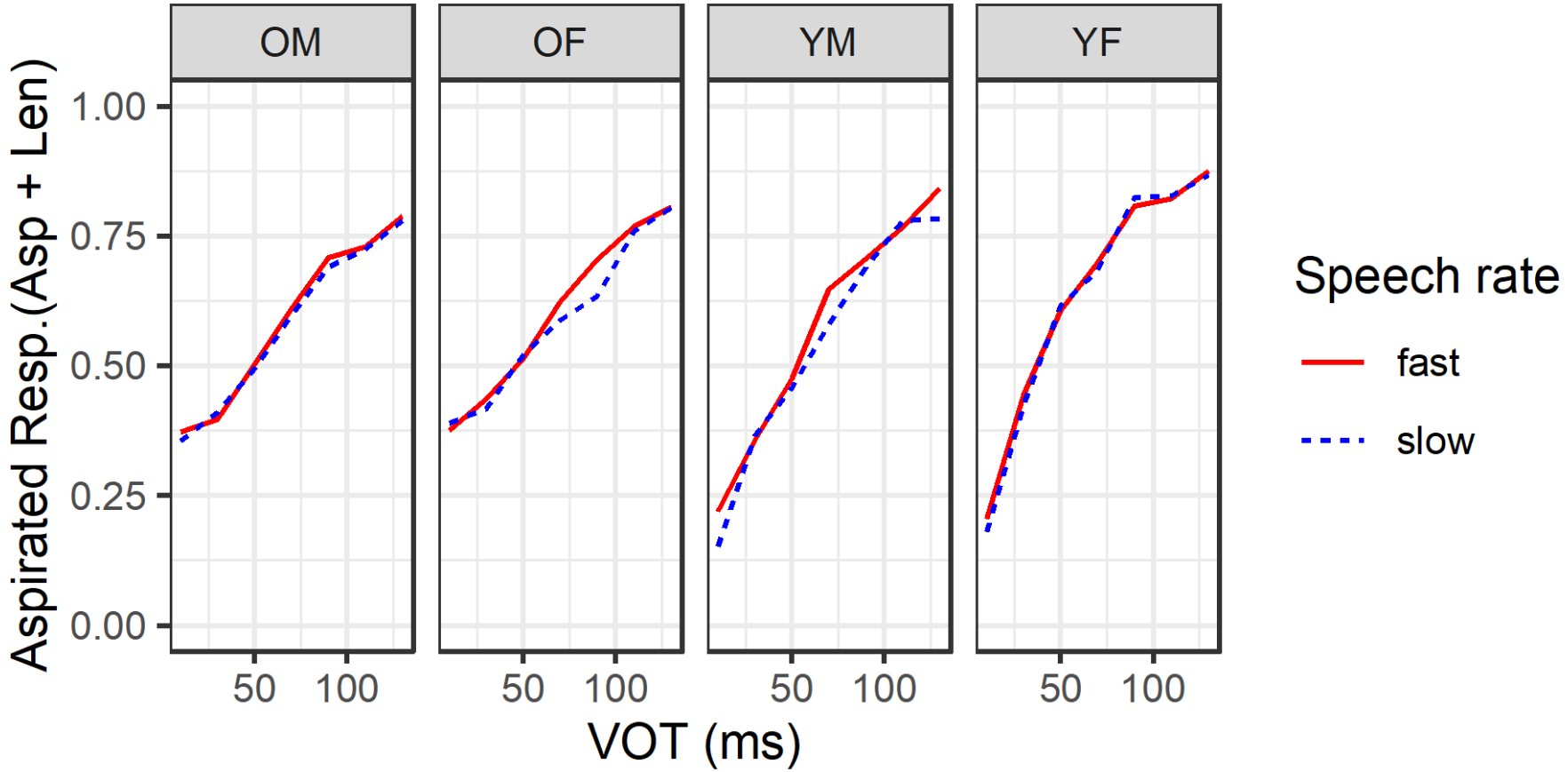
Main effects: VOT***, F0***



