An investigation of Uyghur co-speech gestures: implications for metrical structure

Katie Franich[1], Connor Mayer[2], Travis Major[3], and Gülnar Eziz[1]

Harvard University[1], UC Irvine[2], USC[3]

Tu+9

Cornell University

3/23/2024

Background on Uyghur stress and intonation

Uyghur

- ~10 million speakers
- Spoken primarily in Xinjiang, China and neighboring regions
- Southwestern Turkic language, most closely related to Uzbek

Xinjiang Uygur Autonomous Region

Stress in Turkic Languages

Turkish:

- A stress language with mostly final stress? (e.g. Kabak & Vogel 2001)
- A lexical pitch-accent language? (e.g. Levi 2005)
- A stress language with both edge- and head-marking intonation (e.g. Ipek 2015)

Uyghur stress

"[Uyghur stress is] remarkable for its complexity and instability." Nadzhip (1971)

- If ultimate or penultimate syllable is heavy (CVV, CVC, CVVC, etc) it receives stress, otherwise final (Hahn 1991a, 1991b)
- Numerous exceptions, especially for loanwords.
- Stress falls on the first heavy syllable, otherwise final:
  - Duration is the only correlate of stress. (Engesaeth et al., 2009/2010)
- It is unclear how stress judgments were determined in each of the above resources.

Yakup (2013), Yakup & Sereno (2016)

Identified sets of minimal or near-minimal stress pairs with consistent stress judgments from Uyghur speakers.

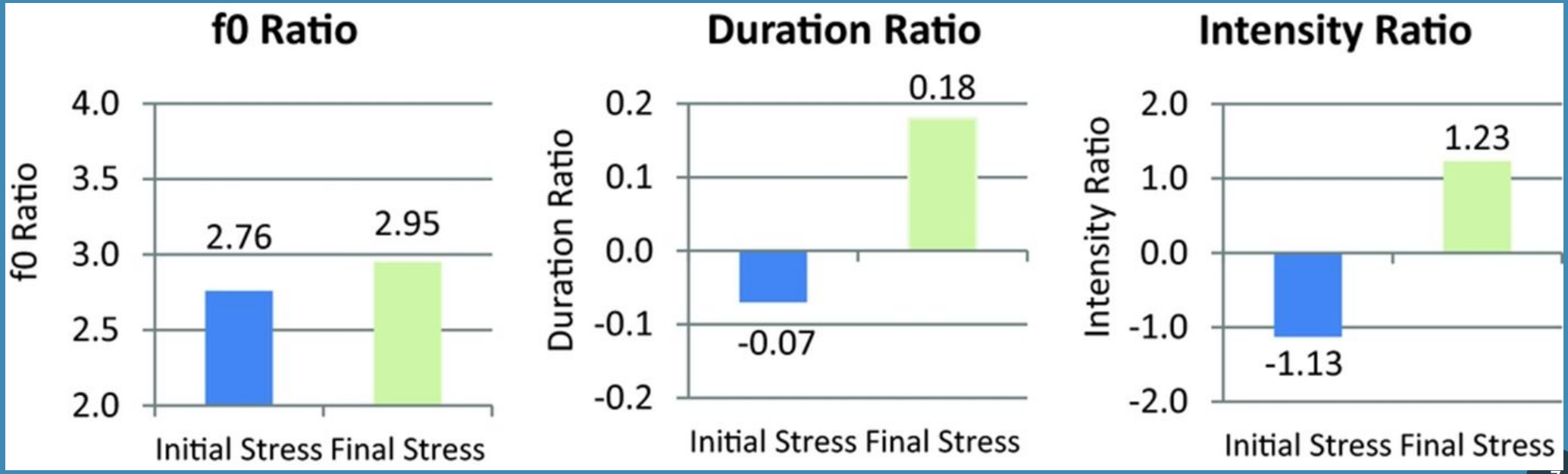
- There were MANY other words where speakers disagreed on stress location!

#	Initial stress	IPA	English gloss	Final stress	IPA	English gloss
1	Acha	/ˈatʃa/	'elder sister'	aCHA	/aˈtʃa/	'branching'
2	Ara	/ˈara/	'fork'	aRA	/aˈra/	'between'
3	TÖshük	/ˈtøʃyk/	'kitchen'	töSHÜK	/tøˈʃyk/	'hole'
4	BAla	/ˈbala/	'child'	baLA	/baˈla/	'disaster'
5	CHAtaq	/ˈtʃataq/	'bad branch of tree'	chaTAQ	/tʃaˈtaq/	'problem'
6	PAchaq	/ˈpatʃaq/	'leg'	paCHAQ	/paˈtʃaq/	'piece'

Yakup (2013), Yakup & Sereno (2016)

Speakers produced disyllabic words with initial or final stress in a carrier phrase.

Duration and **intensity** differed based on stress location; **f0 did not**!



Major & Mayer (2018, in press)

Replicated and extended Yakup (2013) and Yakup & Sereno (2016)

Elicited disyllabic stress (near-)minimal pairs in sentence-initial AND sentence-medial position from eight speakers:

Initial: _____ bek yaxshi söz "____ is a very good word"

Medial: Mahinur _____ deydu "Mahinur will say _____"

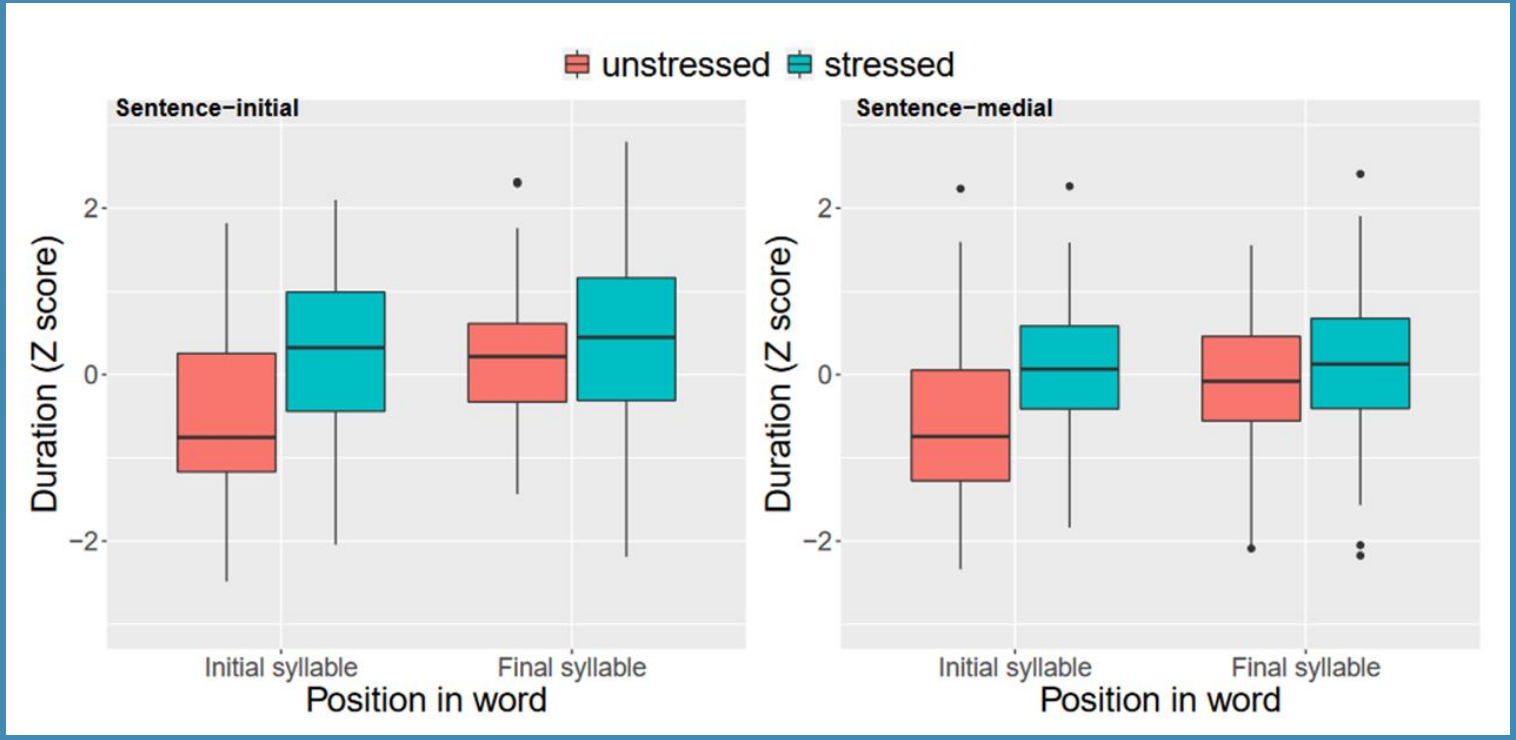
Measured duration, intensity, and f0 of both syllables

Major & Mayer (2018, in press)

Duration results

Syllables are longer when they are:

- Stressed
- Word-final
- Sentence-initial



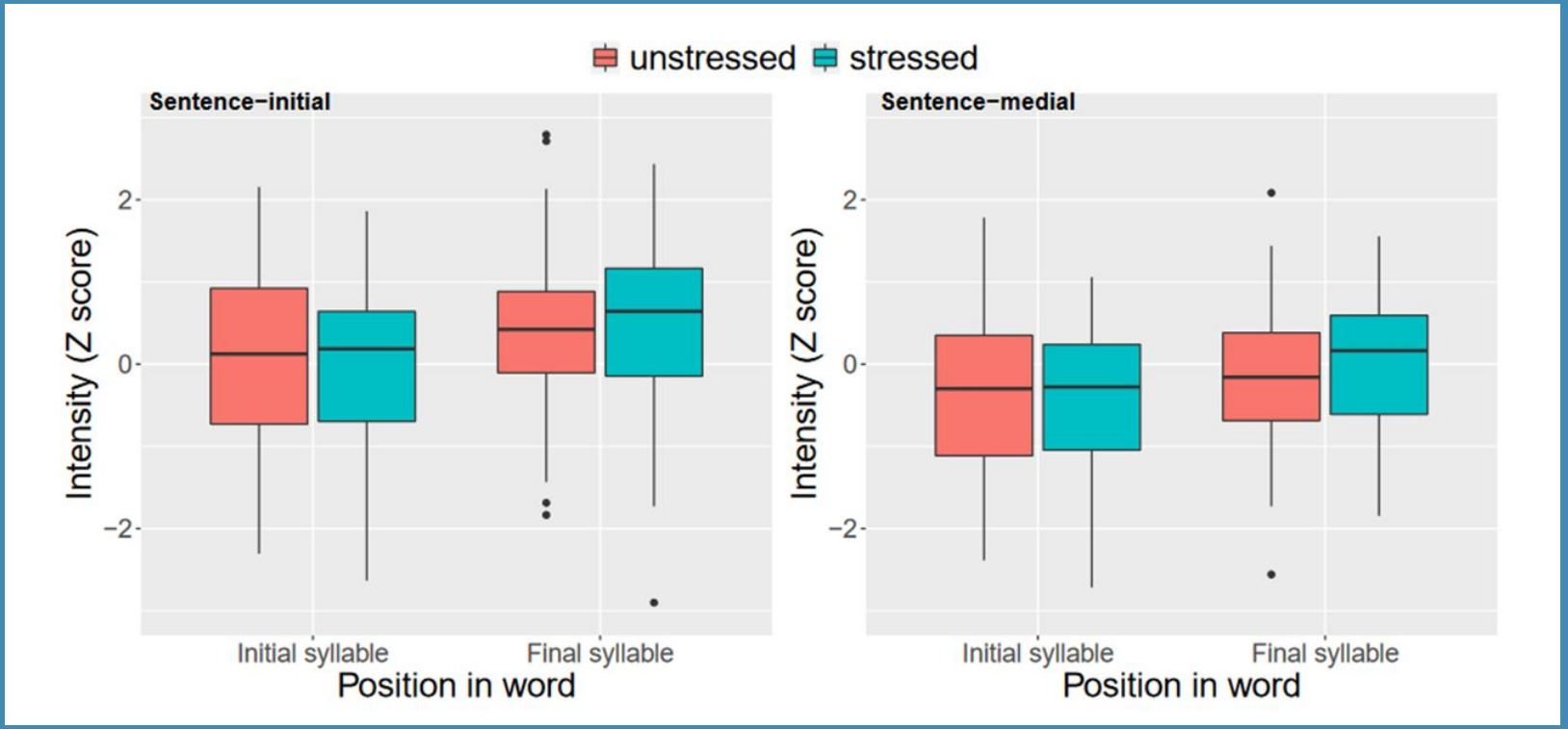
Major & Mayer (2018, in press)

