Unlaboratory Phonology III

Enriching Laboratory Phonology with Rich Language Data: From Documentary Fieldwork to Corpus Phonetics

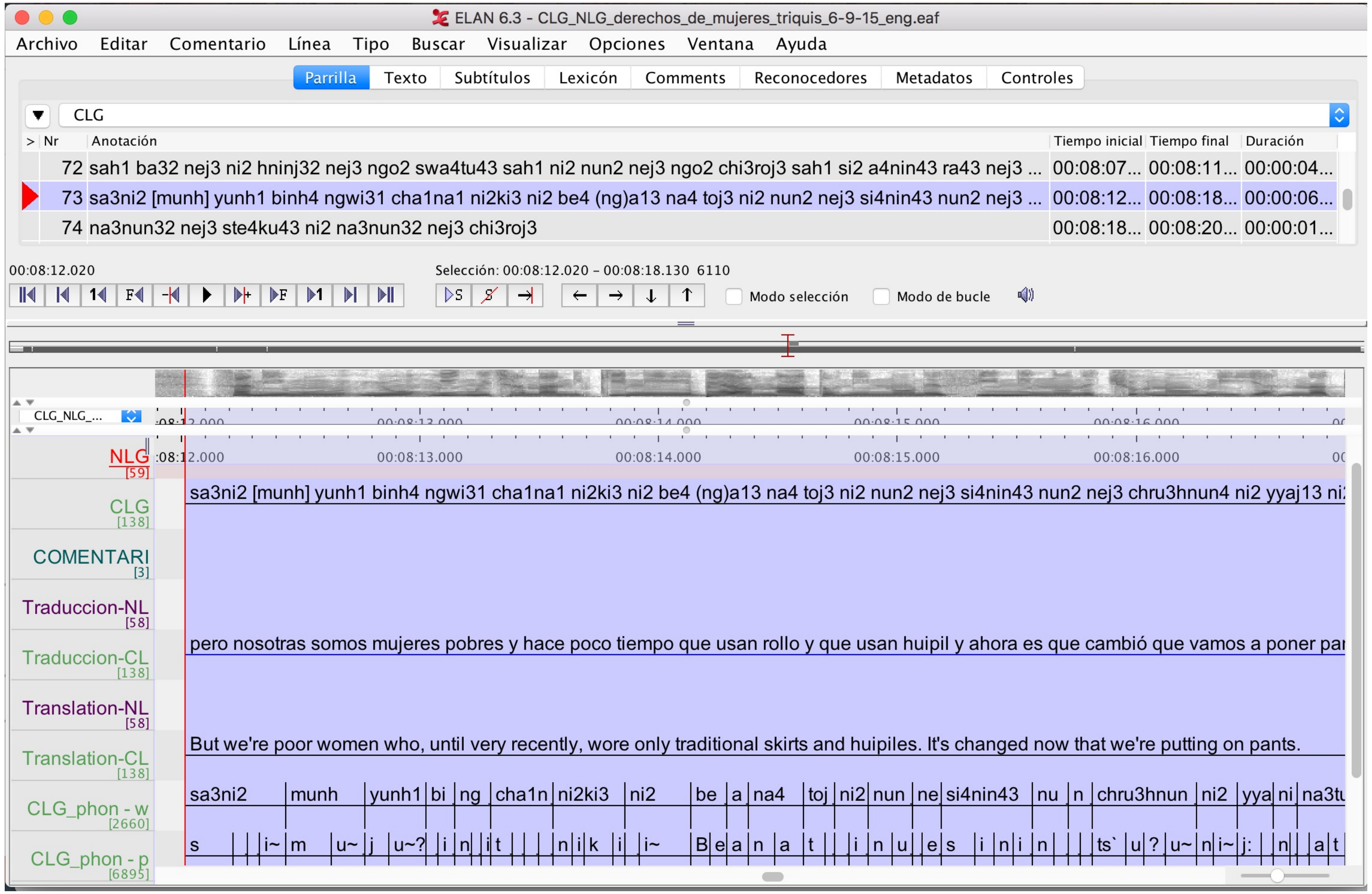
Dr. Christian DiCanio

University at Buffalo

Department of Linguistics

What is **rich** language data?

