Speech rate accommodation in Korean vowel production and perception

Yoonjung Kang, Suyeon Yun, and Na-Young Ryu

University of Toronto, Chungnam National University, Pennsylvania State University

NWAV-AP 7

December 14 2022

Chulalongkorn University

Speech rate variation

- Compared to slow/normal speech, fast speech is marked by shortening of segments, gestural overlap, articulatory reduction, prosodic reorganization, as well as 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 variation is ubiquitous. It is one of the common sources of synchronic variation and a potential source of diachronic sound change.

Speech rate – interspeaker variation

- Speech rate not only varies within speaker but also across speakers and speech rate is associated with the speaker's age, dialect, gender, ethnicity, as well as perceived personality traits.
- In particular, older speakers generally speak slower than younger speakers, and speech rate is a salient cue for speakers' age.
  - Harnsberger et al. 2008; Jacewicz et al. 2009; Jacewicz et al. 2010; Kendall 2013; Schnoebelen 2009; Shipp et al. 1992; Skoog Waller and Eriksson 2016; Skoog Waller et al. 2015

Perceptual compensation

- Speech is highly variable: the same target structure can be realized very differently depending on the linguistic context and the speaker.
- Speech communication is largely successful because listeners take into account the effect of the context, and calibrate their perception to arrive at the intended message 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

Perceptual compensation – interspeaker variation

- Speakers vary in the degree of perceptual compensation depending on their age (Drager 2011), gender (Yu 2010) and their own production pattern (Kang et al. 2018)
- Such mismatches in perceptual compensation are argued to be a key step toward a rise of novel variants and eventual sound change (Ohala 1994, Garrett and Johnson 2012).

Goals of our study

- How does speech rate variation affect vowel quality production and perception (and lead to eventual sound change)?
- Production:
  - We expect vowels to be generally raised or centralized in fast speech.
  - (Do speakers differ in how speech rate affects vowel production?)
- Perception
  - If vowels are systematically higher/centralized in fast speech, do listeners adjust their perception in fast vs. slow speech accordingly?
  - (Do listeners compensate for speech rate in perception differently depending on the age and gender of the talkers and the listeners?)

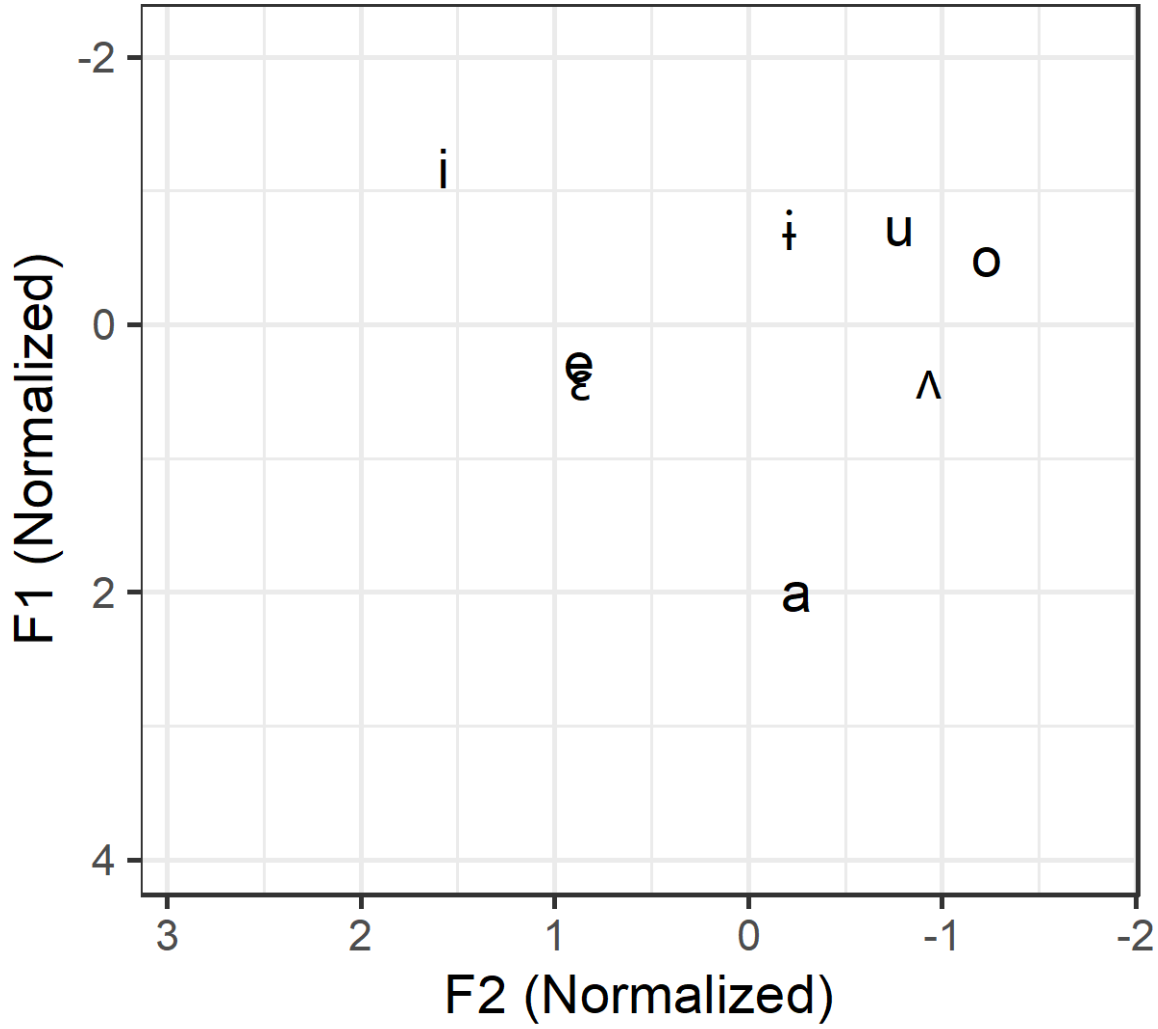
Language and participants

- Daejeon Korean
  - Spoken in the city of Daejeon in the central region of South Korea.
  - Chungnam dialect, associated with a stereotype of “slow” speech
- Participants
  - 81 speakers of Daejeon Korean