Intensity results

Intensity is higher in syllables when they are:

- Word-final
- Sentence-initial

No effect of stress



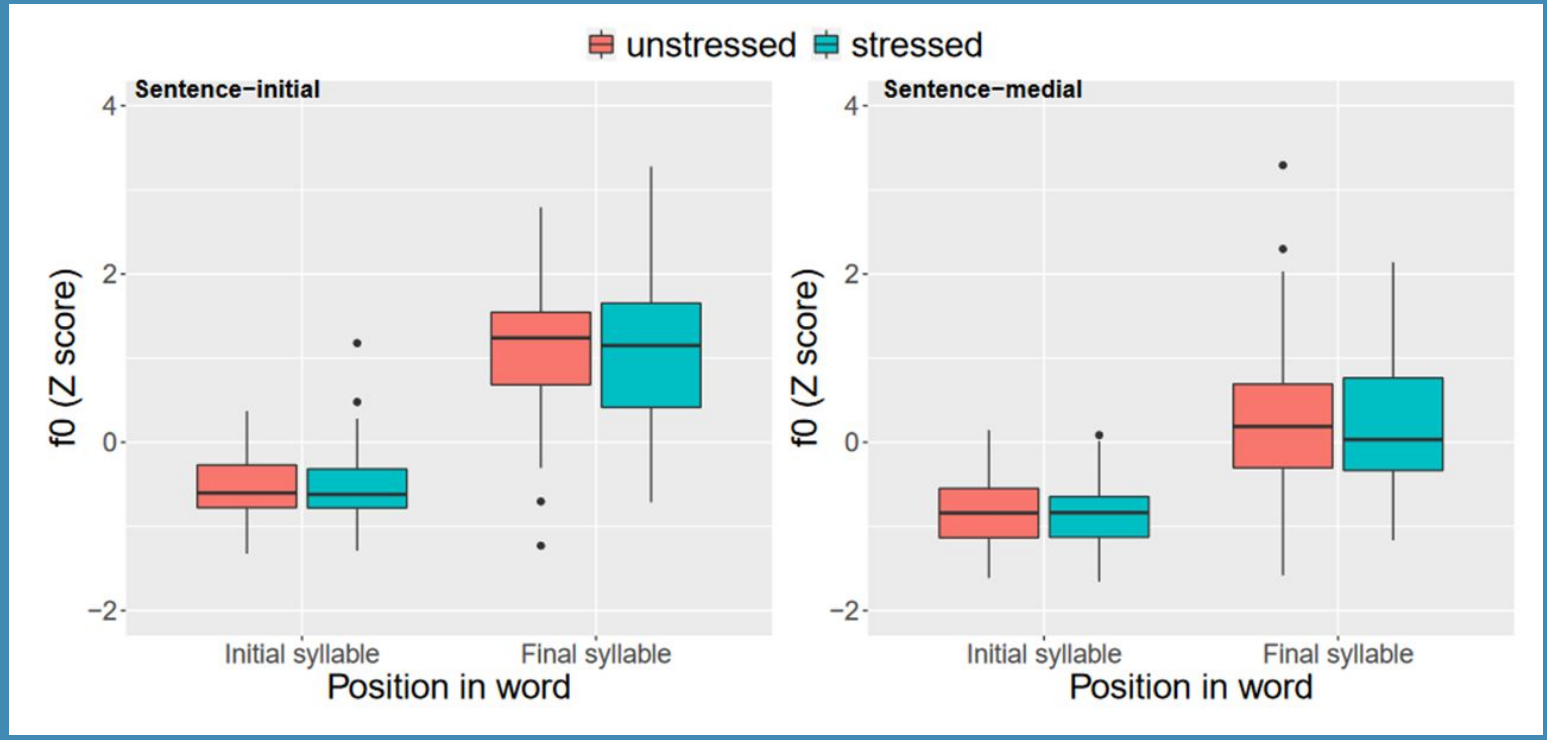
Major & Mayer (2018, in press)

f0 results

f0 is higher in syllables when they are:

- Word-final
- Sentence-initial

No effect of stress



Summary of Quantitative Studies

Duration is the only acoustic cue consistently associated with stress

No studies have found that f0 correlates with stress

- Different sentence types lead to differences in f0
- Word-final syllables have higher f0
- It is unclear what can be said about intensity

Summary of Impressionistic Findings

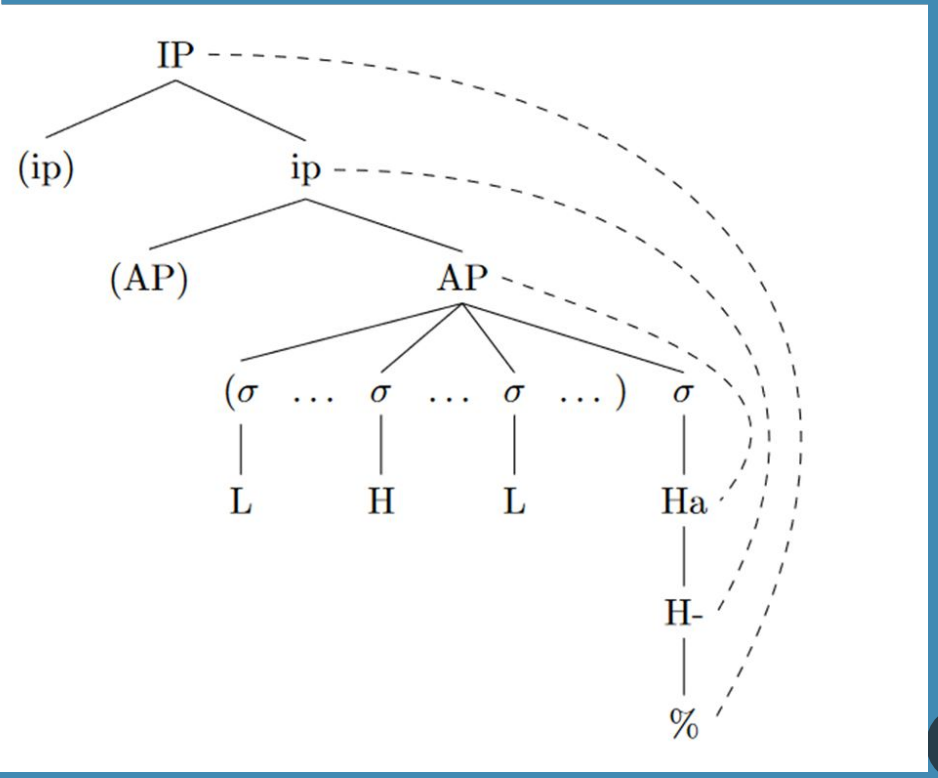
Stress is really complicated in Uyghur

- Stress is often associated with final syllables.
- There is a tendency for non-final stress to associate with heavy syllables.
- When asked, speakers show a lot of variation with respect to where they identify stress.

Uyghur Intonation

Major & Mayer (2018; to appear) present an autosegmental metrical model (Pierrehumbert 1980, Beckman & Pierrehumbert 1986) of Uyghur intonation:

Phrase-internal tones	
AP-internal tones	
(LHL)	Optional AP-internal tones. The first tone that is realized is typically L, and is generally aligned with the first syllable of the AP. The alignment of the other tones is variable.
Boundary tones	
AP boundary tones	
Ha	Realized on AP-final syllable
ip boundary tones	
H-	Realized on ip-final syllable
IP boundary tones	
L%	Used in declaratives; realized on IP-final syllable
H%	Used in questions and continuation rises; realized on IP-final syllable
HL%	Used in declaratives; realized on IP-final syllable
LH%	Used in questions and continuation rises; realized on IP-final syllable



Major & Mayer (2018, in press)

- Both AP and ip end with a high tone (Ha and H-, respectively)
  - Both undergo phrase-final lengthening
  - H- final syllables are longer than Ha-final syllables.
  - **A surprising conclusion: Uyghur has stress, but the intonational system does not interact with stress.**

Remaining issues

Both AP and ip end with a high tone and show final lengthening:

- Stress-conditioned duration in word-final syllables is obfuscated by phrase-final lengthening
- Informally, speakers often select non-final syllables as "the most prominent" despite those syllables being shorter
- Speaker judgments are often inconsistent about prominence

Back to square one:

What is stress in Uyghur and how can we measure it?!?!

Co-speech gestures

WHAT ARE CO-SPEECH GESTURES?

Movements of the hands/arms, head, shoulders, etc. which accompany spoken language and which are **temporally coordinated to speech**

- Occur even in the physical absence of an interlocutor (Wei 2006)
- Are acquired by both blind and sighted speakers (Özçalışkan et al. 2018)

WHAT ARE CO-SPEECH GESTURES?

'Beat gestures': a subset of co-speech gesture which are tied closely to prosodic structure

- 'Non-meaningful'
- Conditioned by prosodic structure (Kendon 1980; McNeill 1992)
- Influence speech perception (Leonard & Cummins 2011; Bosker & Peeters 2021)

WHAT ARE CO-SPEECH GESTURES?

Other types of gestures:

- Iconic / depictive
- Deictic
- Conventional

WHAT ARE CO-SPEECH GESTURES?

Other types of gestures:

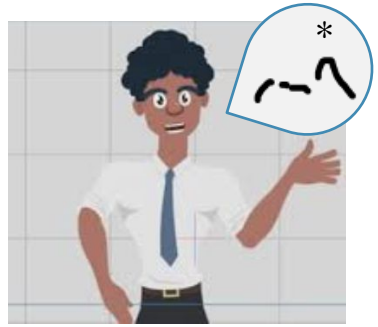
- Iconic / depictive
- Deictic
- Conventional

- All gestures have prosodically important timing (McClave, 1998; Krahmer & Swerts, 2005, 2007; Loehr, 2004, 2007; Prieto 2018)

PROSODIC LINKS IN SPEECH AND GESTURE

- Gestures preferentially aligned with **metrically-prominent** syllables

- The *apex* (max. extension) of a gesture aligns closely with **pitch peaks** of pitch-accented syllables in several Western European languages

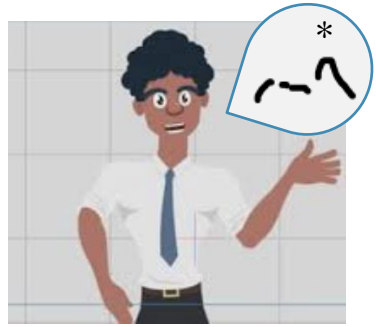


Kendon 1980; McNiell 1992; Esteve-Gibert & Prieto 2013; Leonard & Cummins 2010; Loehr 2012; Esteve-Gibert et al. 2017

HOW DOES ALIGNMENT WORK IN TURKIC LANGUAGES?

