Modeling timing in speech sounds

Christopher Geissler
Department of Linguistics
Carleton College

April 15, 2024
Slides available on cageissler.github.io/resources

Roadmap

- Phonology, phonetics, and time
- Types of evidence
  - Gestures & coupling: Tibetan C-V timing
  - Levels of timing: Sámi consonant lengths
    German-English accommodation
  - Fitting articulatory trajectories
- Conclusion

Roadmap

- **Phonology, phonetics, and time**
- Types of evidence
  - Gestures & coupling: Tibetan C-V timing
  - Levels of timing: Sámi consonant lengths
    German-English accommodation
  - Fitting articulatory trajectories
- Conclusion

Situating temporal articulation

- Linguistics: “What do you know when you know a language?”
  - Phonetics: how sounds are produced & perceived
  - Phonology: how sounds vary by context
- Today’s topic: coordination of speech articulators in time

Phonology: basic

Categorical behavior

- In German, voiced consonants are voiceless when they occur at the end of words (but not elsewhere):
  - *Mau̲s̲* ‘mouse’ [maʊ̲s̲], but plural *Mäu̲s̲e* [mɔʏ̲z̲ə]
  - *Ra̲d̲* ‘wheel’ [ʀɑ̲t̲], but plural *Rä̲d̲er* [ʀɛ̲d̲ɐ]
  - compare:
    Ra̲t̲ ‘council’ [ʀɑ̲t̲], but plural *Rä̲t̲e* [ʀɛ̲t̲ə]

Phonology: basic

Categorical behavior

- In German, voiced consonants are voiceless when they occur at the end of words (but not elsewhere):
  - *Mau̲s̲* ‘mouse’ [maʊ̲s̲], but plural *Mäu̲s̲e* [mɔʏ̲z̲ə]
  - *Ra̲d̲* ‘wheel’ [ʀɑ̲t̲], but plural *Rä̲d̲er* [ʀɛ̲d̲ɐ]
  - compare:
    Ra̲t̲ ‘council’ [ʀɑ̲t̲], but plural *Rä̲t̲e* [ʀɛ̲t̲ə]

Linguists are really good at this

Phonology: advanced

Probabilistic behavior

- In English, t/d at the end of a word sometimes isn't there
  - *rift* = [ɹɪft̚] or [ɹɪf_]; *build* = [bɹɬd] or [bɹɬ]
  - More likely among some groups
  - More likely in some social contexts
  - More likely around some sounds
  - More likely in *mist* than in *missed*

Phonology: advanced

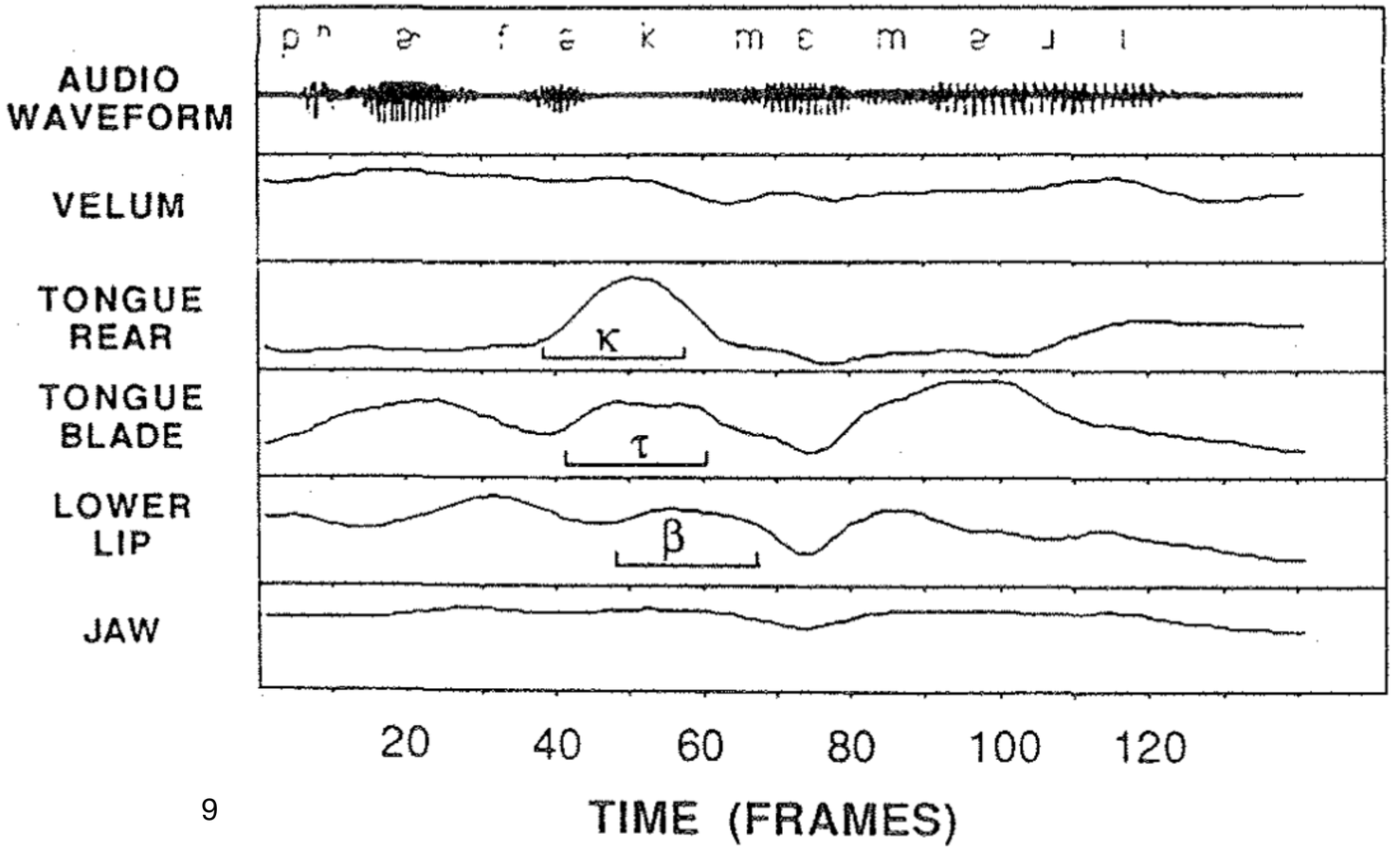
Probabilistic behavior

- In English, t/d at the end of a word sometimes isn't there
  - *rift* = [ɹɪft] or [ɹɪf_]; *build* = [bɹɬd] or [bɹɬ]
  - More likely among some groups
  - More likely in some social contexts
  - More likely around some sounds
  - More likely in *mist* than in *missed*

Linguists get excited about this

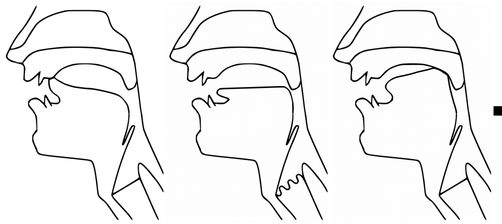
…uh-oh

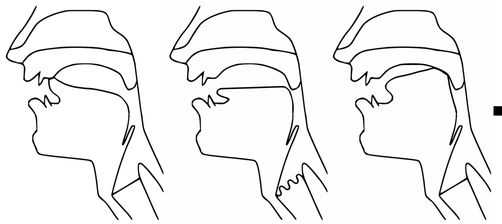
- *Perfect memory*
- At least some “deleted” t’s/d’s are visible in articulation, but not in acoustics
  - (Actually it’s most)



Midsagittal sections

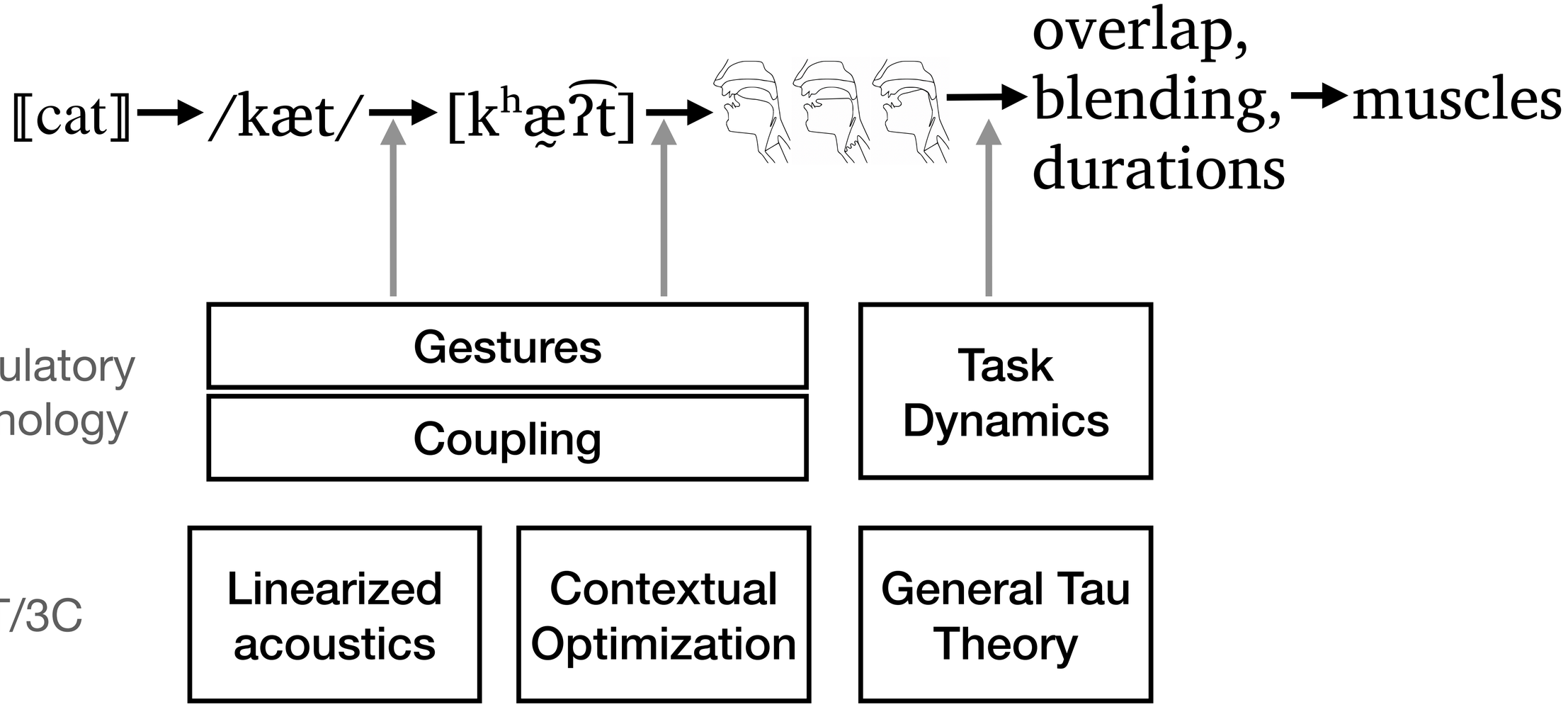
(Browman & Goldstein 1988, Purse 2019)

⟦cat⟧ → /kæt/ → [kʰæ̃ʔt͡] →  → overlap, blending, durations → muscles

⟦cat⟧ → /kæt/ → [kʰæ̃ʔt͡] →  → overlap, blending, durations → muscles

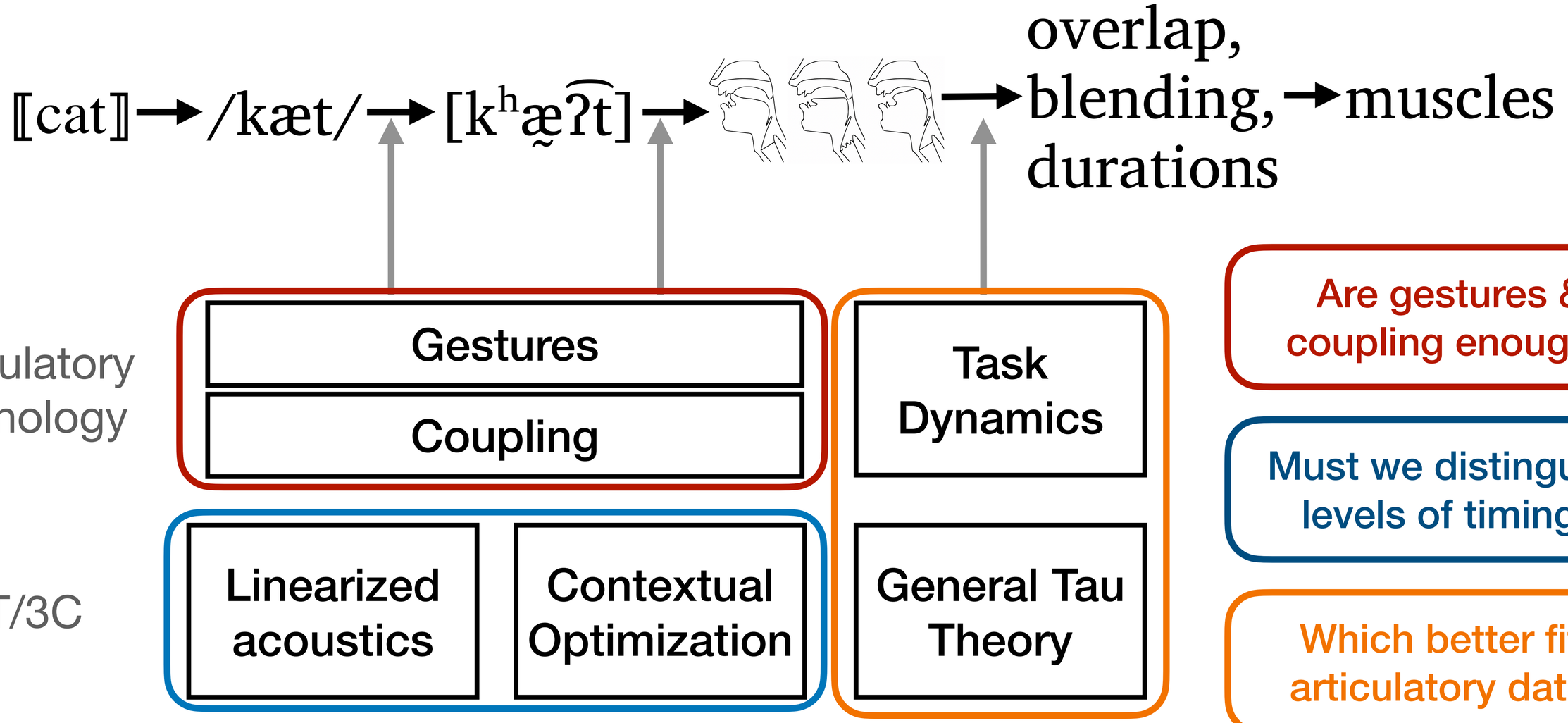
When does temporal information enter?

Representational units



AP: Browman & Goldstein (1986) et seq.; TD: Saltzman & Munhall (1989)
XT/3C: Turk & Shattuck-Hufnagel (2020); Tau: Lee (1998)

Representational units

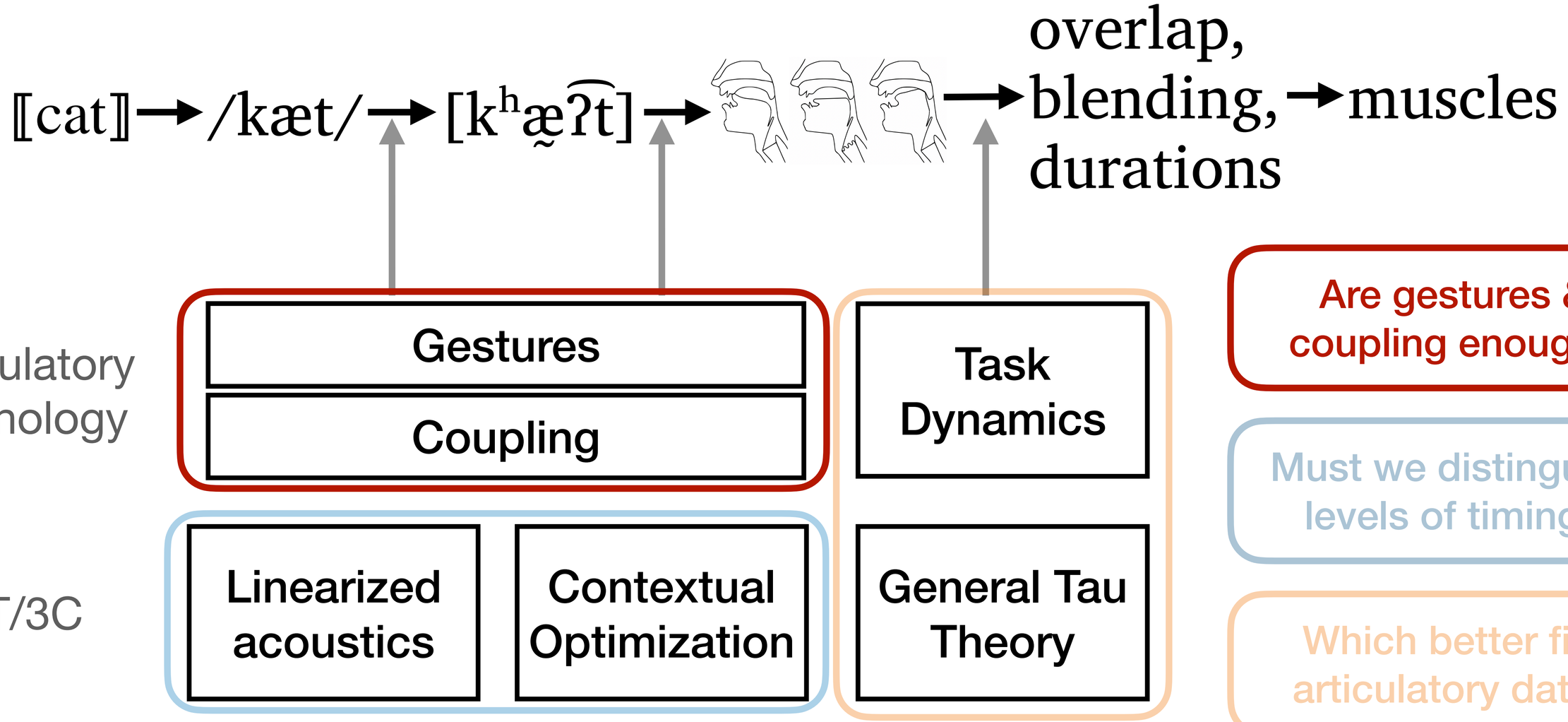


AP: Browman & Goldstein (1986) et seq.; TD: Saltzman & Munhall (1989)
XT/3C: Turk & Shattuck-Hufnagel (2020); Tau: Lee (1998)

Roadmap

- Phonology, phonetics, and time
- Types of evidence
  - **Gestures & coupling: Tibetan C-V timing**
  - Levels of timing: Sámi consonant lengths
    German-English accommodation
  - Fitting articulatory trajectories
- Conclusion

Representational units



AP: Browman & Goldstein (1986) et seq.; TD: Saltzman & Munhall (1989)
XT/3C: Turk & Shattuck-Hufnagel (2020); Tau: Lee (1998)

Bimanual tapping interlude

Oscillators

- Synchronization in non-speech and speech movements:
  - “pa… pa… pa… pa.pa[…]pa.pa.pa.pa”
  - “ap… ap… ap… ap.ap.[…]pa.pa.pa.pa”
- Tapping: “in-phase” more stable than “anti-phase”
  (both more stable than any other phasing)
  … in speech too?

