The role of segment and pitch accent in Japanese spoken word recognition

Hironori Katsuda[1] & Jeremy Steffman[2]

[1]UCLA [2]Northwestern University

LSA 2021

The UCLA Phonetics Lab

Background

- Spoken word recognition is generally assumed to involve two phases:
  1. **lexical activation** (i.e., evoke multiple candidates that are consistent with the incoming speech)
  2. **lexical competition** (i.e., those candidates compete for recognition)

- It is well-established that both segmental and suprasegmental properties are important for **lexical activation**

- But the role of suprasegmental properties **differs across languages** heavily depending on how informative it is to identify words in the language

Background: the role of stress differs across languages

- **English:**
  - Stress usually correlates with vowel quality (e.g., OBject vs. obJECT)
  - Minimal stress pairs (e.g., FORbear vs. foreBEAR) are rare
  - Cutler's (1986) cross-modal priming study:

    E.g., FORbear primes both FORbear and foreBEAR

  → Stress is not used to restrict word activation

- **Spanish:**
  - Stress does not correlate with vowel quality
  - Soto-faraco et al.'s (2001) Experiment 1 (cross-modal priming):

  stress mismatch (segmental match) → an inhibitory effect

  e.g., prinCI- does not prime PRINcipe

  → Stress is used to restrict word activation

Background: relative contributions of segmental and suprasegmental properties

- **Spanish:**
  - Soto-faraco et al.'s (2001) Experiment 2 (cross-modal priming):
  - Segmental mismatch (stress match) → an inhibitory effect

  e.g., abun- does not prime abanDOno
  - Comparable inhibitory effects in both Exp1 and Exp2

  → Stress and segment are equally used to restrict word activation

- **Mandarin Chinese:**
  - Lexical tone: four distinct tones
  - Sereno & Lee's (2015), auditory priming study:
    - Segmental match (tonal mismatch) → a facilitatory effect (e.g., ru3 primes ru4)
    - Tonal match (segmental mismatch) → an inhibitory effect (e.g., sha4 does not prime ru4)

  → Segment plays more important role in restricting word activation

Background: the role of pitch accent

- **Japanese:**
  - Lexical pitch accent: three accent patterns are possible for disyllables (e.g., háshi ‘chopsticks’, hashí ‘bridge’, hashi ‘edge’)
  - Cutler & Otake’s (1999) auditory priming task:
    - Segmental match (pitch accent mismatch) → No facilitatory effect (e.g., háshi does not prime hashí)

    → **Pitch accent is used to restrict word activation**

Background

- **Some potential issues (Cutler & Otake 1999):**

1. No distinction between finally-accented words and unaccented words
   - They are prosodically neutralized in isolation (both are pronounced with a LH pitch pattern) (Poser 1984, Sugiyama 2006)
   - Both types were treated as the LH words (as opposed to the HL words)
   - Assuming that both types are represented equally
2. No accent-match (segment-mismatch) condition
   - e.g., áme 'rain' → tóshi 'city'

   →Cannot compare the role of segment and pitch accent directly
3. Word frequencies were not matched
   - the role of accent in word activation is modulated by word familiarity (Sekiguchi 2006)

Present study

- An attempt to replicate and extend Cutler & Otake's (1999) study in a more comprehensive experimental design

- Directly compare the role of segment and pitch accent in word recognition

Experimental design

- Auditory lexical decision task
  - Disyllables

- Four types of primes:
  - Identity: káme ‘turtle’ → káme
  - Segment(-match): kamé ‘pot’ → káme
  - **Accent(-match):** ríka ‘science’ → káme (no shared segments between prime and target)
  - Control: eŕi ‘collar’ → káme (no shared segments between prime and target)

Experimental design

- Three accent patterns:
  - Initial accent**:** káme
  - **Final accent:** erí
  - **Unaccented:** chiri ‘dust’

- Frame sentence: X-da ‘(It) is X.’

Experimental design

Materials:

- 48 word targets; 48 non-word targets (96 trials)
- 4 prime types for each target word (**identity, segment, accent, control**)
  - 4 counterbalanced lists by prime (identical target words)
  - Prime types frequency-matched (Maekawa 2003)

Stimuli:

- Recorded by Tokyo Japanese speaker
- Frame (X-da) appropriate for each accent type (three in total) were spliced after each word
  - Invariant frame within accent type

Experimental design

Procedure: 8 practice trials → 96 test trials

Trial
Prime → (ISI: 250 ms) → Target → Lexical decision

- 42 native Tokyo Japanese speaking participants
- In-person experiment (in the UCLA phonetics lab) (N = 19)
  - A button box response
- Online experiment (N = 23)
  - Key press ("F": no, "J": yes); precise time-syncing using Labvanced platform

Experimental design

Two Analyses: Bayesian mixed-effects regression

1. Reaction time: Linear regression for log-transformed RT ~ prime

```
rt ~ prime * setting ~ (1 + prime | participant)
```

2. Accuracy : Binomial logit regression (correct/incorrect) ~ prime

```
response ~ prime * setting ~ (1 + prime | participant)
```


[1] `brms`; Bürkner 2018 & `emmeans`; Lenth et al. 2018

Predictions

1. Role of accent:

Identity = Segment

→ Accent is **not used** to restrict word activation (like English stress, Cutler 1986)

Identity < Segment

→ Accent is **used** to restrict word activation (like Cutler and Otake 1999)

2. Relative contributions of segment and accent:

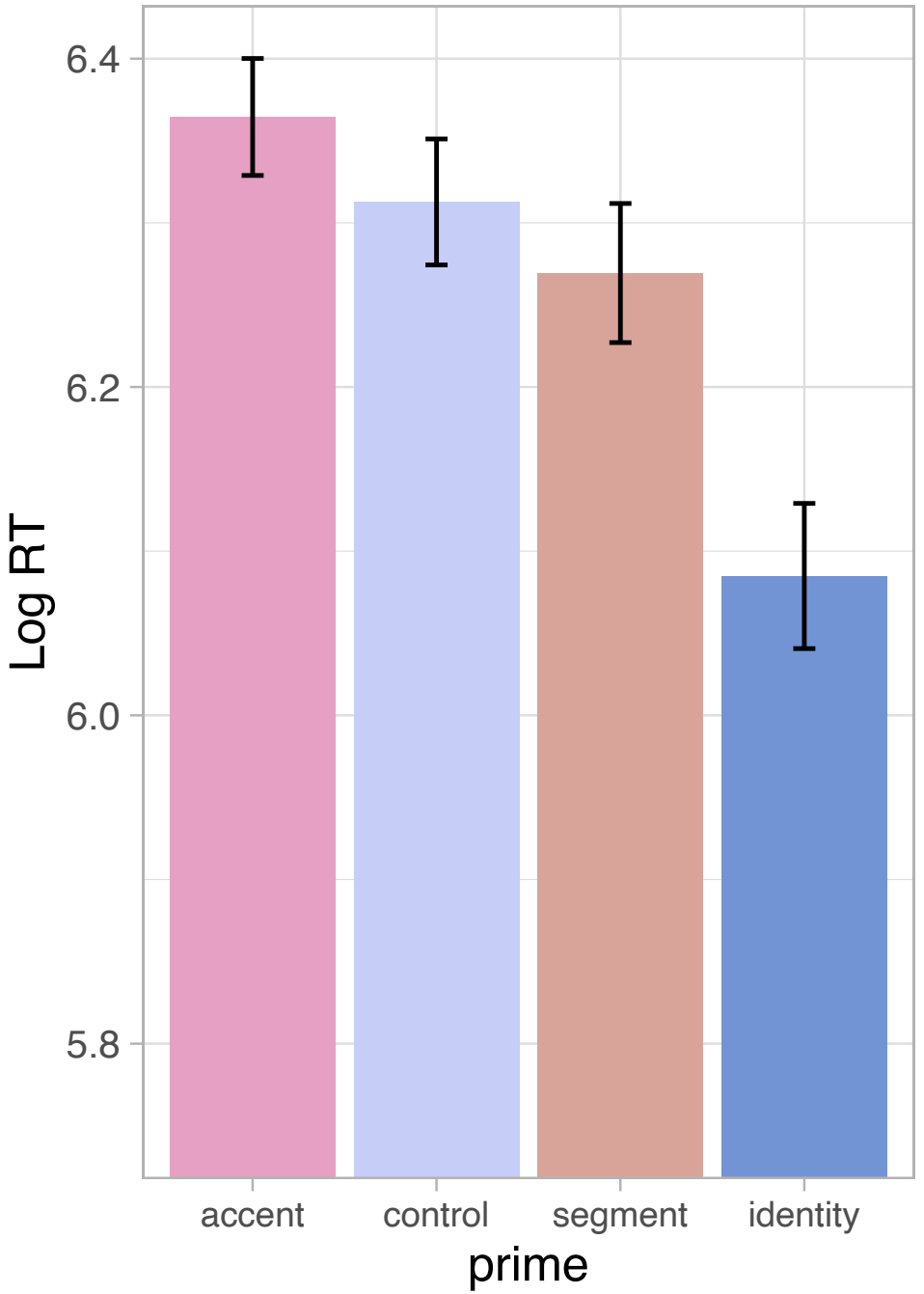
Segment = Accent

→ Their roles are **comparable** (like Spanish, Soto-Faraco et al. 2001)