- Gestures preferentially aligned with **metrically-prominent** syllables

- The *apex* (max. extension) of a gesture aligns closely with **pitch peaks** of pitch-accented syllables in several Western European languages



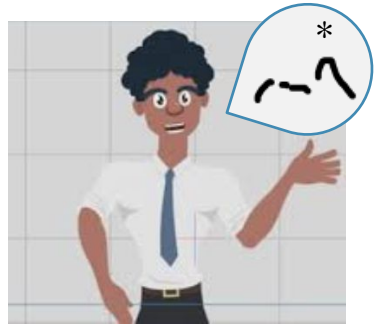
Kendon 1980; McNiell 1992; Esteve-Gibert & Prieto 2013; Leonard & Cummins 2010; Loehr 2012; Esteve-Gibert et al. 2017

HOW DOES ALIGNMENT WORK IN TURKIC LANGUAGES?

- In Uyghur, we have proposed that stress and intonation **do not interact** (as they do in English)

- Gestures preferentially aligned with **metrically-prominent** syllables

- The *apex* (max. extension) of a gesture aligns closely with **pitch peaks** of pitch-accented syllables in several Western European languages



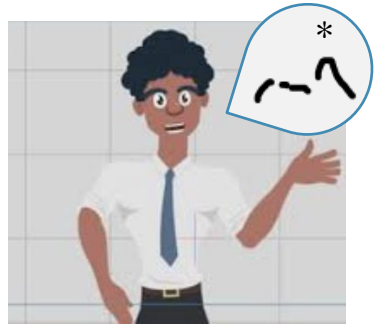
Kendon 1980; McNiell 1992; Esteve-Gibert & Prieto 2013; Leonard & Cummins 2010; Loehr 2012; Esteve-Gibert et al. 2017; Franich & Keupdjio 2022

HOW DOES ALIGNMENT WORK IN TURKIC LANGUAGES?

- In Uyghur, we have proposed that stress and intonation **do not interact** (as they do in English)
- Gestures tend to gravitate to rhythmically-prominent syllables cross-linguistically

- Gestures preferentially aligned with **metrically-prominent** syllables

- The *apex* (max. extension) of a gesture aligns closely with **pitch peaks** of pitch-accented syllables in several Western European languages



Kendon 1980; McNiell 1992; Esteve-Gibert & Prieto 2013; Leonard & Cummins 2010; Loehr 2012; Esteve-Gibert et al. 2017; Franich & Keupdjio 2022

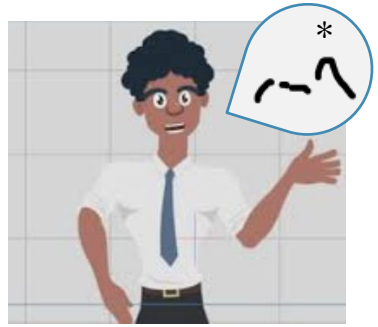
HOW DOES ALIGNMENT WORK IN TURKIC LANGUAGES?

Prediction:

Gestures in Uyghur **will be attracted to stressed syllables** (as reflected in duration), rather than pitch peaks (which mark phrase boundaries, **rather than rhythmic prominence**)

- Gestures preferentially aligned with **metrically-prominent** syllables

- The *apex* (max. extension) of a gesture aligns closely with **pitch peaks** of pitch-accented syllables in several Western European languages



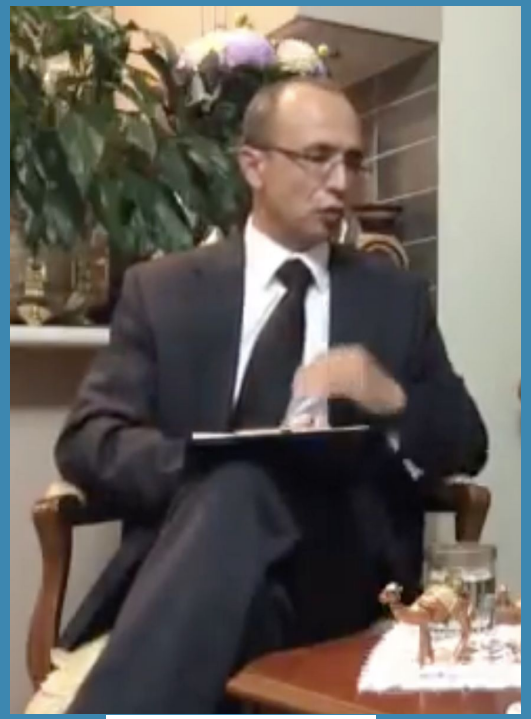
Kendon 1980; McNiell 1992; Esteve-Gibert & Prieto 2013; Leonard & Cummins 2010; Loehr 2012; Esteve-Gibert et al. 2017; Franich & Keupdjio 2022

GESTURE CORPORA

- Three Uyghur speakers video- and audio-recorded (accessed via wikitongues through YouTube) describing aspects of Uyghur language and culture
- Audio data transcribed and force-aligned at word-, syllable- and phone-levels using Montreal Forced Aligner (McAuliffe et al. 2017) and subsequently checked for alignment accuracy
- ~3200 phones total
- 165 gesture-accompanied phones
  (more data collection currently in progress!)