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

Monophthongs of Daejeon Korean

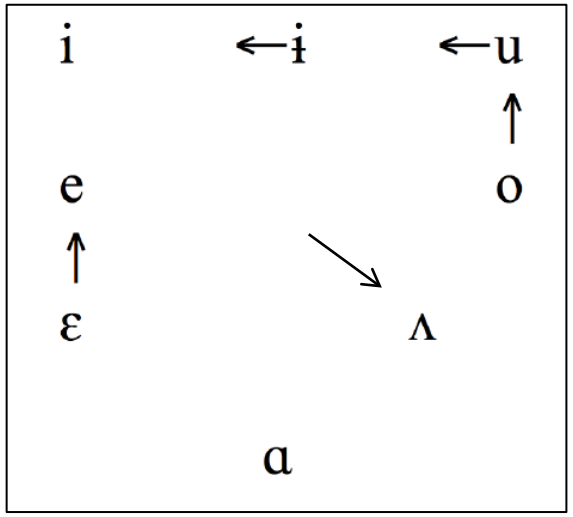
i	ɨ	u
e	ʌ	o
ɛ	a	

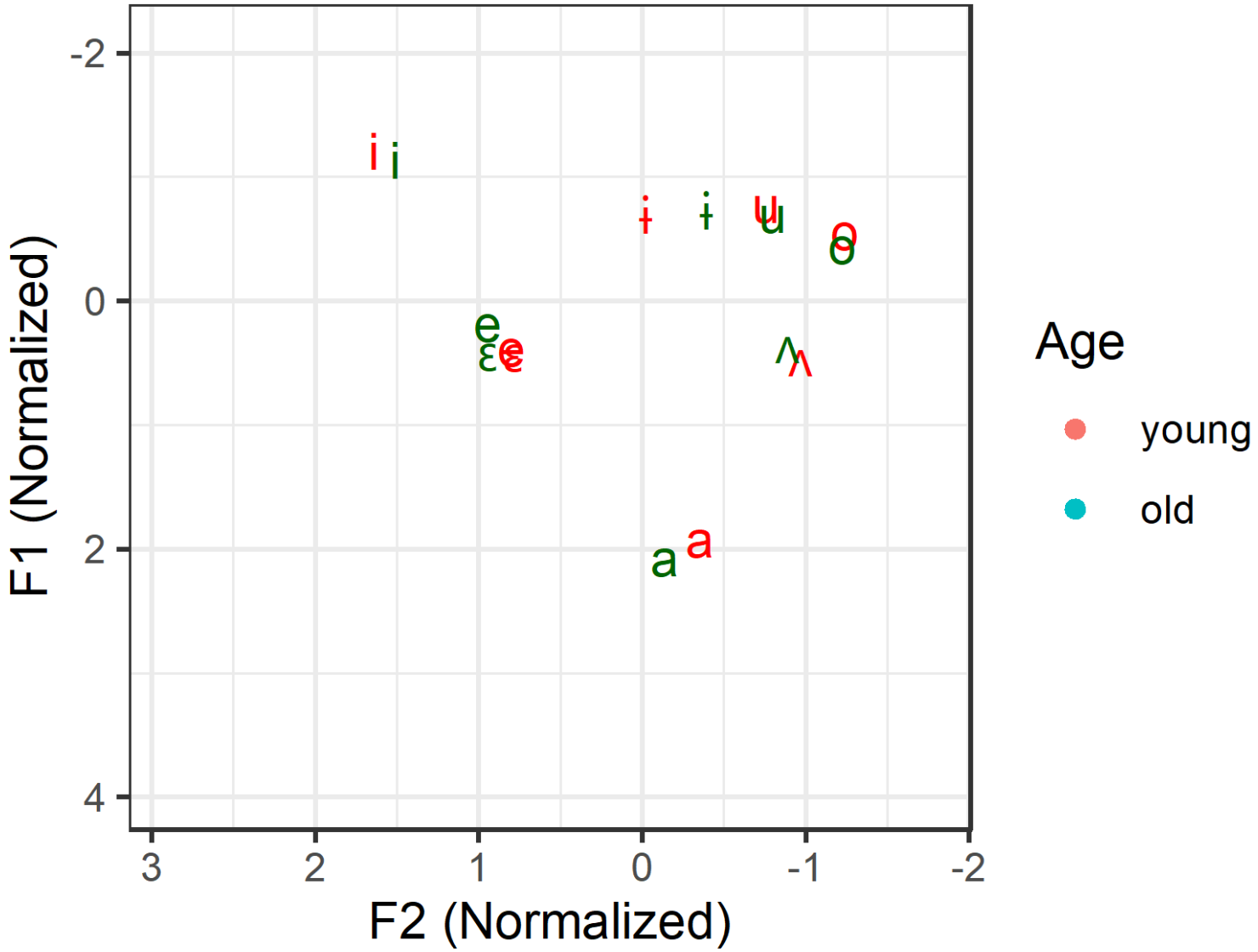


Monophthongs of Daejeon Korean

- Back vowel shift
- /e/-/ɛ/ merger

similar to Seoul Korean (cf. Kang 2016)





Production

- Speech materials
  - 8 monophthongal vowels
  - Carrier sentence:
    - 문장 맨 마지막 말은 __다. “The last word of the sentence is __.”
- Model talker prompt
  - A male speaker of Daejeon Korean in his 40s
  - A question form of the carrier sentence:
    - 문장 맨 마지막 말은 **뭐**다? “The last word of the sentence is **what**?”
  - Manipulated to vary in speech rate:
    - Slow: 120% of the mean duration of the talker’s natural production
    - Fast: 80% of the mean duration of the talker’s natural production

Production

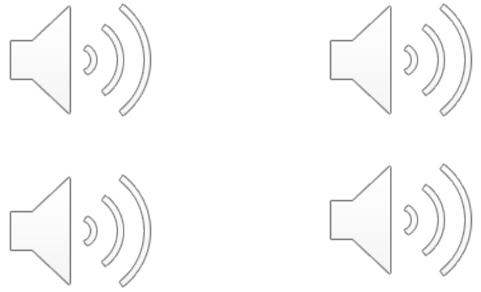
- Shadowing
  - The participants were instructed to produce the target sentence, displayed on screen with the target word filled in, trying to match the model talker's speech rate as best as they can.

문장 맨 마지막 말은 **오**다.

"The last word of the sentence is [o]."