Segment ≠ Accent

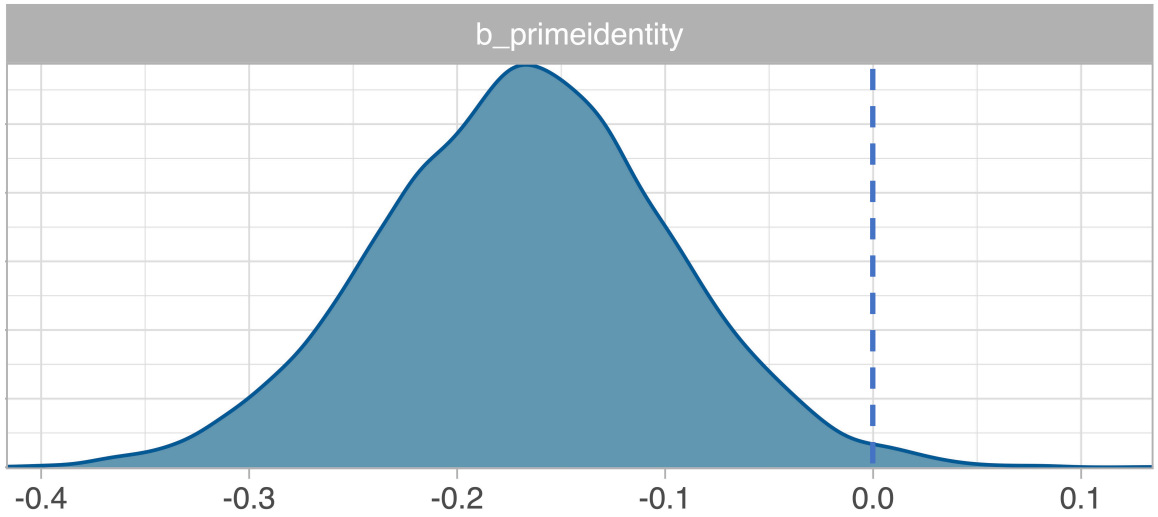
→ They have **different** roles (like Mandarin, Sereno and Lee's 2015)

Results: reaction time



Results: reaction time

Segment > Identity β = -0.17 95%CI = [-0.31,-0.03]



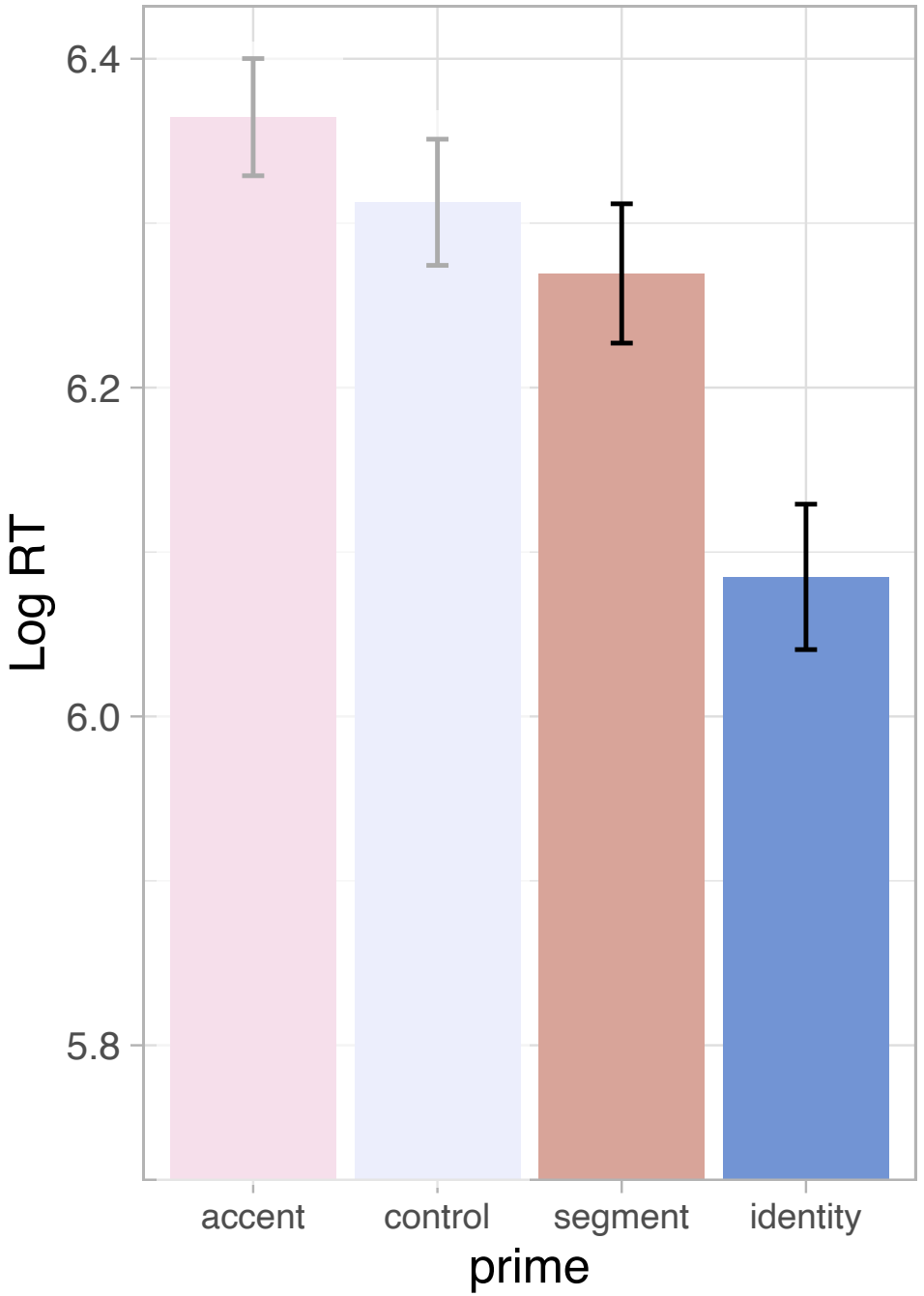
Role of accent:

~~Segment = Identity~~

~~→ Accent is not used to restrict word activation~~

Segment > Identity

→ Accent is used to restrict word activation



Results: reaction time

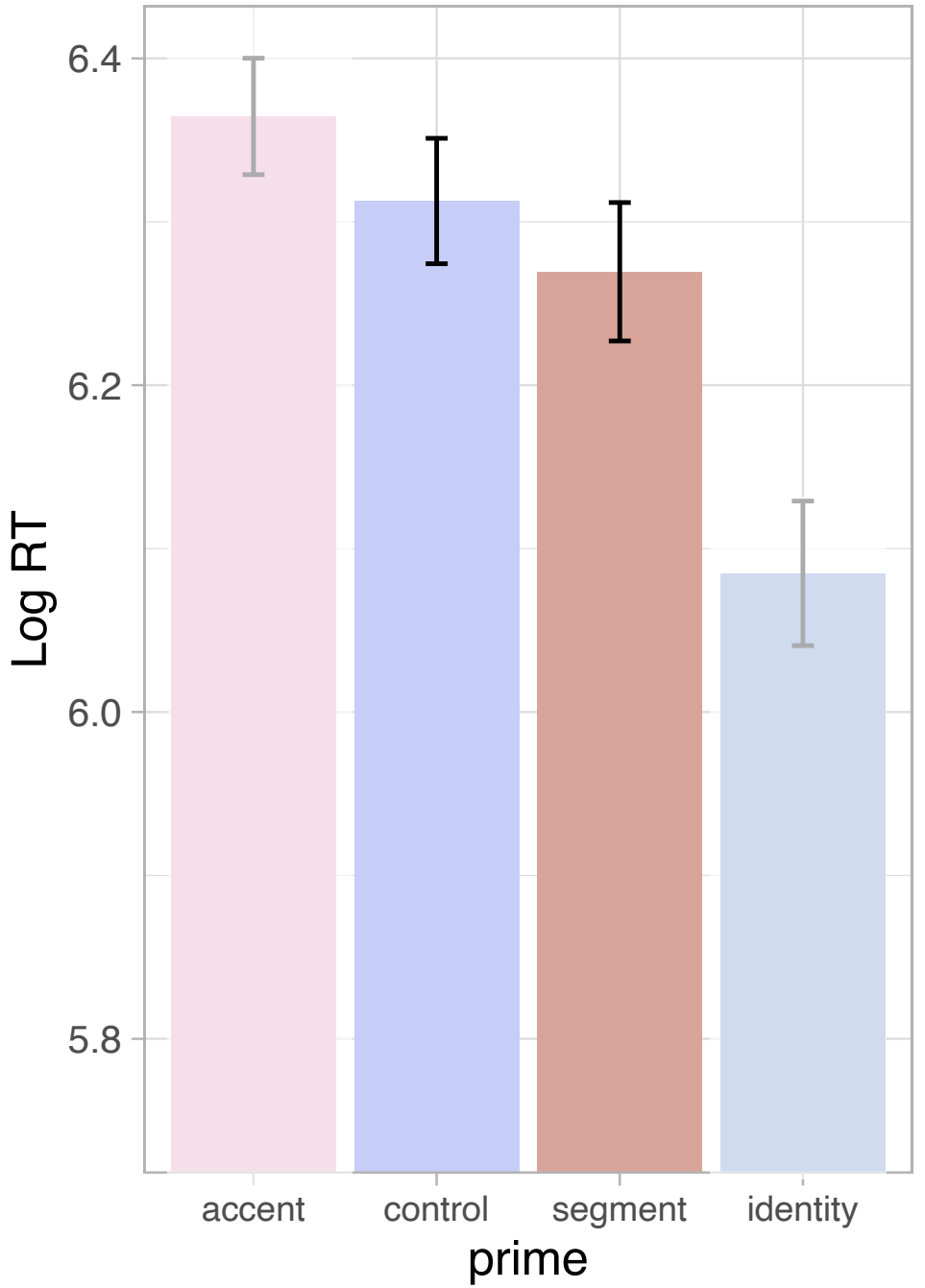
no reliable difference for Control vs. Segment