CV vs. VC syllables

in-phase

[pa]	
LIPS	Labial closure
TONGUE TIP	
TONGUE BODY	pharyngeal wide

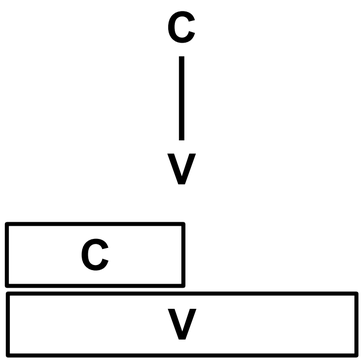
anti-phase

[ap]	
LIPS	labial closure
TONGUE TIP	
TONGUE BODY	pharyngeal wide

CV vs. VC syllables

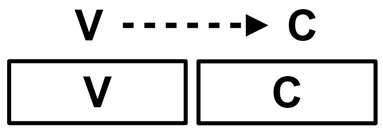
in-phase

[pa]	
LIPS	Labial closure
TONGUE TIP	
TONGUE BODY	pharyngeal wide



anti-phase

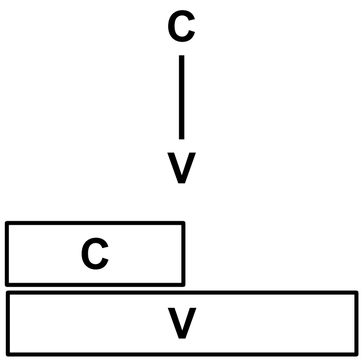
[ap]	
LIPS	labial closure
TONGUE TIP	
TONGUE BODY	pharyngeal wide



CV vs. VC syllables

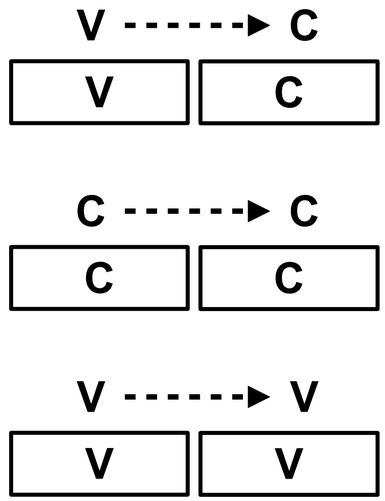
in-phase

[pa]	
LIPS	Labial closure
TONGUE TIP	
TONGUE BODY	pharyngeal wide



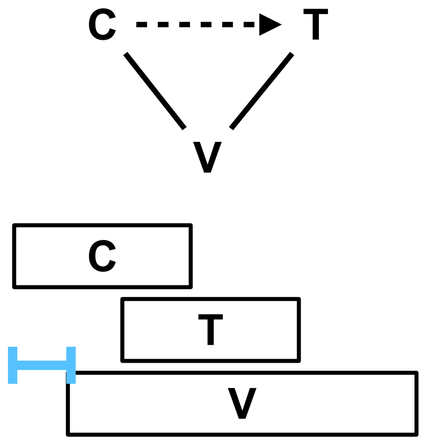
anti-phase

[ap]	
LIPS	labial closure
TONGUE TIP	
TONGUE BODY	pharyngeal wide

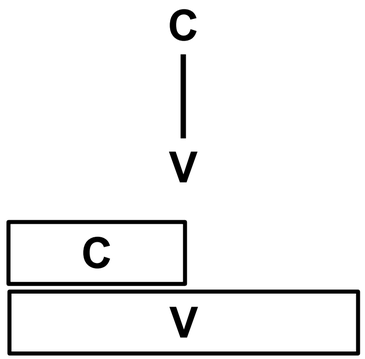


Predictions

- If there is a tone gesture in a syllable:
  - C-V timing like in clusters:
    C-V lag positive, ~50ms



- If there is no tone in that syllable:
  - Simultaneous C & V:
    C-V lag ~0ms



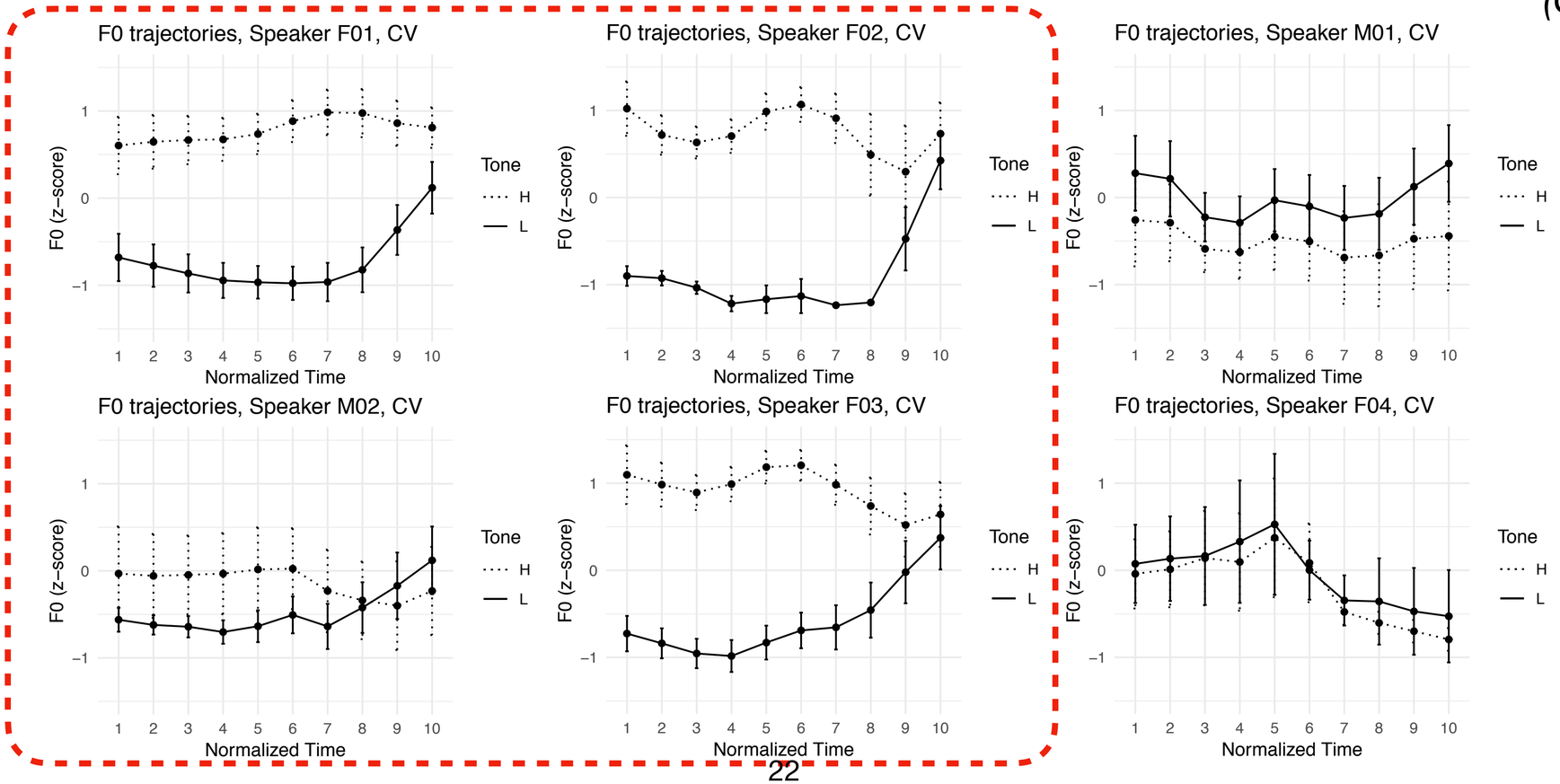
The perfect test case?

A language where some speakers produce tone and others don't

(Geissler 2019, 2021)

- 4 speakers produce a tone contrast, two do not (images: /mV/)

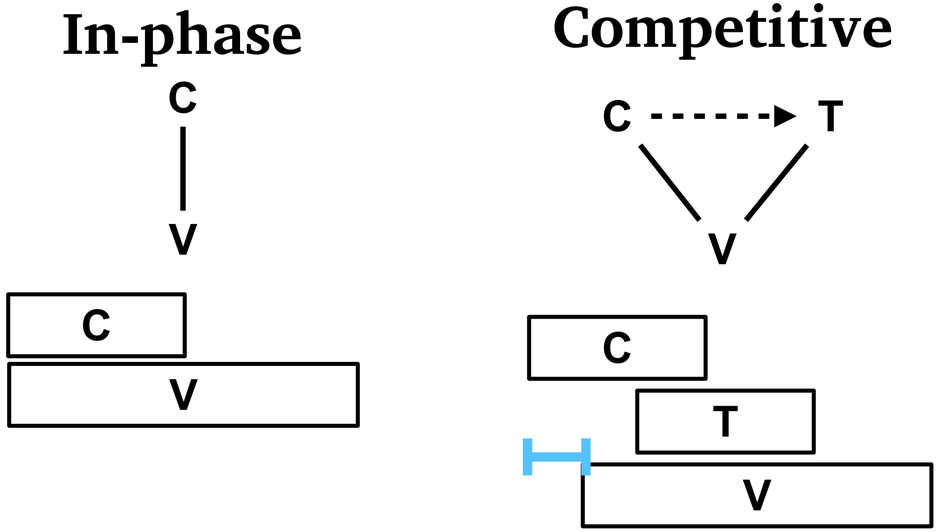
(Geissler et al. 2021)

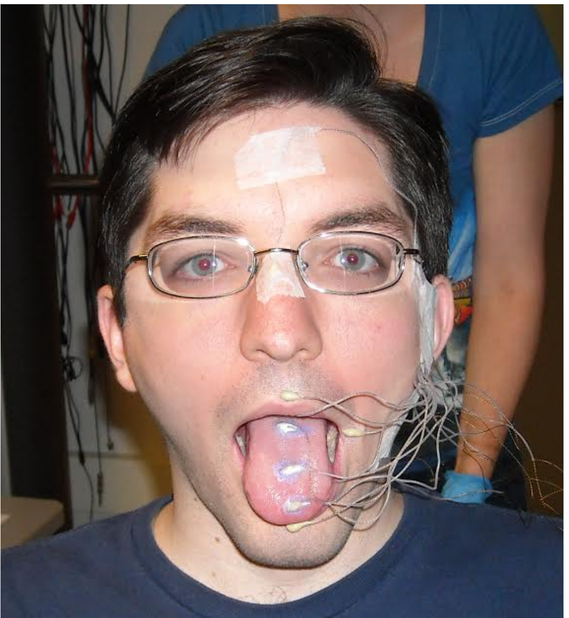


EMA study

articulatory trajectories

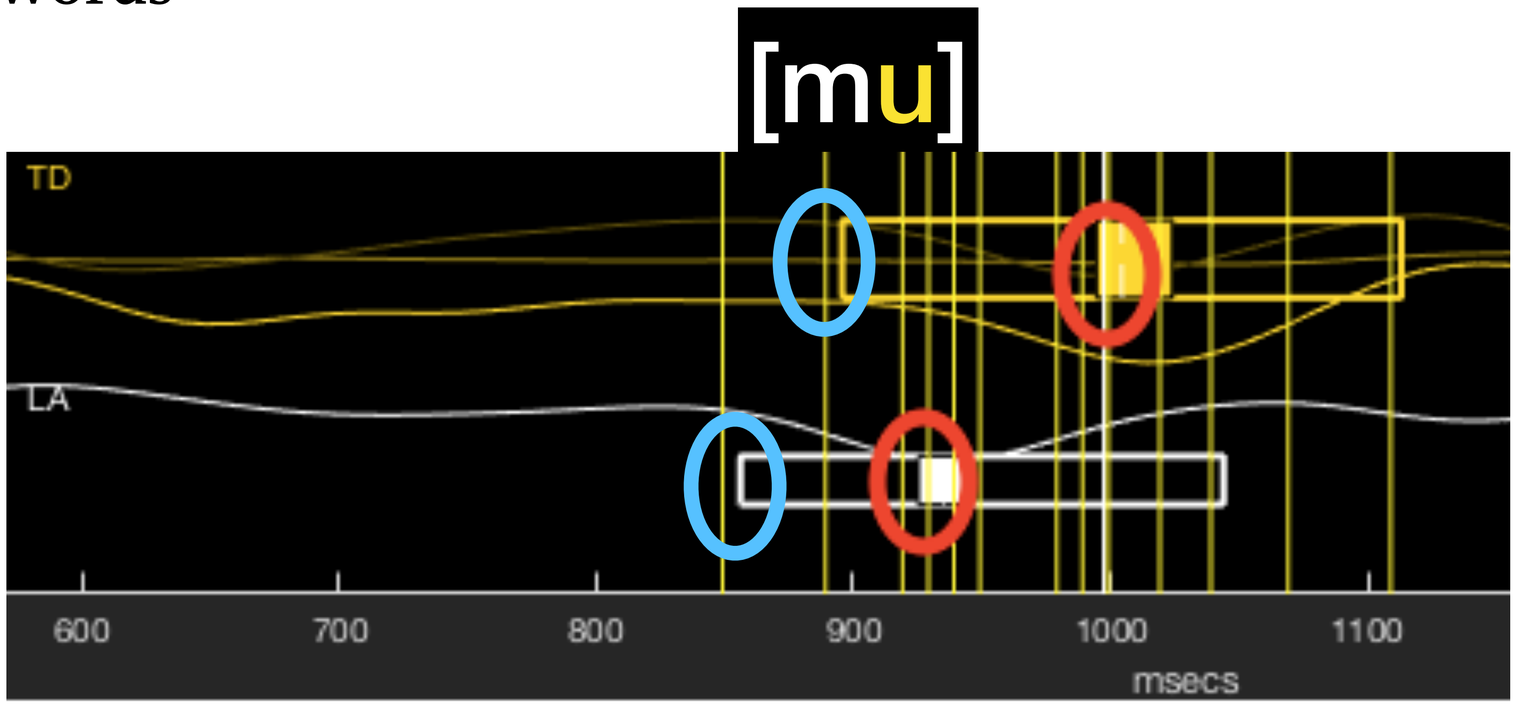
- [p p^h m]: distance between lip sensors
- [i]→[u o a]: tongue dorsum retraction
- H, L tones; 1- and 2-syllable words
- C-V lag as diagnostic of tone





Tongue Dorsum: front ↓ back

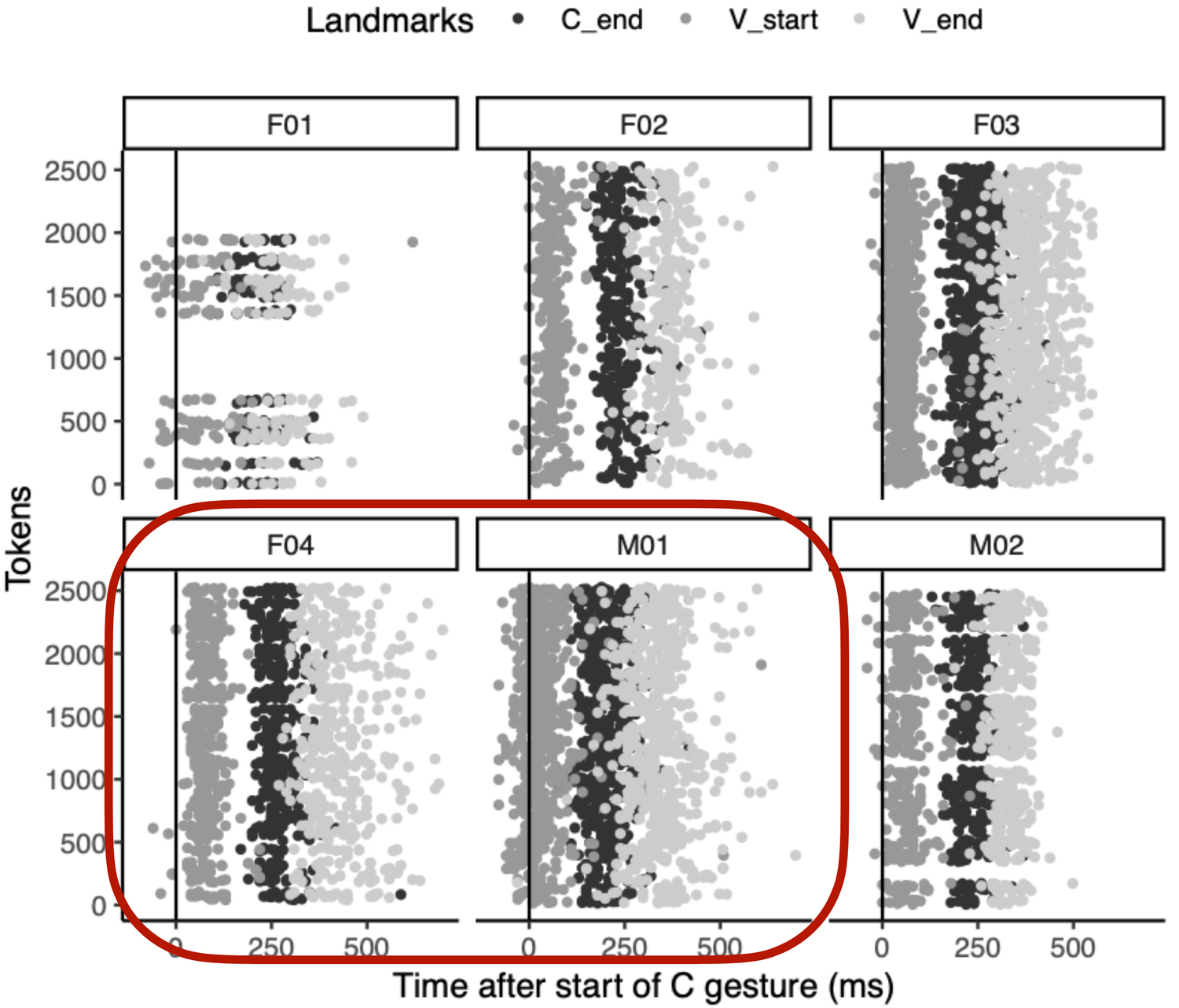
Lip Aperture: open ↓ closed



(Data: Zhang, Geissler, & Shaw 2019)
(Mview software: Tiede 2005)