Slow

Fast



Production

- Repetition
  - Each vowel target had 10 trials, 5 fast and 5 slow.
  - Fast and slow trials were mixed and presented in random order.
  - Each vowel was presented in a separate block.
- 8 vowels * 2 speech rates * 5 reps * 81 speakers = 6,480
- 43 tokens excluded due to mispronunciation or disfluency.

Production data processing

- Recordings were cut into individual sentence files using a Praat script.
- For each sentence file, a TextGrid with segmentation was created using the Korean Forced Aligner (https://tutorial.tyoon.net/).
- Target vowel segmentation was manually checked for segmentation.

Formant measurements

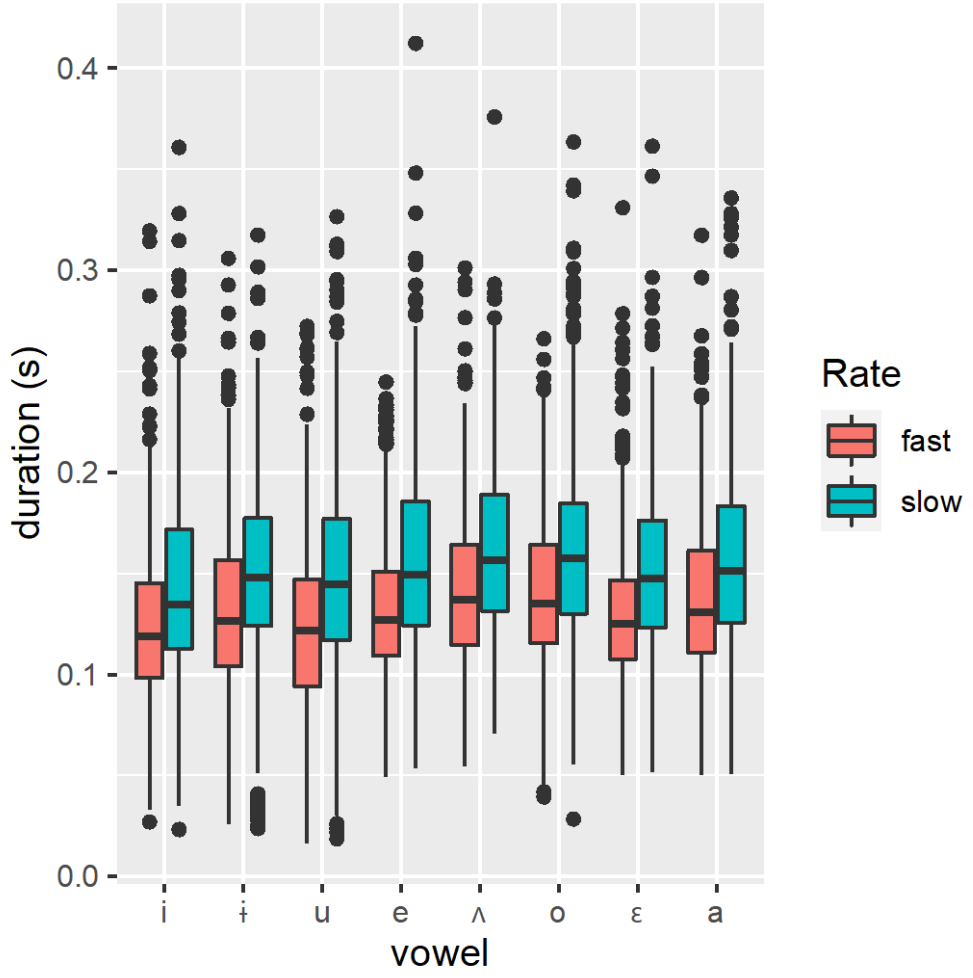
- Formant measurements were taken over the mid 20% of the vowel duration with a gender-specific formant ceiling (M = 5000 Hz, F = 5500 Hz).
- Lobanov normalization (by speaker z-score transformation)
- Outlier formant values were removed (> 2.5 SD of each vowel's distribution)
- To quantify the degree of peripheralization/centralization, Euclidean distance from the center of the vowel was calculated for each vowel.

Statistical analysis

- Linear mixed-effects models with maximal random effect structure, where possible.