β = 0.05 95%CI = [-0.06,0.15]

Role of accent:

Segment = Control

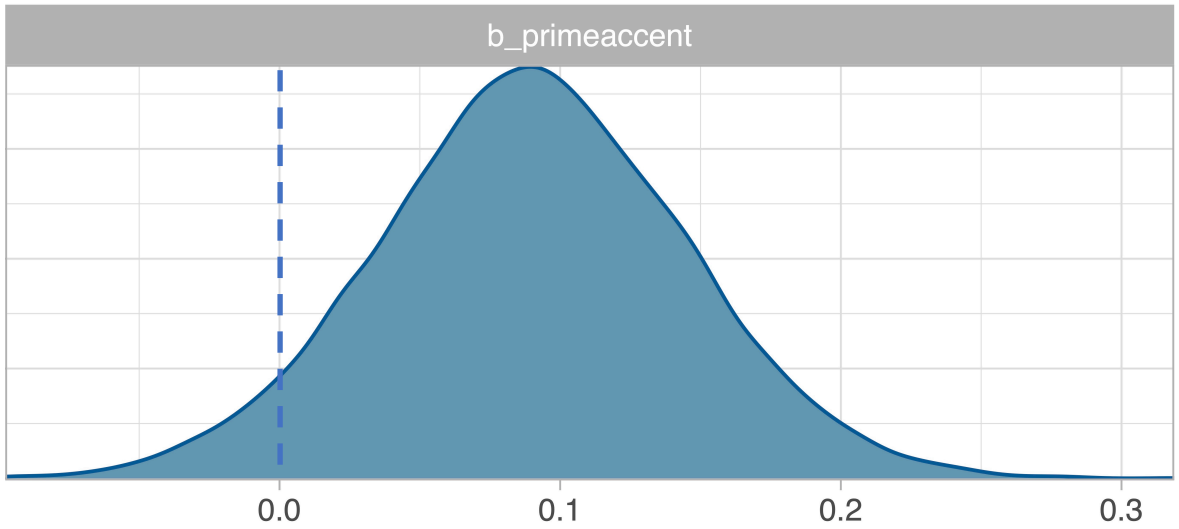
→ Accent is used to restrict word activation



Results: reaction time

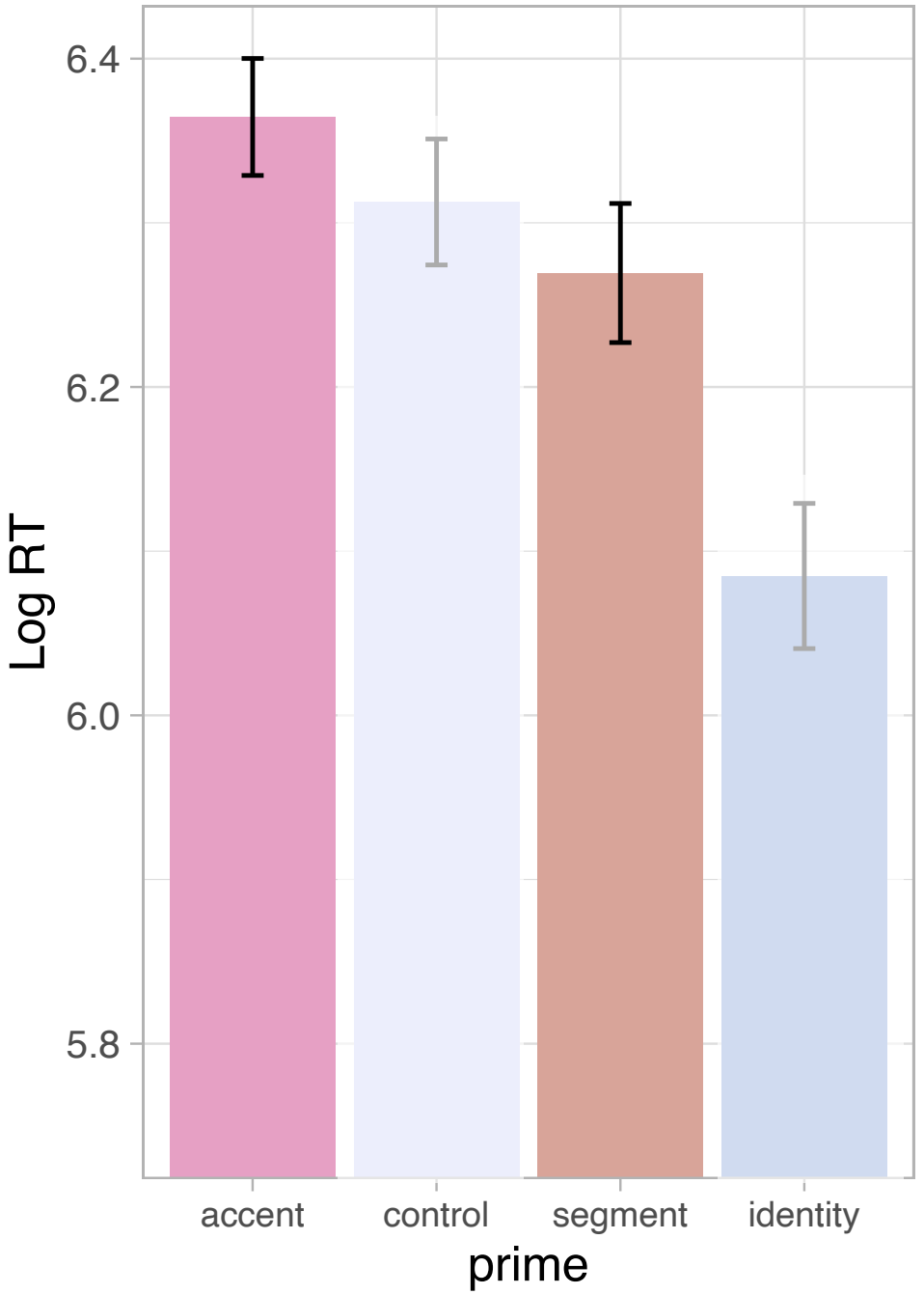
Moderate evidence for Accent > Segment

β = 0.09, 95%CI = [-0.01,0.20]

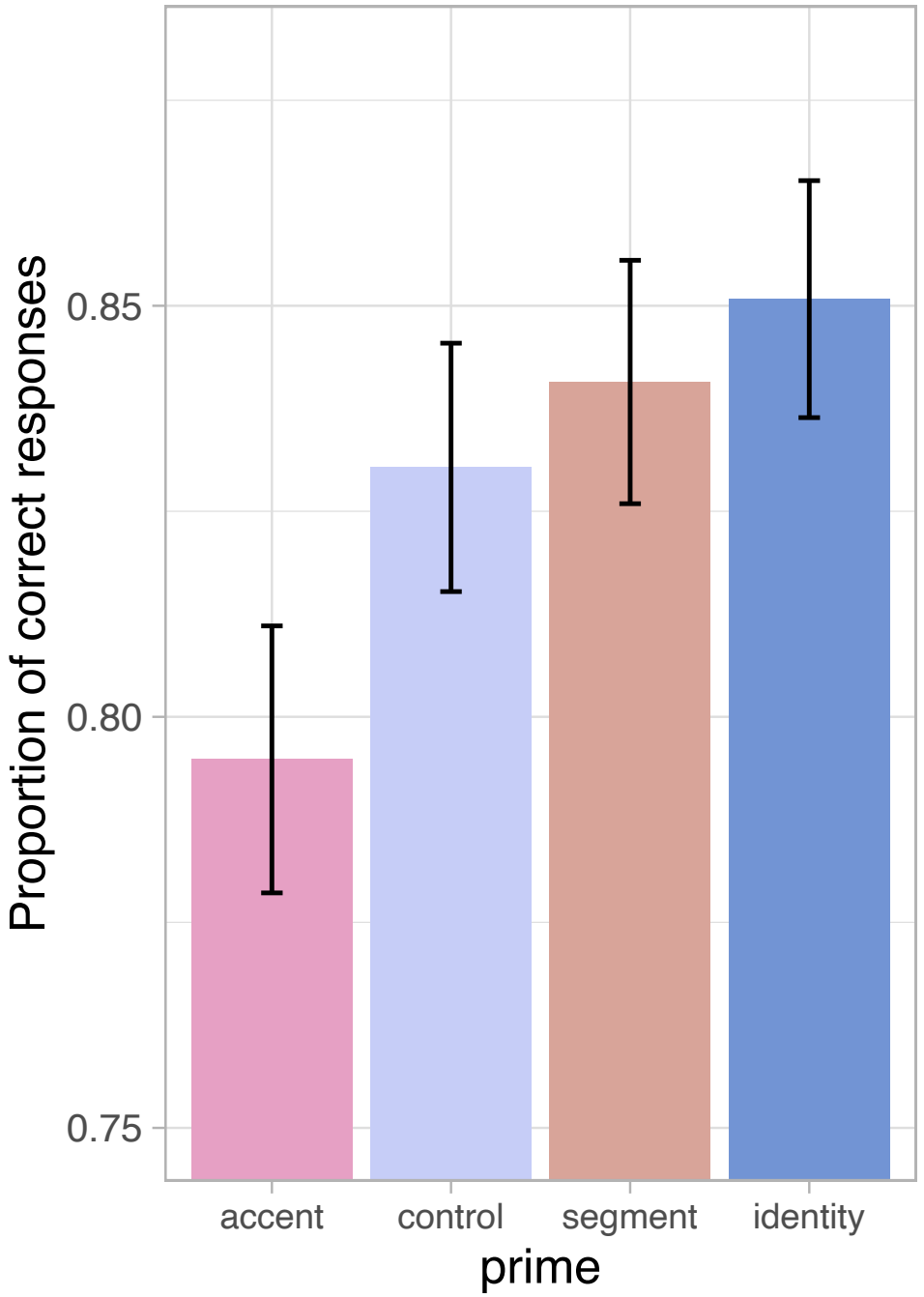


Relative contributions of segment and accent:

Segment ≠ Accent ?