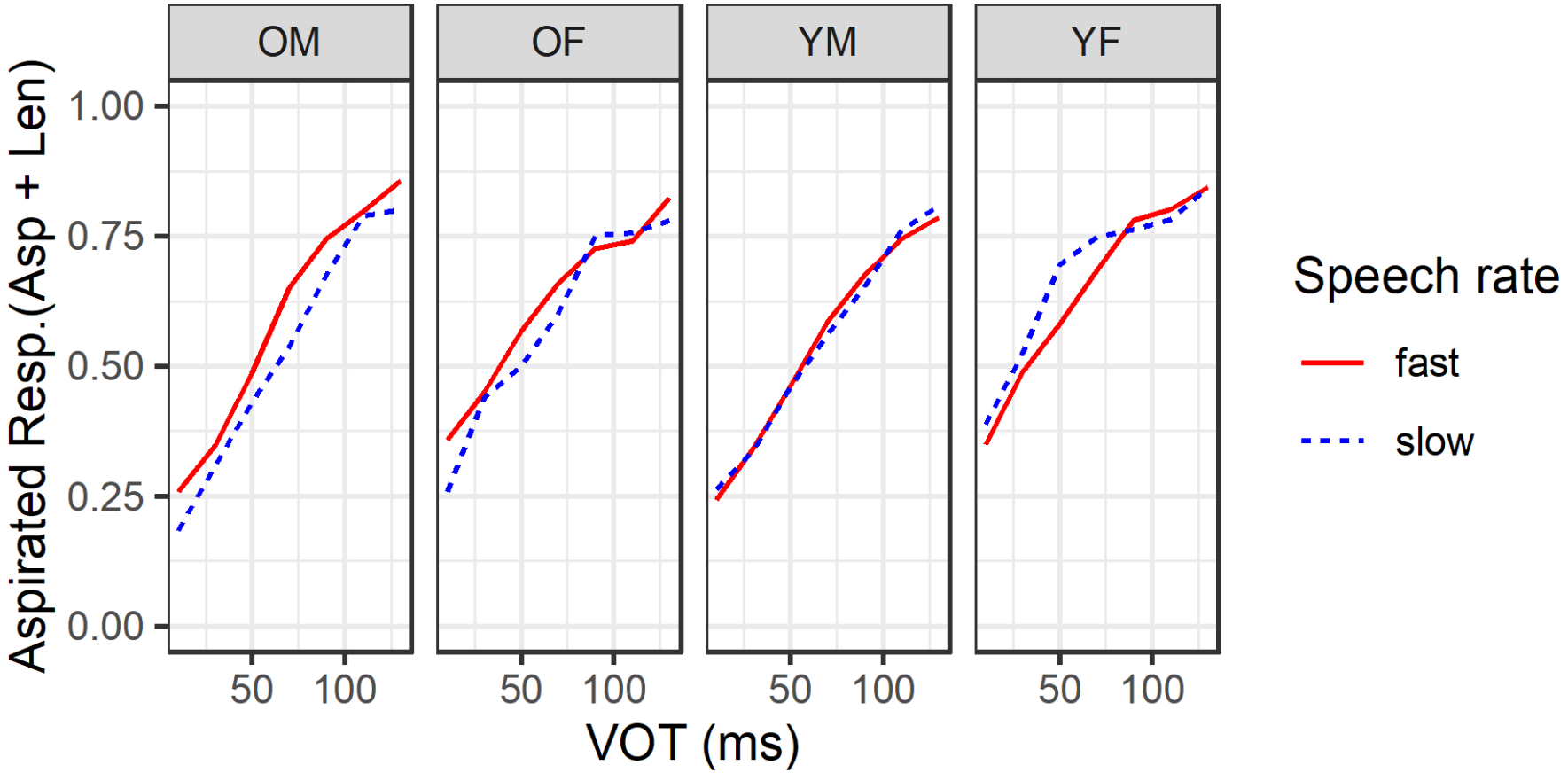
Main effect: Speech rate (p = 0.135)