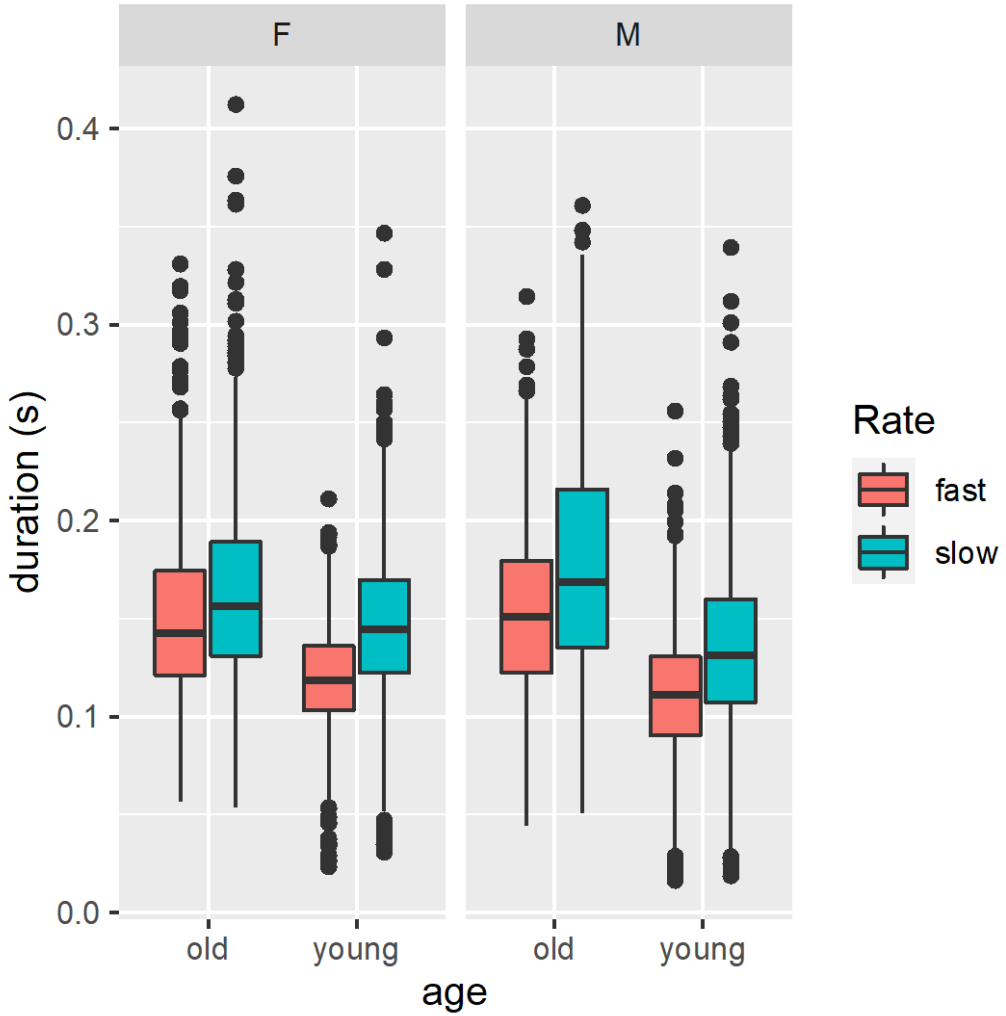
Speech rate and vowel duration

- Vowels are consistently produced with a longer duration in the slow condition, across all vowels.
- The task was successful in inducing the speech rate variation.



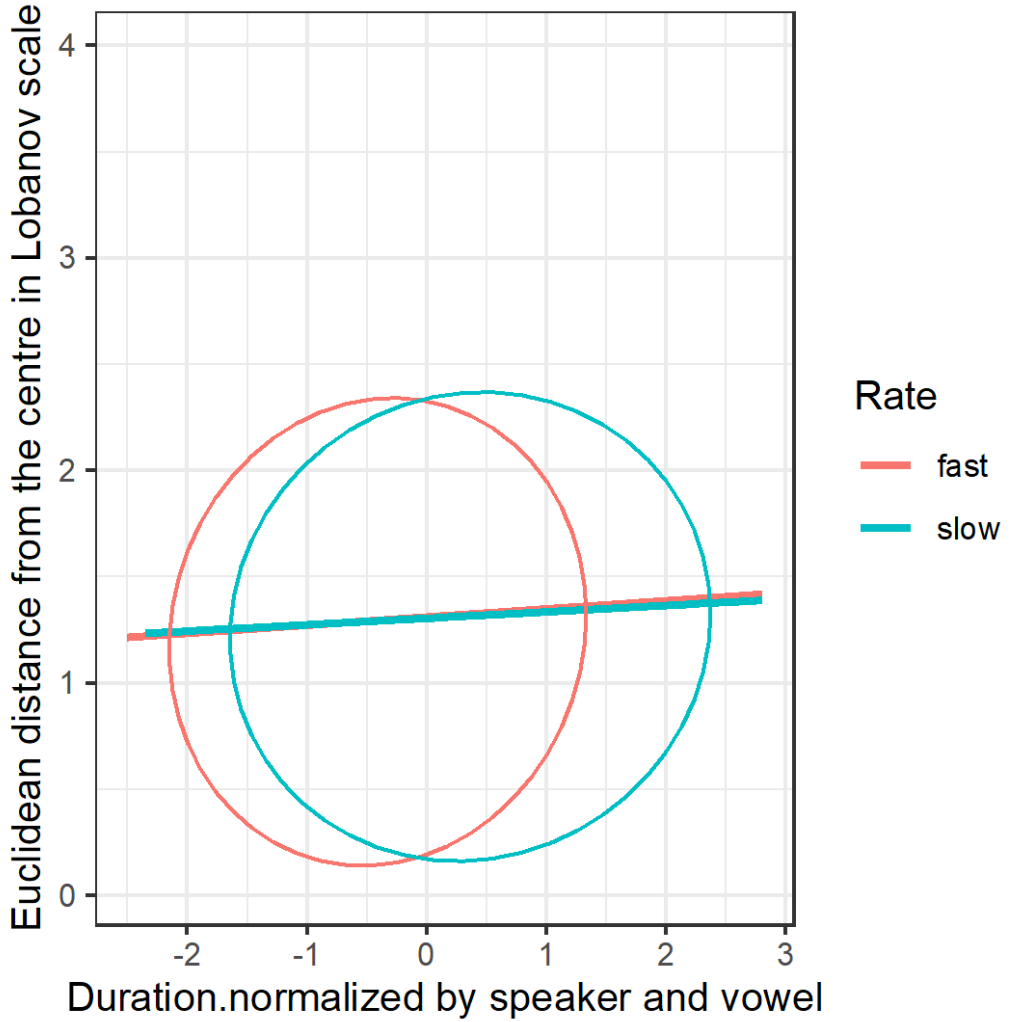
Speech rate and vowel duration

- Older speakers' vowels were overall longer than younger speakers'.
- Even in the shadowing task, the age-based rate variation is retained.
- There was no gender difference.