- Language documentation recordings/archives are rich since they include **ecological**, **linguistic**, and **stylistic** diversity. They are therefore ***variationally rich***.
- Phonetic and phonological patterns are best examined in relation to a typologically diverse set of languages. Otherwise, we can never generalize about *language*.



Larger goal: How do we get to a laboratory phonology that considers all languages equally?

- More phonetic fieldwork needed!
- Utilize existing archival data/recordings.
- Prioritize ***exploratory approaches*** since there are fewer theoretical motivations based on past literature that pertain to understudied languages (cf. Roettger 2019)

How do we access this rich data?
(a roadmap)

I. How do you move from a documentation corpus to a phonetic corpus?

II. How do we address theoretical, exploratory, and descriptive questions utilizing documentation corpus data? Case studies.

Topics not specifically covered here

- Praat scripting (see Eleanor's tutorial).
- Statistical modeling of phonetic data (see Morgan's tutorial).
- Technical aspects of how to write transducers or run forced aligners on speech corpora

I'm happy to discuss any of these topics, but they are not the focus of this presentation.

What is a documentation corpus?

- Serves more holistic documentary goals (multi-genre, text transcription & translation, lexicographic & grammatical materials).

- Recordings and associated transcriptions (Transcriber, **ELAN**, *even Word documents*)

- Idealized users of materials are *not* phoneticians, but fieldworkers/linguists and community members/speakers.

Documentation corpora: processing

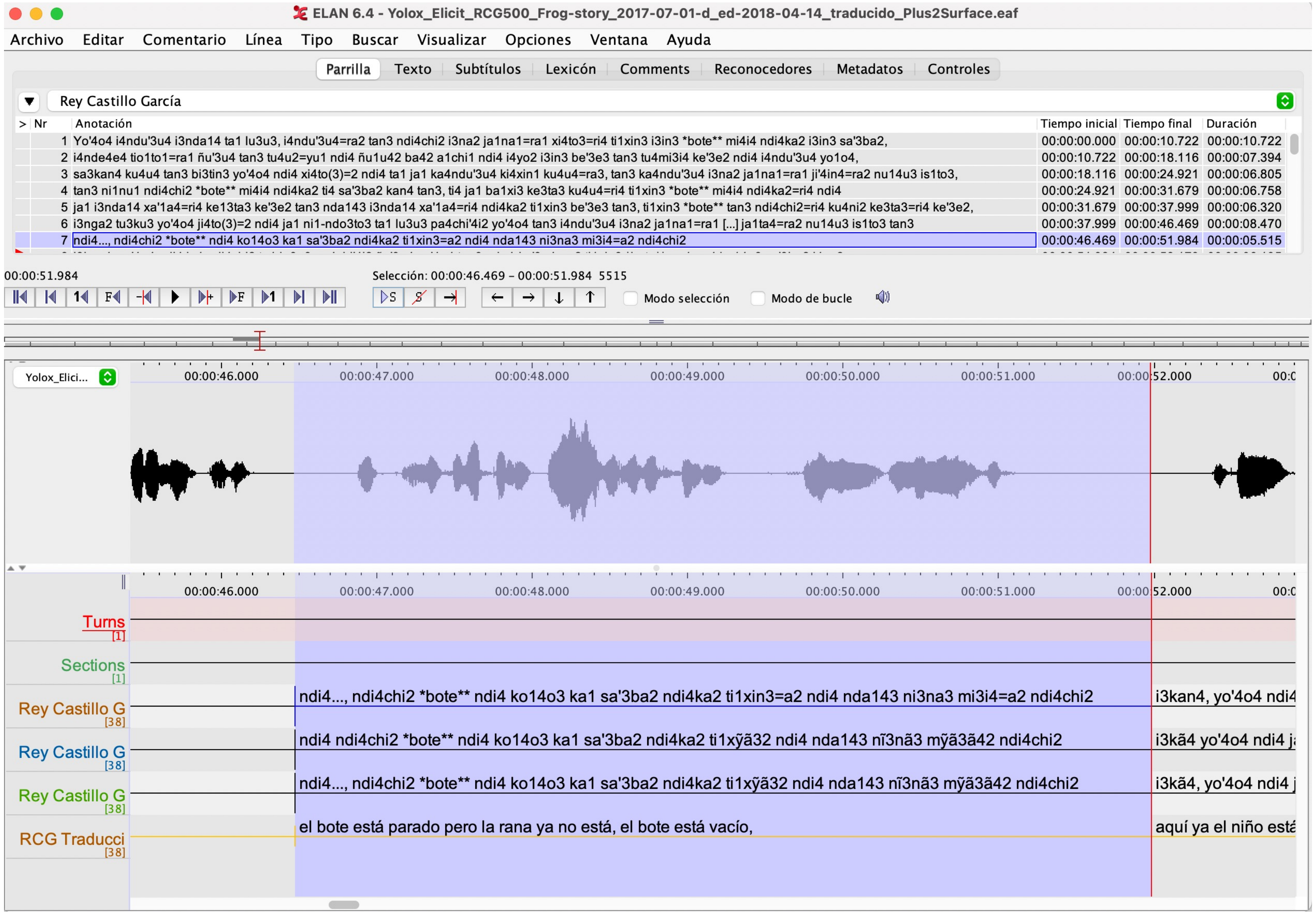
Typical documentation pipeline:

Recording of speakers > Transcription/translation of texts > Archiving

Typical phonetic corpus pipeline:

Transcription (ELAN) → Phonological transducer → Aligner creation and testing → Word and phone level segmentation

Documentation corpora: processing via transducers



If transcriptions are morphological (no morphophonology has been applied), a phonological transducer is necessary – example from Yoloxóchitl Mixtec (Otomanguean).

transcription
transducer output (no punctuation)
transducer output (with punctuation)
translation (Spanish)

Input	Output
ti1xin3 = a2	ti1xyã32
/ti^{1}ʃĩ^{3} = a^{2}/	[ti^{1}ʃjã^{32}]
under = 3.INANIM	
‘under it’	
mi3i4 = a2	myã3ã42
/mi^{3}i^{4} = a^{2}/	[mjã^{3}ã^{42}]
only = 3.INANIM	
‘only it’	

The work of a phonological transducer is to transform ascii-encoded input into ascii-encoded output which reflects surface phonetic patterns.

If you do not do this, forced alignment will fail.

Note that nasalization has to be adjusted to spread onto the fused clitic.

Documentation corpora: processing via transducers

- Surface-level, time-aligned transcriptions are ideal for initial stages of phonetic and phonological corpus analysis.

- Phonological transducers essentially just apply rewrite rules to the transcriptions on an additional tier (in ELAN or in Praat textgrids).

e.g.	...*in=a*	>	...yã	yu3ba4=on4
	...*i=a*	>	...ya	father=2s
				yu3bon4 [ju^{3}βõ^4]
	(palatalization rule)			(vowel replacement rule)

NB: You have to know the phonology of the language to correctly apply the transducer to the transcription.

Assessment of the output with speakers is usually necessary.

The Yoloxóchitl Mixtec corpus required 4-6 months of work on the transducer (2016 – 2017) to make sure it was accurate and we already had a phonological grammar written (Castillo García 2007).

But we now have a 100+ hour corpus that is force aligned.

Documentation corpora: processing via *alignment*

Forced alignment is a method of automatically segmenting a corpus utilizing a surface-level transcription, a pronunciation dictionary, and associated sound files (Babinski et al 2019, DiCanio et al. 2013, Johnson et al. 2018, Tang & Bennett 2019)