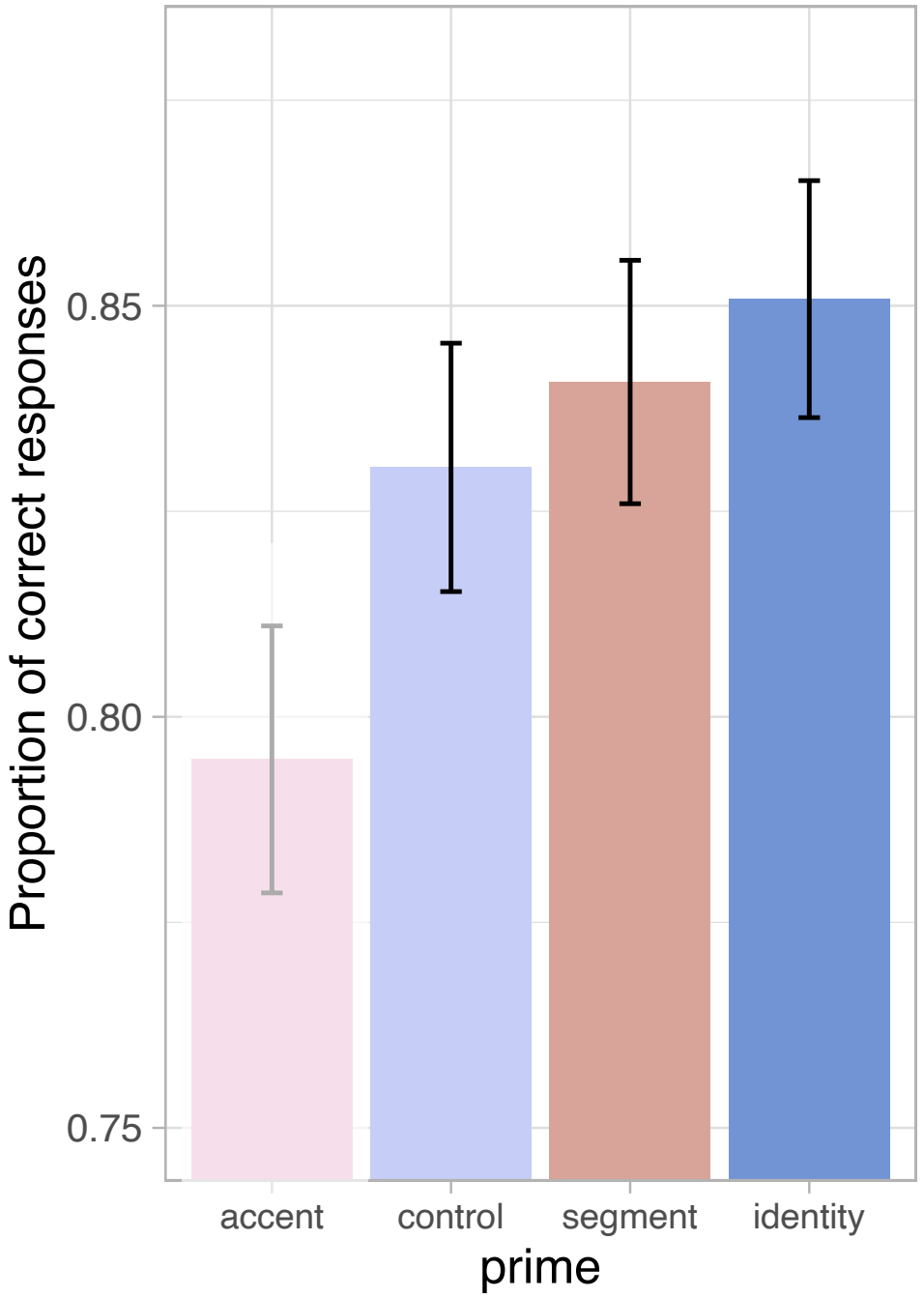


Results: accuracy



Results: accuracy

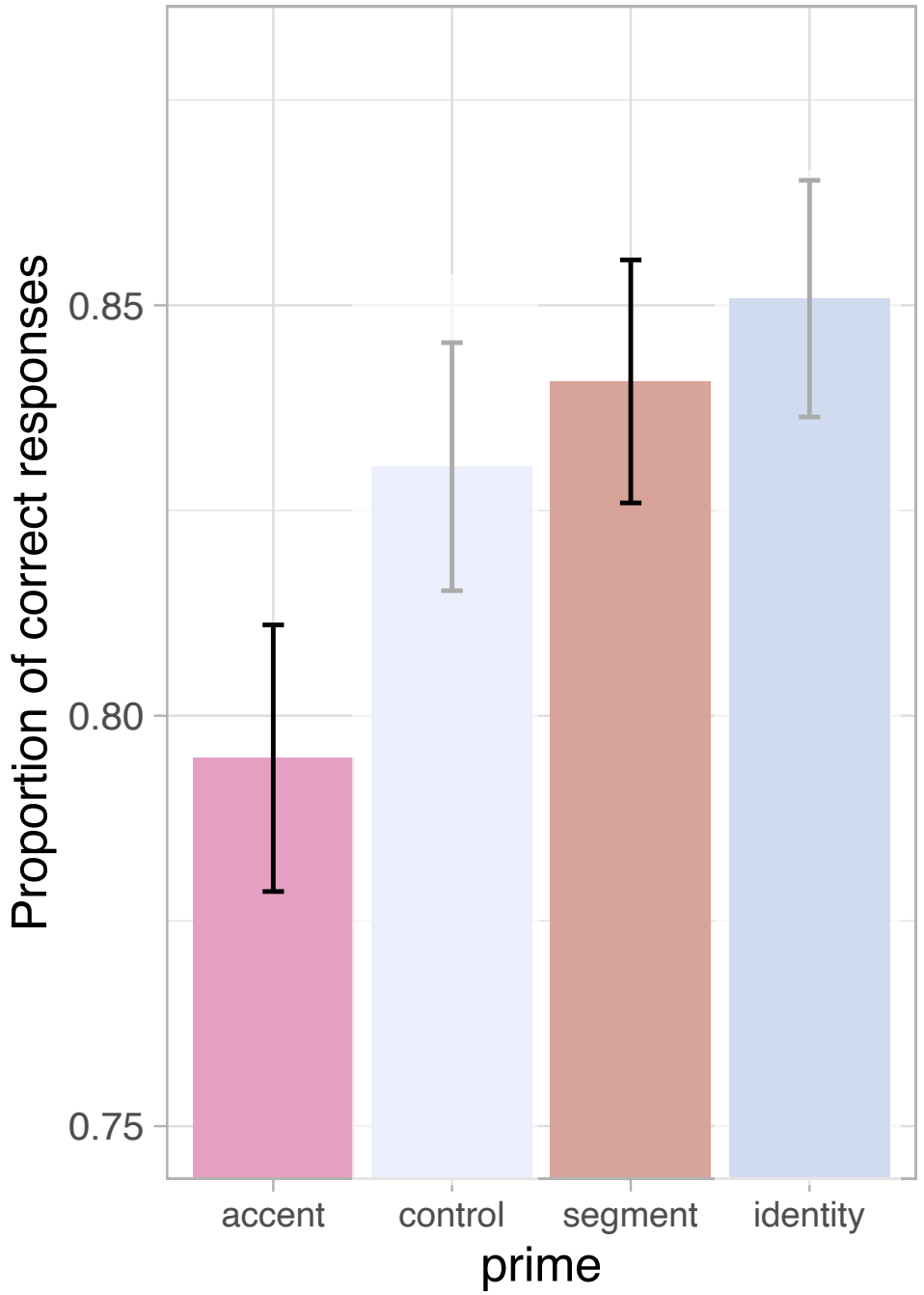
Segment vs Control β = 0.05 95%CI = [-0.33,0.43]
Control vs Identity β = -0.14 95%CI = [-0.57,0.36]
Segment vs Identity β = -0.09 95%CI = [-0.50,0.30]



Results: accuracy

Accent < Segment

β = 0.32 95%CI = [0.01, 0.71]