GESTURE PHASES AND ANNOTATION

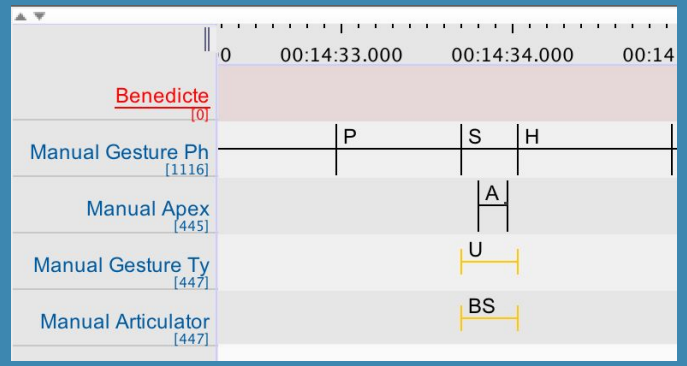
Pre-Gesture



Preparation

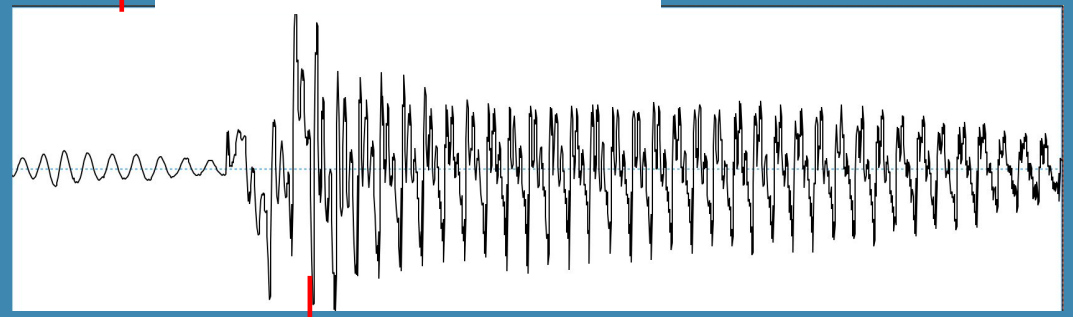
Stroke Onset

Stroke Apex



Consonant-aligning apex

b a



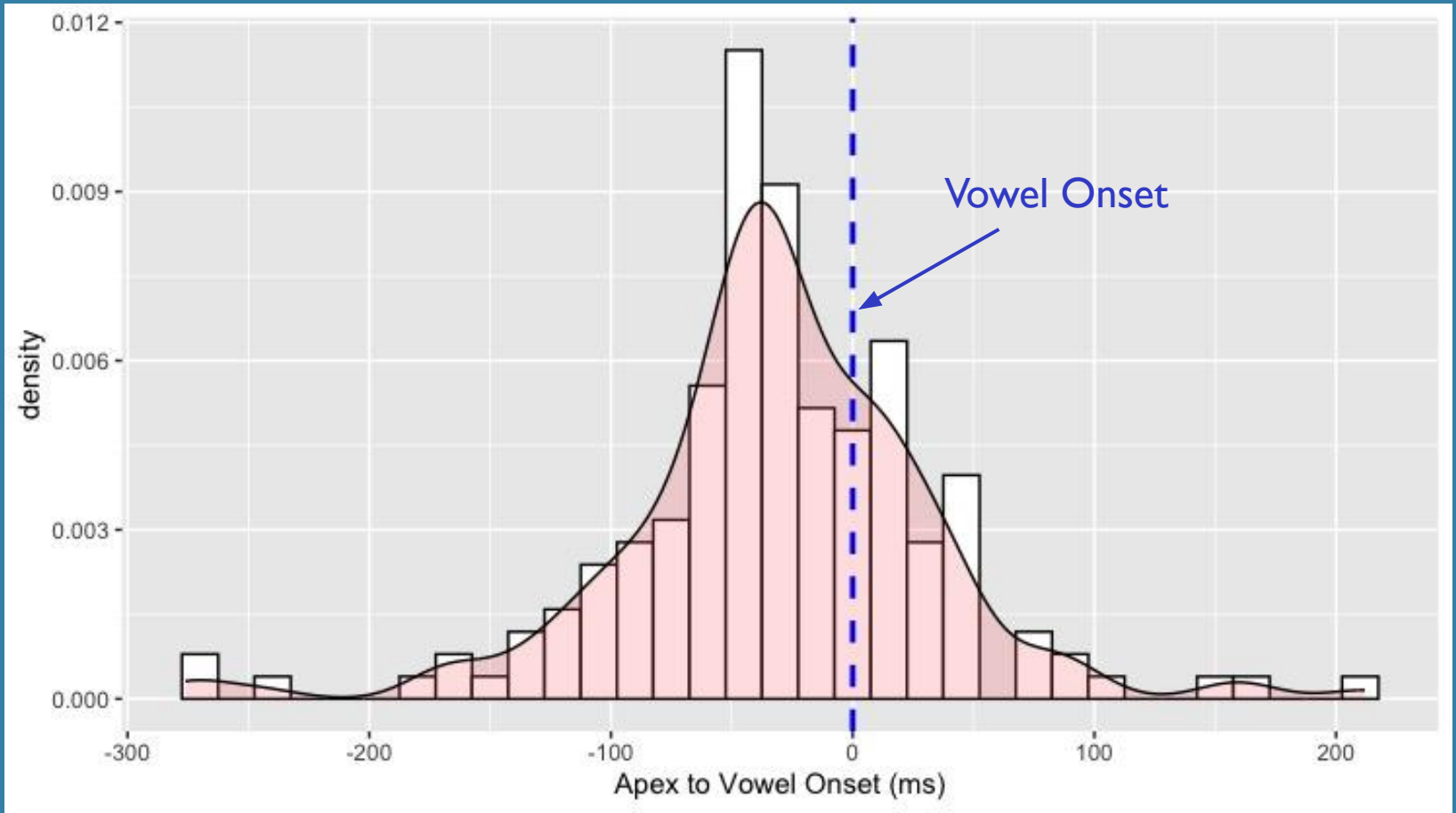
Time

Vowel Onset

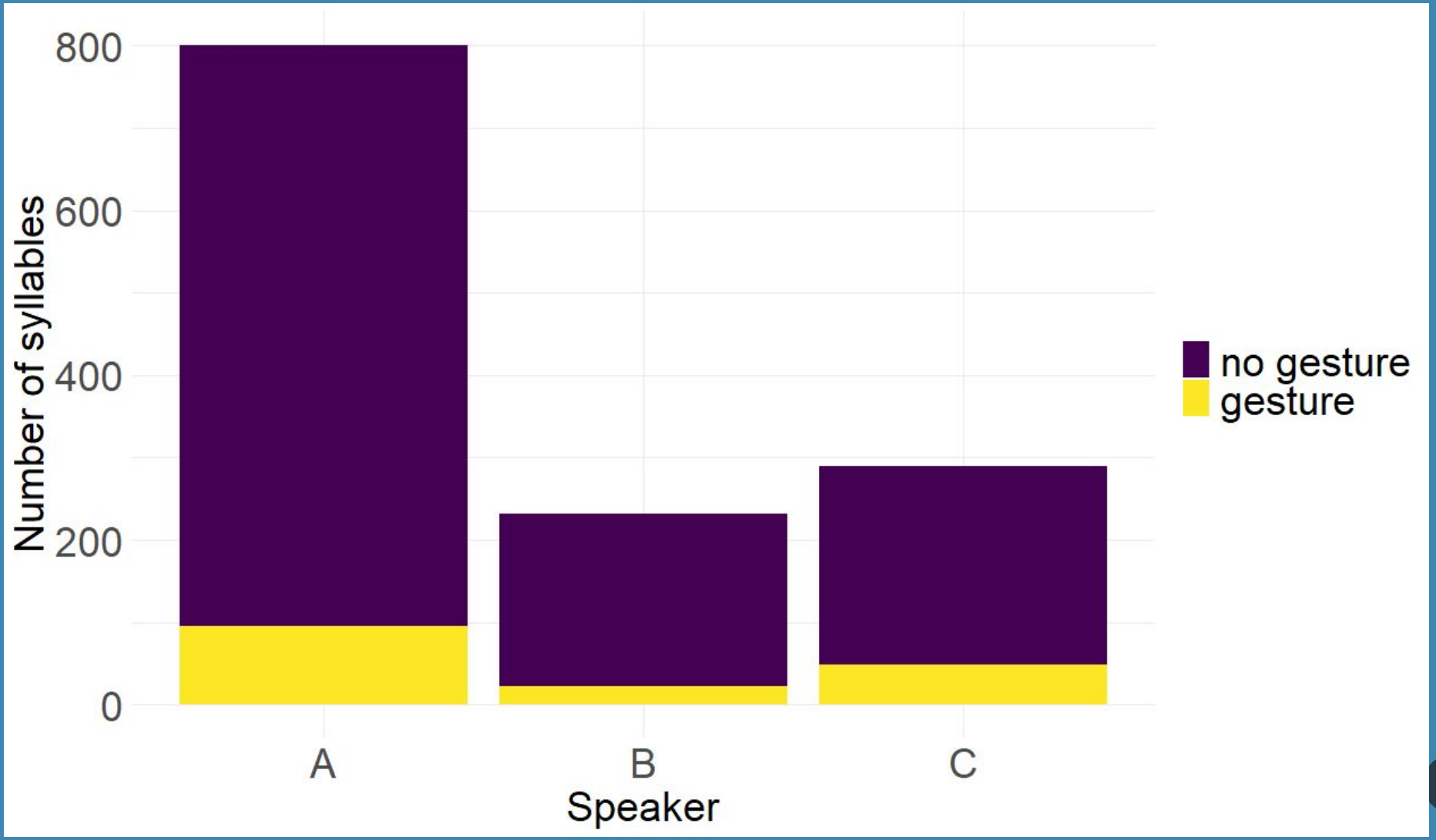
Vowel-aligning apex

Co-speech gestures in Uyghur

Results: Apex Timing



Gesture rates across speakers



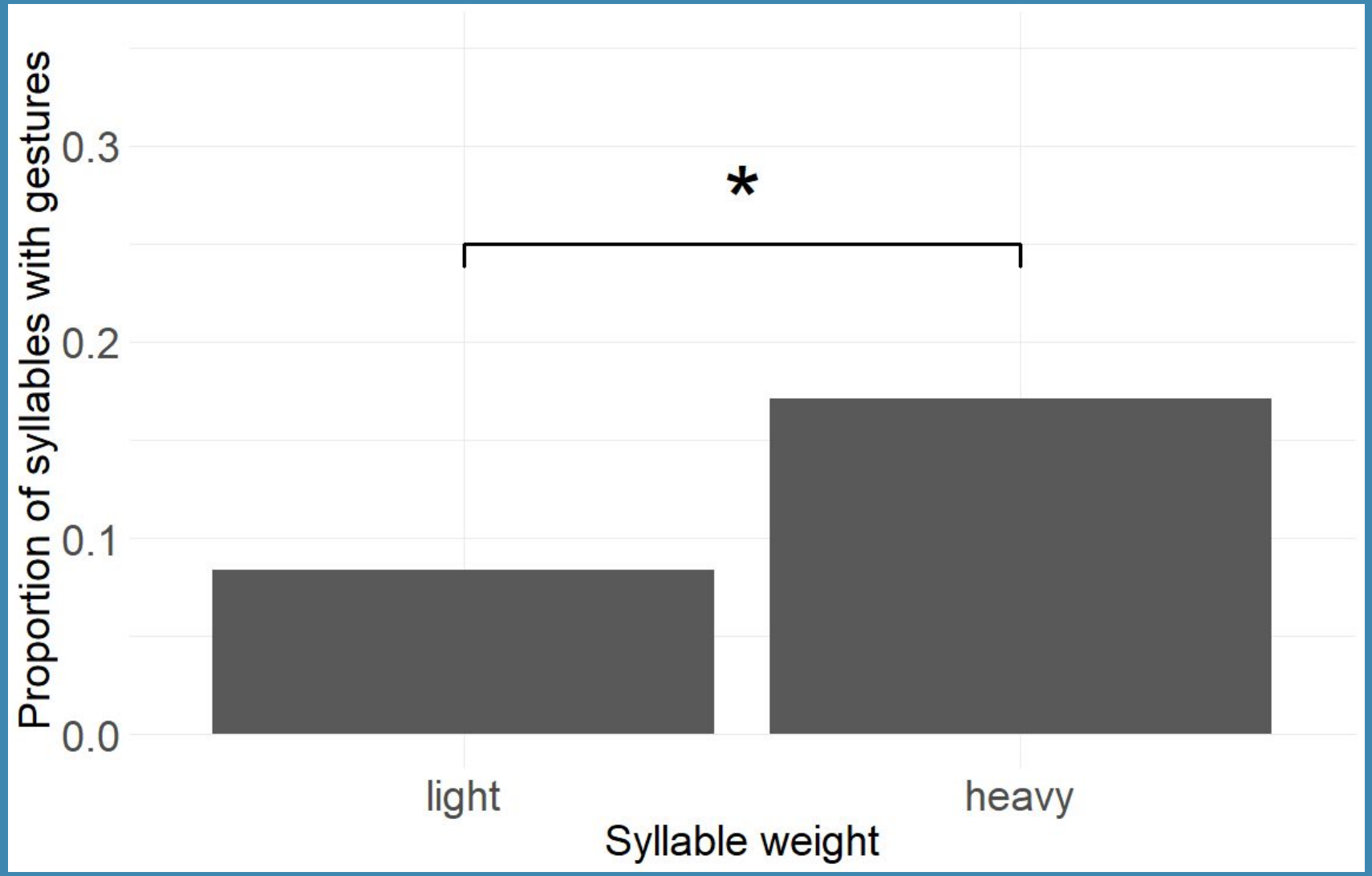
Predicting gestures

Question: What syllable features predict a co-occurring gesture?

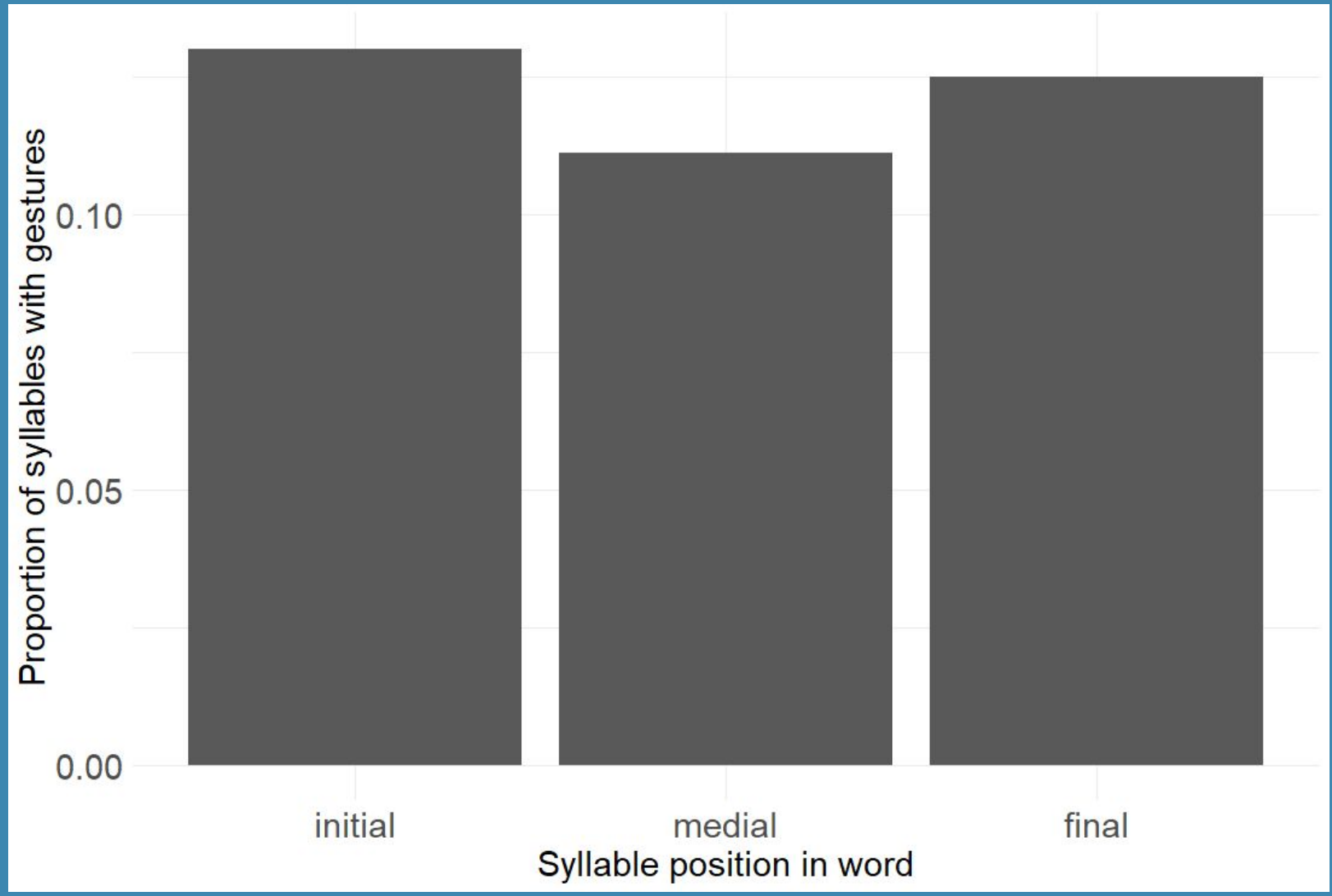
Logistic regression model fit to all syllables

gesture ~ syl. weight * position + maxf0 + intensity +duration + (1|speaker) + (1|syllable)