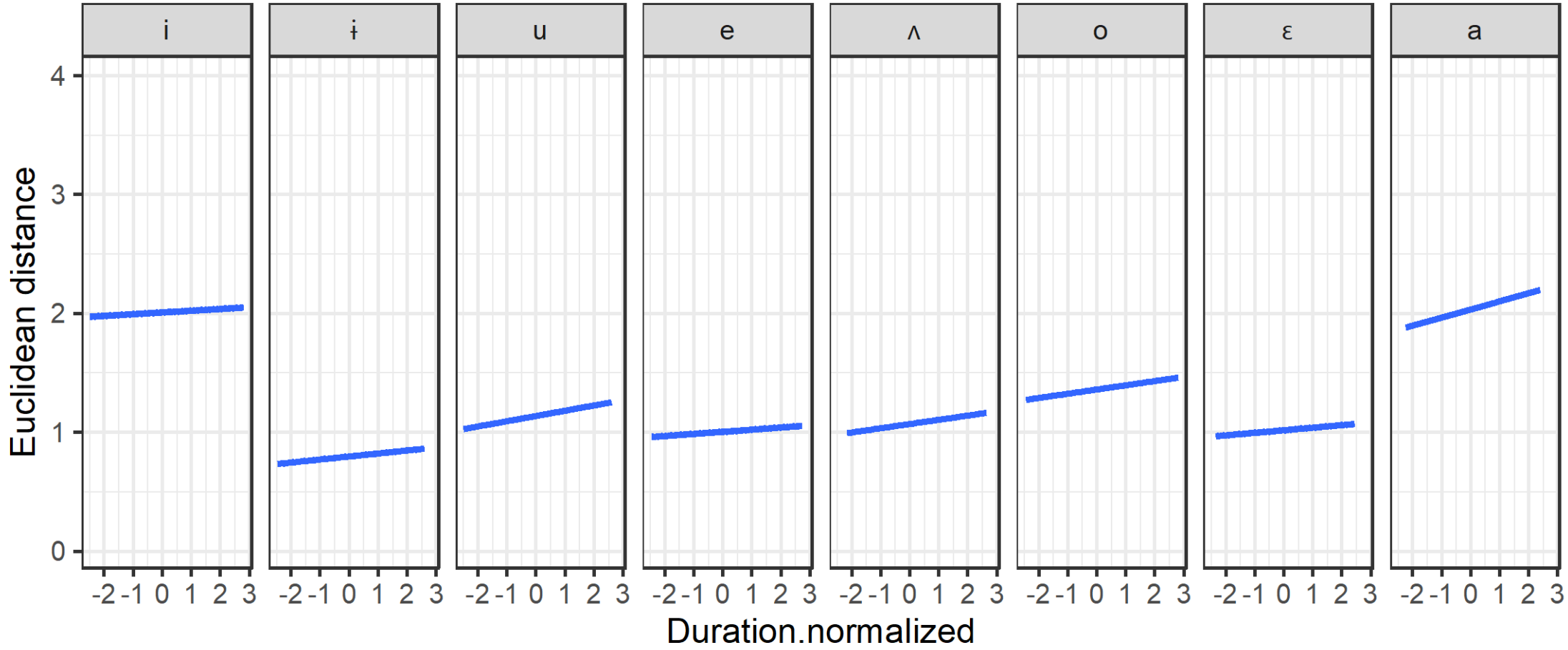
Duration and peripheralization

- Other things being equal, shorter vowels (~ fast speech) are produced more centralized and longer vowels are produced more peripheral.



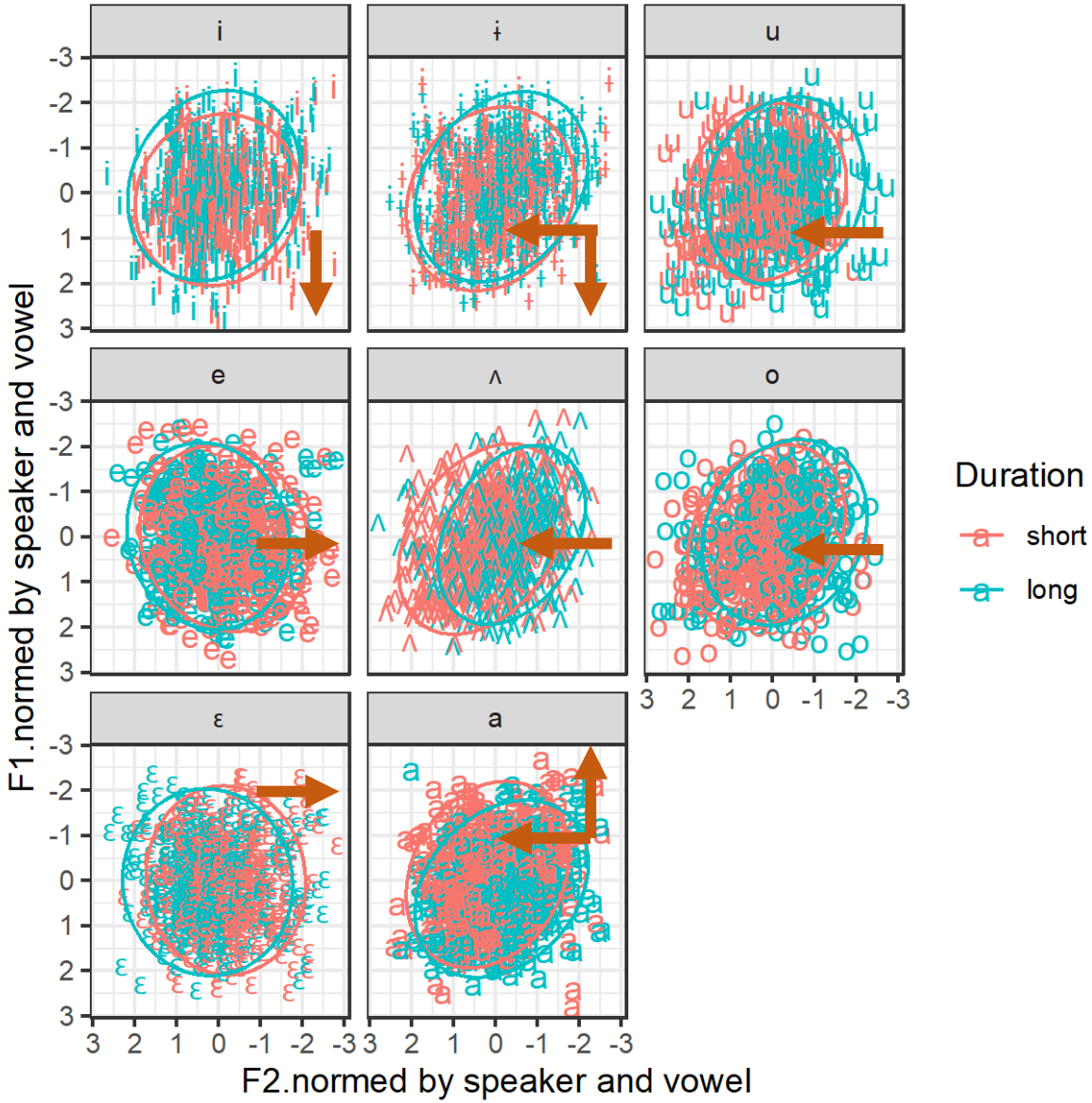
Duration and peripheralization

- This effect is significant for each individual vowel, except for /i/.



Duration and formants

- When F1 and F2 are examined separately.
- In short duration:
  - F1: high vowels [i,ɨ] lower (F1 raises) and low vowel [a] raises.
  - F2: back vowels [ɨ, ʌ, a, u, o] move front (F2 raises) and front vowels [e,ɛ] move back.