Results: C-V lag

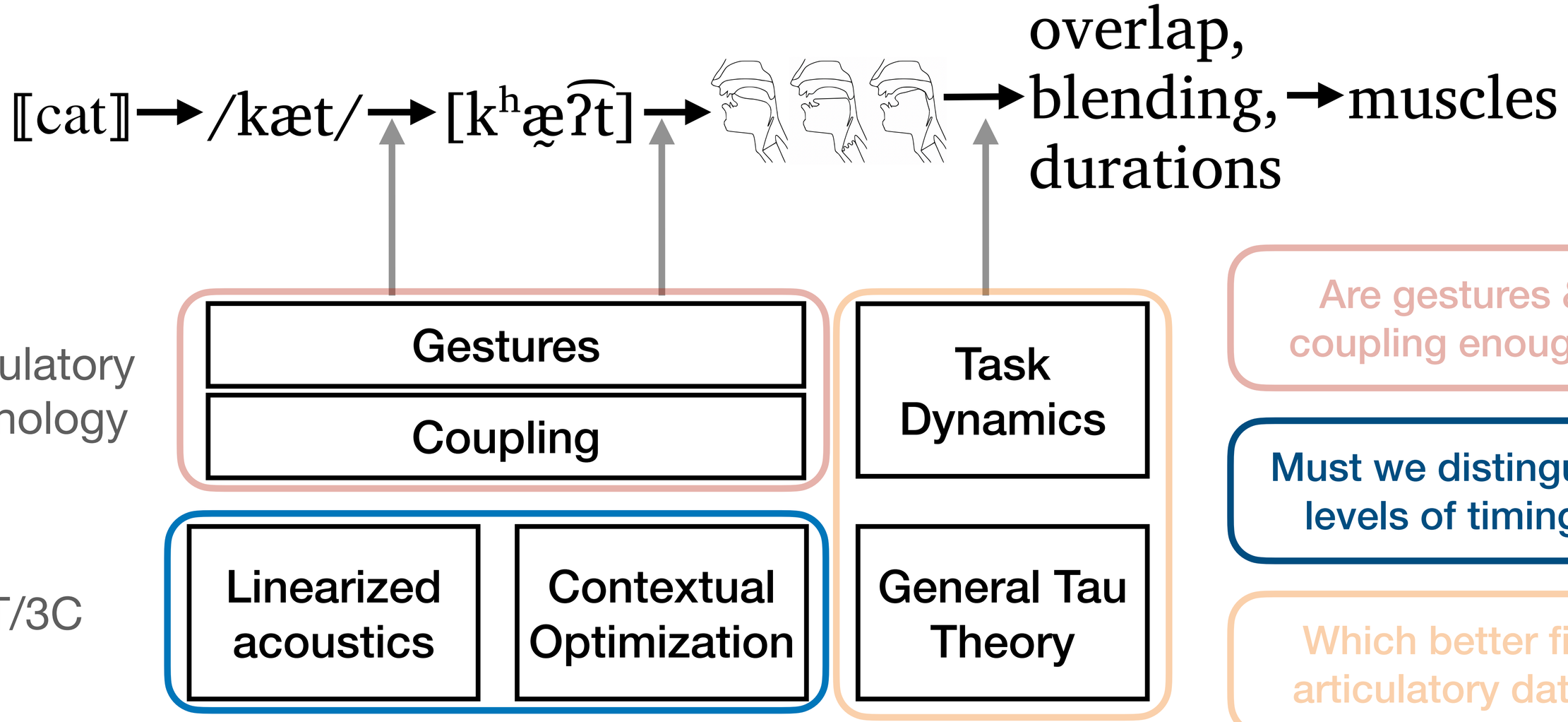
- There is a positive C-V lag… for speakers with ***and*** without the tone contrast (and in both tones)
- Competitive Coupling has no explanation for the 50ms lag



Roadmap

- Phonology, phonetics, and time
- Types of evidence
  - Gestures & coupling: Tibetan C-V timing
  - **Levels of timing: Sámi consonant lengths**
    German-English accommodation
  - Fitting articulatory trajectories
- Conclusion

Representational units



AP: Browman & Goldstein (1986) et seq.; TD: Saltzman & Munhall (1989)
XT/3C: Turk & Shattuck-Hufnagel (2020); Tau: Lee (1998)

Northern Sámi quantity distinctions

Posson (2024); Posson & Geissler (2024)

- 2 vowel lengths
- 3 (!!!) phonological consonant lengths:
  - Q1: [vi͡esu] ‘house (acc sg)’
  - Q2(~Q1): [vi͡esːu] ‘house (nom sg)’
  - Q2(~Q3): [rū͡oːsːa] ‘cross (acc sg)’
  - Q3: [rū͡osːːa] ‘cross (nom sg)’
- Notice the [ū͡o]~[ū͡oː]?

Northern Sámi quantity distinctions

- 2 vowel lengths
- 3 (!!!) phonological consonant lengths:
  - Q1: [vi͡esu] ‘house (acc sg)’
  - Q2(~Q1): [vi͡esːu] ‘house (nom sg)’
  - Q2(~Q3): [rū͡oːsːa] ‘cross (acc sg)’
  - Q3: [rū͡osːːa] ‘cross (nom sg)’
- Notice the [ū͡o]~[ū͡oː]? [nom sg] has a floating mora

Confirm phonetically

- Predict:

$$Q1 < Q2(\sim Q1) = Q2(\sim Q3) < Q3$$

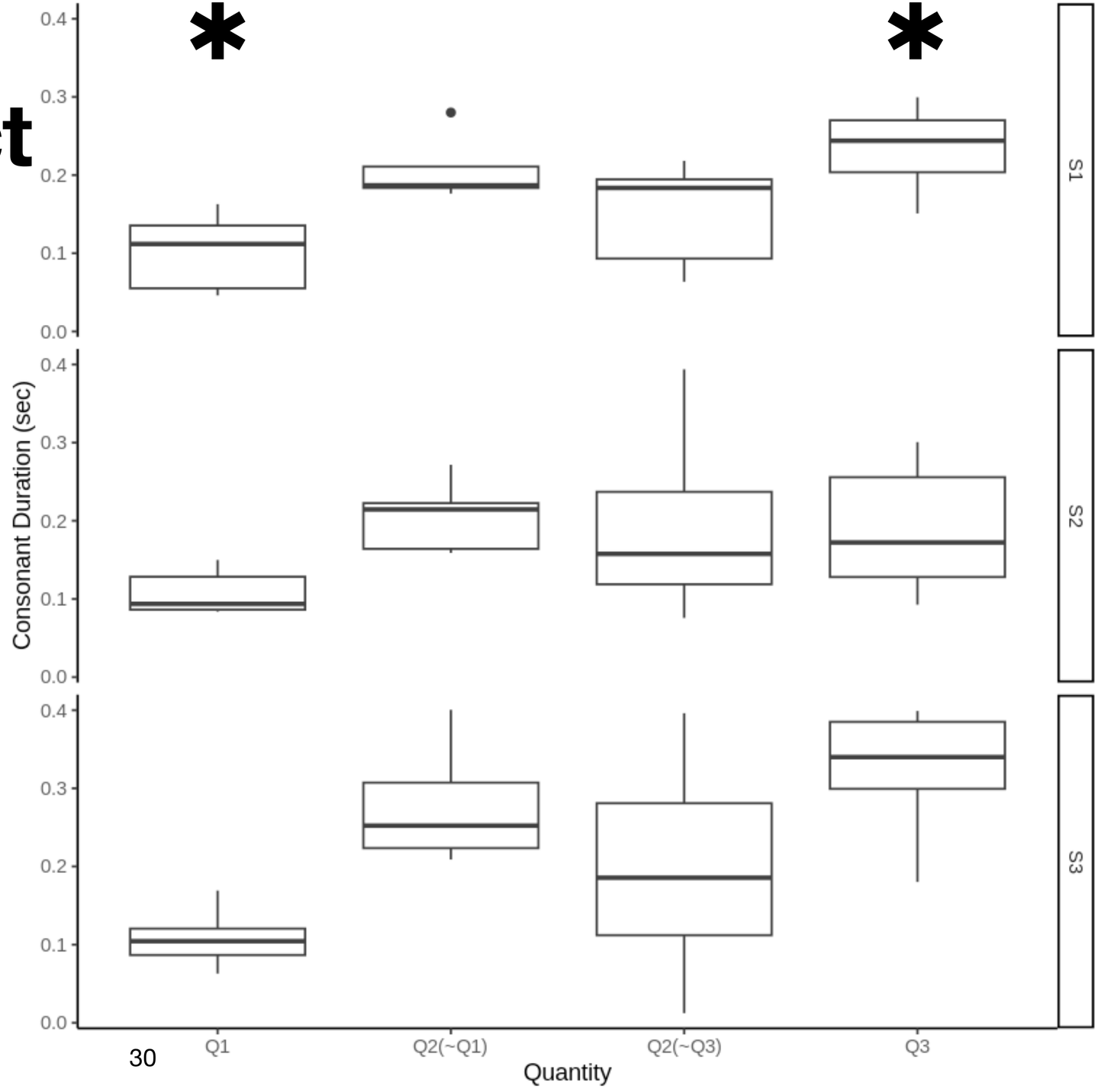
shortest——————————————————————longest

Phonological effect

3 lengths? 3 lengths.

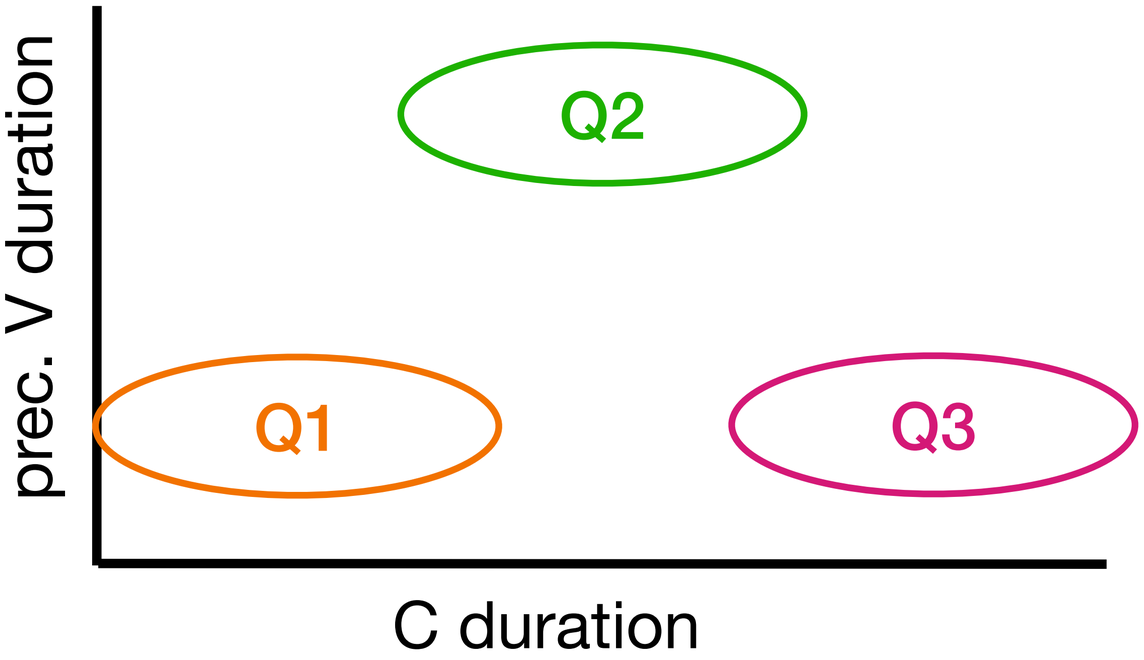
Posson (2024); Posson & Geissler (2024)

- Overall:
  Q3
  longer than
  Q2(~Q1) = Q2(~Q3)
  longer than
  Q1
- S2 might have only two lengths; insufficient data

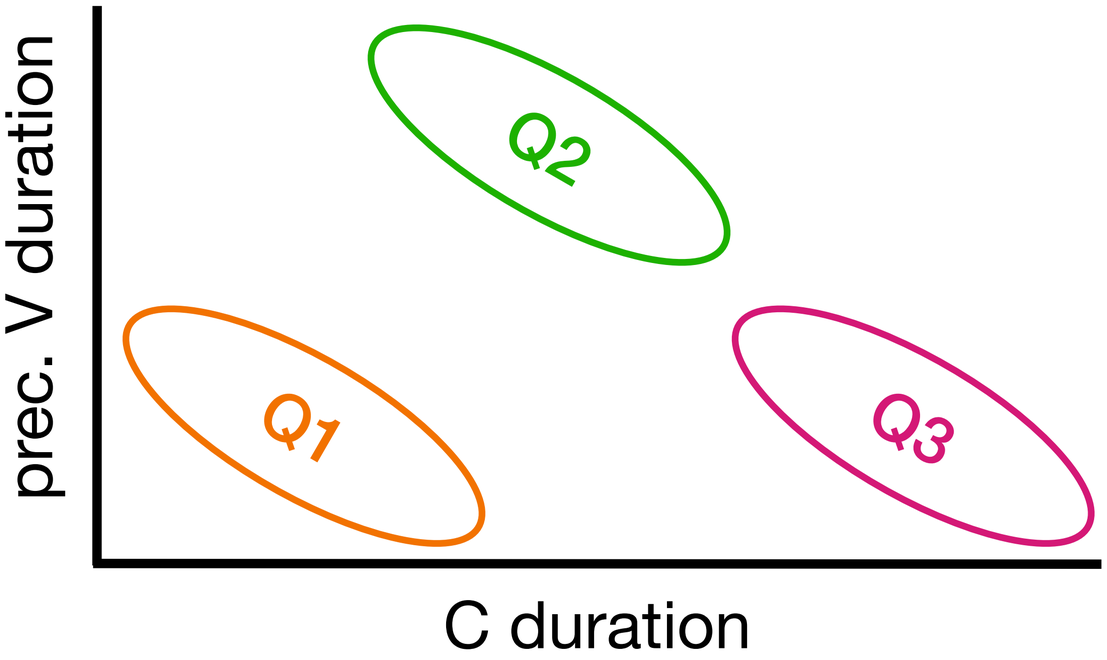


V.C relationship: phonological or phonetic?

- If phonological:



- If phonetic:

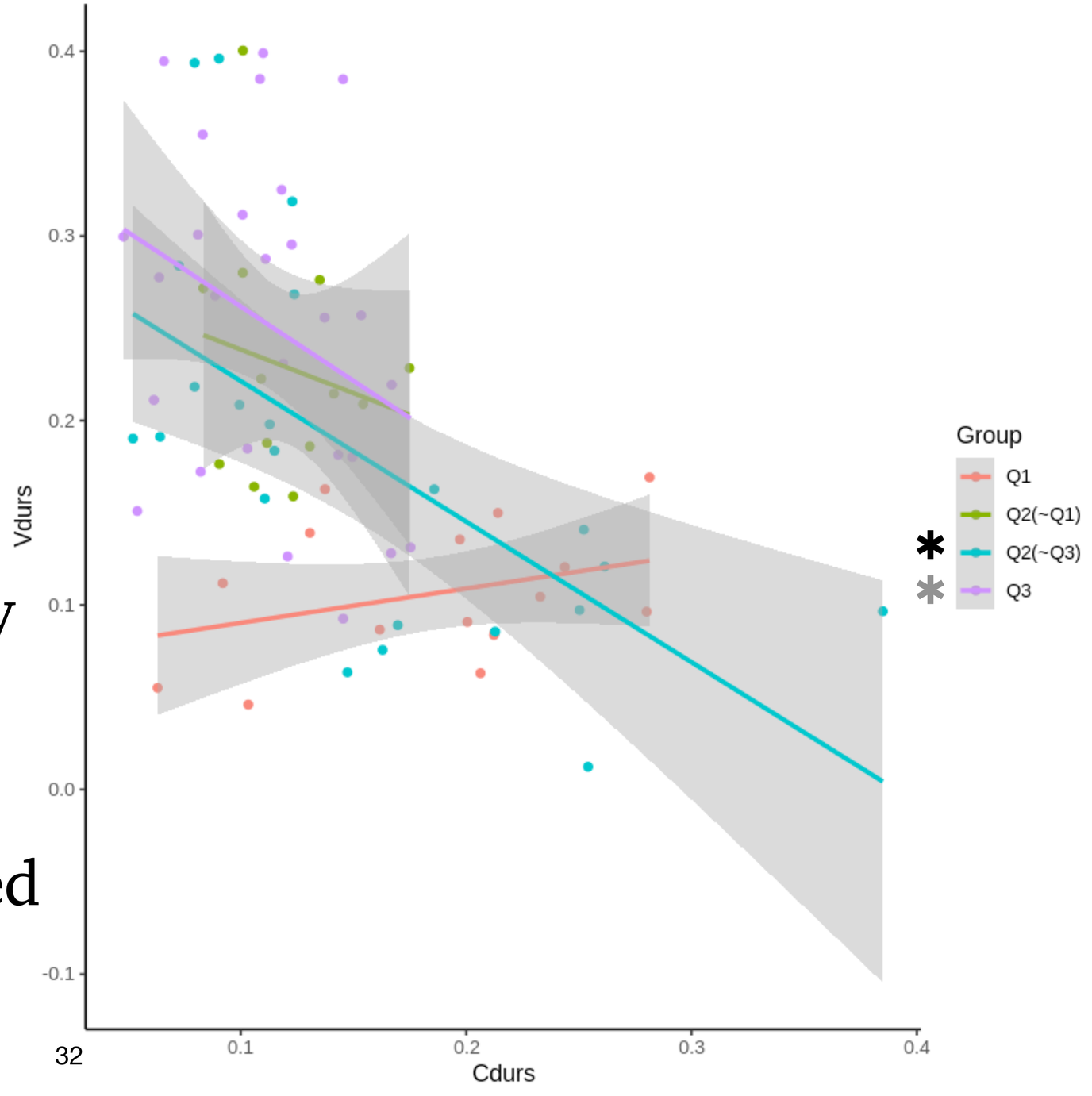


Phonetic effect

Inverse correlation

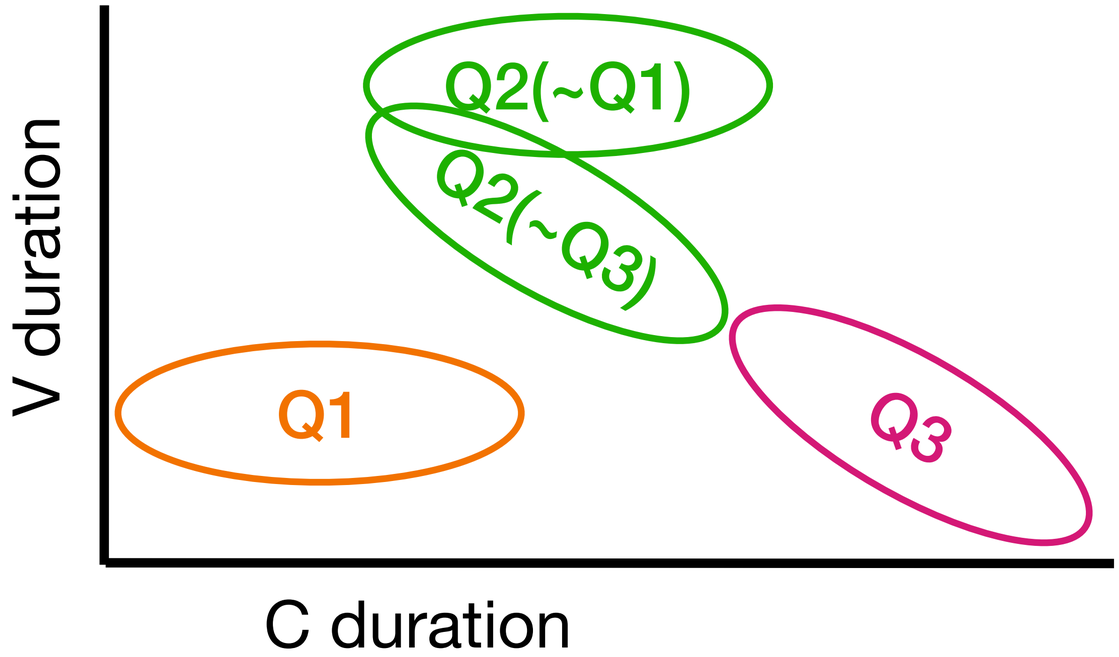
Posson (2024); Posson & Geissler (2024)

- Significant inverse relationships (V decrease when C increase) only in underlying Q3 Cs; driven by one speaker
- Trends in expected directions; more data needed



Phonological or phonetic?