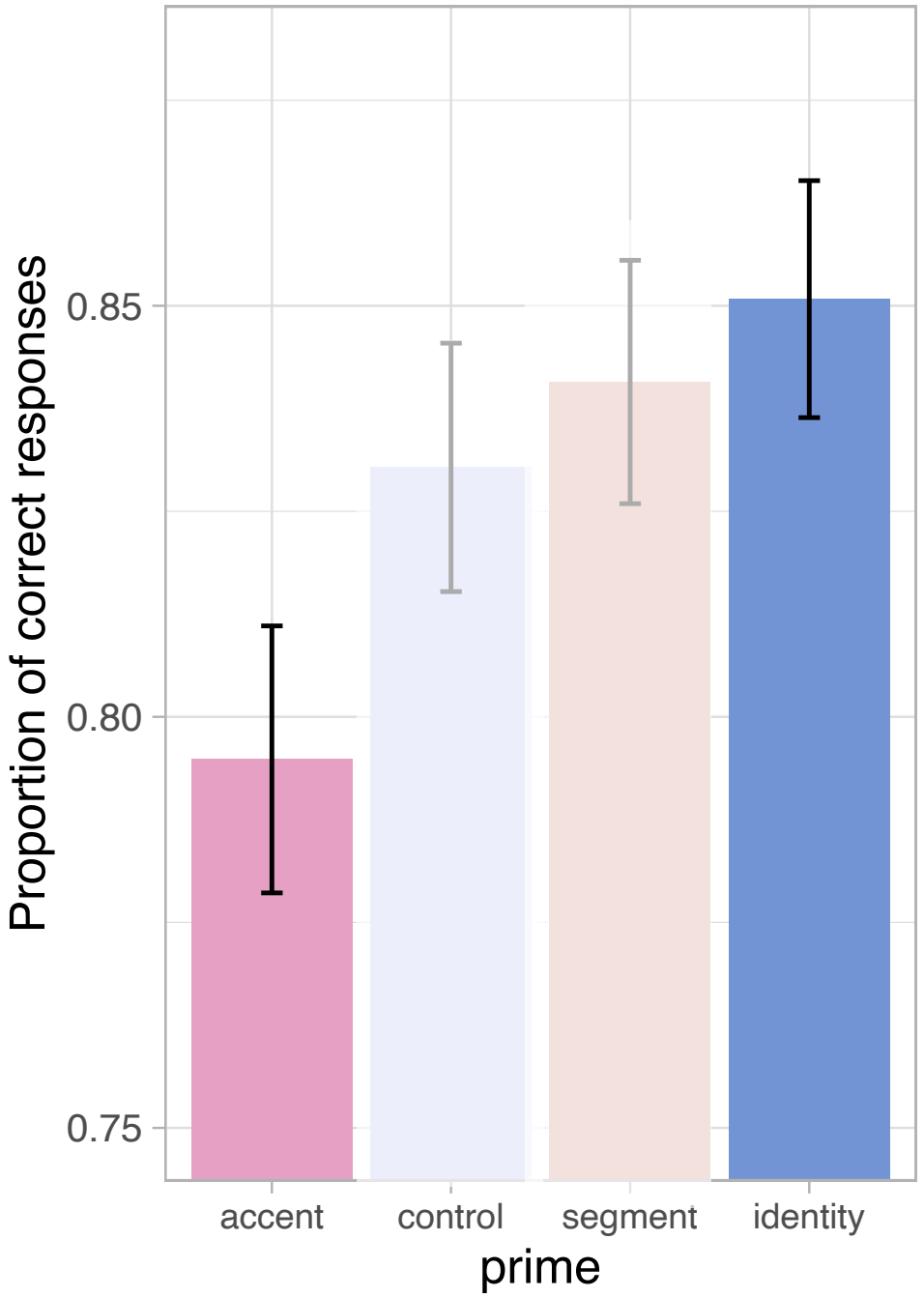


Results: accuracy

Accent < Segment
β = 0.32 95%CI = [0.01, 0.71]

Accent < Identity (unlike Segment)
b = 0.44 95%CI = [0.04, 0.85]

Relative contributions of segment and accent:
~~Accent = Segment~~
~~→ Their roles are comparable~~
Accent ≠ Segment
→ They have different roles



Discussion

Role of accent:

- Identity < Segment (i.e., facilitation in Identity, no facilitation in Segment)
  - E.g., kamé does not prime káme

→ Accent is used to restrict word activation (replicate Cuter & Otake 1999)

Relative contributions of segment and accent:

- Segment < Accent (moderately strong evidence), but Segment = Control
- The difference is due to the **inhibition** in the Accent condition (consistent with the accuracy data too)

Discussion

Why inhibition in the Accent condition?

- A group activation (Poss et al. 2008 in Mandarin tone): tone evokes lexical items with the same tone
- Pitch accent likewise evokes lexical items with the same pitch accent profile
- Competition among those lexical items results in the inhibitory effect observed in the Accent condition

Further directions

- Conduct experiments with varying degrees and locations of segmental overlap (e.g., Norris et al. 2002, Dufour & Peereman 2003)
  - [**k**ári] ‘hunt’ → [**h**ári] ‘needle’ (1st onset mismatch)
  - [**só**ri] ‘sled’ → [**há**ri] ‘needle (1st syllable mismatch)
  - [hár**u**] ‘spring’ → [hár**i**] ‘needle’ (final vowel mismatch)
  - [há**da**] ‘skin’ → [há**ri**] (final syllable mismatch)

- Investigate the role of phrasal prosody
  - E.g., the effect of final accent neutralization on processing
  - Does surface [hana] (= /hana/ ‘nose’ or /haná/ ‘flower’) prime only /hana/ ‘nose’ or both /hana/ ‘nose’ and /haná/ ‘flower’?
  - How is word-final accent represented in the lexicon?

Take aways

Replication

- Accent is used to restrict word activation (Cuter & Otake 1999)

Novel findings

- Accent and segment have **different** roles in word recognition

→ Contribute to a better understanding of the role of suprasegmental properties in spoken word recognition and build a foundation for follow-up studies

Acknowledgement

- Many thanks to Sun-Ah Jun, Claire Moore-Cantwell, and members of the UCLA Phonetics lab for valuable feedback on this project.

References

Bürkner, P. (2018) Advanced Bayesian Multilevel Modeling with the R package brms. *The R Journal,* **10**(1), 395-411.

Cutler, A. (1986). Forbear is a homophone: Lexical prosody does not constrain lexical access. *Language and Speech*, **29**, 201-220.

Cutler, A., & Chen, H.-C. (1997). Lexical tone in Cantonese spoken-word processing. *Perception and Psychophysics*, **59**, 165-179.

Cutler, A., & Otake, T. (1999). Pitch accent in spoken-word recognition in Japanese. *Journal of the Acoustic Society of America*, **105**, 1877-1888.

Cutler, A., & van Donselaar, W. (2001). Voornaam is not a homephone: Lexical prosody and lexical access in Dutch. *Language and Speech*, **44**, 171-195.

Dufour, S., & Peereman, R. (2003). Inhibitory priming effects in auditory word recognition: When the target's competitors conflict with the prime word. *Cognition*, **88**, B33-B44.

Lee, C.-Y. (2007). Does horse activate mother? Processing lexical tone in form priming. *Language and Speech*, **50**, 101-123.

Lenth, R., Singmann, H., Love, J., Buerkner, P., & Herve, M. (2018). emmeans: Estimated Marginal Means, aka Least-Squares Means. Retrieved from https://CRAN.R-project.org/package=emmeans

Maekawa, K. (2003). Corpus of spontaneous Japanese: Its design and evaluation. In: Proceedings ISCA and IEEE workshop on spontaneous speech processing and recognition (pp. 7–12).

Norris, D., McQueen, J. M., & Cutler, A. (2002). Bias effects in facilitatory phonological priming. *Memory and Cognition*, **30**, 399-411.

Poser, W. (1984). *The Phonetics and Phonology of Tone and Intonation in Japanese.* Ph.D. dissertation, MIT.

Poss, N., Hung, T. H., & Will, U. (2008). The effects of tonal information on lexical activation in Mandarin. In M. K. M. Chan & H. Kang (Eds.), *Proceedings of the 20th North American Conference on Chinese Linguistics* (pp. 205-211). Columbus: Ohio State University.

Sekiguchi, T. (2006). Effects of lexical prosody and word familiarity on lexical access of spoken Japanese words. *Journal of Psycholinguistic Research, 35*, 369–384.

Sereno, J. A., & Lee, H. (2015). The contribution of segmental and tonal information in Mandarin spoken word processing. *Language and Speech*, **58**(2), 1-21.

Soto-Fraco, S., Sebastian-Galles, N., & Cutler, A. (2001). Segmental and suprasegmental mismatch in lexical access. *Journal of Memory and Language*, **45**, 412-432.

Suguyama, Y. (2006). Japanese pitch accent: Examination of final-accented and unaccented minimal pairs. *Toronto Working Papers in Linguistics 26*, pp. 73-88.