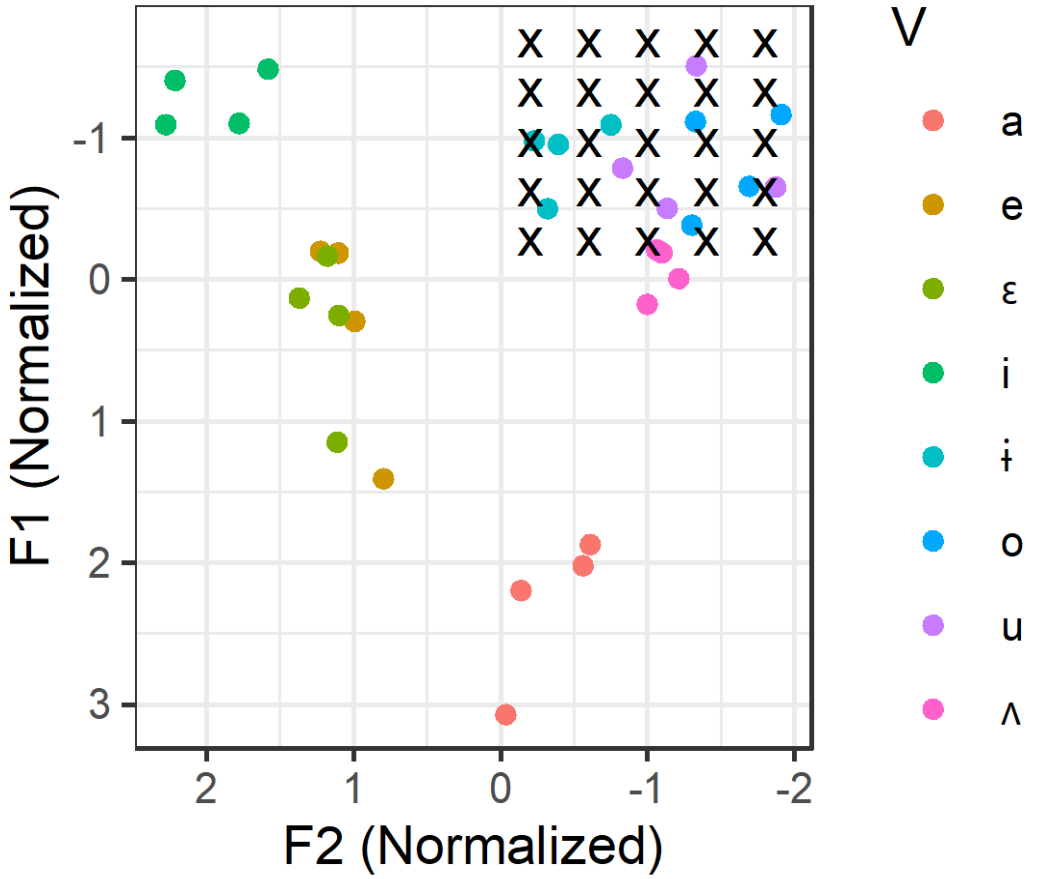


Perception

- Speech materials
  - Vowels embedded in the same carrier sentence as production.
  - Produced at normal speech rate, with multiple repetitions.
- Stimuli talkers
  - One each of younger (20s) female, younger male, older (50s+) female, and older male speaker (YF, YM, OF, OM)
  - Different from the production model talker
- Phoneme identification task
  - The participants heard stimuli (carrier sentence + target word) and chose the vowel heard, out of the four vowel options (o u ɨ ʌ).

Perception

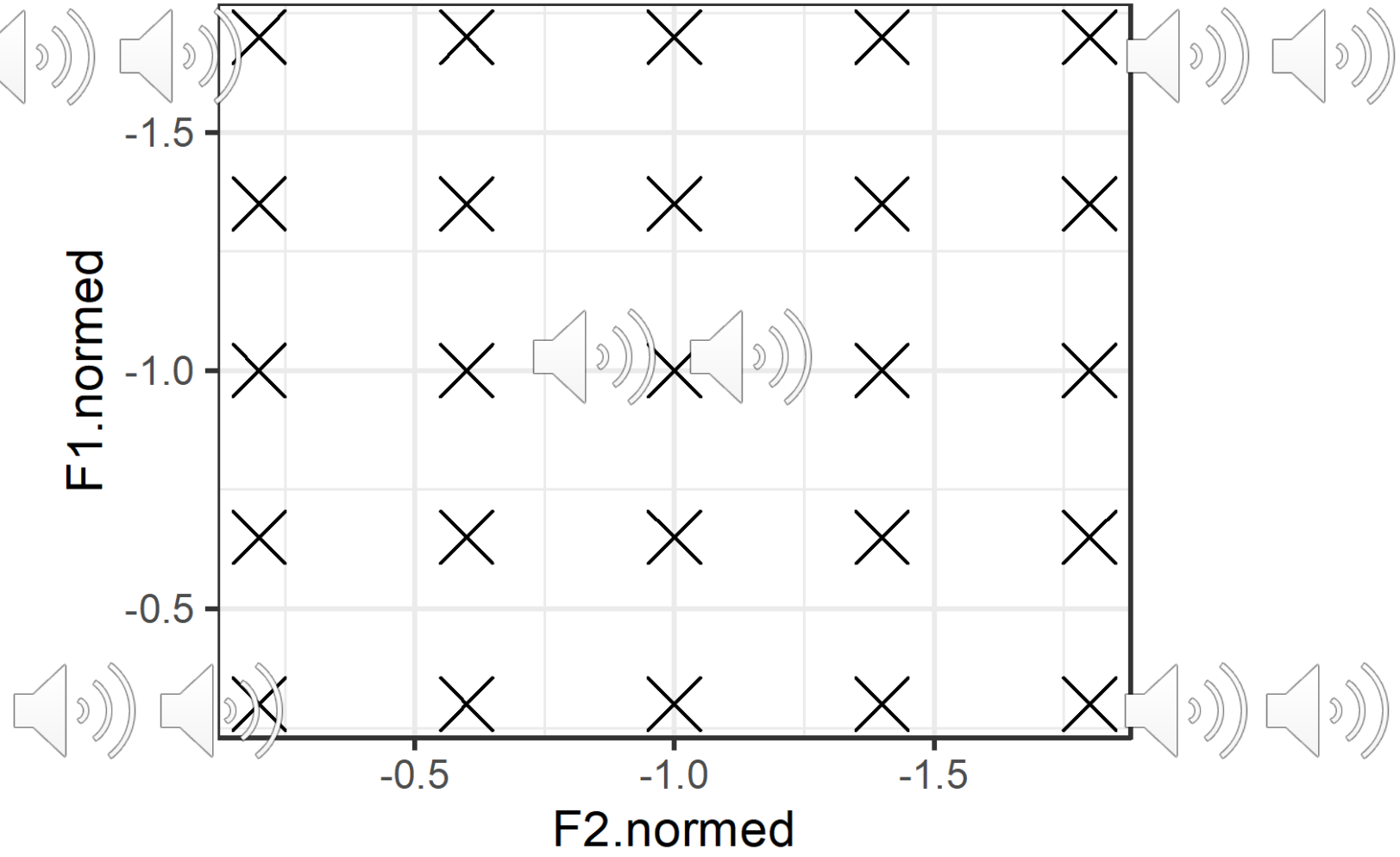
- Target vowel manipulation
  - A token of /o-ta/ was chosen for each speaker and the formants were manipulated to create acoustic space covering the four back non-low vowels (ɨ, u, ʌ, o).
  - The parameters were determined based on the stimuli speakers' production range in the Lobanov-normalized F1 * F2 space.
  - F1= [-0.3,-0.65,-1.0,-1.35,-1.7]
  - F2= [-0.2,-0.6,-1.0,-1.4,-1.8]