- Phonological:
  - Q1
  - Q2(~Q1)
- Phonetic
  - Q3
  - Q2(~Q3)



Sámi summary

- For phonologically longest C's,
  longer C's → shorter preceding V's
  - *this is over and above the phonological effect*
- For phonologically shortest C's,
  no phonetic effect
  - *there is only the phonological effect*
- … need more data…

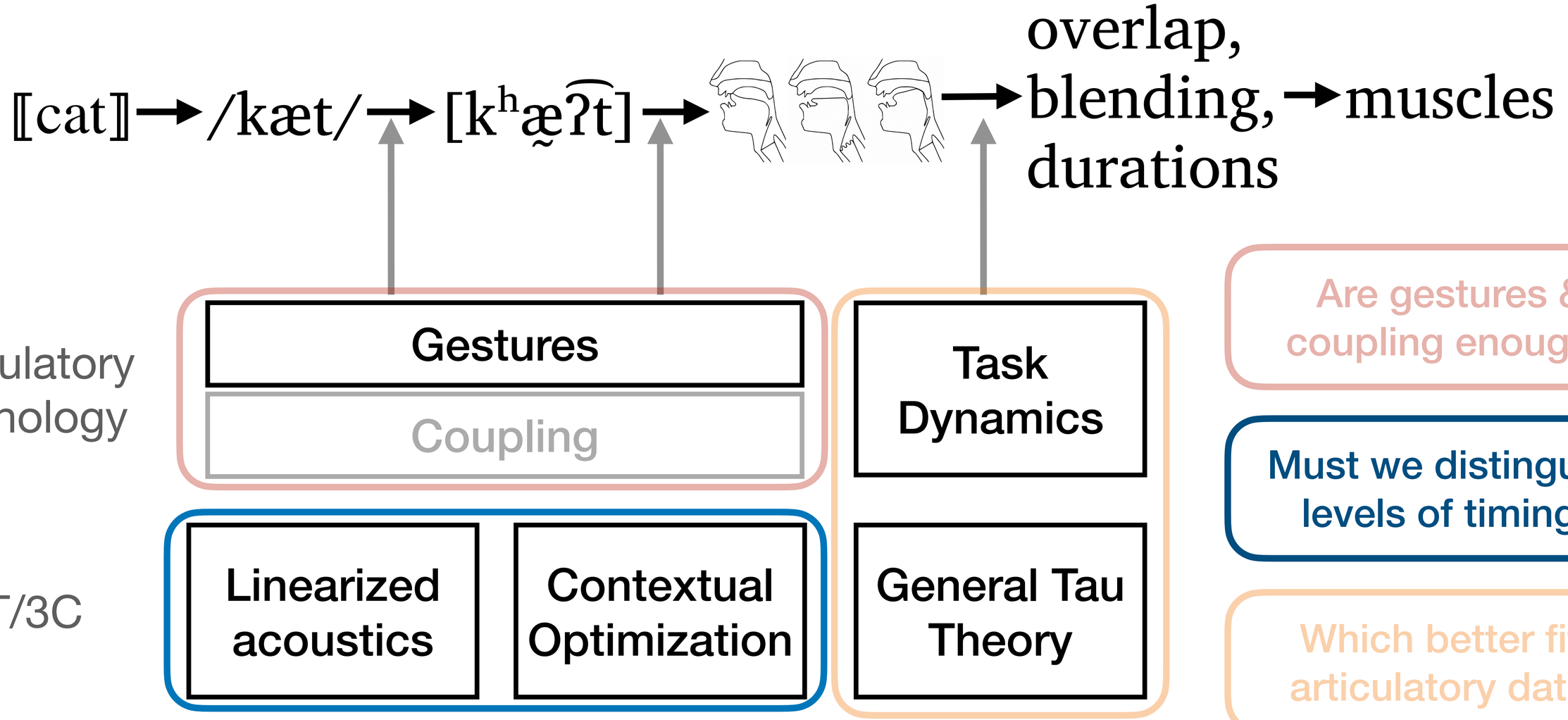
categorical *and* continuous timing

only categorical timing

Roadmap

- Phonology, phonetics, and time
- Types of evidence
  - Gestures & coupling: Tibetan C-V timing
  - **Levels of timing:** Sámi consonant lengths
    German-English accommodation
  - Fitting articulatory trajectories
- Conclusion

Representational units



AP: Browman & Goldstein (1986) et seq.; TD: Saltzman & Munhall (1989)
XT/3C: Turk & Shattuck-Hufnagel (2020); Tau: Lee (1998)

Passive L2 accommodation

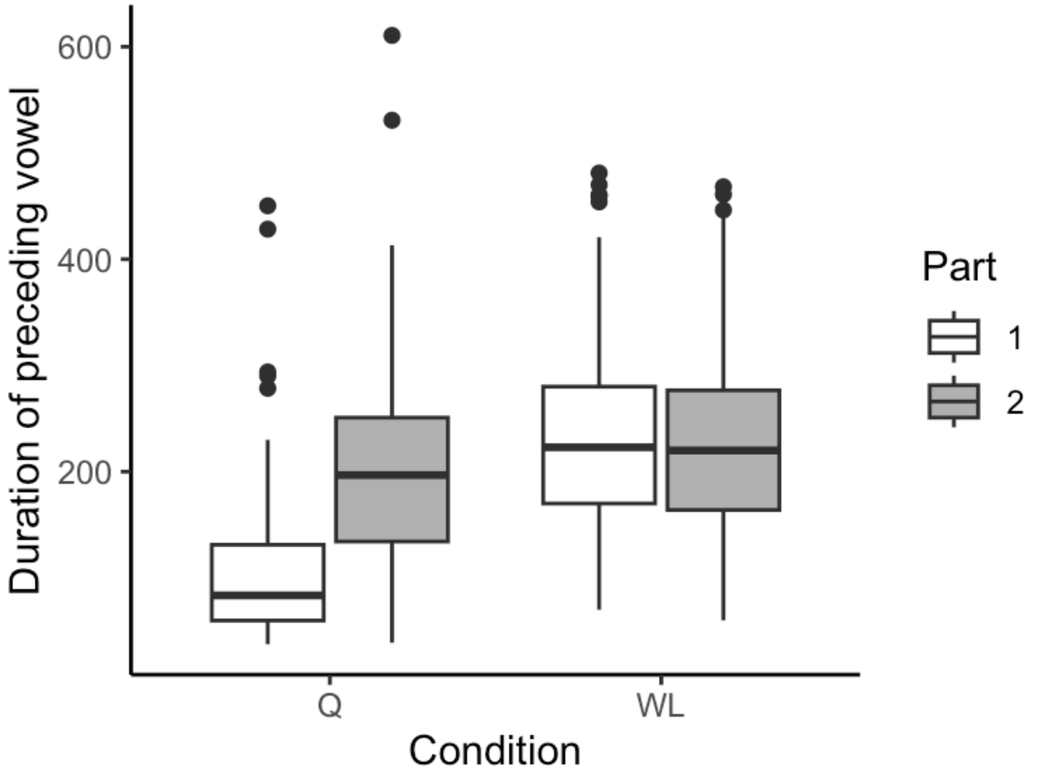
Weber (2023); Weber & Geissler (2023)

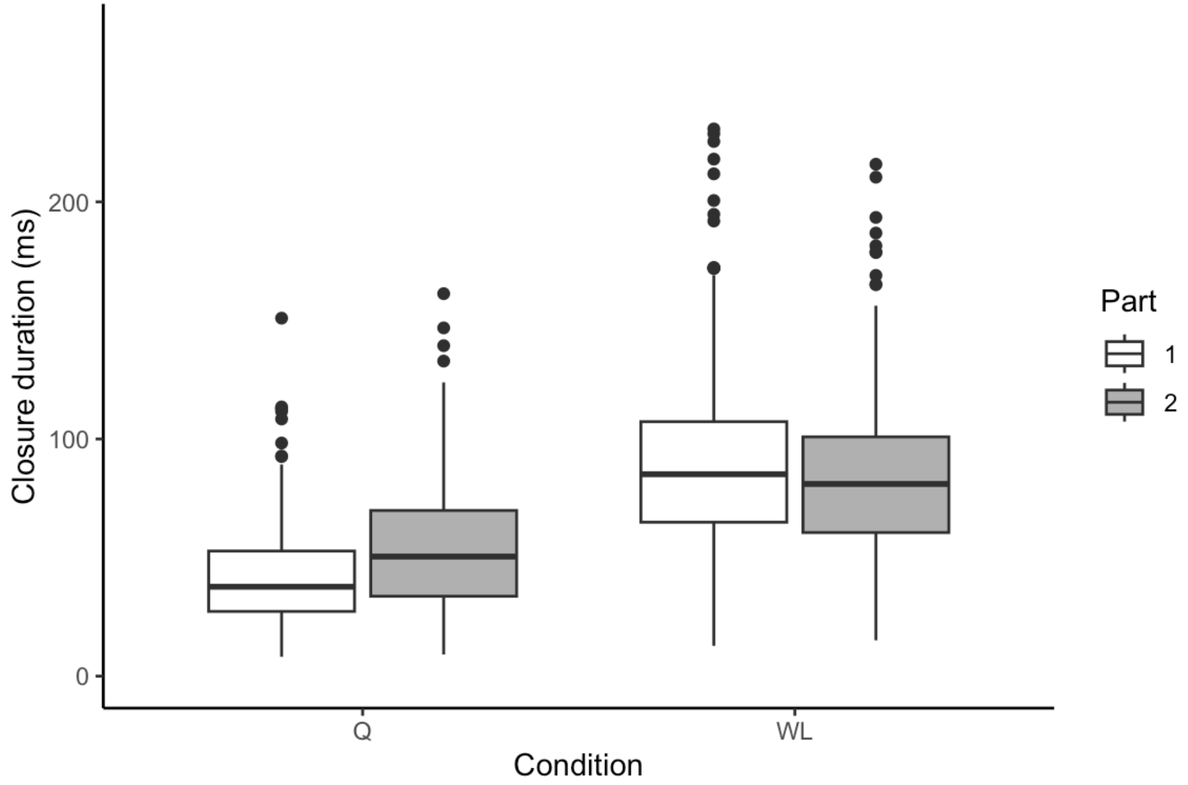
- L1: German; L2: English
- Record before and after watching one episode of *Friends*
- Do speakers shift to a more native-like pronunciation of:
  - TRAP vowel (German L1 speakers often merge with DRESS)
  - Word-final voicing
- Passive activity—no interlocutor—though still social aspect

Temporal accommodation

Weber (2023); Weber & Geissler (2023)

- Longer—more voiced-like—final consonants





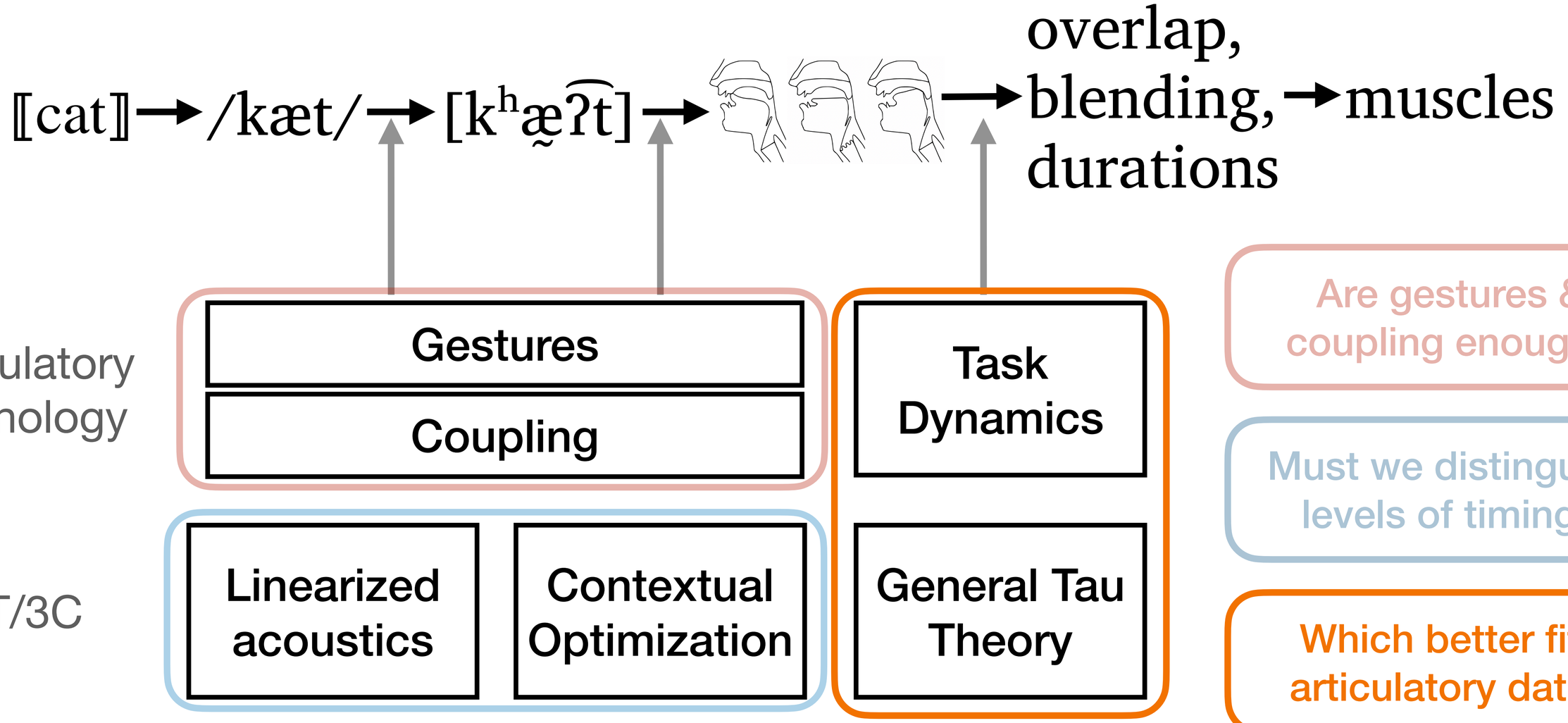
Passive L2 accommodation

- Passive activity—no interlocutor!
  - Still some social attitudes, etc.
- Temporal data mirrors vowel quality
- L2 speakers already had contrast in wordlist, but enhanced it in the questions

Roadmap

- Phonology, phonetics, and time
- Types of evidence
  - Gestures & coupling: Tibetan C-V timing
  - Levels of timing: Sámi consonant lengths
    German-English accommodation
  - **Fitting articulatory trajectories**
- Conclusion

Representational units



AP: Browman & Goldstein (1986) et seq.; TD: Saltzman & Munhall (1989)
XT/3C: Turk & Shattuck-Hufnagel (2020); Tau: Lee (1998)

Task Dynamics (Articulatory Phonology)

(Haken et al. 1985, Saltzman & Munhall 1989, Nam & Saltzman 2003)

- Model movement as critically-damped mass-spring oscillator
- Timing is *internal to the gesture* (sine waves are circles)

$$ma + bv + k(x - C) = 0$$

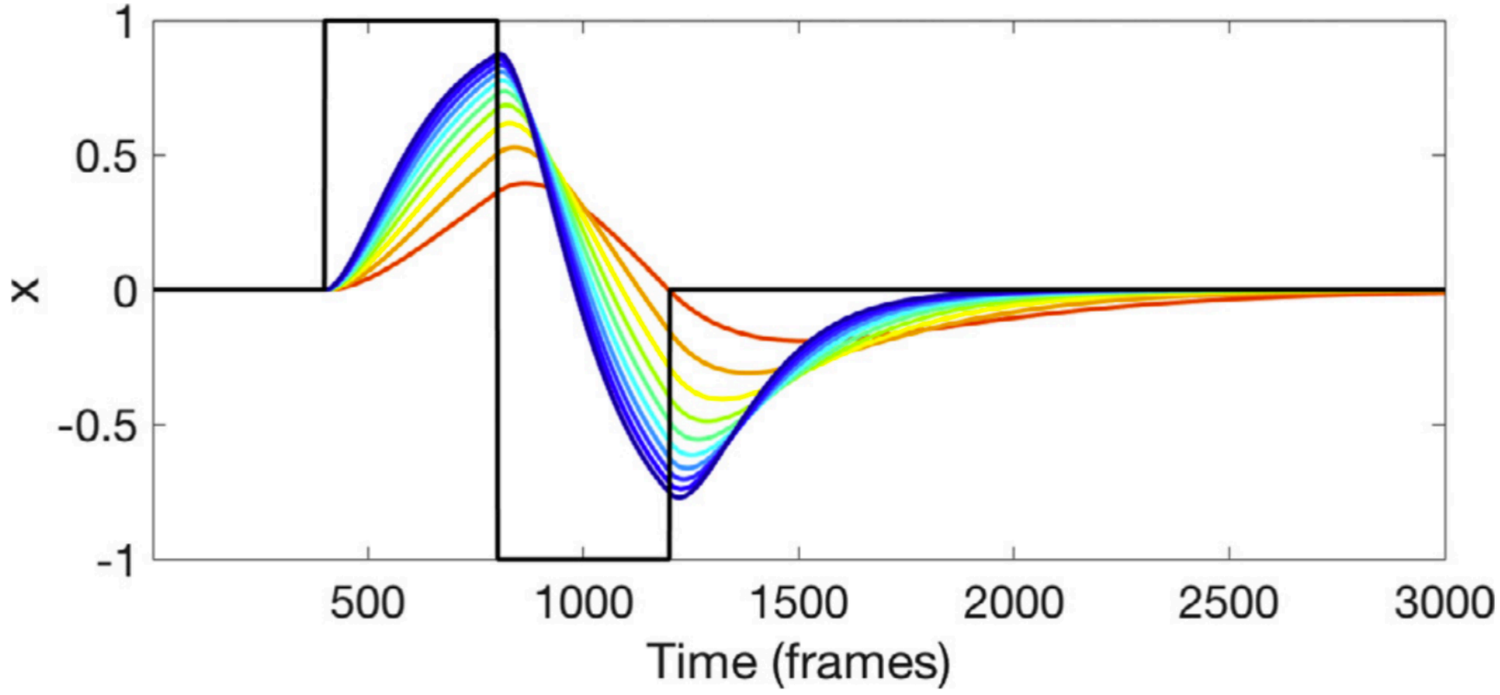
acceleration

velocity

position

stiffness

target



(Iskarous 2013)

General Tau Theory (XT/3C)

(Lee & Reddish 1981, Turk & Shattuck-Hufnagel 2020)