Appendix: target words and primes 1

Target	Identity	Segment	Accent	Control
chiri	chiri	chi'ri	same	ka'so
mushi	mushi	mu'shi	gake	ka'mo
shiro	shiro	shi'ro	kubi	ne'gi
kashi	kashi	ka'shi	himo	so'hu
kabu	kabu	ka'bu	mizo	chi'ji
aku	aku	a'ku	shiwa	I'ta
sake	sake	sa'ke	eda	yu'zu
shika	shika	shi'ka	ume	ka'gu
hana	hana	hana'	tsubo	imo'
take	take	take'	goma	ima'
nami	nami	nami'	kita	doku'
tsume	tsume	tsume'	gaka	koya'
kimi	kimi	kimi'	sara	wata'
mochi	mochi	mochi'	nazo	aza'
kiri	kiri	kiri'	taka	hama'
kaki	kaki	kaki'	shiba	iwa'

Appendix: target words and primes 2

Target	Identity	Segment	Accent	Control
syo'ki	syo'ki	syoki	na'ya	tana
ku'wa	ku'wa	kuwa	ta'mi	sode
ka'ku	ka'ku	kaku	so'ra	nuno
ha'ri	ha'ri	hari	ki'bo	keta
tsu'ya	tsu'ya	tsuya	ge'ki	kizu
mo'mo	mo'mo	momo	hu'ne	kabi
ta'ki	ta'ki	taki	gi'mu	yome
ka'go	ka'go	kago	yo'ka	huta
I'shi	I'shi	ishi'	ka'ge	tsuno'
tu'yu	tu'yu	tuyu'	sa'gi	haji'
ka'me	ka'me	kame'	ri'ka	eri'
I'ji	I'ji	iji'	ku'ro	kusa'
ka'sa	ka'sa	kasa'	ro'ji	siri'
su'mi	su'mi	sumi'	ho'ko	tera'
to'shi	to'shi	toshi'	ka'do	sewa'
ka'mi	ka'mi	kami'	chi'yu	doro'

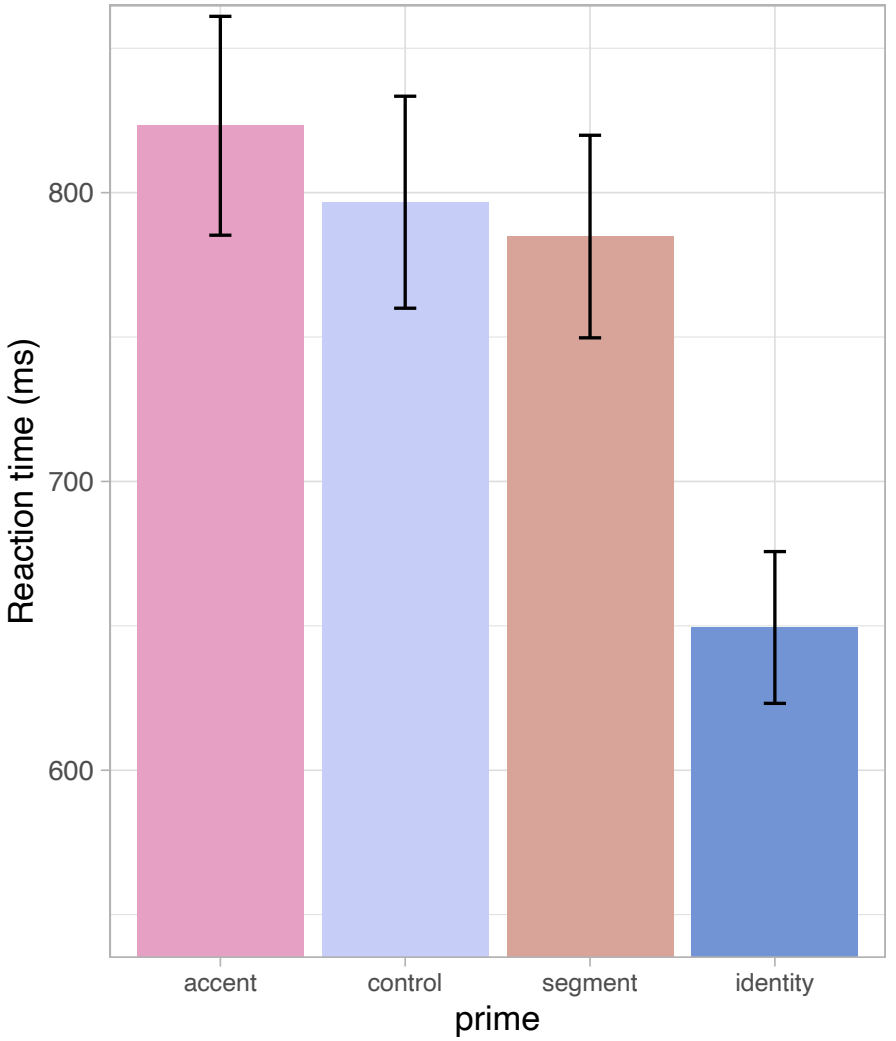
Appendix: target words and primes 3

Target	Identity	Segment	Accent	Control
hati'	hati'	hati	ike'	oka
mura'	mura'	mura	hiji'	hane
hata'	hata'	hata	yumi'	migi
hashi'	hashi'	hashi	toge'	muda
tori'	tori'	tori	ago'	hige
heta'	heta'	heta	dashi'	mane
kake'	kake'	kake	osu'	semi
mame'	mame'	mame	netsu'	kugi
umi'	umi'	u'mi	nawa'	si'ya
hibi'	hibi'	hi'bi	ura'	ke'sa
saku'	saku'	sa'ku	oni'	I'to
seki'	seki'	se'ki	numa'	ki'nu
kata'	kata'	ka'ta	niku'	tu'mi
awa'	awa'	a'wa	jimi'	ji'ko
tabi'	tabi'	ta'bi	kome'	o'ke
aka'	aka'	a'ka	seme'	o'no