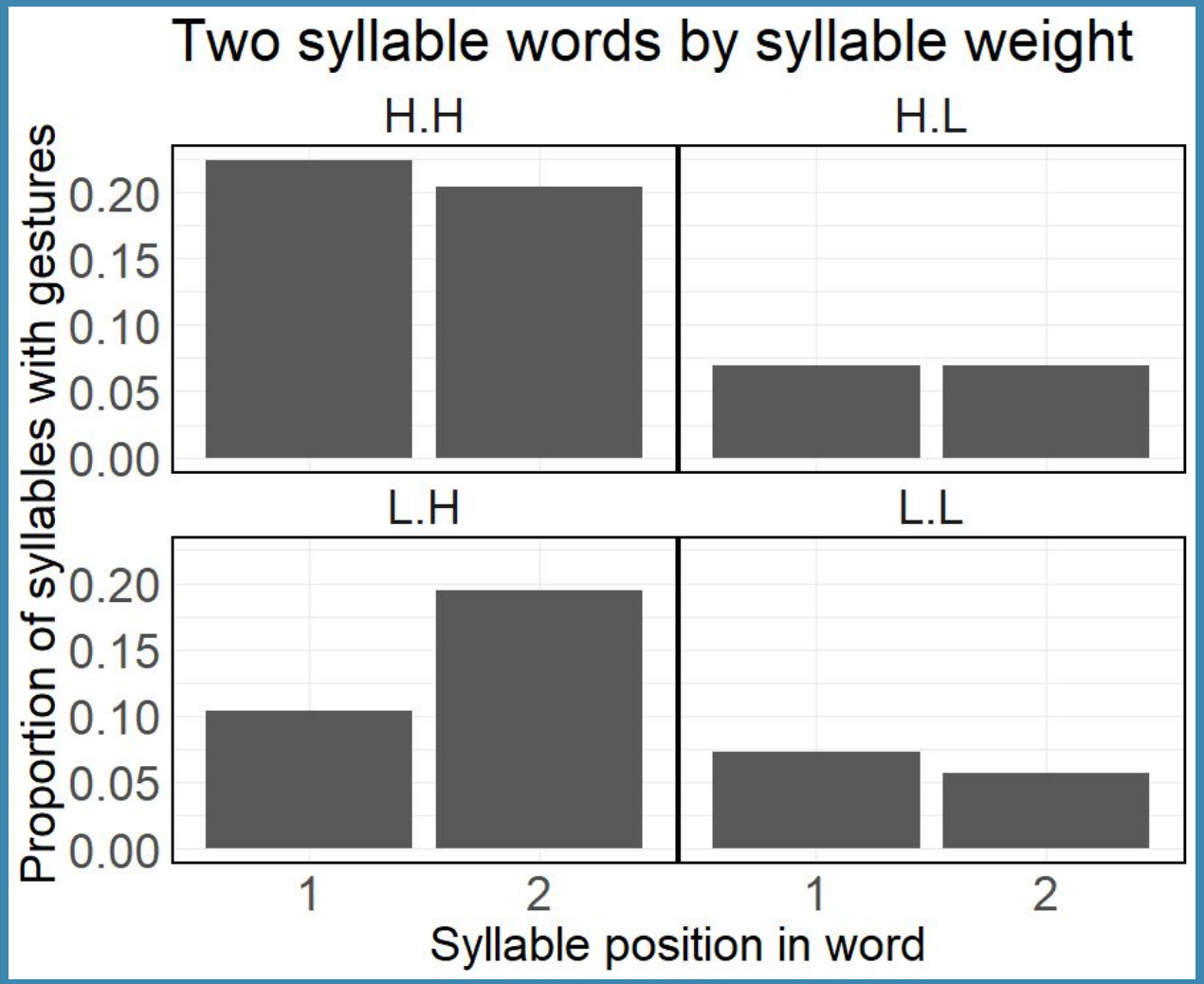
Gestures prefer heavy syllables

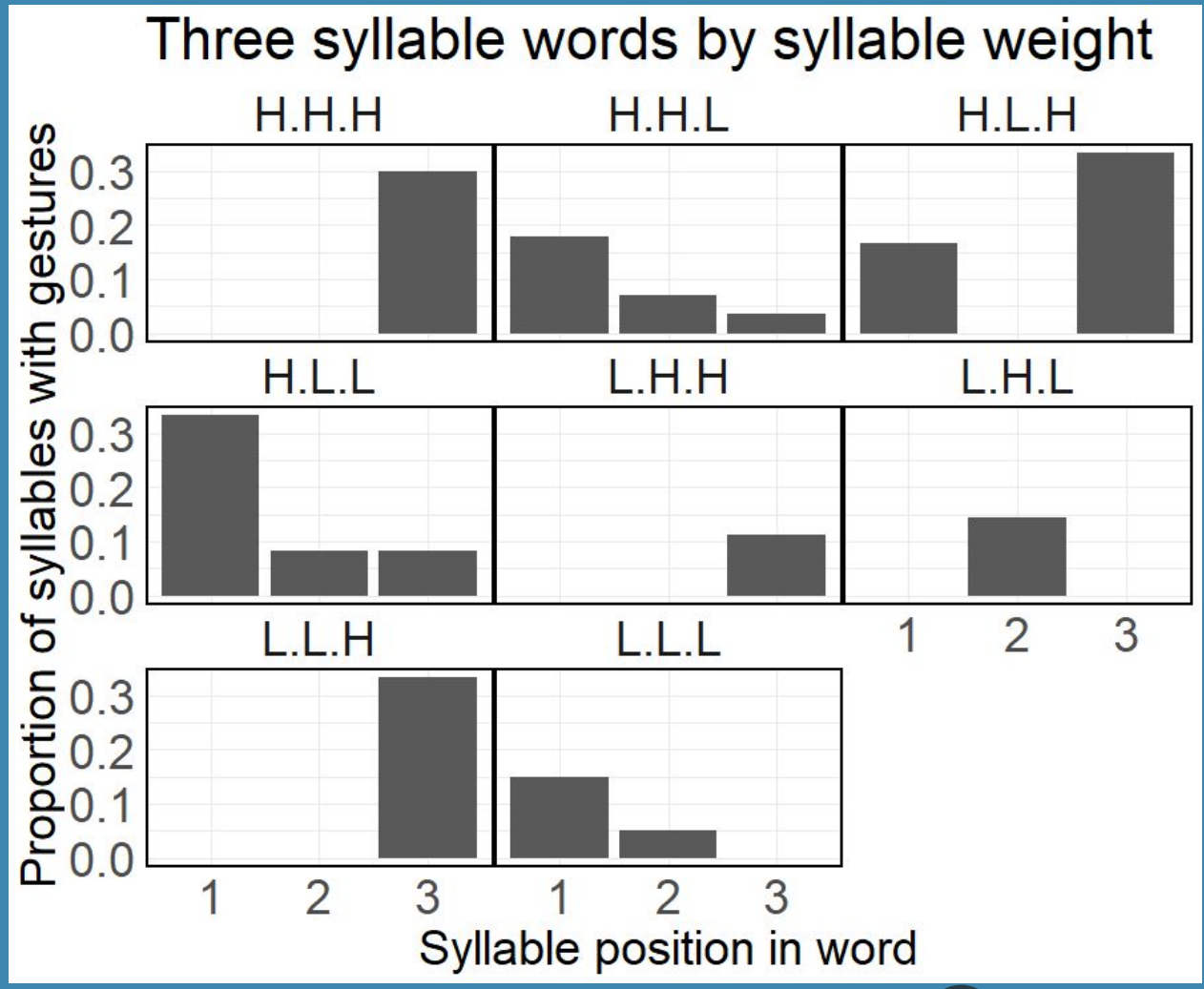


Gestures do not prefer final position

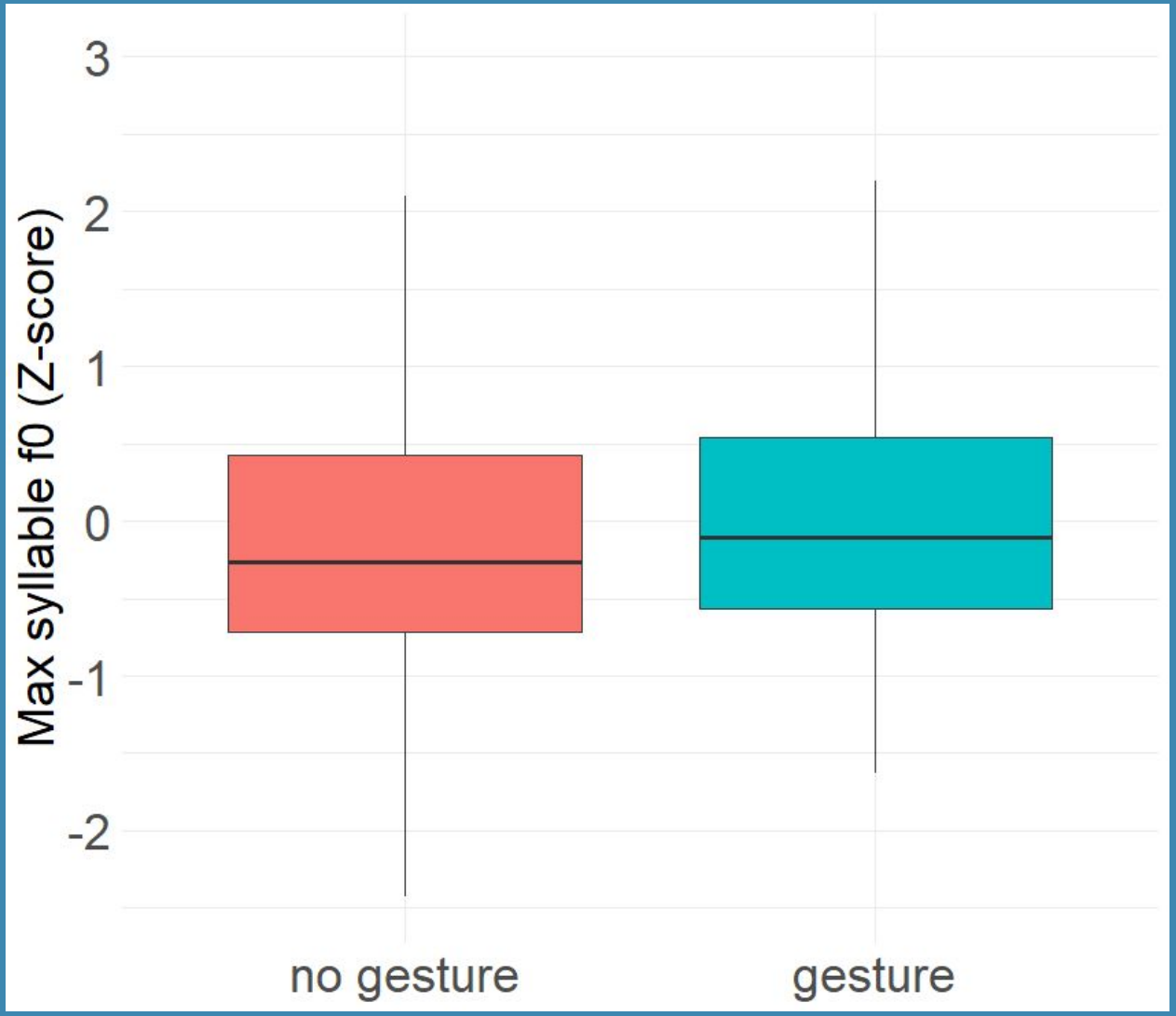


Position x weight

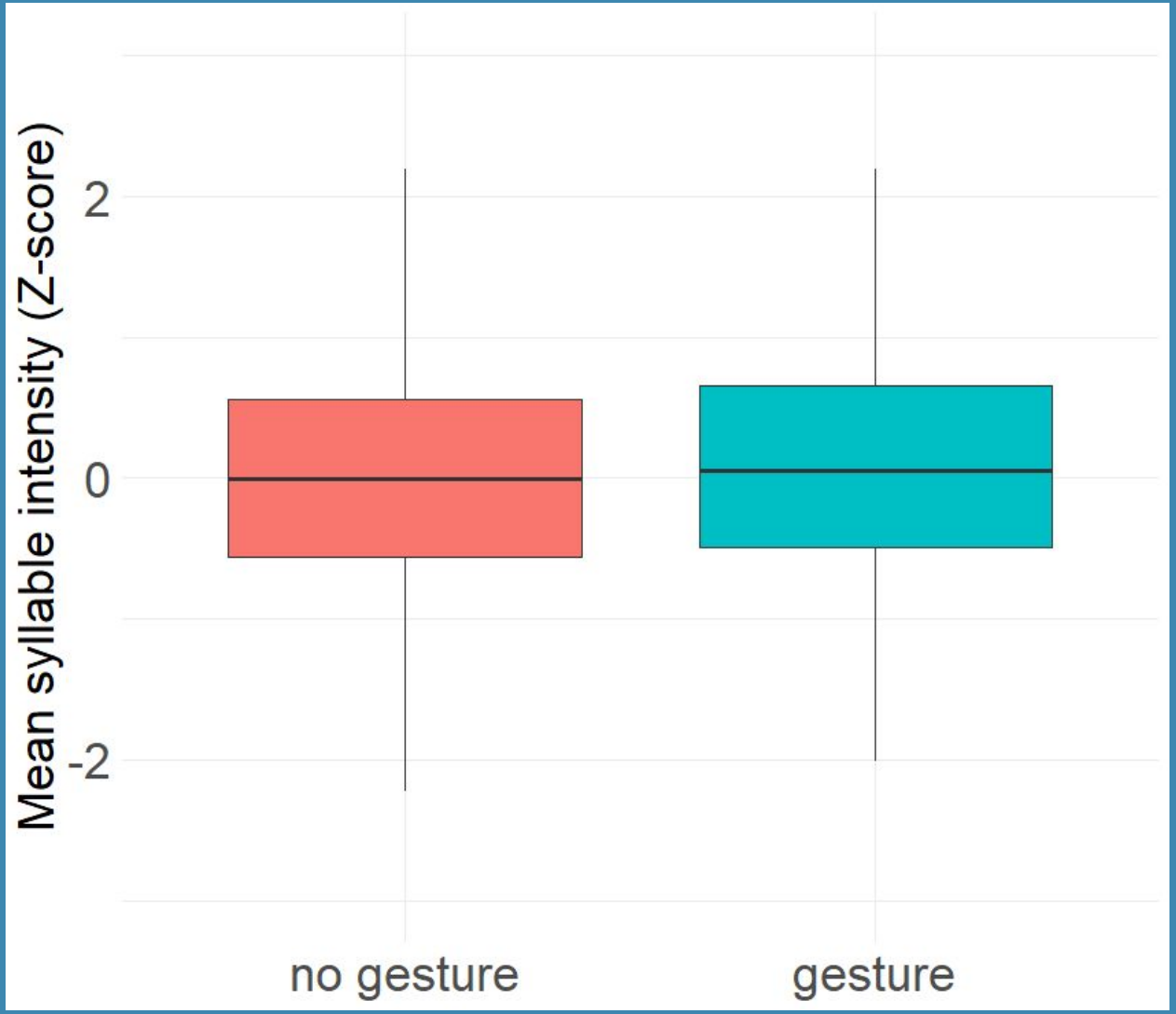




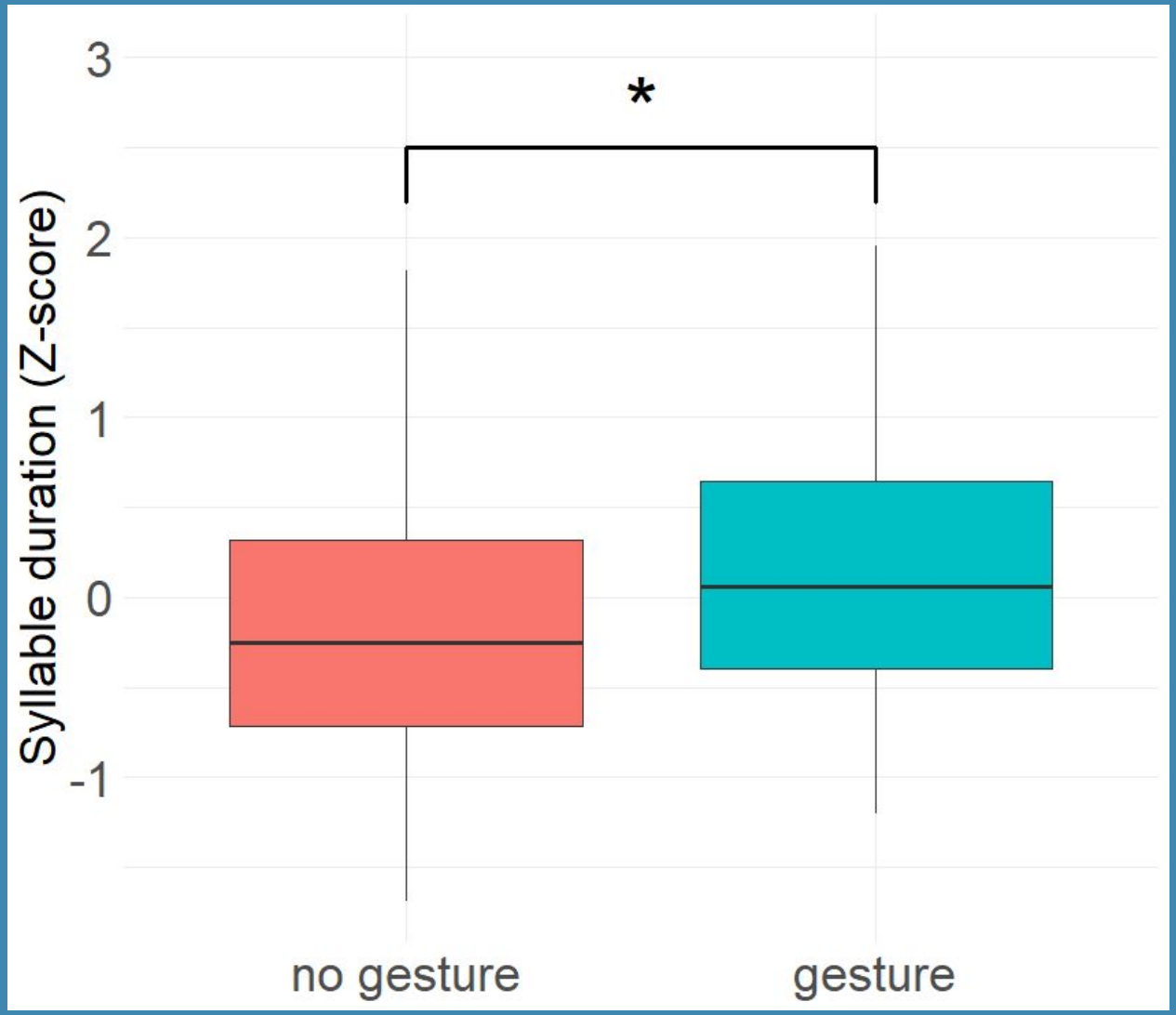
Gestures are not predicted by f0…