- Model kinematics as gap-closing function
- Time only in regular, system-external time

$$X(t) = X_0 \left(1 - \frac{t^2}{T^2}\right)^{\frac{1}{\kappa}}$$

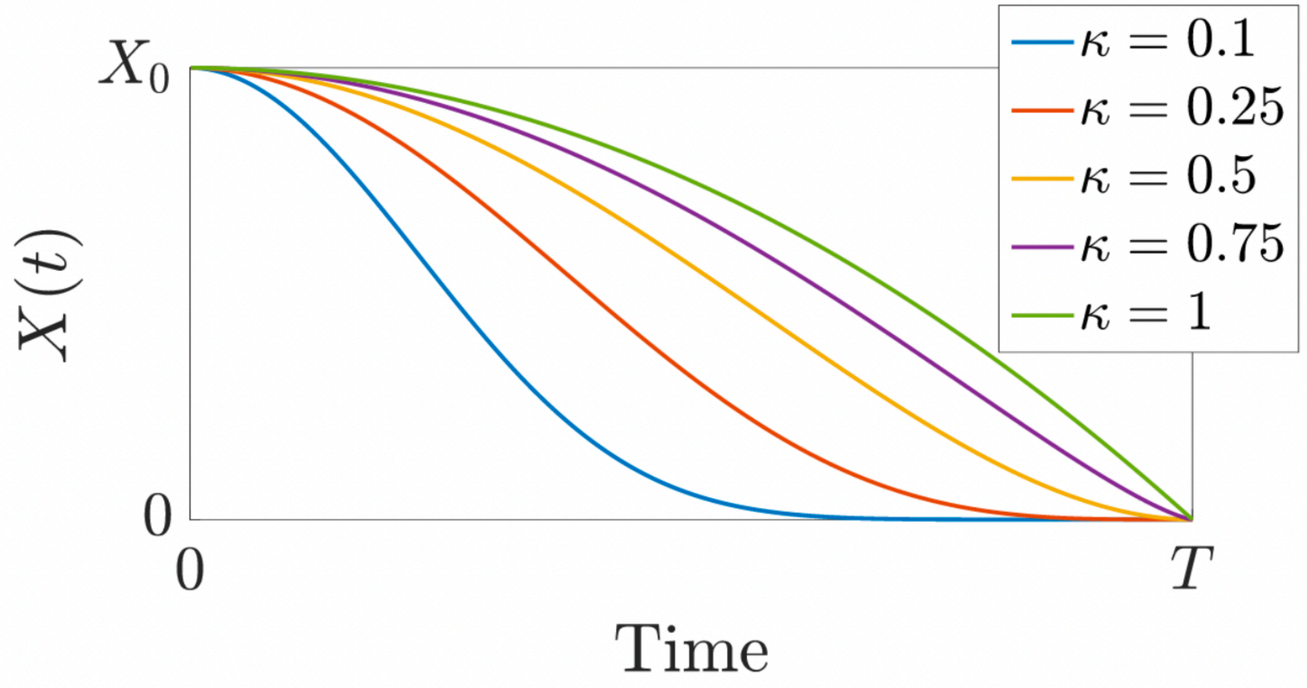
position ← $X(t)$

time ← t

start position → X_0

end time → T

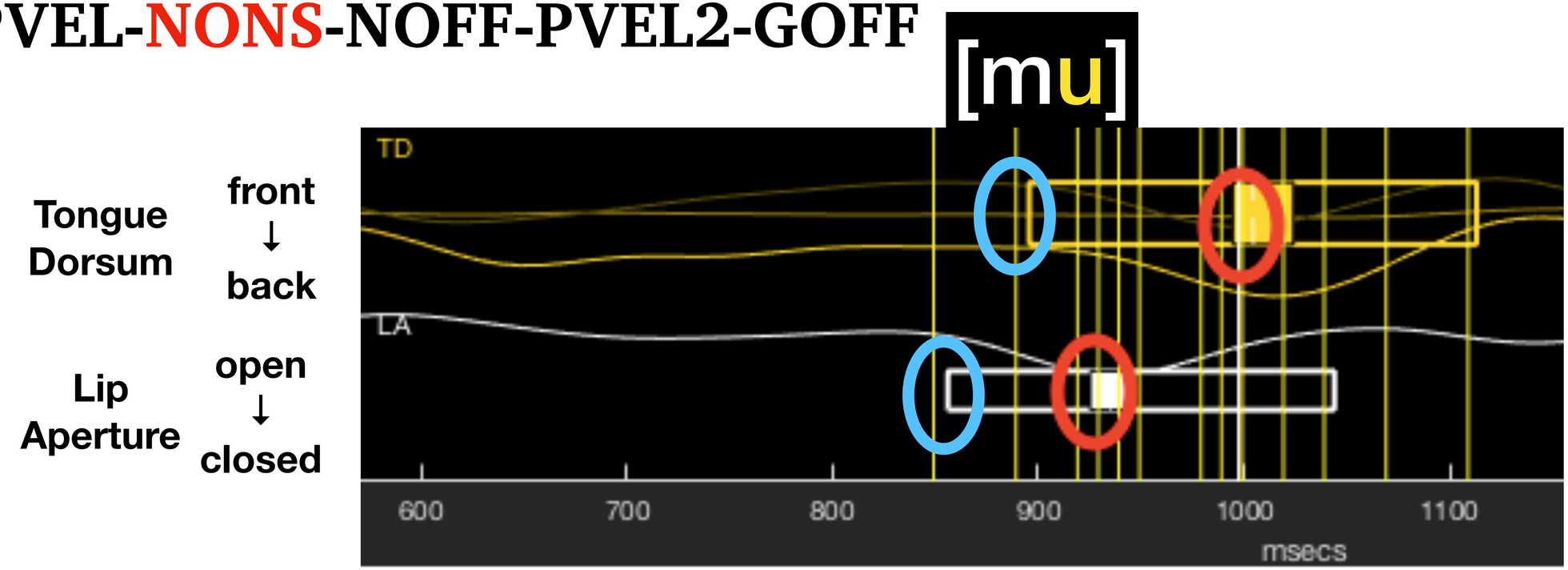
constant ← κ



(Elie et al. 2023)

Which fits data better?

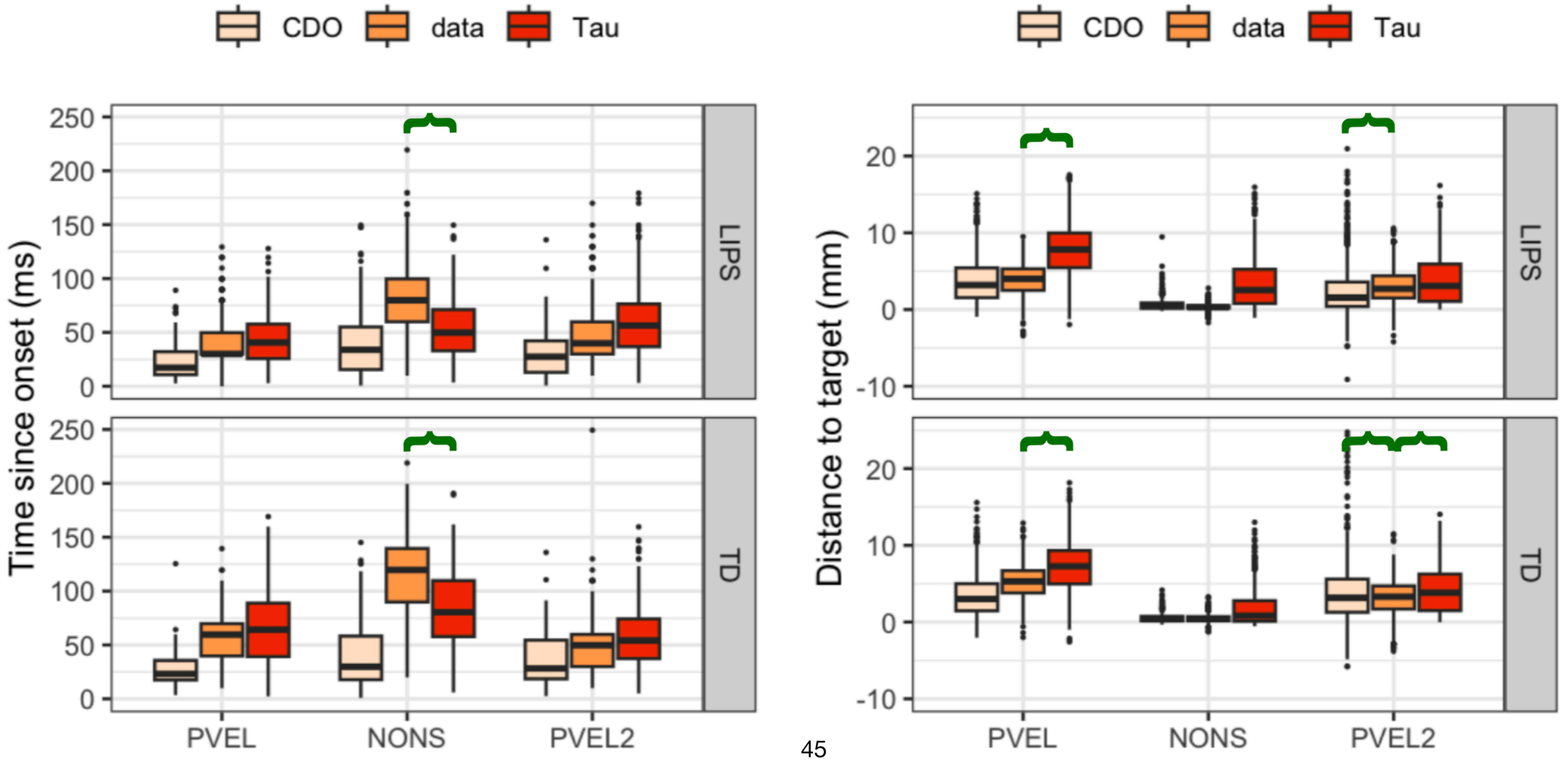
- Predicting landmarks from other landmarks:
  - **GONS-PVEL-NONS-NOFF-PVEL2-GOFF**



Which fits data better?

Time? Tau. Position? Unclear.

Geissler & Nellakra (2024)



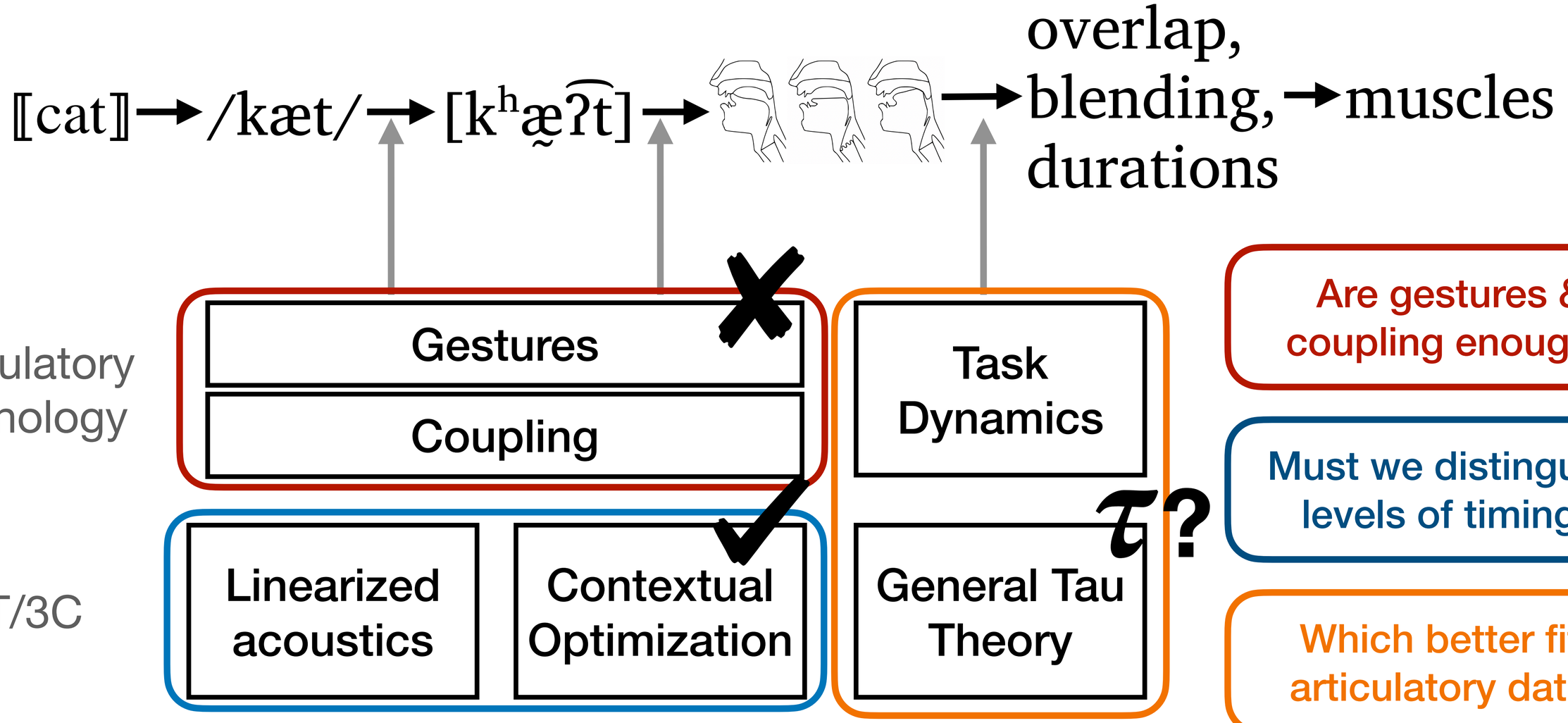
TD/Tau: Conclusion

- Work in progress!
  - Current: full trajectories, not just landmarks
- Results mixed
  - Tau better at *when* landmarks take place
  - Neither much better at *where* landmarks take place
  - This is weird

Roadmap

- Phonology, phonetics, and time
- Types of evidence
  - Gestures & coupling: Tibetan C-V timing
  - Levels of timing: Sámi consonant lengths
    German-English accommodation
  - Fitting articulatory trajectories
- **Conclusion**

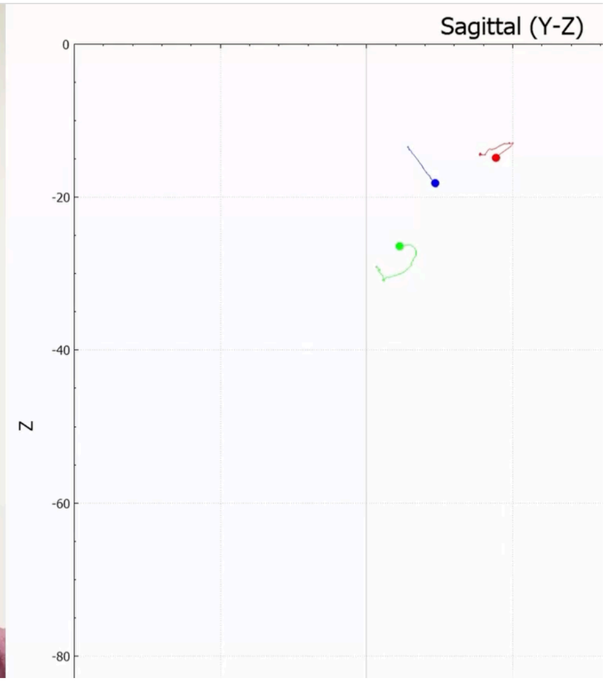
Representational units



AP: Browman & Goldstein (1986) et seq.; TD: Saltzman & Munhall (1989)
XT/3C: Turk & Shattuck-Hufnagel (2020); Tau: Lee (1998)

Upcoming projects

- Inexpensive, student-friendly methods for studying articulatory timing
  - Acoustic ratios in clusters
  - Evaluating MagTrack



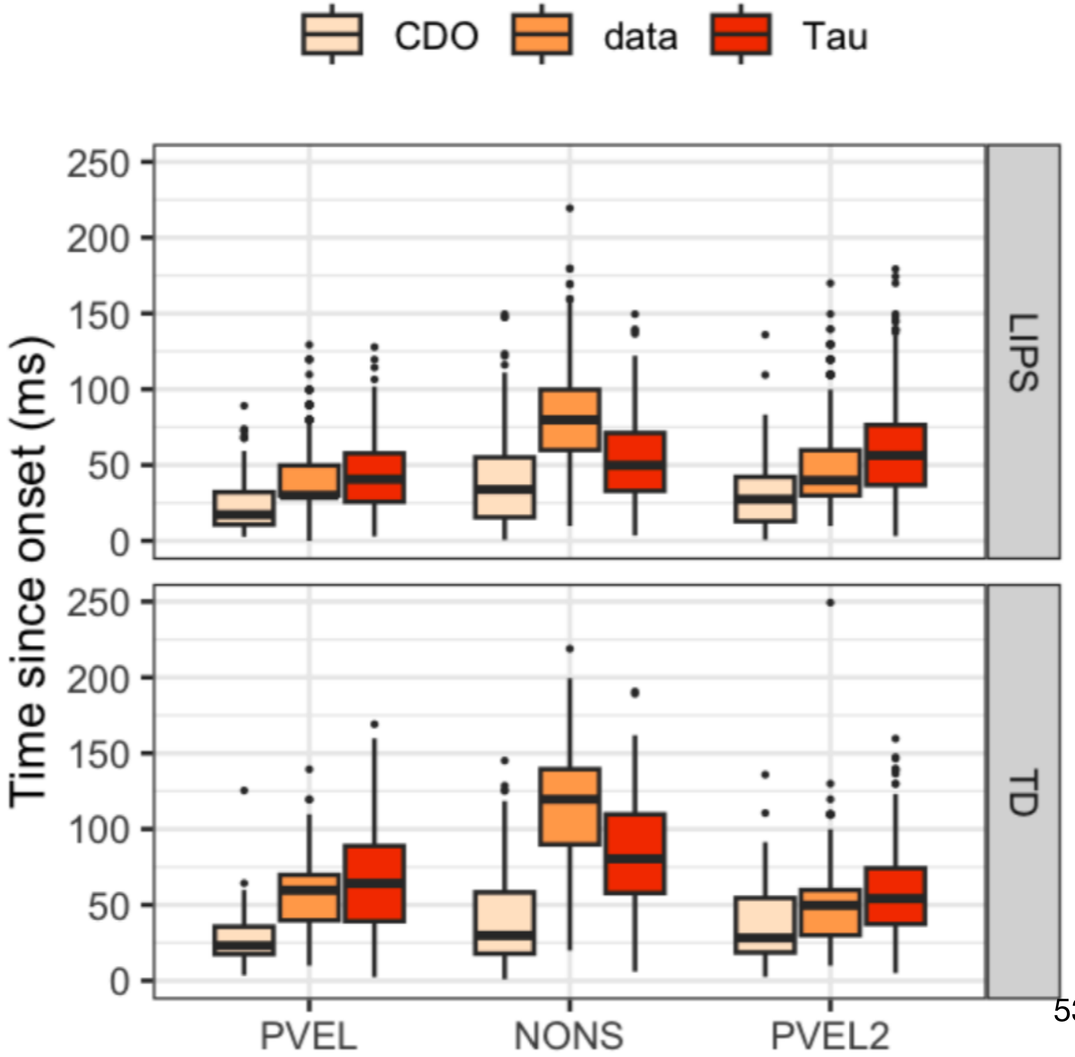
ཐུགས་རྗེ་ཆེ།

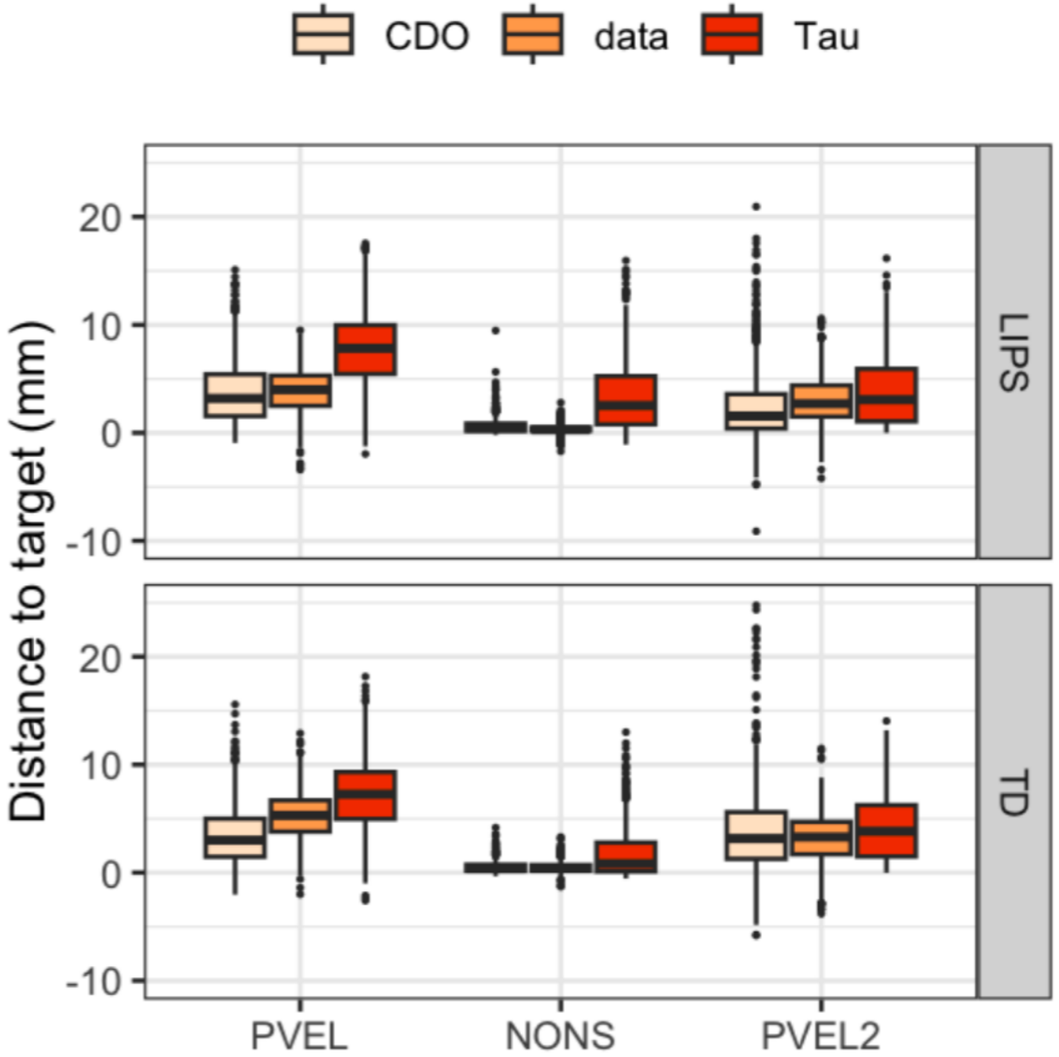
Thank you!