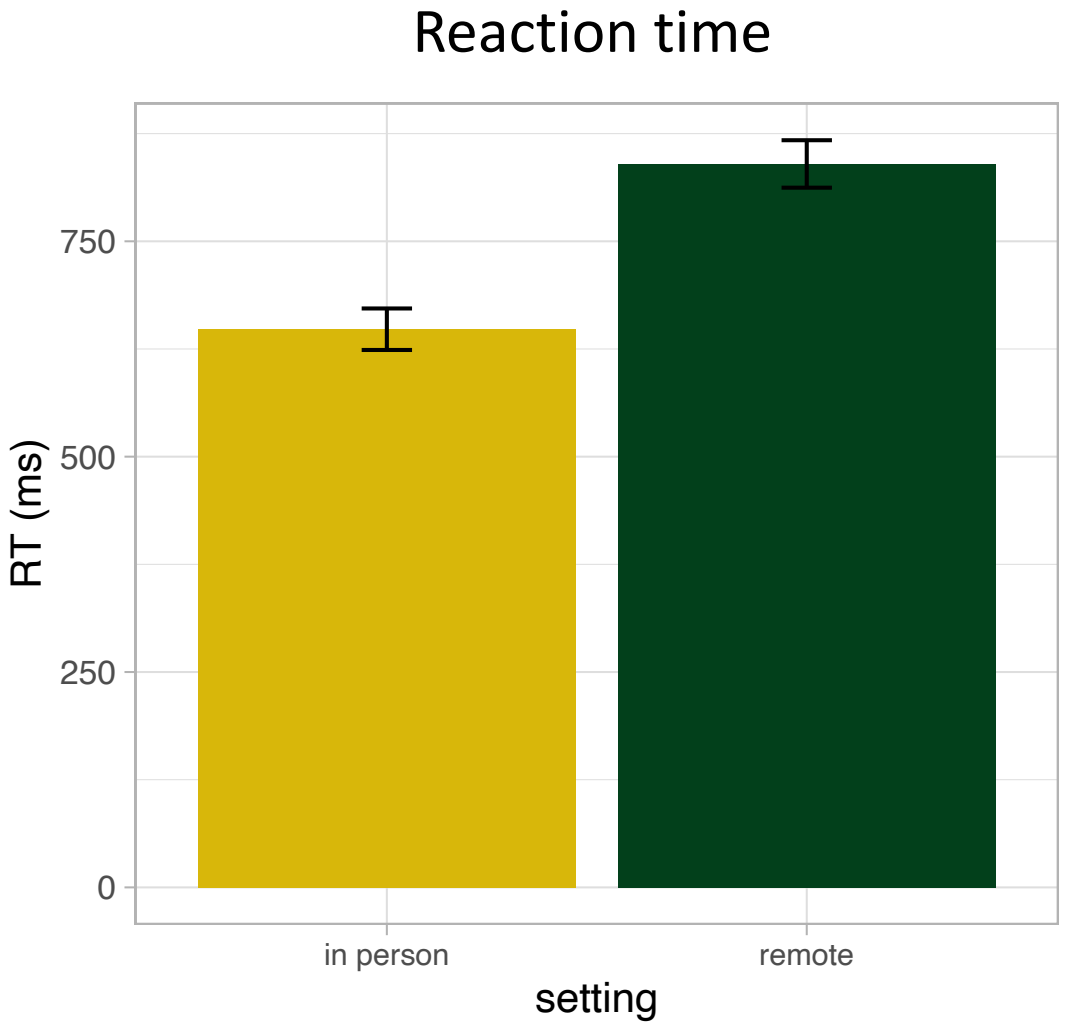
Appendix: target nonwords

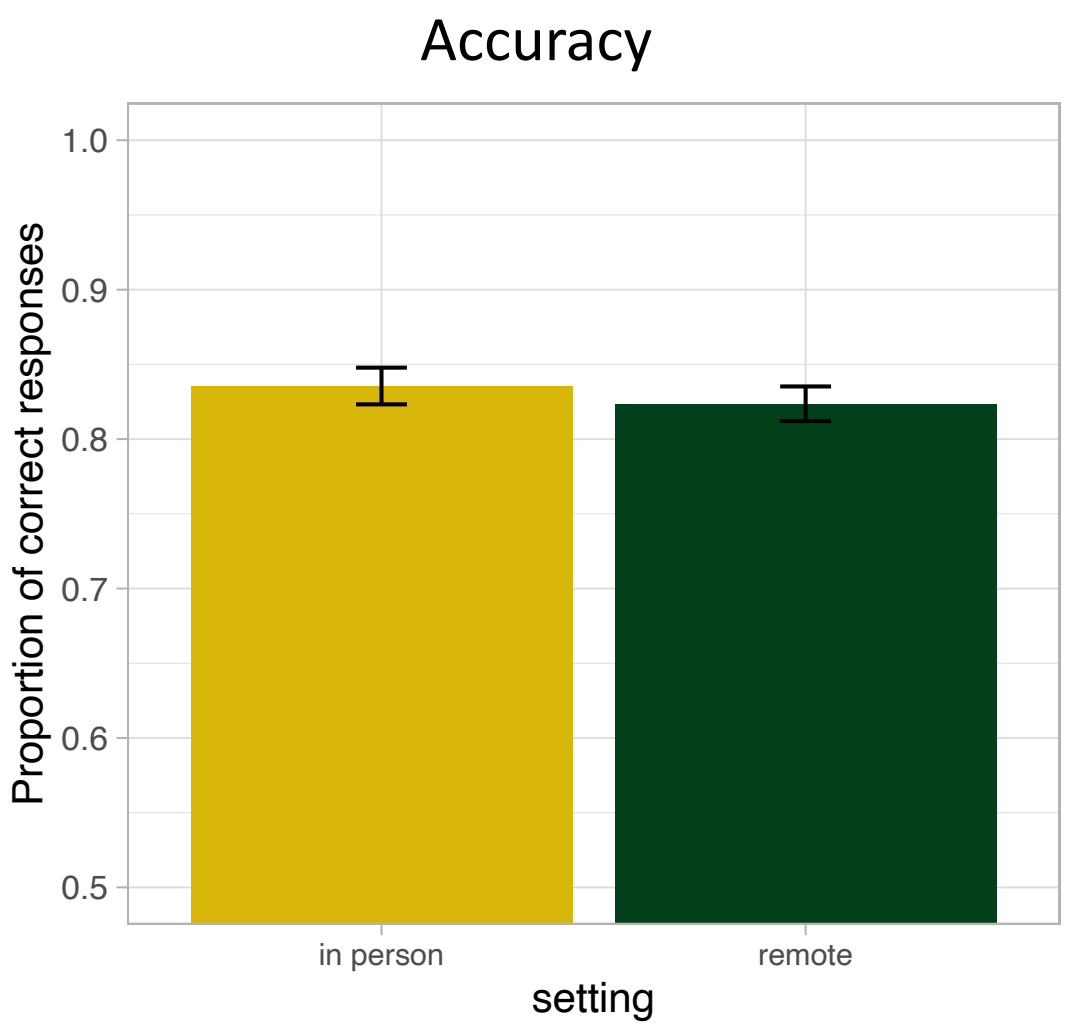
Unaccented	1st accent	2nd accent
mishi	do'me	mani'
moni	ne'ka	koru'
toma	sa'ba	nota'
basa	ta'mu	kazo'
tase	so'mo	tosa'
meku	u'de	uko'
kose	no'shi	sano'
sene	a'bo	yare'
sona	ri'ba	gozo'
taso	ko'nu	kiho'
mabo	ha'te	naso'
yosu	mi'shi	yako'
toni	ya'mi	Isa'
somi	ma'ge	iza'
nagu	o'ze	miha'
keto	go'ra	tanu'