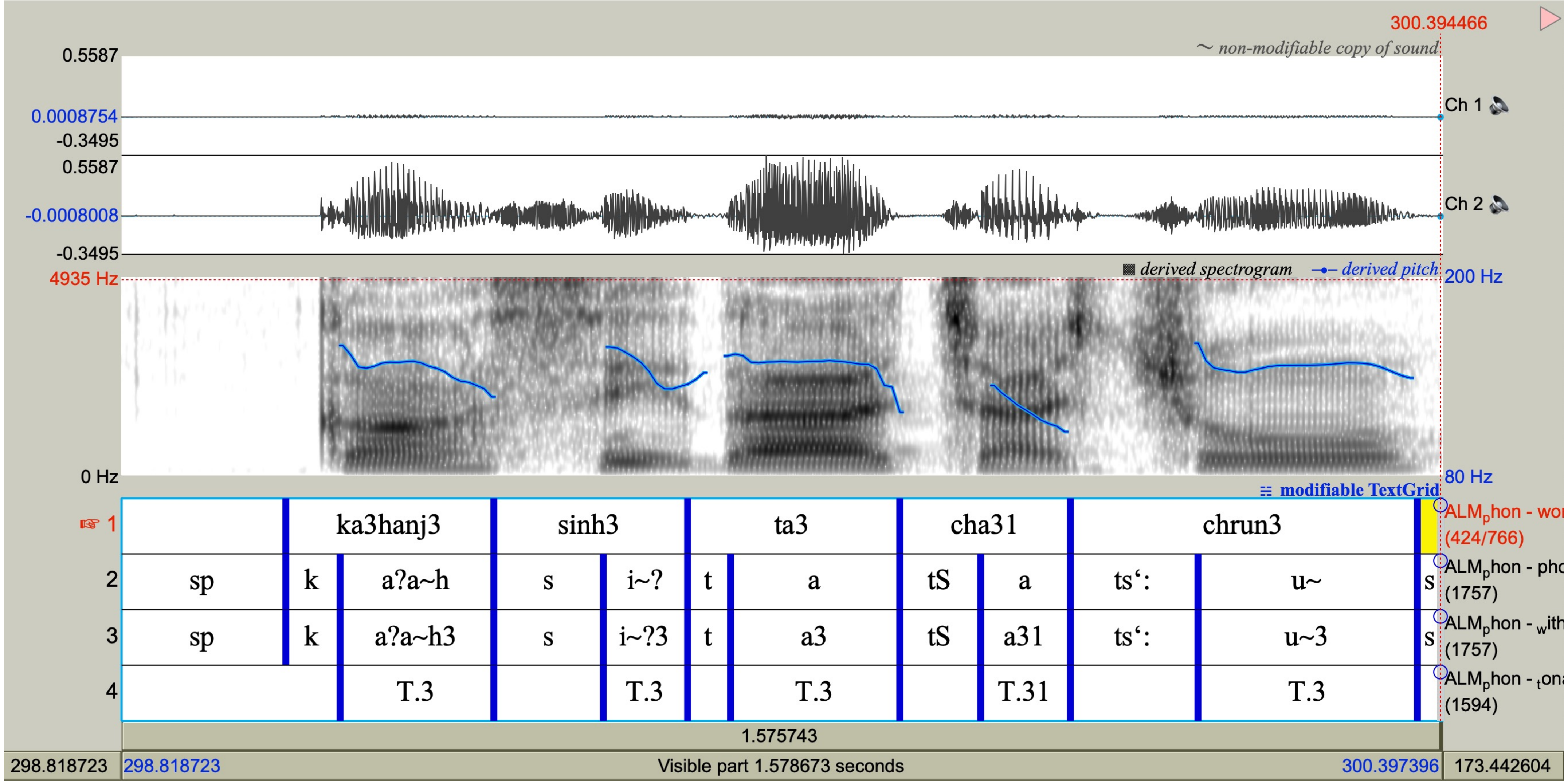
The goal is to create a multi-layered speech corpus that is phonologically-annotated and that can be used for phonetic data analysis. How do we get there?

a. What is the structure of the corpus?

b. What issues arise in organizing a corpus so that one can create a language-specific alignment system?

c. What are the steps in processing a corpus to create a forced alignment system?

Alignment of Itunyoso Triqui (Otomanguean)



The Itunyoso Triqui corpus

- Otomanguean:Mixtecan
- Complex segmental and tonal inventory with complex morphophonology related to person (DiCanio 2010, 2016, DiCanio et al 2020).
- Collected between 2014 - 2019 on an NSF DEL/DLI grant.
- Dictionary of approximately 4,000 entries.
- Approximately 29 hours of narrative and conversational speech by 34 speakers.
- Mostly Triqui, but Spanish loanwords appear throughout the recordings.
- Trained a Triqui model using the **Montreal Forced Aligner** (McAuliffe et al 2017)

Issues to consider while aligning documentation corpora

- What is your level of transcription? (morphophonological, phonemic, surface phonological, phonetic)
- Many endangered language corpora are **multilingual,** but most aligners are assumed to be monolingual. How do you separate different languages?
- How are loanwords encoded?
- How are disfluencies encoded?
- How are elided tokens encoded?
- How are complex segments encoded?