Pocket slides

Which fits data better?

Time? Tau. Position? Oscillator?

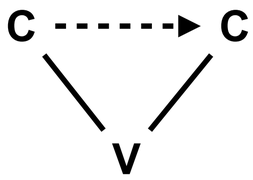


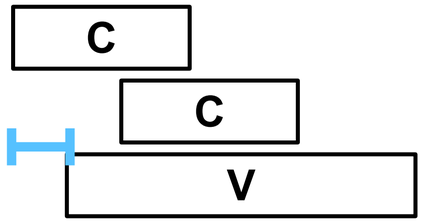


What about clusters?

- Empirically, onset clusters overlap

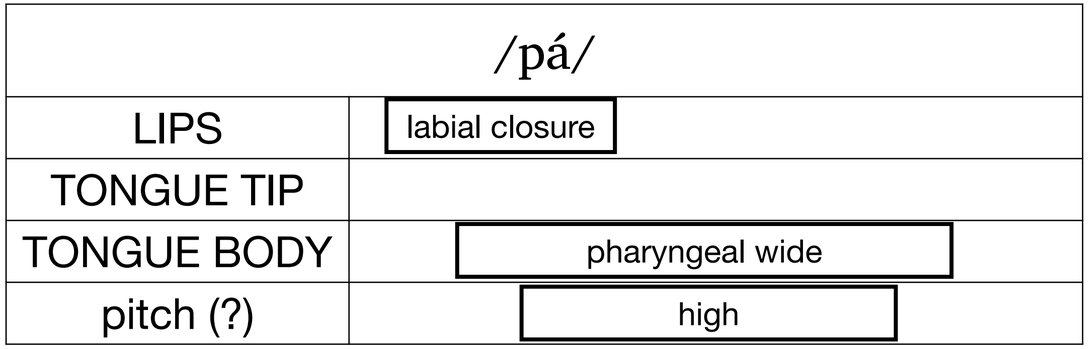
/spa/ ‘spa’	
LIPS	labial closure
TONGUE TIP	alveolar critical
TONGUE BODY	pharyngeal wide





What about tone?

- Empirically, V lags following C
  - (In *lexical tone* languages only)

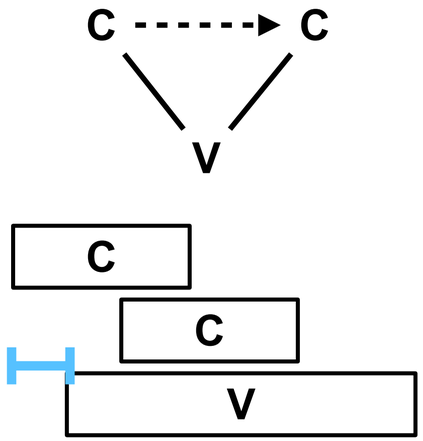


(Gao 2008, Niemann et al. 2011, Karlin 2014)

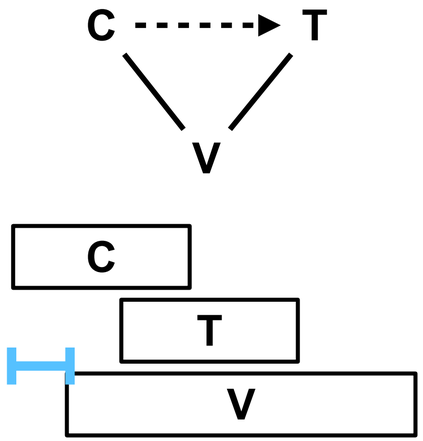
Competitive coupling account 🎉

- Unifies clusters and tone (neat for typology)
- Unifies syllables (and up?), contrast, and planning

/spa/ ‘spa’	
LIPS	labial closure
TONGUE TIP	alveolar critical
TONGUE BODY	pharyngeal wide

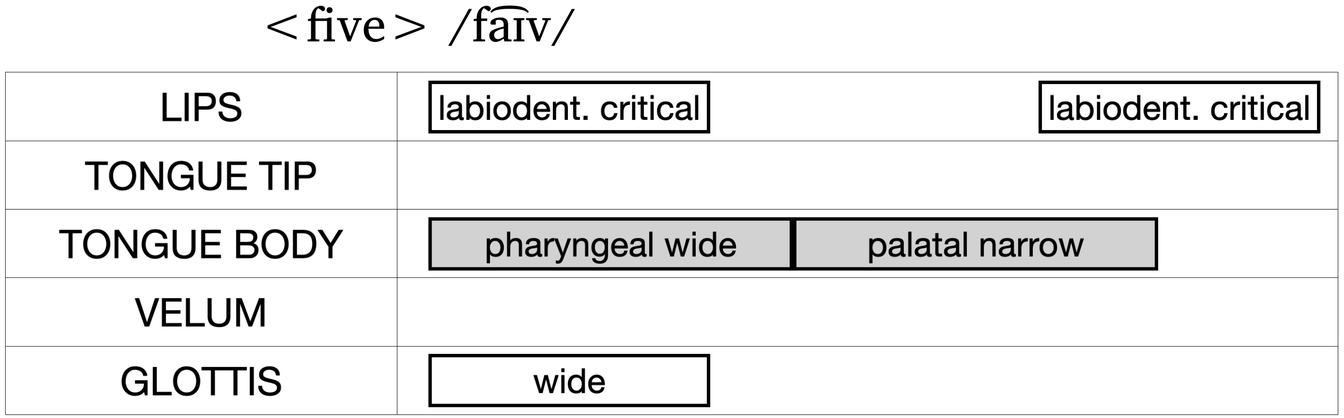


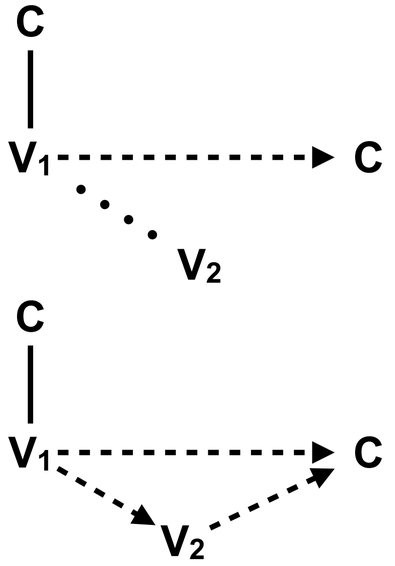
/pá/	
LIPS	labial closure
TONGUE TIP	
TONGUE BODY	pharyngeal wide
pitch (?)	high



What about diphthongs?

- Can approximately describe with in-phase/anti-phase
- How do diphthongs change when they get shorter?

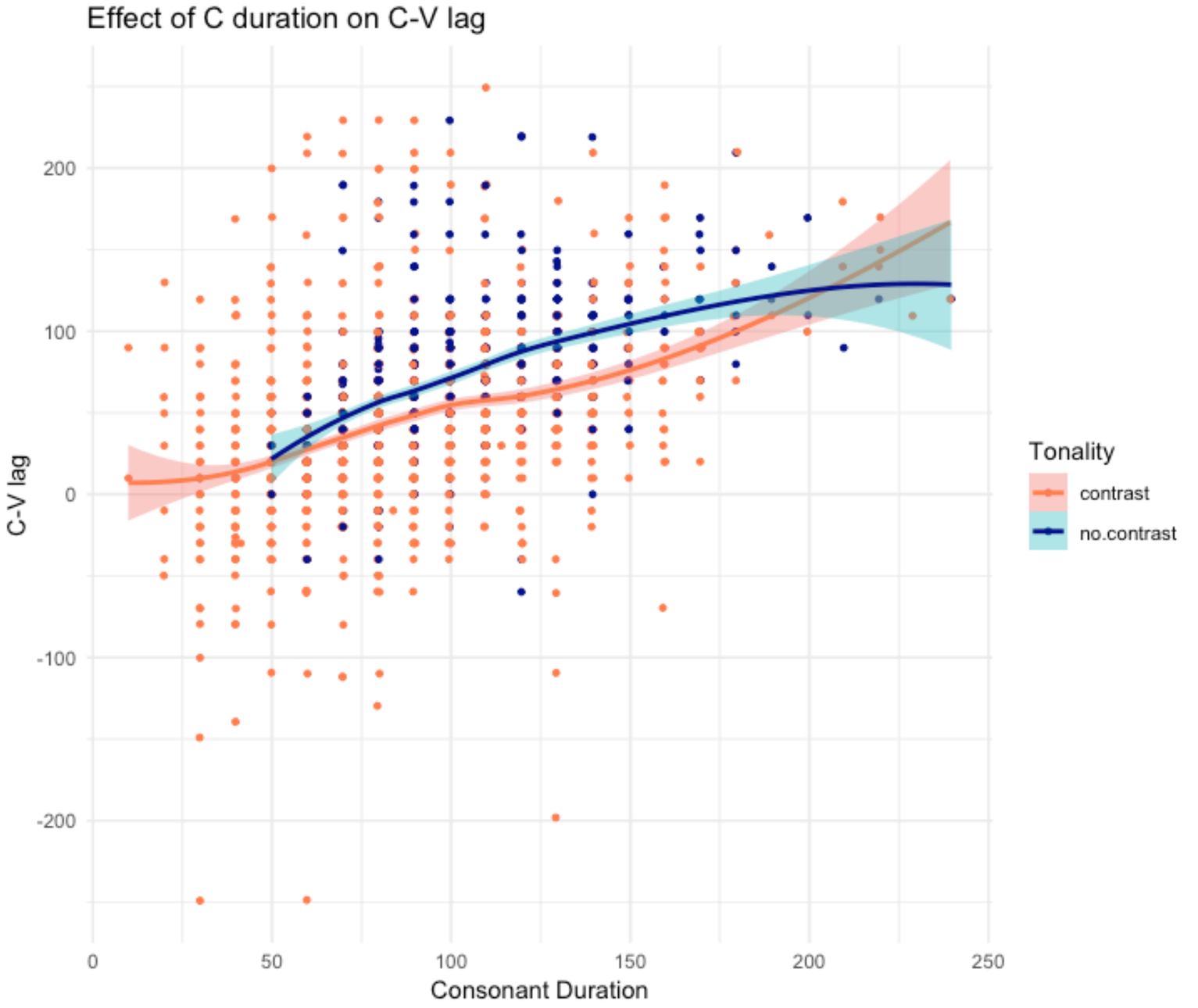




(Goldstein et al. 2000)

Results: C-V lag

- C-V lag *does* increase with C duration
- so, the 50ms lag isn't just a fixed value
- intrinsic account: all speakers anti-phase (ish)
- extrinsic account: gestures and coordination both affect by speech rate



Articulatory study

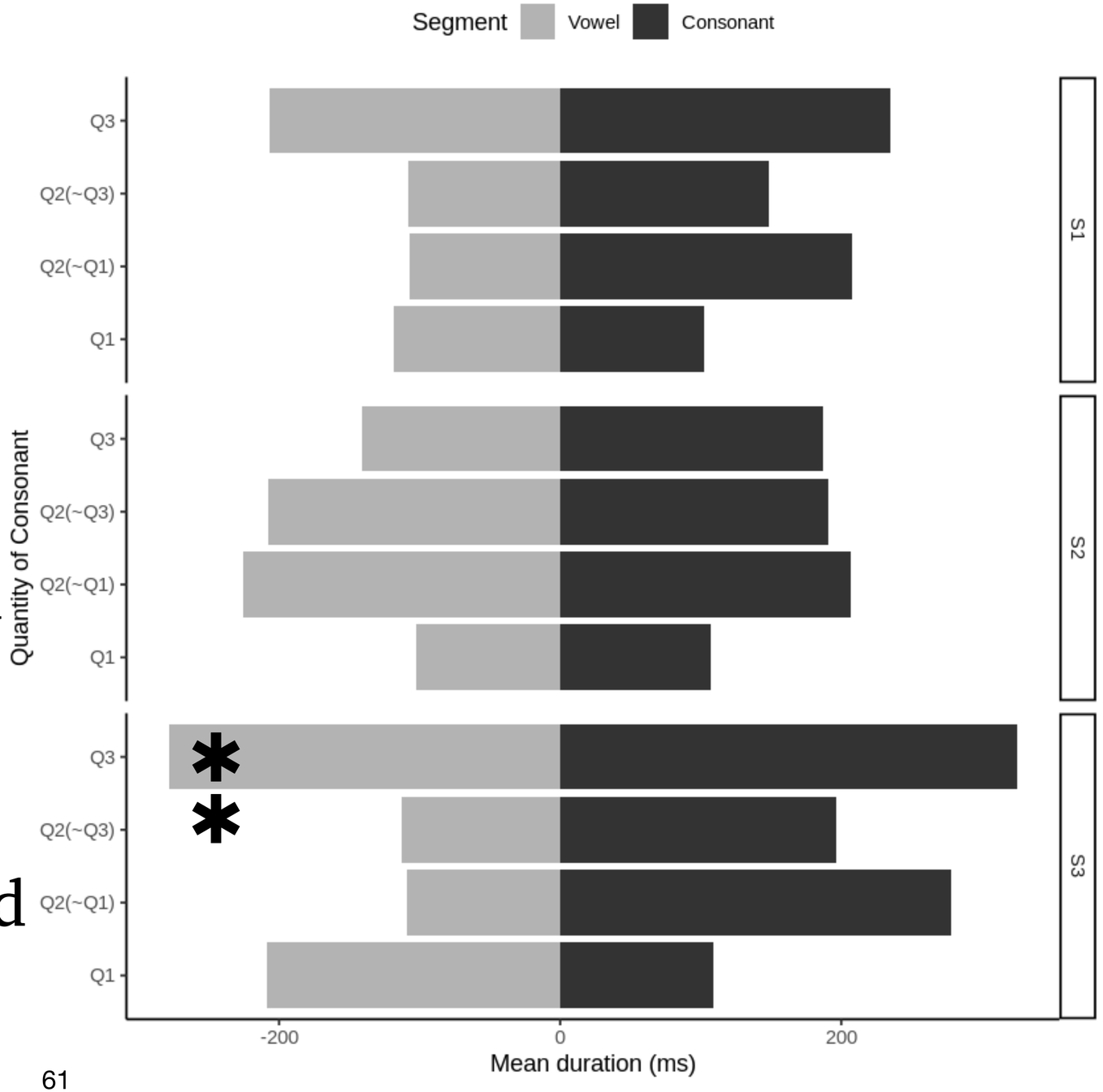
Geissler et al. (2021), Geissler (2021ch4)

- H1: variation in timing conditioned by presence/absence of lexical tone
  - speakers with tone contrast will have competitive coupling (pos. C-V lag)
  - speakers without tone contrast will have in-phase C-V timing (no C-V lag)
- H2: timing convergence:
  - all speakers will have similar coordination patterns despite interspeaker variation in presence/absence of tone
- What kind of tone contrast is there?
    - If H-∅, then difference will be visible in high vs. low tone words
    - If H-L, then no difference in timing by tone.

EMA Study conclusions

- H1: variation in timing conditioned by presence/absence of lexical tone
  - speakers with tone contrast will have competitive coupling (pos. C-V lag)
  - speakers without tone contrast will have in-phase C-V timing (no C-V lag)
- ✔ **H2: timing convergence:**
  - all speakers have similar coordination patterns despite interspeaker variation in presence/absence of tone
- What kind of tone contrast is there?
  - If H-∅, then difference will be visible in high vs. low tone words
  - ✔ If **H-L**, then no difference in timing by tone.

- Significant inverse relationships (V decrease when C increase) only in underlying Q3 Cs; driven by one speaker
- Trends in expected directions; more data needed



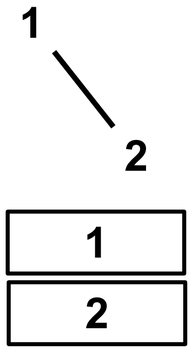
The temporal basis of complex segments

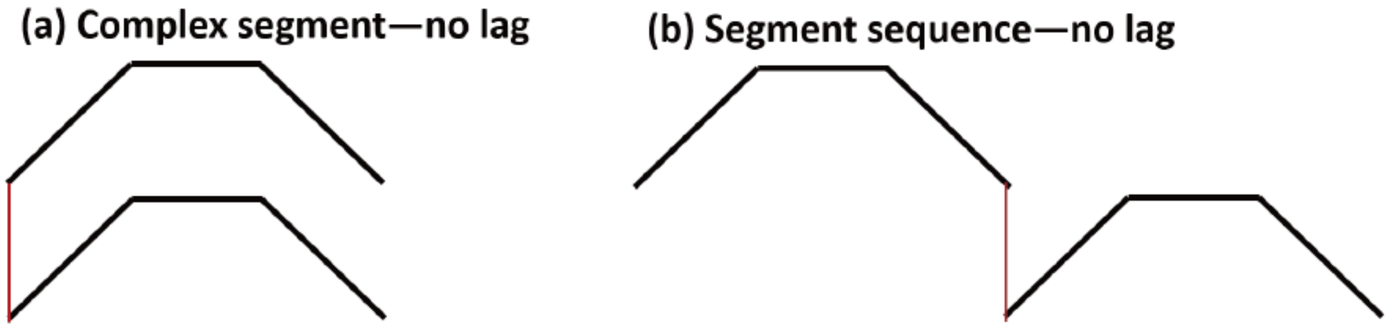
Shaw et al. 2019

The temporal basis of complex segments

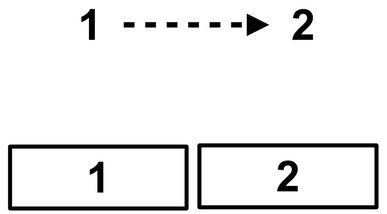
Shaw (2019): predictions

In-phase

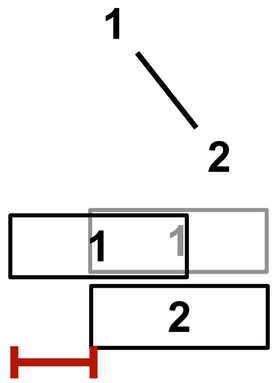


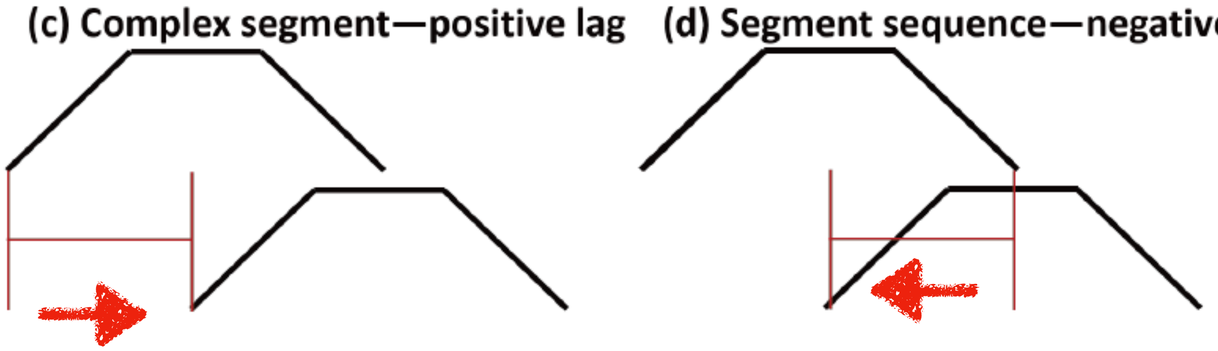


Anti-Phase



In-phase + lag (offset)