Gestures are not predicted by intensity…



But syllables with gestures are longer



Takeaway points

Gestures prefer to be associated with **heavy syllables**

Gestures have **no preference for final position**

Syllables with gestures are **longer**

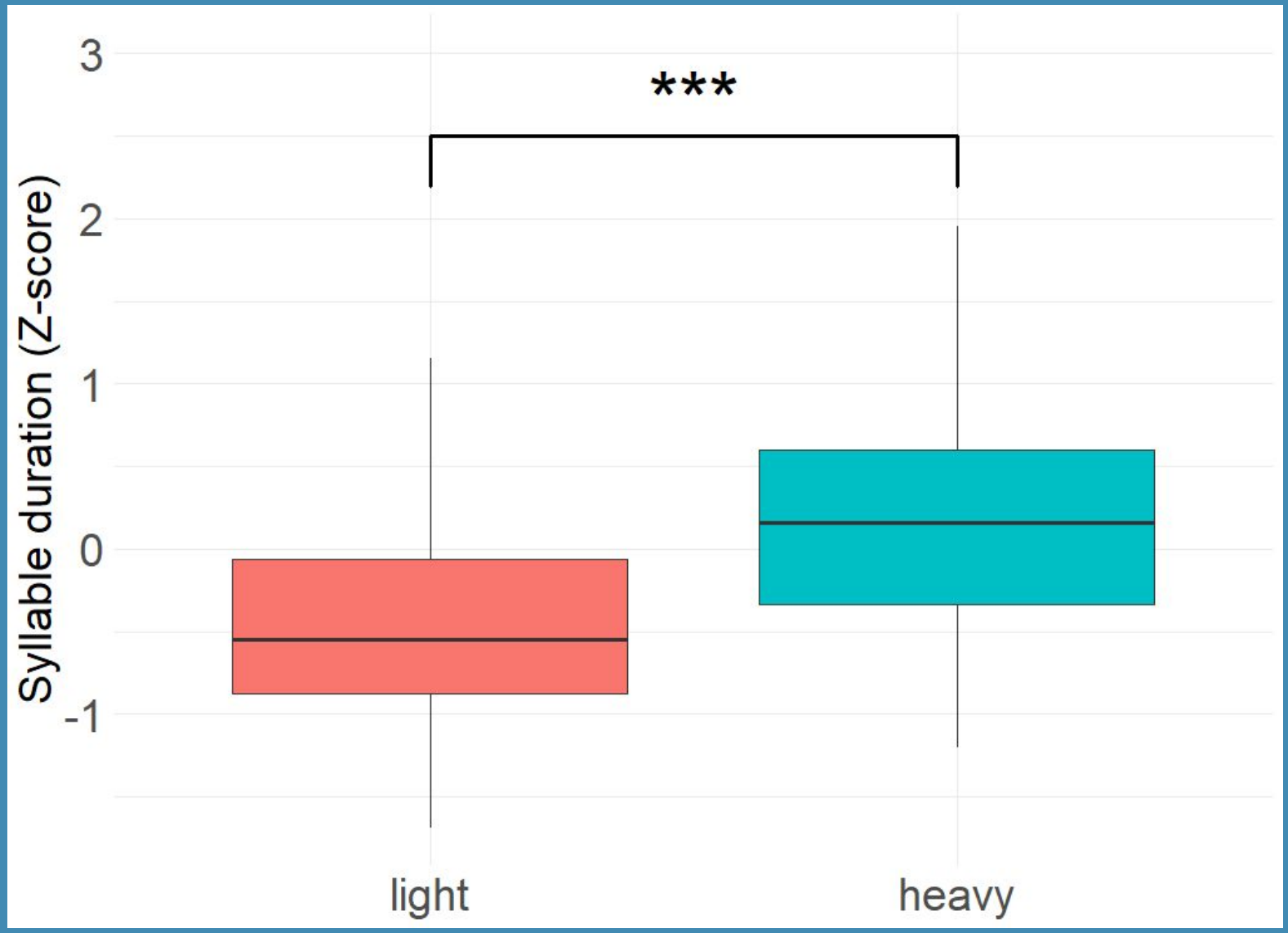
Duration model

Question: What factors predict syllable duration?