Encoding solutions in the Triqui corpus

1. (...) marks elided speech | **You don’t align these by accident.**

taj13	ki3hyaj3	nni4=(reh1)	yoj3
like.so	did	mother=2S	then

'Your mother did (it) like that then.'

2. **...* marks another language | **You get pron. via another dict.**

be4=nih2unj4	ku3man4	**sesenta*	ni2	**sesenta y cinco*	bin3
TOP=PL.1P	PERF.exist	sixty	and	sixty five	be

'We were (there) in (19)60 and (19)65, it was...'

3. [...] marks disfluencies | **You flag these as disfluencies.**

ta1ranh3	nej3	sinj5	bin3...	[ranh]
all	3P	people	be	??

'...all of them that were there'

4. Loanwords use Triqui orthography | **You modify transcriptions to fit a single orthography.**

sa4na43	'manzana' (apple)
skwe4la43	'escuela' (school)
mu3ni3si4pio43	‘municipio’ (municipality)
se3si4te43	‘CECYTE’

The pronunciation dictionary

Developed from the Triqui speech corpus with Spanish targets taken from an existing Spanish pronunciation dictionary.

X-SAMPA (full IPA accessible)

Vowel + /ʔ, h/ rime sequences treated as **single units**, e.g. /ni3yaj3/ > n i y **ah**

Glottalized consonants treated as single units, e.g. /ni4hya43/ > n i **?j** a

```
ni3tin4 n i t i~
ni3un32 n i u~
ni3ya3  n i j a
ni3ya32 n i j a
ni3yaj3 n i j ah
ni4aj4  n i ah
ni4hya43   n i ?j a
ni4hyaj1   n i ?j ah
ni4hyaj3   n i ?j ah
ni4hyun43 n i ?j u~
ni4ka43 n i k a
ni4kaj1 n i k ah
ni4kaj3 n i k ah
ni4koh1 n i k o?
ni4man1 n i m a~
ni4man43   n i m a~
ni4manj4   n i m a~h
ni4munh4   n i m u~?
ni4taj3 n i t ah
ni4un43 n i u~
ni4yaj3 n i j ah
```

Glottal stops get aligned terribly
(Data from Yoloxóchitl Mixtec)

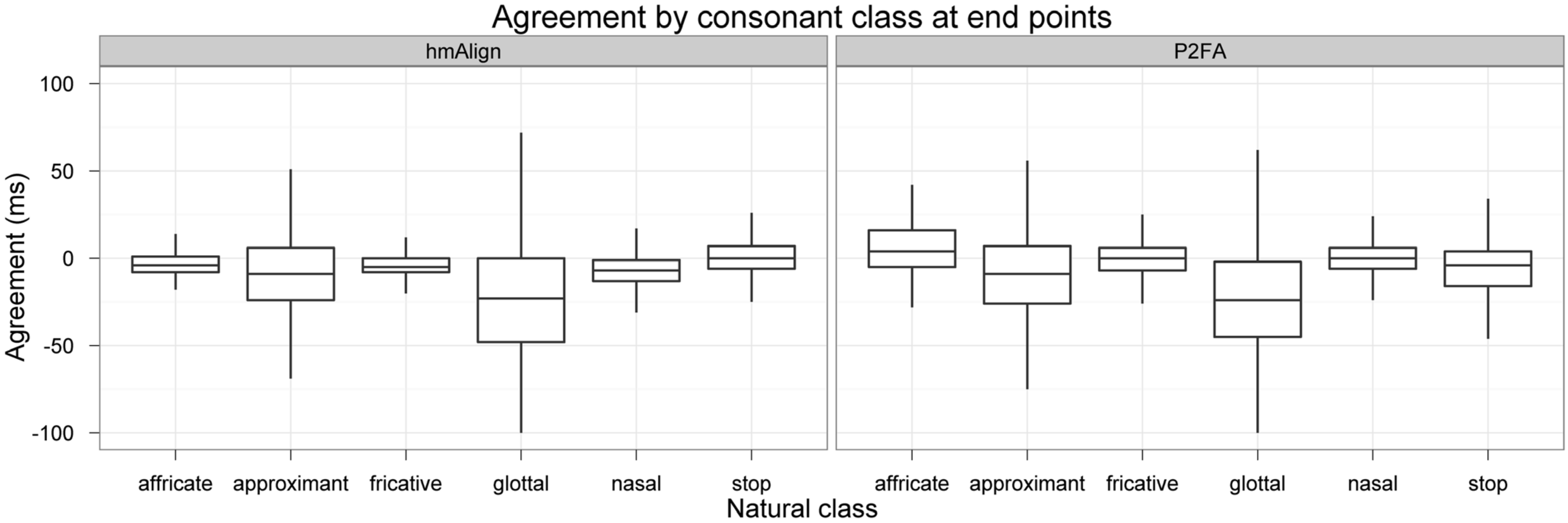


FIG. 2. Agreement for consonant classes across aligners.