Perception

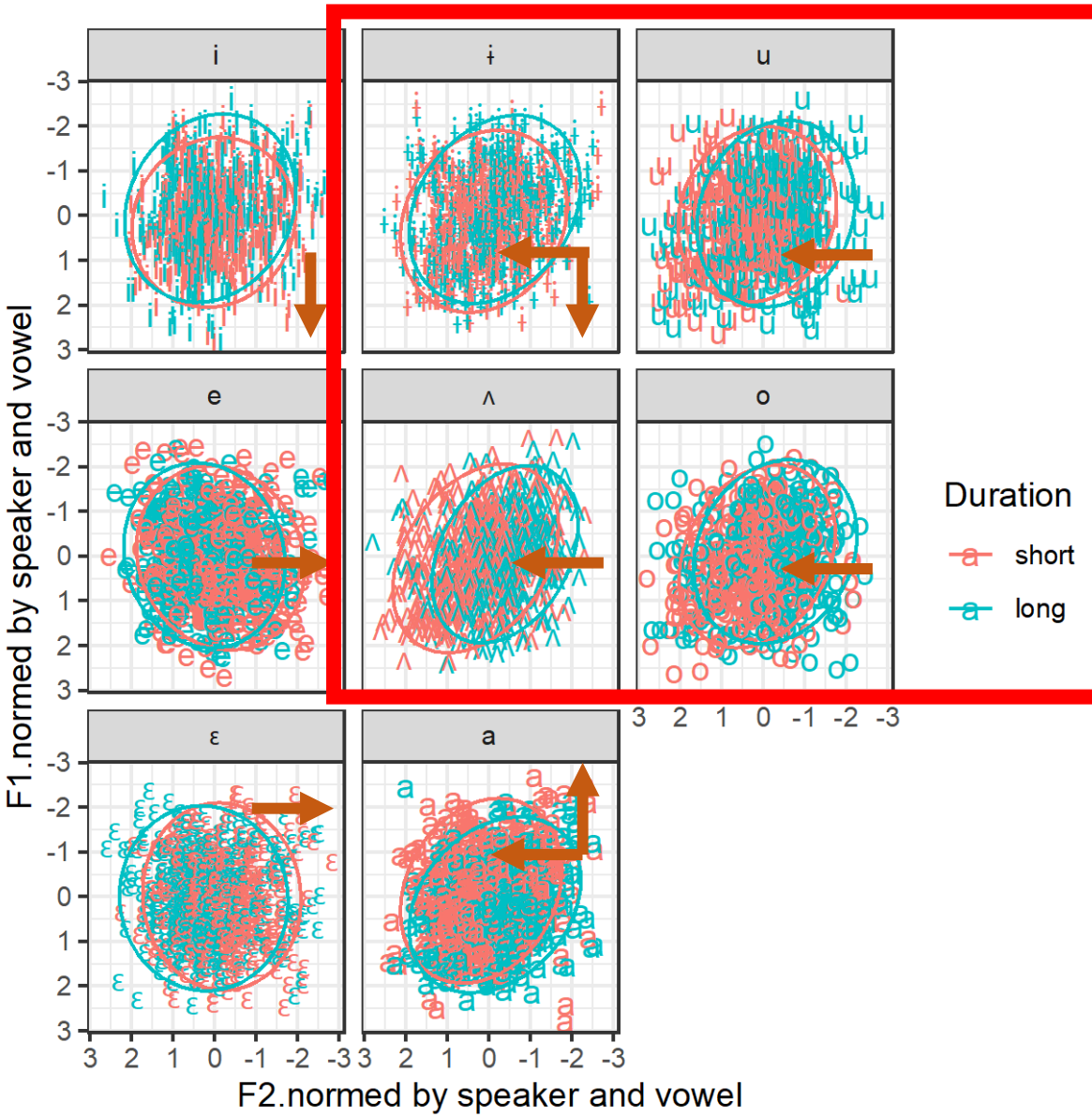
- Carrier sentence manipulation
  - A single carrier sentence token was chosen for each speaker.
  - The carrier sentence was manipulated to create fast (80% of the normal duration) and slow (120%) speech rate versions, exactly matching the corresponding duration of the production prompts.
  - The target word was also manipulated to vary in speech rate matching the carrier frame's rate.
  - The carrier frame and the target word were spliced together.
- 5 F1.steps * 5 F2.steps * 2 speech rates * 4 speakers = 200 tokens
- Presented with randomization in one block

Stimuli (OM, older male)