Anti-Phase - lag (offset)

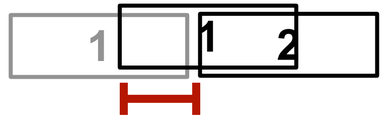


Figure 1: Hypothesized gestural coordination patterns for complex segments (a), (c) and segment sequences (b), (d)

The temporal basis of complex segments

Shaw (2019): results

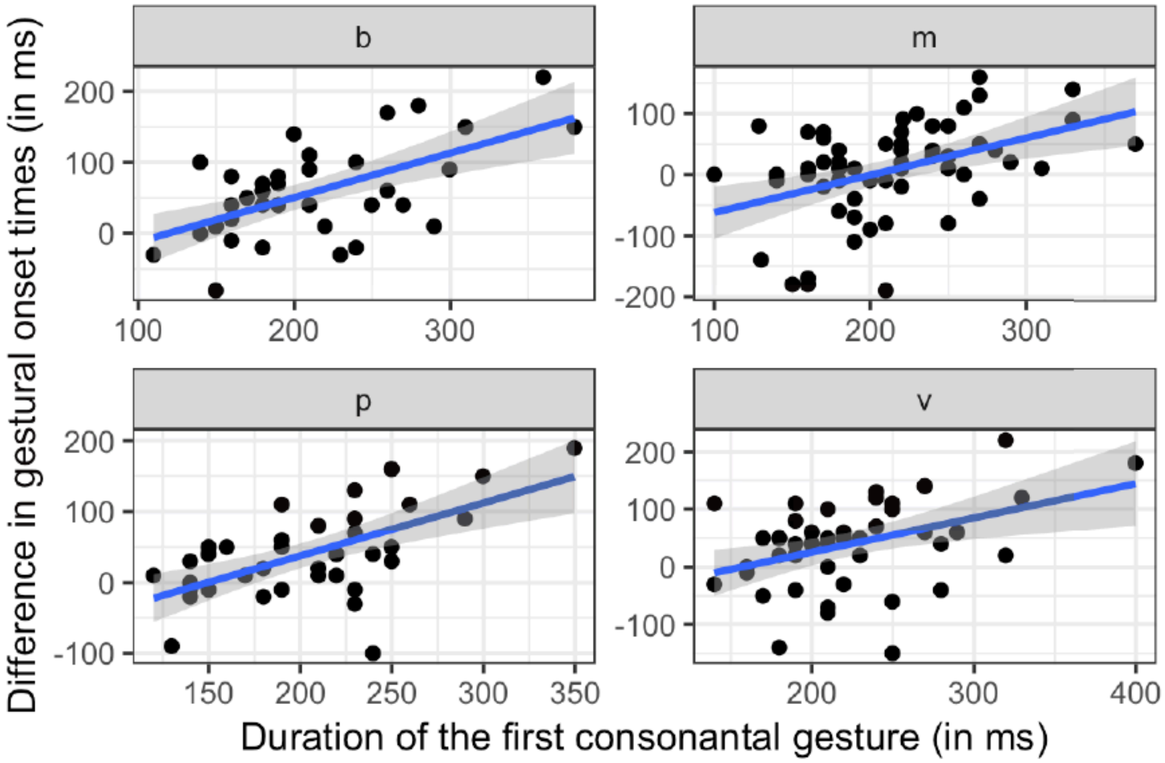


Figure 4: Correlations for the data from the English experiment

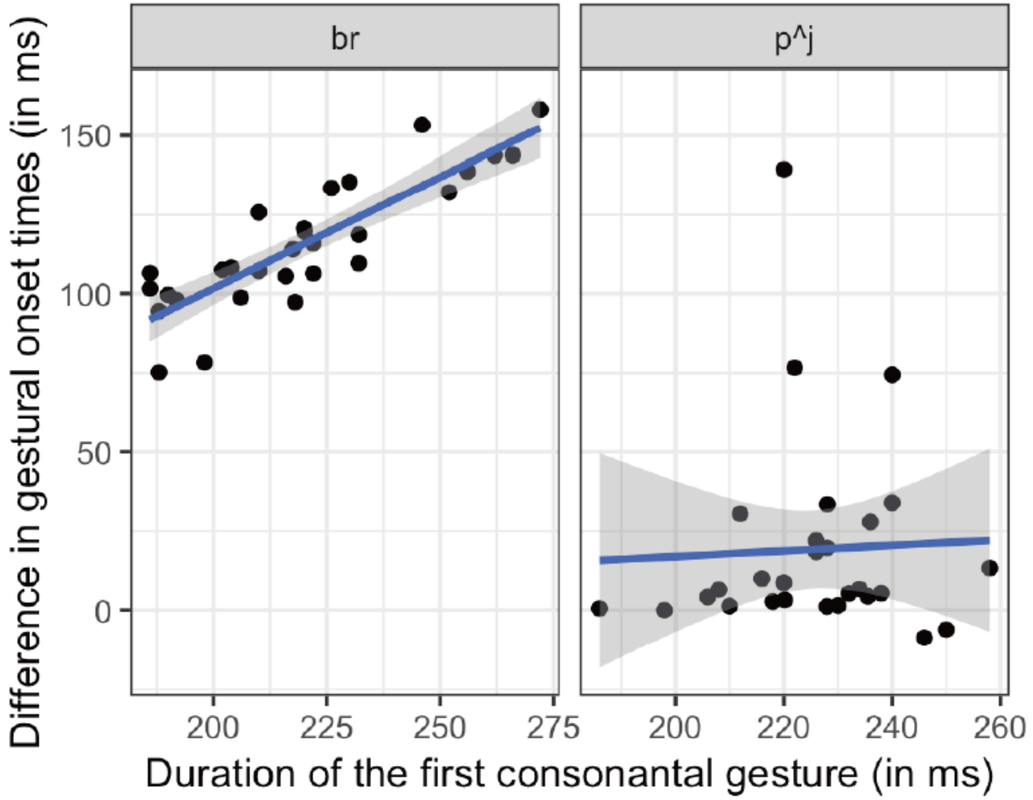


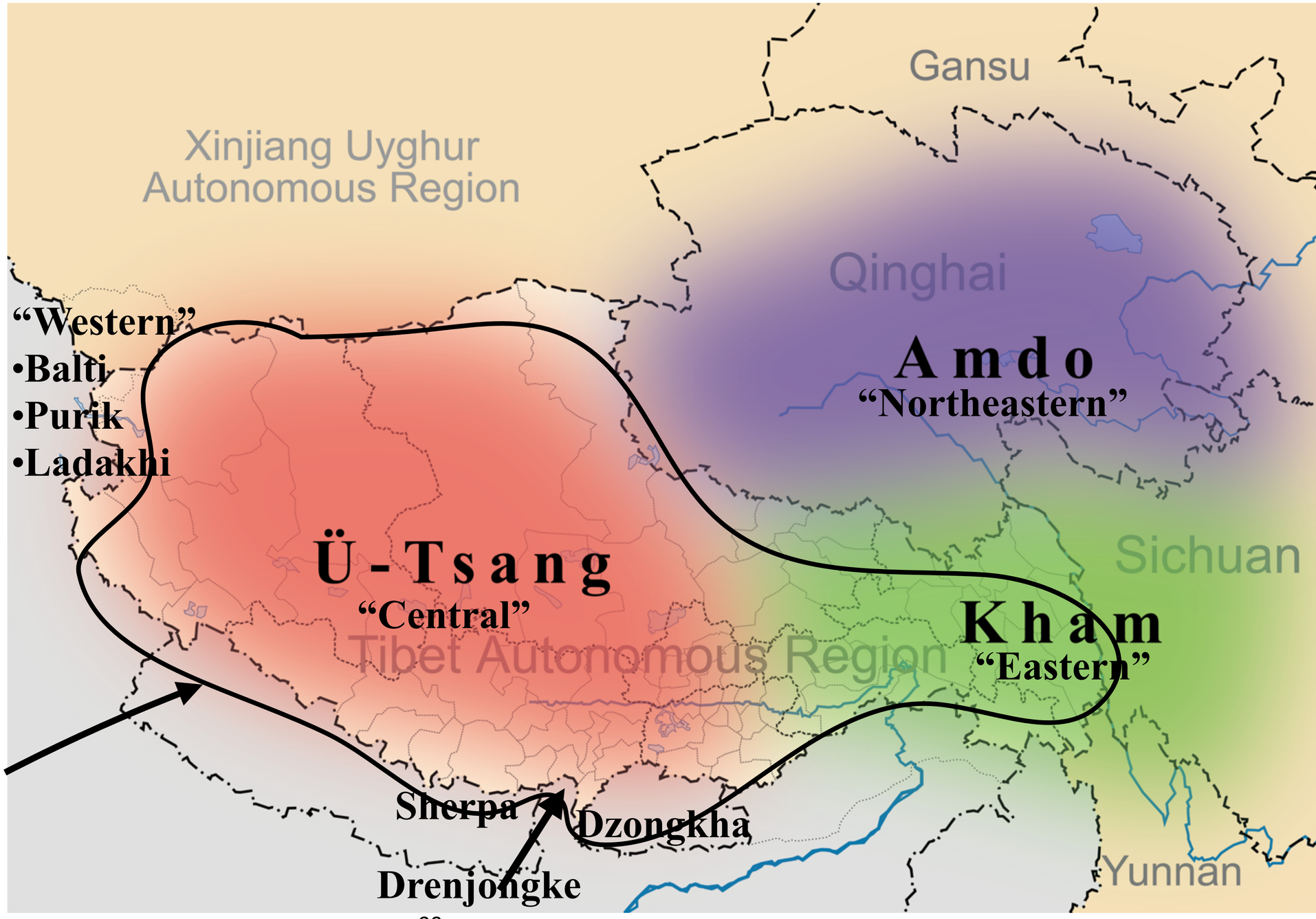
Figure 2: Correlations for the Russian data

Tibetan dialects

Tibetan

བོད་སྐད་

- “archaic”/ “cluster”
- “innovative”/ “non-cluster”
- dialect continuum
- post-1959 diaspora



Approx. extent of tone

Dialects: Natural laboratory

- tonogenesis
- laryngeal variation
- cluster simplification
- vowel shifts, spirantization, retroflexion, palatalization
- evidential, honorifics, modality, etc.

Written (Classical) Tibetan	Balti (Western)	Rebkong (Northeastern)	Tokpe Gola (Central)	Gloss
khrag	[kʂʌk]	[t͡ɕɤɣ]	[tʰʌ́k] ([tʰák])	‘blood’
rtswa	[xst͡soa]	[xt͡sa]	[t͡sá]	‘grass’
spyang ki	[spjaŋ.ˈku]	[xt͡ɕaŋ.ˈkʰɤ]	[t͡ʃáŋ.gú]	‘wolf’
bcu bdun	[t͡ɕub.ˈdʊn]	[t͡ɕɤb.ˈdɤn]	[t͡ʃúp.tő] ([t͡ʃúp.tỹ])	‘seventeen’

(Adapted from Caplow 2013)

Tonogenesis
(tonal dialects only)

- Voiceless onsets > high tone
  - *p^{h}ar ‘over there’ > H
    *sa ‘earth’ > H
- Voiced onsets > low tone
  - *bar ‘between’ > L
    *za ‘eat’ > L
    *mar ‘butter’ > L
- Sonorants with pre-initial > high tone
  - *sman ‘medicine’ > H

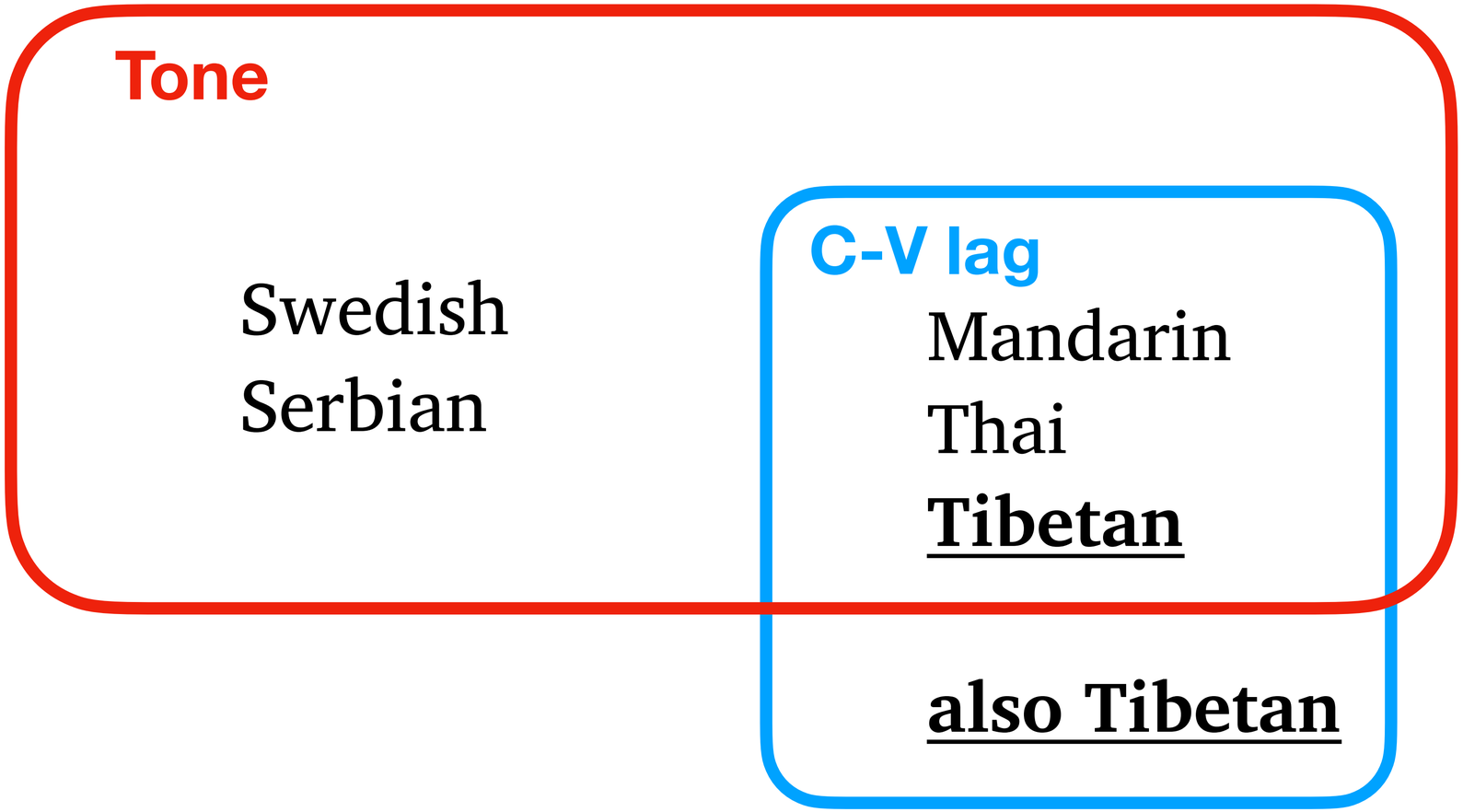
Laryngeal contrasts

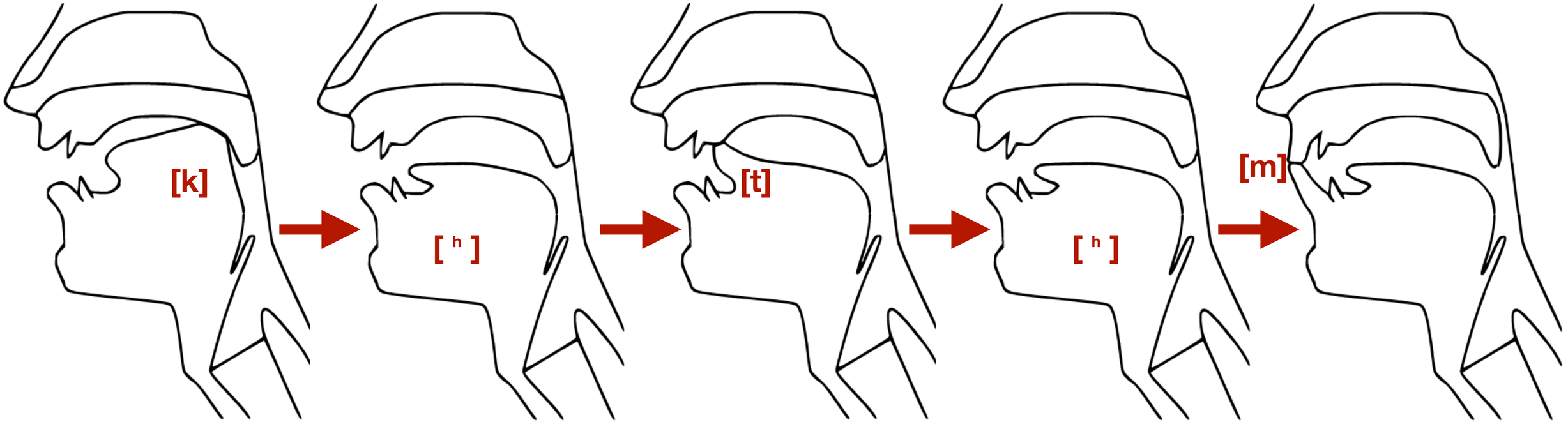
	Etymological onsets							Innovative features
Orthography	སྤ་	ཕ་	བ་	སྦ་	ས་	ཟ་	བཟ་	
Old Tibetan	sᵊpa	pʰa	ba	sᵊba	sa	za	bᵊza	aspiration allphonic
Northeastern and Western dialects	spa	pʰa	ba ~ wa	ʁba	sa	za	za	cluster simplification aspirated/unaspirated contrast
Eastern dialects	pá	pʰá	pà	bà	sá	zà	zà	tonogenesis cluster simplification
Central dialects (Lhasa)	pá	pʰá	pʰà	pà	sá	sà	sà	voiced clusters > voiceless voiced simplex > aspirated