Appendix: raw RT

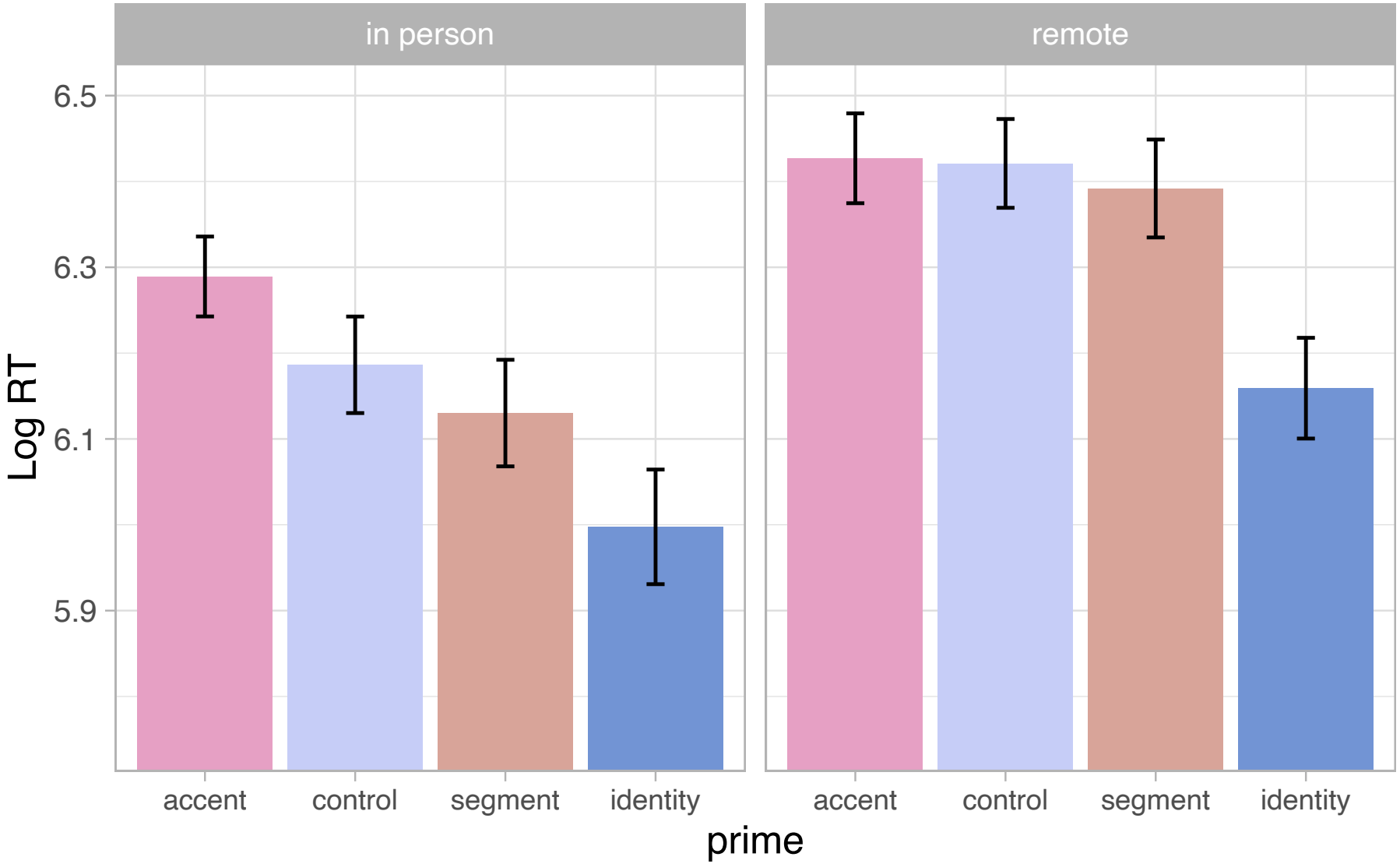


Appendix: effects of setting

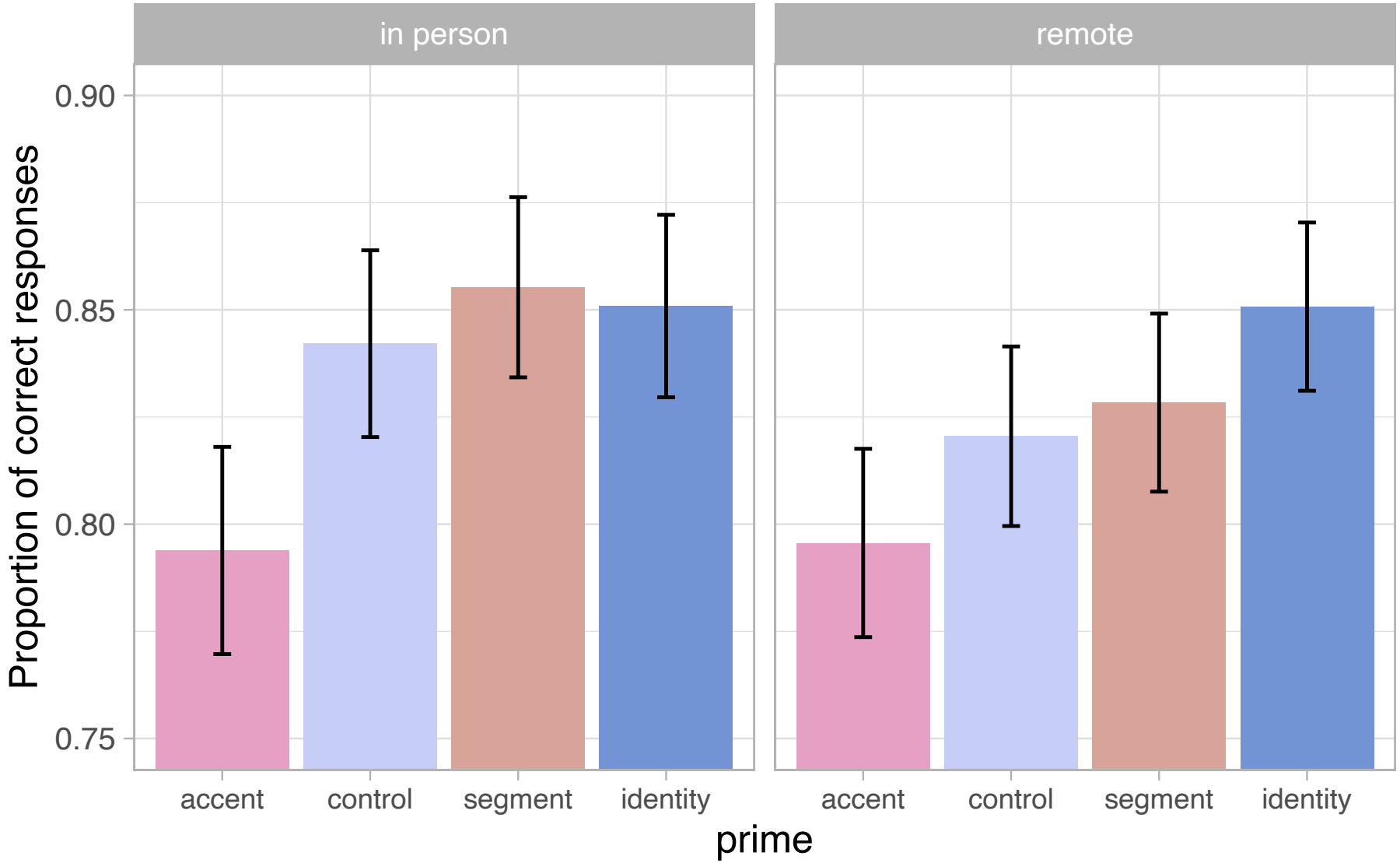




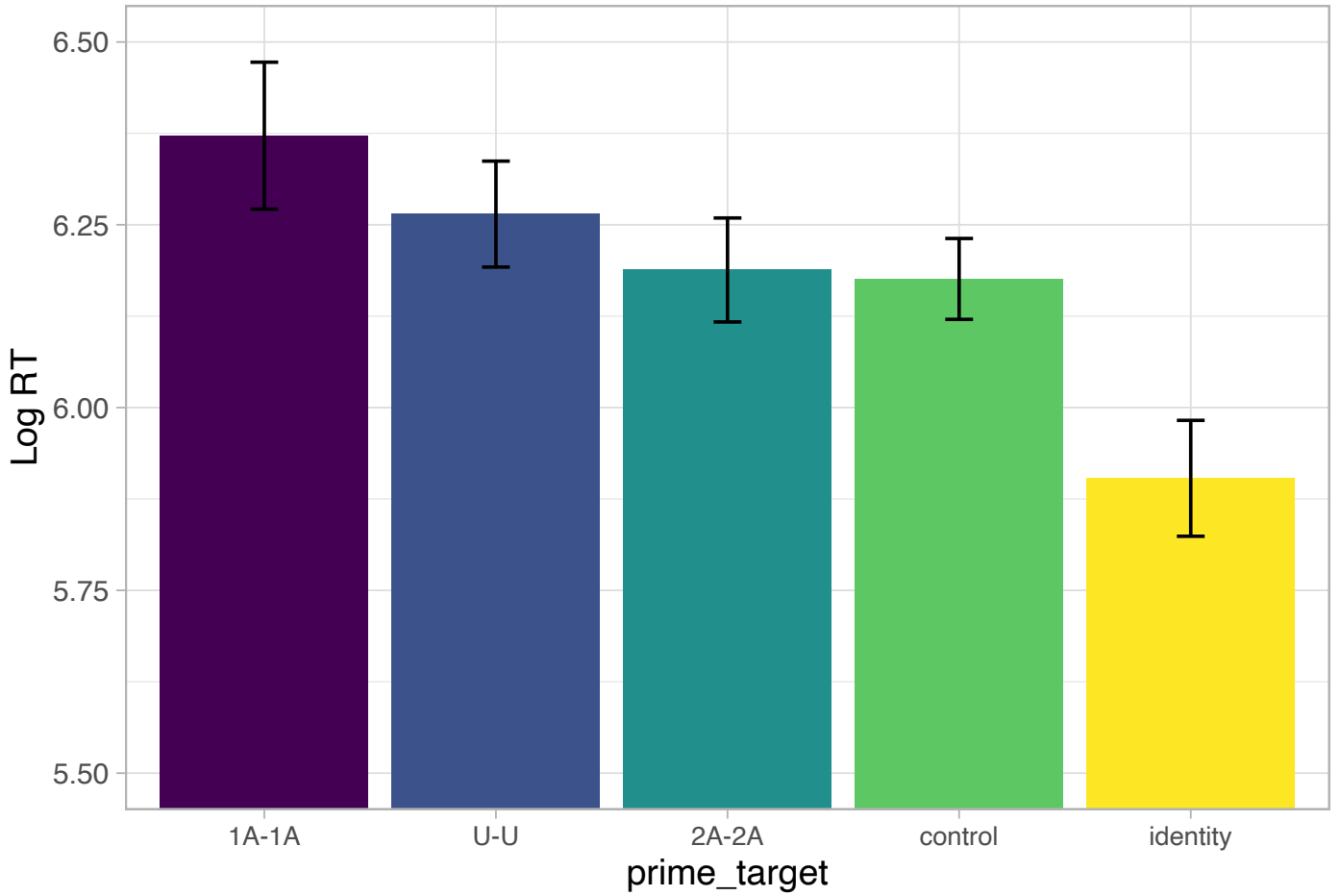
Appendix: RT split by setting



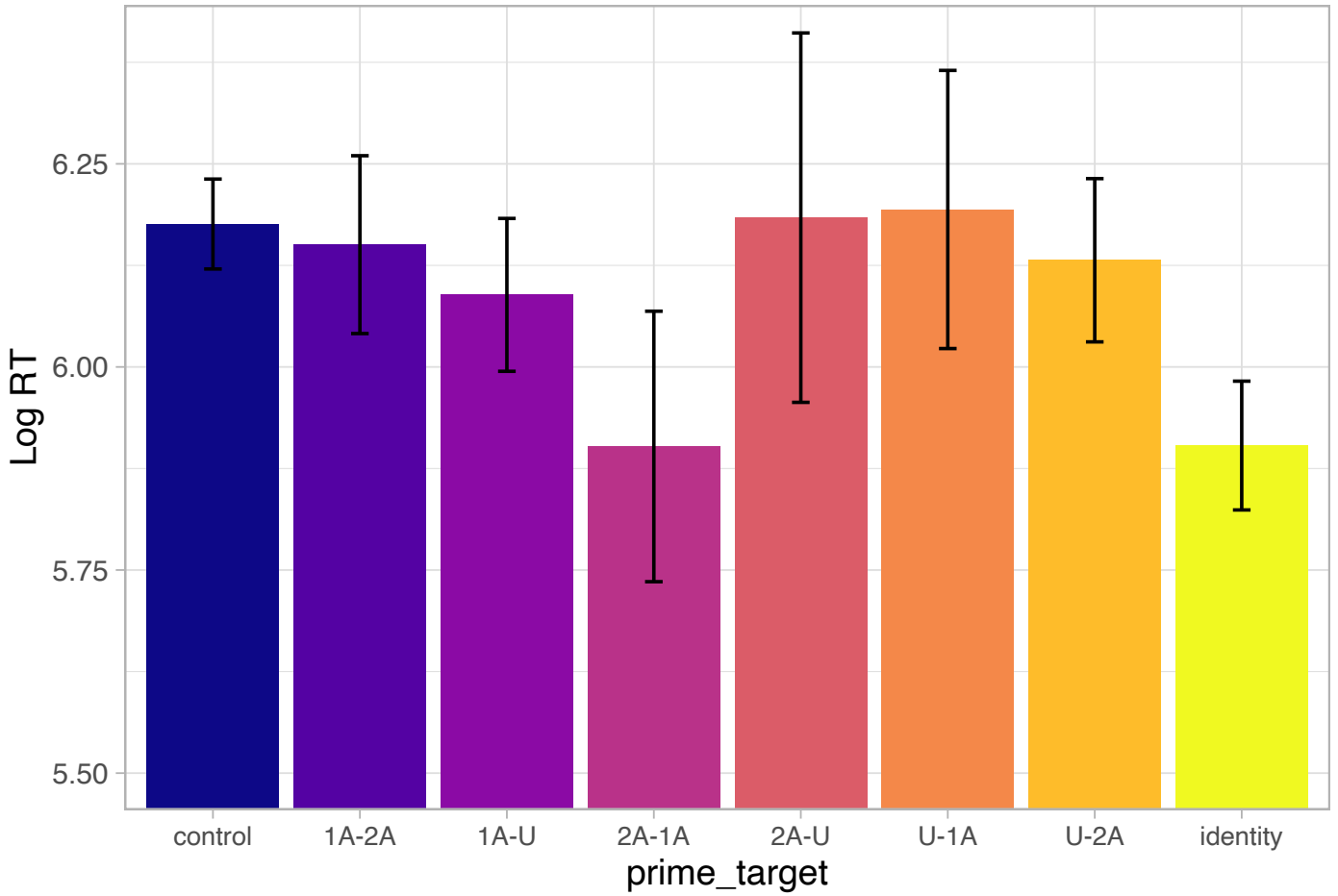
Appendix: Accuracy split by setting



Appendix: Accent primes ~ accent type



Appendix: Segment primes ~ accent type



Frequency information

- identity = 1.68, segment = 1.63, accent = 1.65, control = 1.66
- $F(3,45) = 0.082$, $p = 0.97$