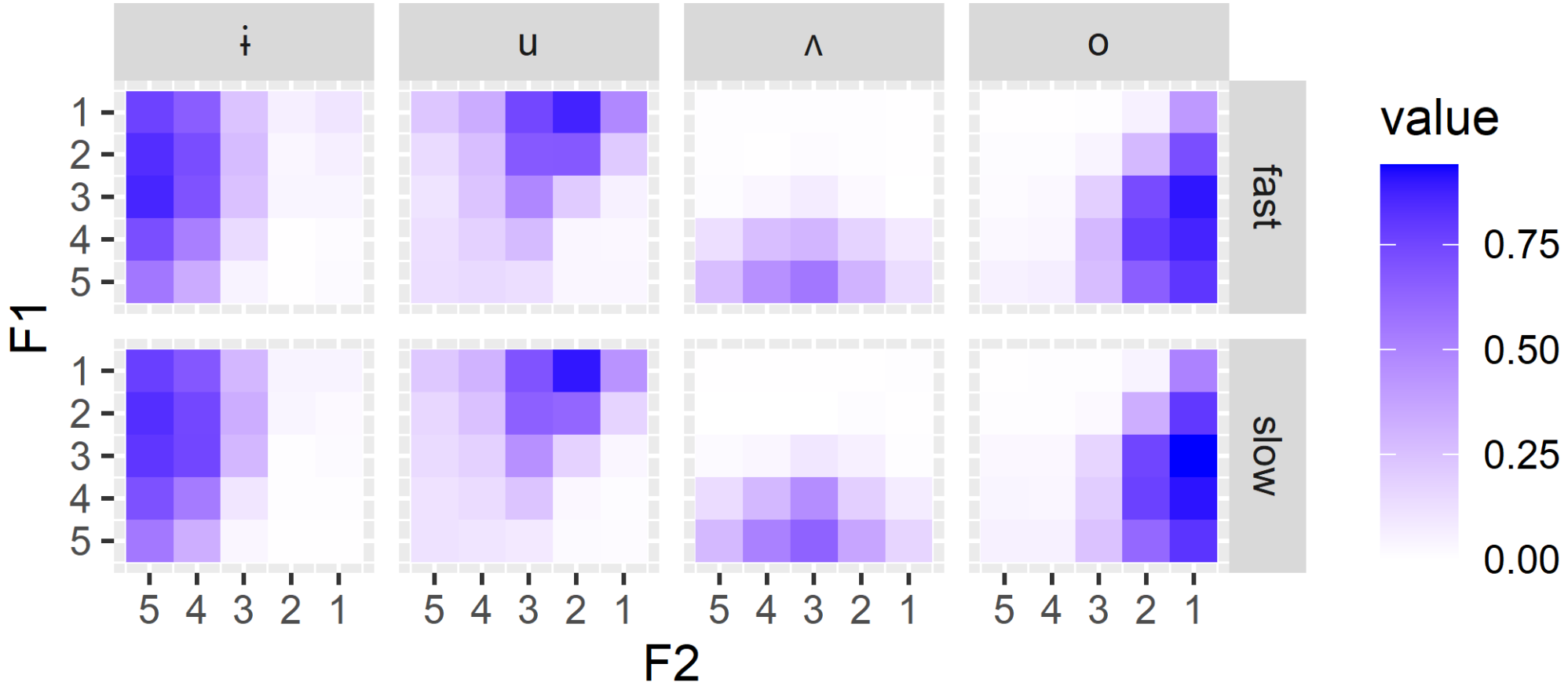


Predictions and Analysis

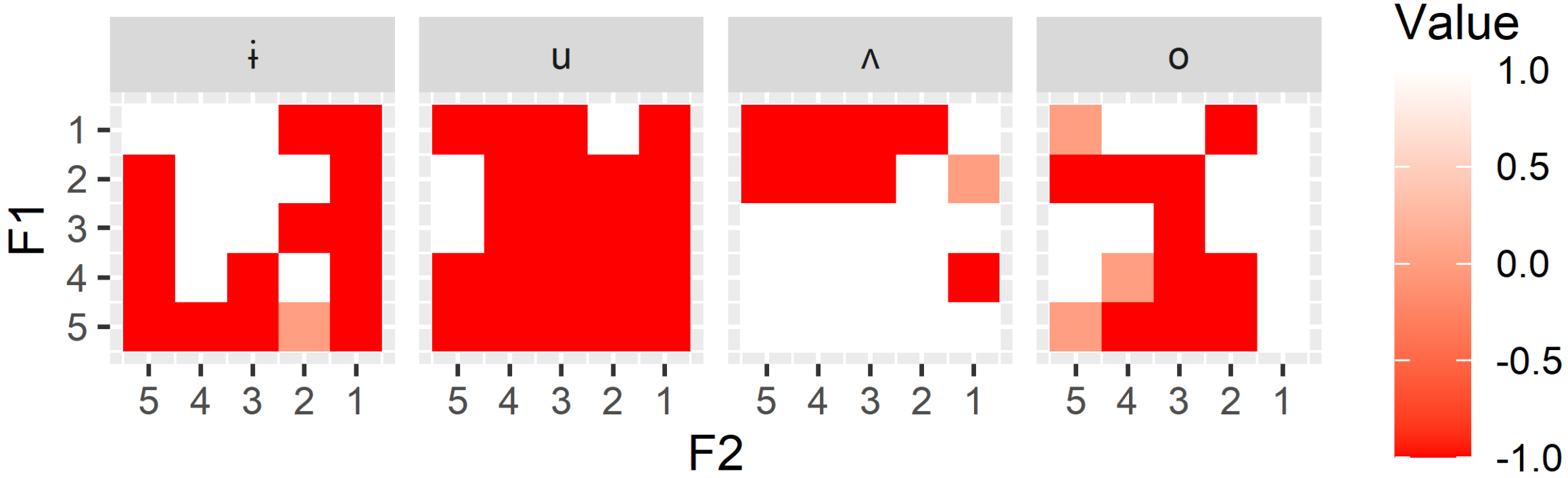
- If listeners compensate for speech rate faithfully mirroring the production pattern, potentially ambiguous stimuli will be more likely heard as more peripheral vowels in fast compared to slow speech conditions.
- To quantify the degree of shift in perception and compare it to the production shift, we calculated the mean of F1 and F2 (1~5 steps) of stimuli heard as each vowel, and the mean of Euclidean distance from the center of vowel space in fast vs. slow conditions.



Response by speech rate



Response difference (red = increase in fast)



Fast speech effects in perception

	ɨ	u	ʌ	o
Euclidean distance	n.s.	n.s.	n.s.	more central
F1	n.s.	lower	lower (YM only)	lower
F2	more back (OM only)	n.s.	n.s.	more front

- Shading – signficant effects found in the production
- Green – rate effect in the expected direction
- Red – rate effect in the opposite direction

Summary and conclusion

- We found listeners perform perceptual compensation of speech rate variation in vowel quality perception in the expected direction for some vowels but not for others.
- Future analysis will examine the interspeaker variation at the group (age and gender) and the individual level to probe where the perceptual compensation “succeeds” and where it “fails”.

Acknowledgements

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