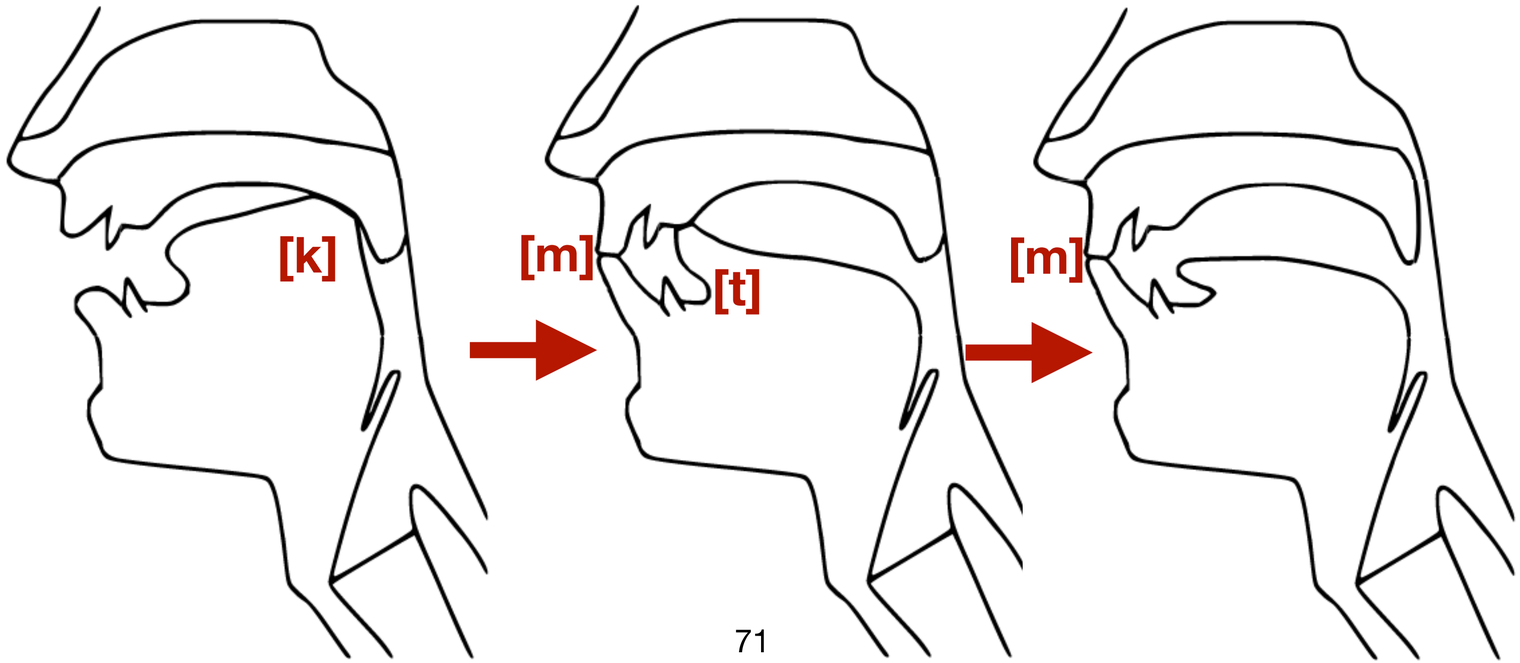
Cross-linguistic evidence (after)

No tone,
no C-V lag

Arabic
Catalan
English
German
Georgian
Italian
Romanian





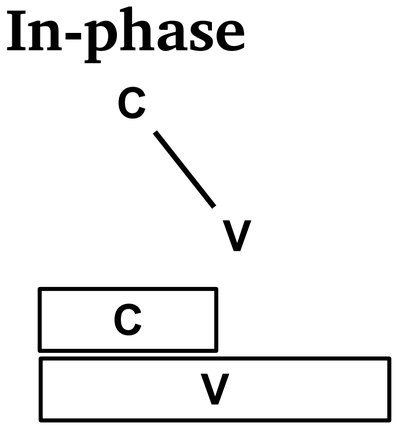


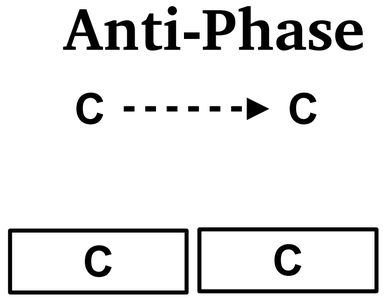
back to slide 7

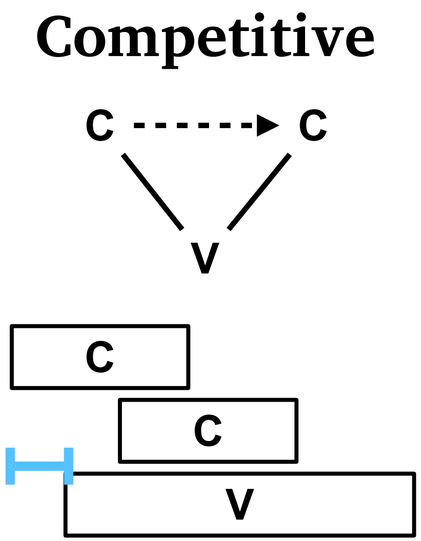
Interactive Sagittal Section

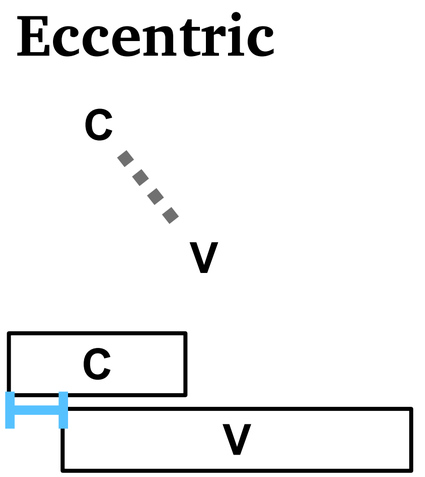
Coordinating gestures in time

- Gestural coupling modes:
  - *In-phase coupling:* (synchronous) and *Anti-phase coupling* (sequential) are most stable
  - *Competitive coupling:* combination of in-phase and anti-phase coupling relations
  - *Eccentric coupling:* one coupling relation, just not intrinsically stable







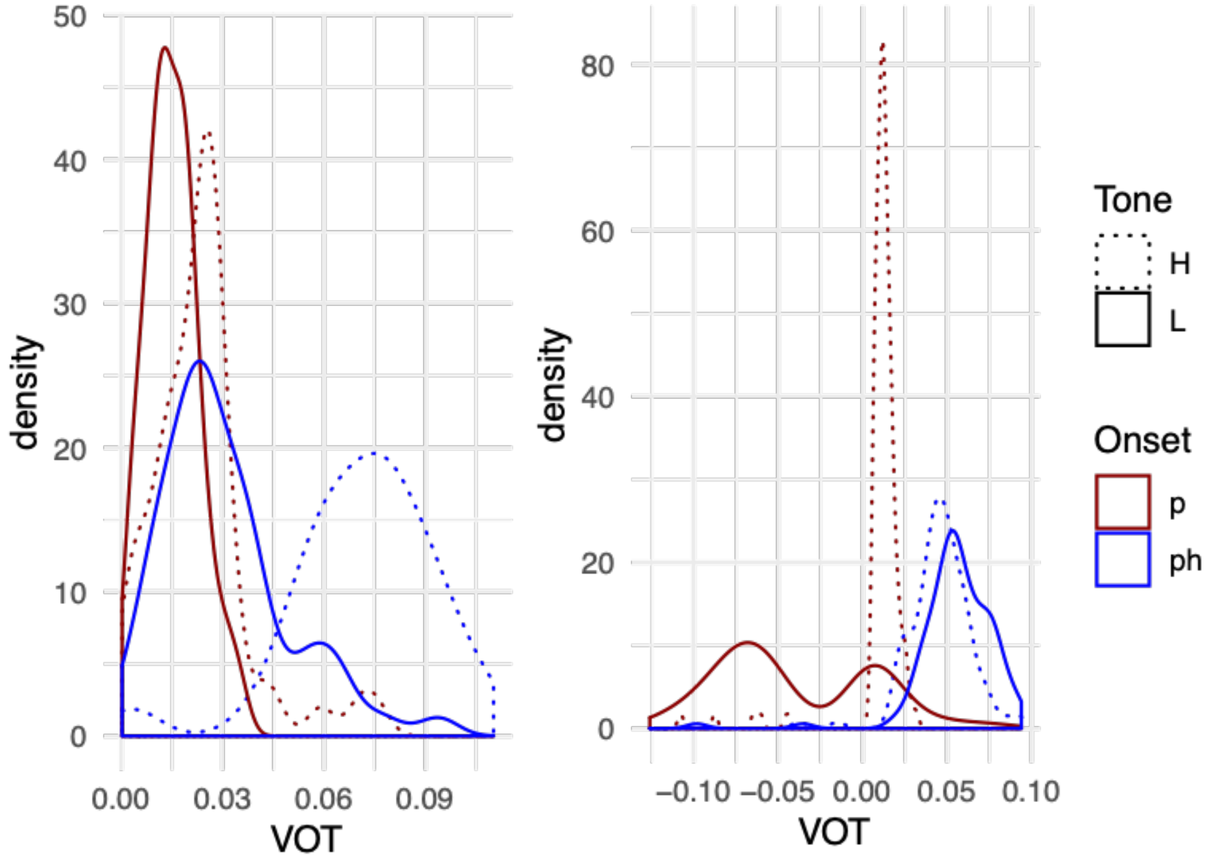


(Nam & Saltzman 2003, Nam et al. 2009, Goldstein 2011)

Two systems of laryngeal contrasts

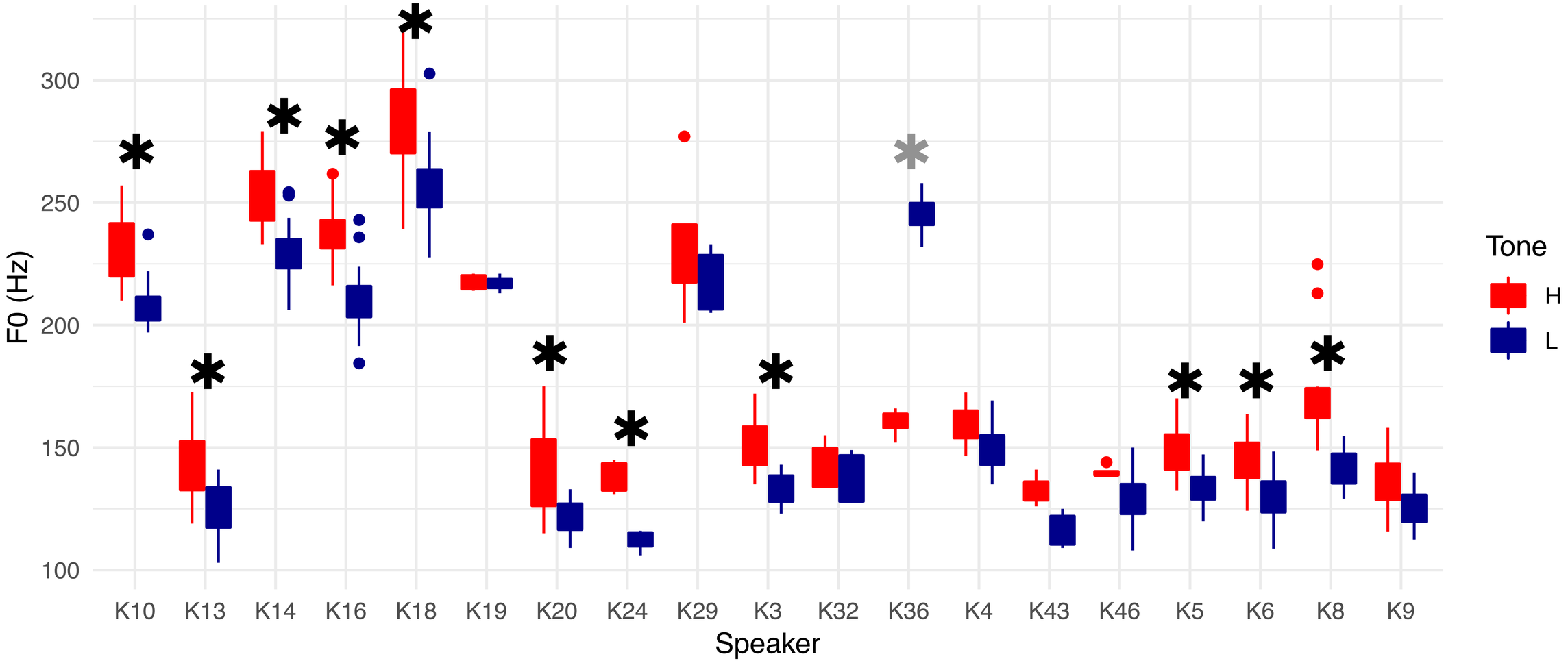
even in speakers with no F0 contrast (!!!)

- Both conditioned by etymological tone category:
- Left speaker
  - no prevoicing
  - long VOT only with H tone
- Right speaker:
  - prevoicing with L tone
  - long VOT with both tones

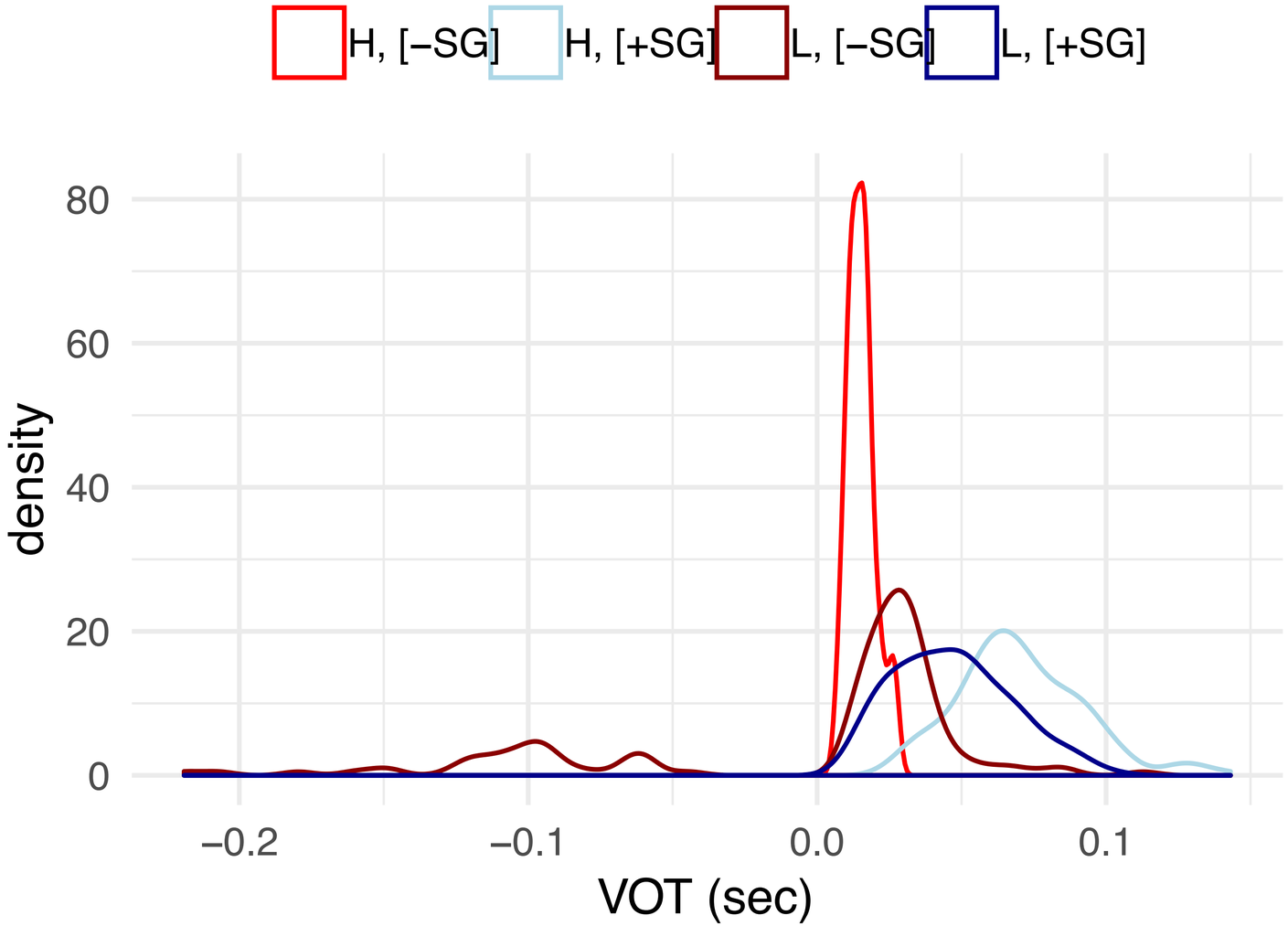


Does H have higher pitch than L?

Yes for 11/19, no for 7/19



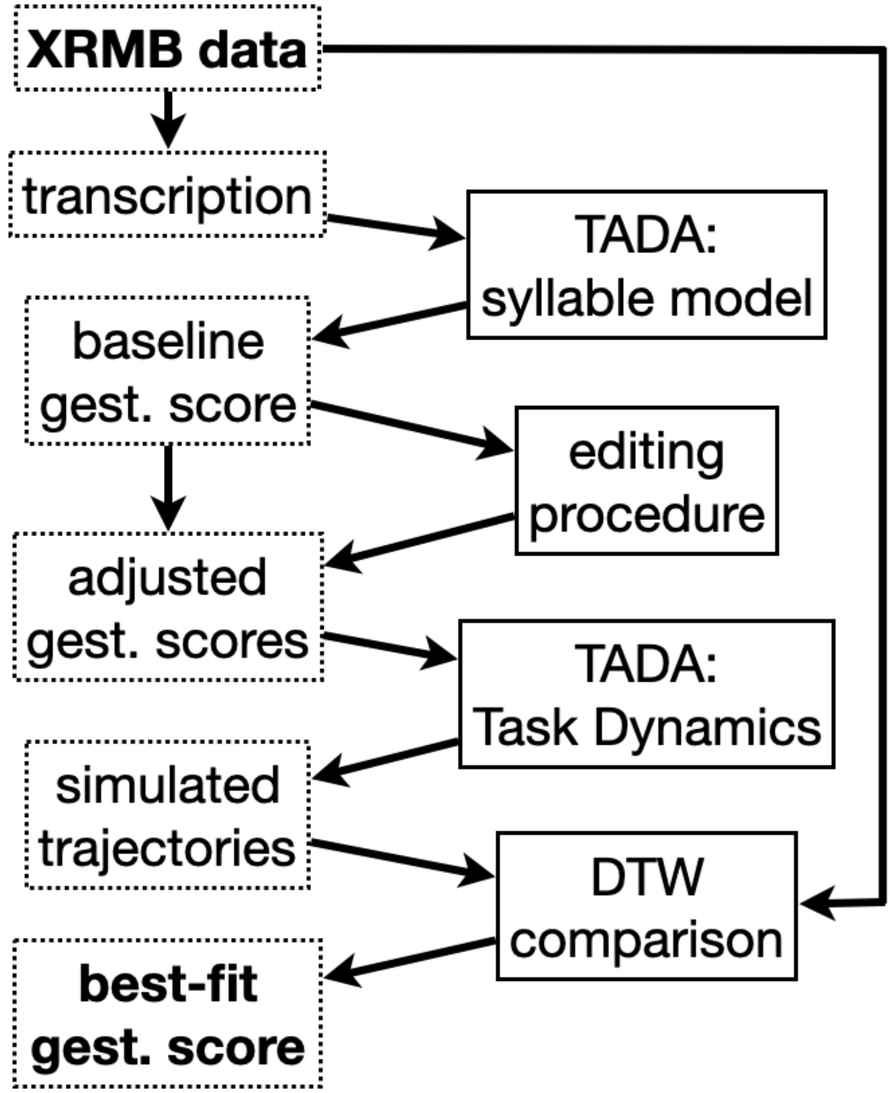
Consonant and tone categories



<five> study: methods

O'Reilly, Geissler, & Tang (2023)

- Ideal test case?
  - diphthongs: all four modes
  - C's with lips, V's with tongue
  - available data



Timing in phonology and/or phonetics?

- "Discrete Phonology" vs. "Gradient Phonetics"
- Speech timing as phonology
  - Is timing *intrinsic* or *extrinsic* to phonology?
  - Are gestures coordinated at *beginning* or *end*?
  - *Symbolic* vs. *phonetically-enriched* representations?

(Turk & Shattuck-Hufnagel 2021)