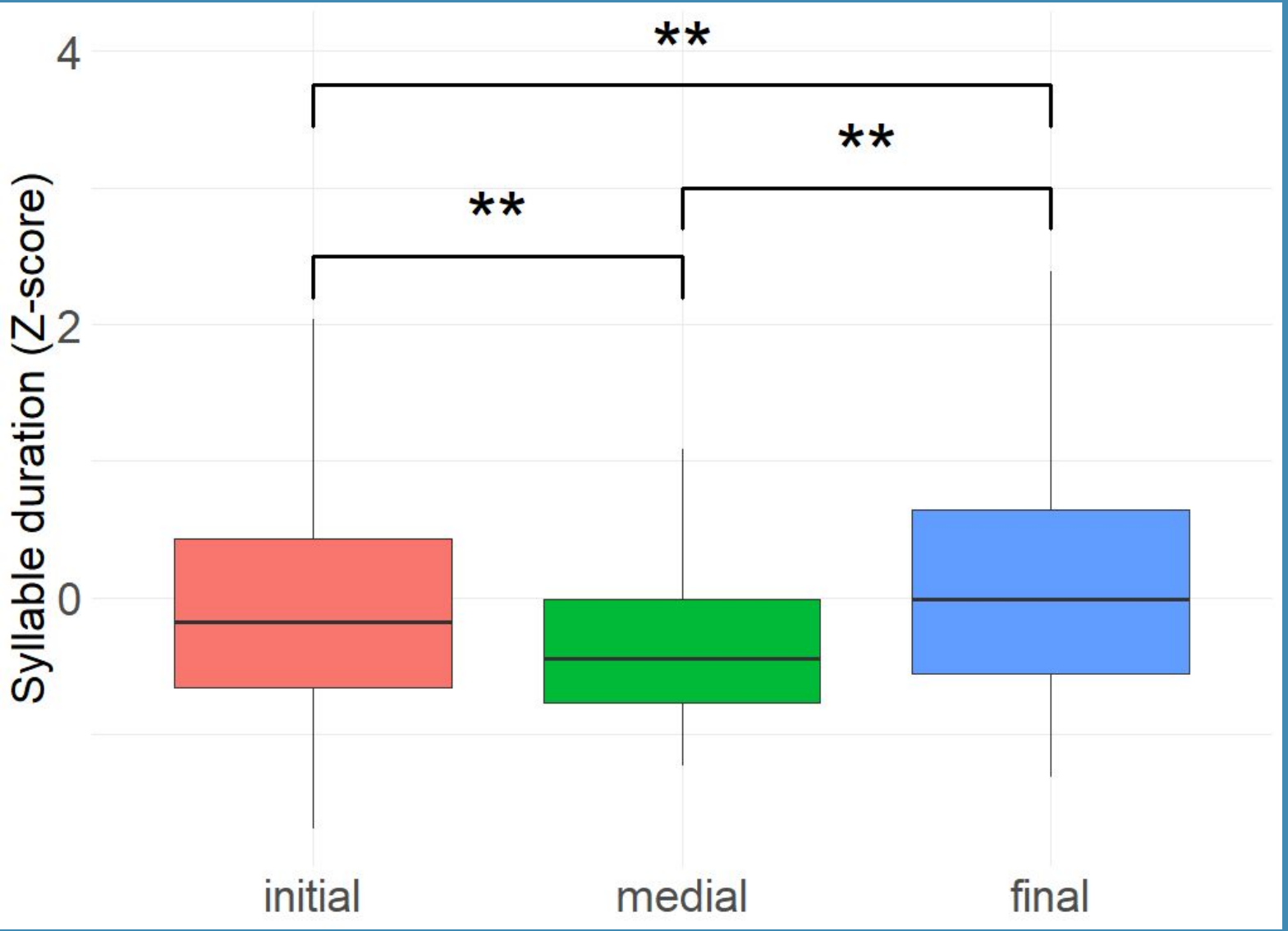
Linear regression model fit to all syllables

duration ~ syl. weight + position * maxf0 + intensity + gesture * position + (1|syllable)

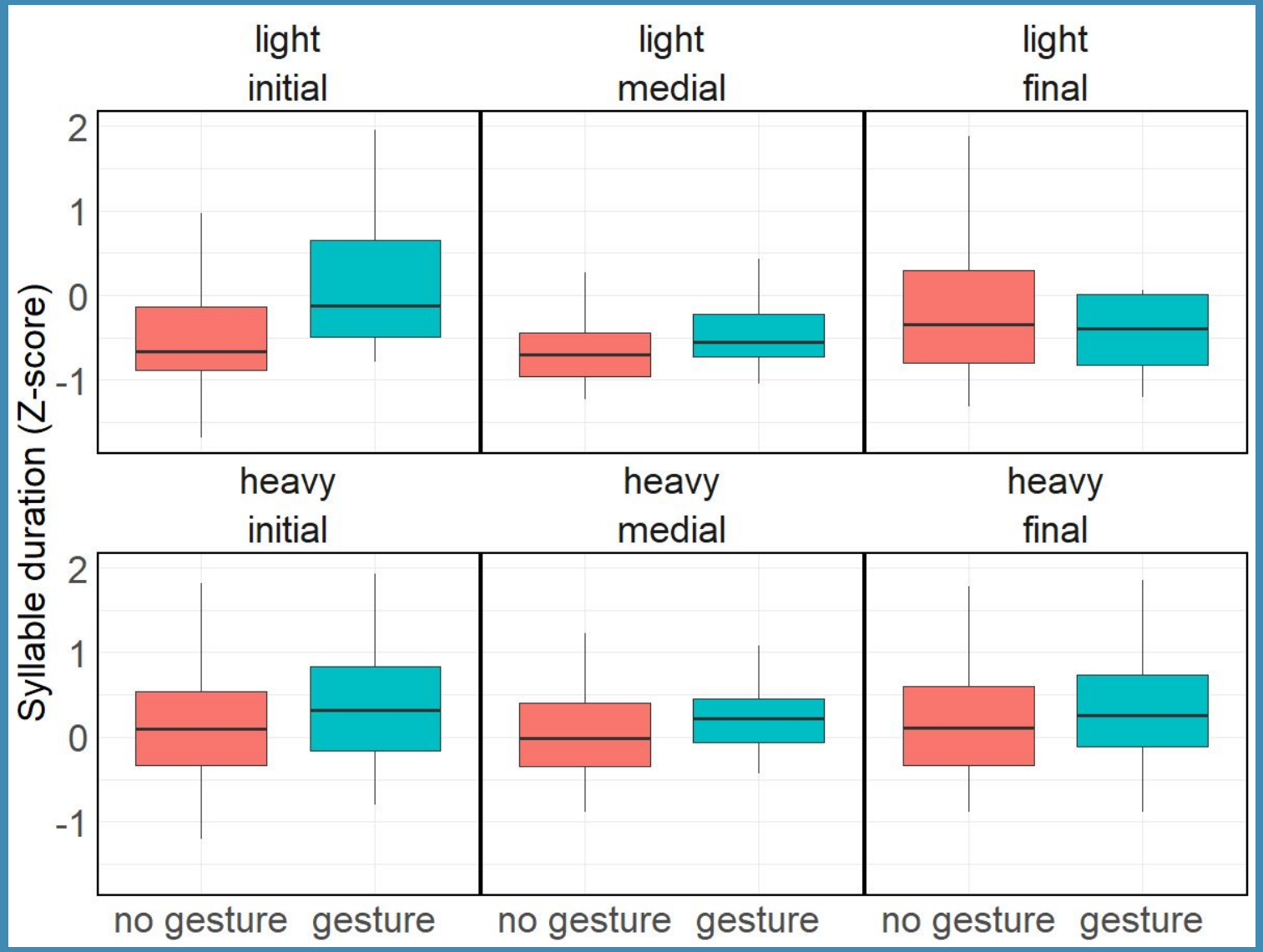
Heavy syllables are longer



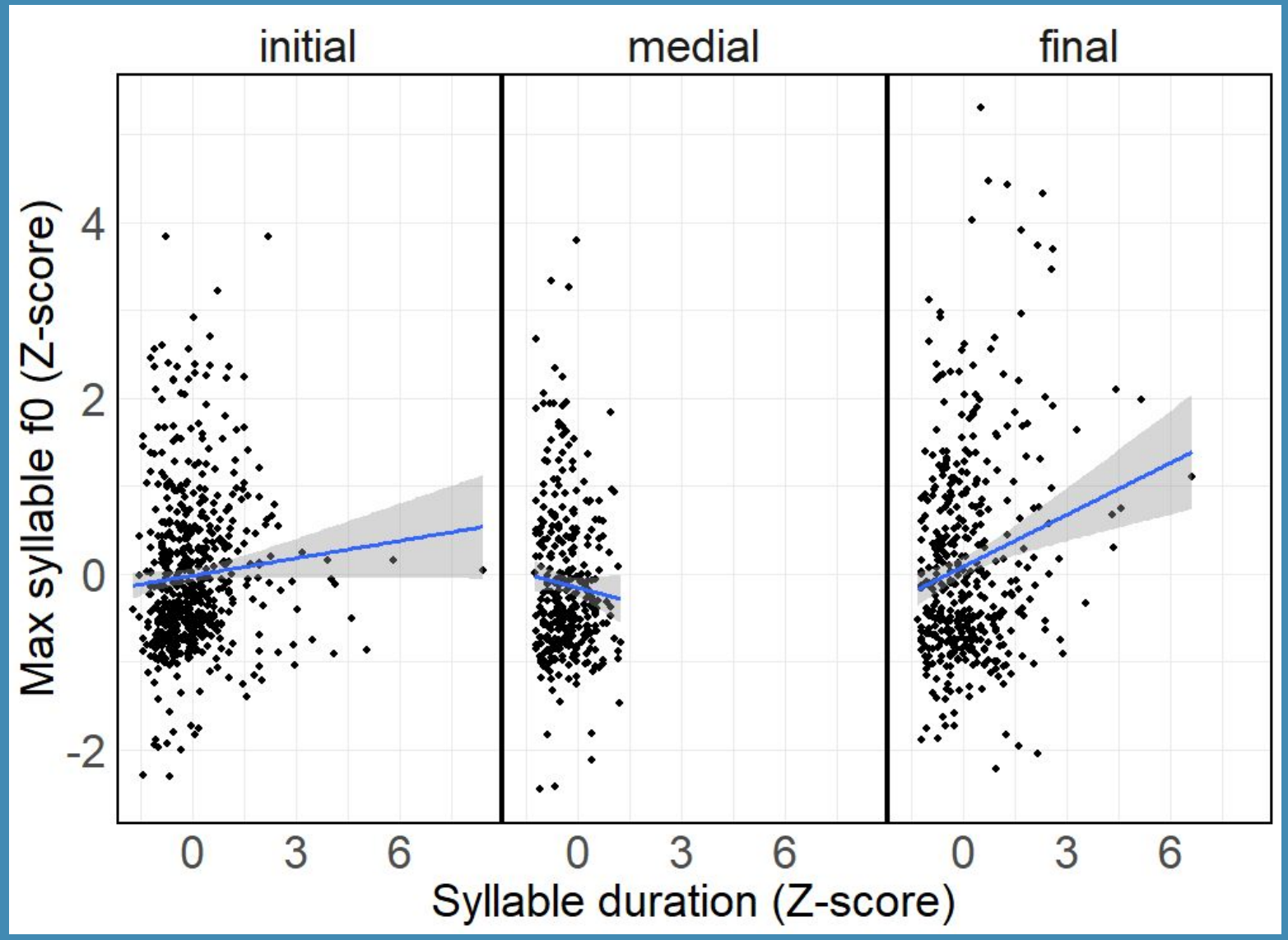
Duration: final > initial > medial



Non-final syllables with gestures are longer



f0 and duration only correlated in final position



Takeaway points

- Heavy syllables are longer
- Final syllables > initial syllables > medial syllables
- Syllables with gestures are longer (weakest in final position)
- f0 and duration covary only at word boundaries
  - Correlates of prosodic boundaries (Major & Mayer 2018, in press)

Why are gestured syllables longer?

Do gestures preferentially align with longer syllables?