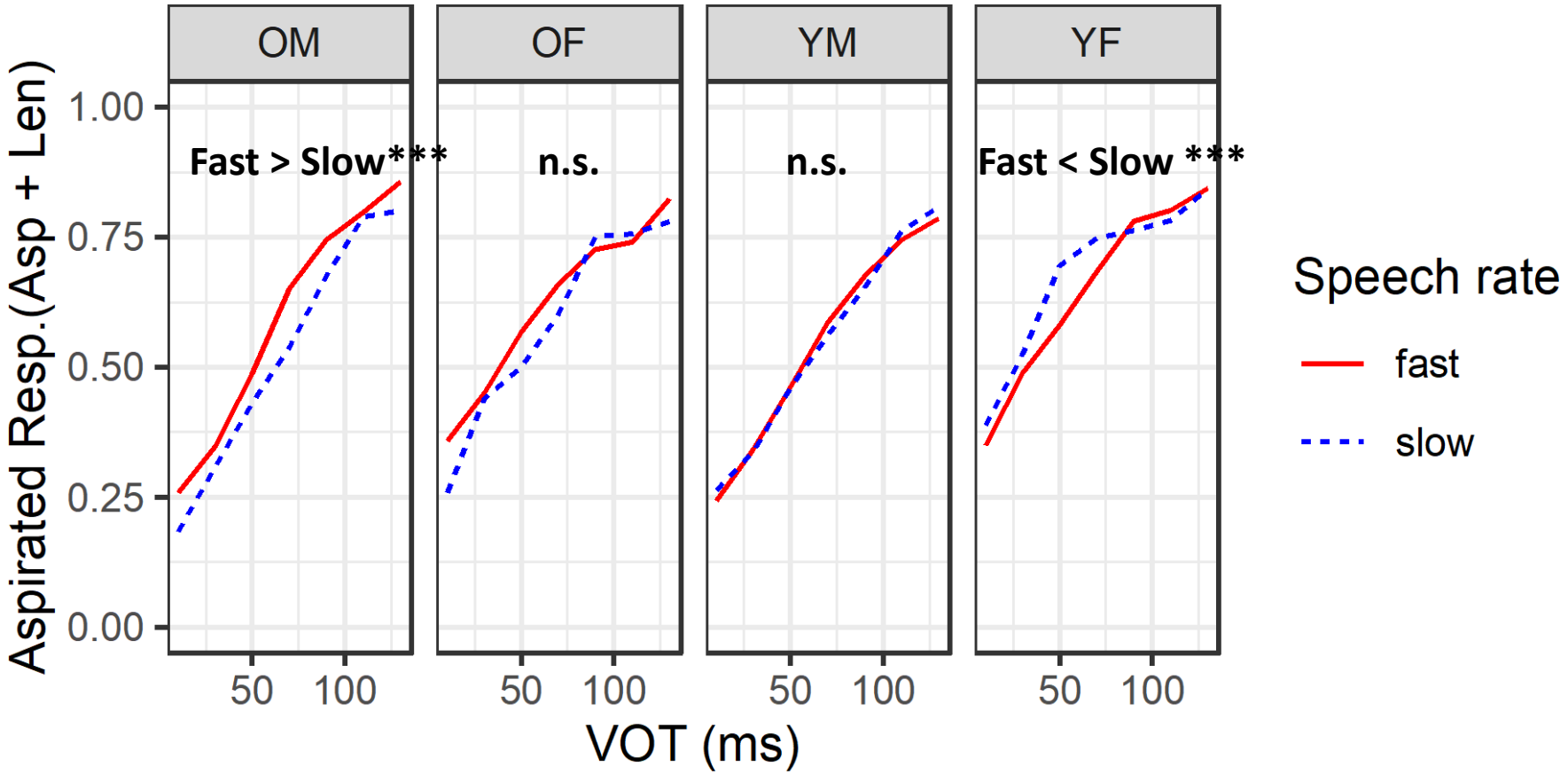
Rate x Listener age : n.s.
Rate x Listener gender: n.s.



Rate x Talker age***
Rate x Talker gender**



Rate * Talker: Post-hoc tests by talker



Summary and Conclusion

- The rate effect was modulated by the talkers' age and gender, but not by the listeners’ own age and gender.
- Post-hoc tests find the expected rate effect (more aspirated stop responses in fast speech) only for the older male talker, a group that retains the VOT contrast more robustly.
- A significant rate effect in the opposite direction was found for the young female talker, who belongs to a group where the VOT distinction has merged and possibly reversed.
- The results are significant in showing the way listeners compensate for speech rate reflects the contrastive status of the duration cue undergoing a sound change in the talkers’ speech.

Acknowledgements

- Participants
- Jung Haechan, Park Jeongin, and Park Beomjoon for help with data collection
- Hyongseok Kwon for help with programming
- SSHRC (Social Sciences and Humanities Research Council of Canada) for funding