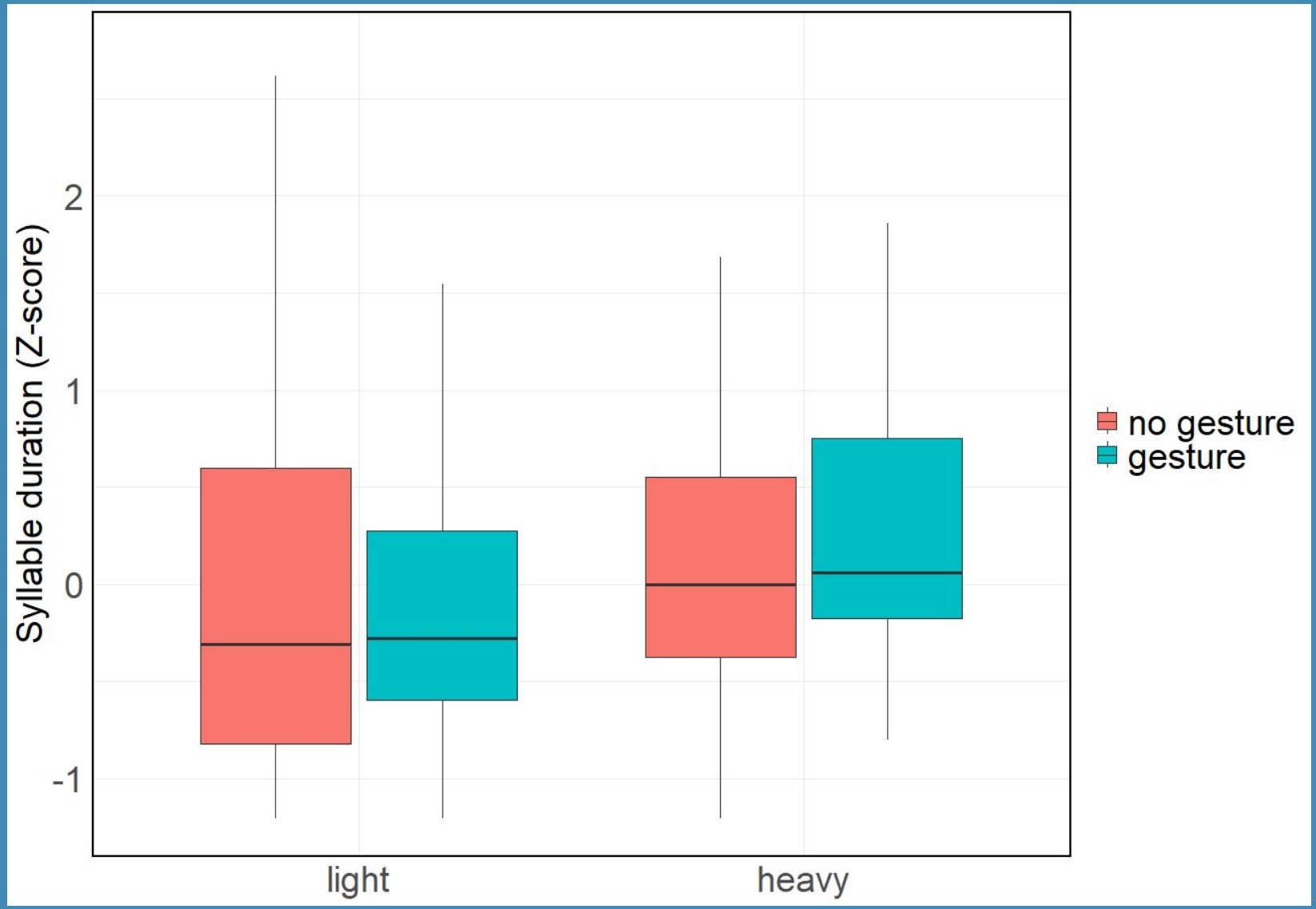
Or does an aligned gesture cause lengthening?

Paired syllables

Compare only syllables that occur in tokens of the same word with and without gestures

Paired t-test between mean duration for each syllable/word in gestured/ungestured tokens shows no difference ($t(59) = |1.32|$, $p = 0.19$)

Paired gestured/ungestured syllables



Takeaway points

No significant difference in duration between syllables in different tokens of the same word with/without gestures

Supports claim that gestures gravitate to long syllables rather than lengthening them

Discussion & Conclusion

Discussion

Consistent with past accounts, f0 and duration do not covary

- Except in word-final position where both are correlates of prosodic boundary strength

Syllables undergo word-final lengthening

Discussion

Gestures are attracted to **heavy syllables** and **longer syllables**

Pitch, intensity, and position in word did not predict gesture location

Consistent with account where stress in Uyghur is

- attracted to heavy syllables
- signaled primary by durational differences

Discussion

Provides further evidence that gesture timing is determined primarily by **rhythmic factors**, rather than pitch 'prominence'

Outstanding questions

Data from more speakers

Evaluate between different accounts of heavy syllable sensitivity?

Method for eliciting better stress judgments?

Future directions

- Look at more data in Uyghur
  - Targeted elicitation of particle constructions
  - Look at pragmatic properties of the gestures we've already coded.
  - Controlled experimental investigation?
- Expand to other Turkic languages

كۆپ رەھمەت سىلەرگە

Thanks to Gulnisa Nazarova and Mustafa Aksu for permission to use the YouTube video and images. We would also like to thank the Harvard PhonLab for feedback.

Extra slides…

‘PROMINENCE’ AND CO-SPEECH GESTURE

US English: Probabilistic relationship between **pitch accent type** and **gesture presence**

Im & Baumann 2020

‘PROMINENCE’ AND CO-SPEECH GESTURE

US English: Probabilistic relationship between **pitch accent type** and **gesture presence**

Gesture Less Likely

Gesture More Likely

Im & Baumann 2020

‘PROMINENCE’ AND CO-SPEECH GESTURE

US English: Probabilistic relationship between **pitch accent type** and **gesture presence**

Why?

Im & Baumann 2020

'PROMINENCE' AND CO-SPEECH GESTURE

US English: Probabilistic relationship between **pitch accent type** and **gesture presence**

Why?

Associated w/ Given Information **Associated Discourse-New/Unused Information**

L* !H* H* L-H*

Gesture Less Likely Gesture More Likely

Pierrehumbert & Hirschberg 1990; Baumann & Riester 2013, Im & Baumann 2020

‘PROMINENCE’ AND CO-SPEECH GESTURE

US English: Probabilistic relationship between **pitch accent type** and **gesture presence**

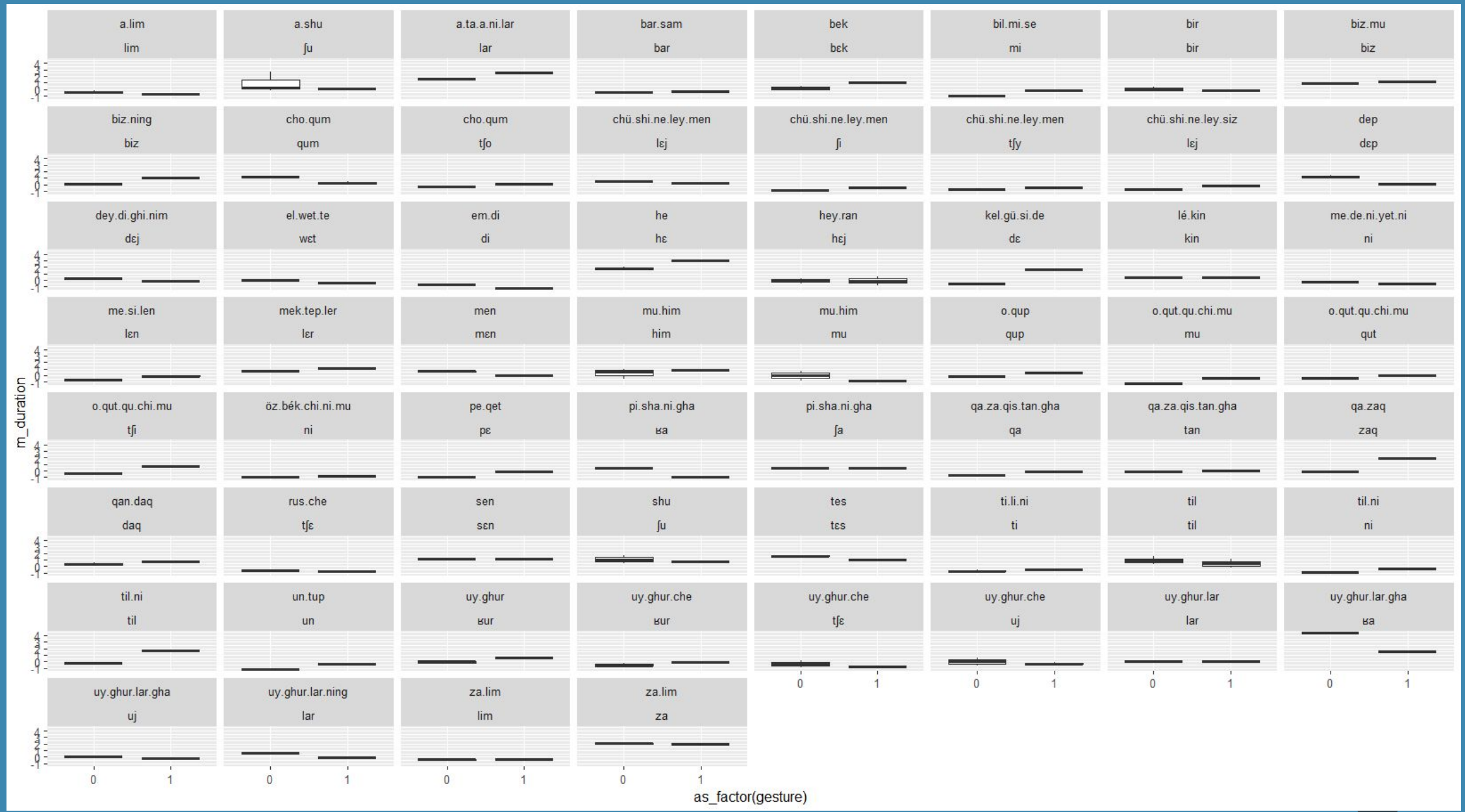
Why?

Less perceptually prominent **More perceptually prominent**

L* !H* H* L-H*

Gesture Less Likely Gesture More Likely

Baumann & Roehr 2015


a.lim
lim
a.shu
ʃu
a.ta.a.ni.lar
lar
bar.sam
bar
bɛk
bɛk
bil.mi.se
mi
bir
bir
biz.mu
biz
biz.ning
biz
cho.qum
qum
cho.qum
tʃo
chü.shi.ne.ley.men
lɛj
chü.shi.ne.ley.men
ʃi
chü.shi.ne.ley.men
tʃy
chü.shi.ne.ley.siz
lɛj
dɛp
dɛp
dey.di.ghi.nim
dɛj
el.wet.te
wɛt
em.di
di
he
hɛ
hey.ran
hɛj
kel.gü.si.de
dɛ
lé.kin
kin
me.de.ni.yet.ni
ni
me.si.len
lɛn
mek.tep.ler
lɛr
men
mɛn
mu.him
him
mu.him
mu
o.qup
qup
o.qut.qu.chi.mu
mu
o.qut.qu.chi.mu
qut
o.qut.qu.chi.mu
tʃi
öz.bék.chi.ni.mu
ni
pe.qet
pɛ
pi.sha.ni.gha
ʁa
pi.sha.ni.gha
ʃa
qa.za.qis.tan.gha
qa
qa.za.qis.tan.gha
tan
qa.zaq
zaq
qan.daq
daq
rus.che
tʃɛ
sen
sɛn
shu
ʃu
tes
tɛs
ti.li.ni
ti
til
til
til.ni
ni
til.ni
til
un.tup
un
uy.ghur
ʁur
uy.ghur.che
ʁur
uy.ghur.che
tʃɛ
uy.ghur.che
uj
uy.ghur.lar
lar
uy.ghur.lar.gha
ʁa
uy.ghur.lar.gha
uj
uy.ghur.lar.ning
lar
za.lim
lim
za.lim
za
m_duration
4
3
2
1
0
-1
0
1
as_factor(gesture)