DiCanio et al. 2013

Takeaways while building your own aligner

1. Consider your units of analysis down the line – you might not need to segment difficult things like vowel-vowel sequences or glottalization. *You need not think in terms of segments.*

2. Addressing encoding issues early on before you train an aligner will improve overall alignment of the speech signal. This reduces time spent manually adjusting alignment in Praat.

Questions?

Documentation corpora: processing via *additional annotation*

Assume you've run a transducer and trained an aligner. How else can we add rich data to a speech corpus?

I. Praat scripts can extract syntagmatic information (adjacent segments, encodings) to provide contextual information.

II. Dictionaries can be accessed to match lexical entries in the corpus and provide grammatical information.

III. Additional scripts for syllabification.

Praat scripts access rich data

Vowel_Acoustics_for_corpus_data_2.praat

Extracts formant data from corpus data alongside information about adjacent segments, syllable affiliation, word position, etc.

Vowel_data_texts_tidy

Speaker	Seg_num	File_nam	Gender.x	Seg.x	Dur	Seg_prev	Seg_subs	Syll.x	Syll_dur	Word.x	Word_du	File_nam	Gender.y	Seg.y	Syll.y	Word.y	t	t_m	COG
R1	1001	R1_0001_	M	aa	73	l	n	laan	163	la	75	R1_0001_	M	aa	laan	la	1	44	327.911
R1	1001	R1_0001_	M	aa	73	l	n	laan	163	la	75	R1_0001_	M	aa	laan	la	1	44	327.911
R1	1001	R1_0001_	M	aa	73	l	n	laan	163	la	75	R1_0001_	M	aa	laan	la	1	44	327.911
R1	1001	R1_0001_	M	aa	73	l	n	laan	163	la	75	R1_0001_	M	aa	laan	la	1	44	327.911
R1	1001	R1_0001_	M	aa	73	l	n	laan	163	la	75	R1_0001_	M	aa	laan	la	2	52	425.531
R1	1001	R1_0001_	M	aa	73	l	n	laan	163	la	75	R1_0001_	M	aa	laan	la	2	52	425.531
R1	1001	R1_0001_	M	aa	73	l	n	laan	163	la	75	R1_0001_	M	aa	laan	la	2	52	425.531
R1	1001	R1_0001_	M	aa	73	l	n	laan	163	la	75	R1_0001_	M	aa	laan	la	2	52	425.531
R1	1001	R1_0001_	M	aa	73	l	n	laan	163	la	75	R1_0001_	M	aa	laan	la	3	60	1074.31
R1	1001	R1_0001_	M	aa	73	l	n	laan	163	la	75	R1_0001_	M	aa	laan	la	3	60	1074.31
R1	1001	R1_0001_	M	aa	73	l	n	laan	163	la	75	R1_0001_	M	aa	laan	la	3	60	1074.31
R1	1001	R1_0001_	M	aa	73	l	n	laan	163	la	75	R1_0001_	M	aa	laan	la	3	60	1074.31
R1	1001	R1_0001_	M	aa	73	l	n	laan	163	la	75	R1_0001_	M	aa	laan	la	4	68	1897.61
R1	1001	R1_0001_	M	aa	73	l	n	laan	163	la	75	R1_0001_	M	aa	laan	la	4	68	1897.61
R1	1001	R1_0001_	M	aa	73	l	n	laan	163	la	75	R1_0001_	M	aa	laan	la	4	68	1897.61
R1	1001	R1_0001_	M	aa	73	l	n	laan	163	la	75	R1_0001_	M	aa	laan	la	4	68	1897.61
R1	1001	R1_0001_	M	aa	73	l	n	laan	163	la	75	R1_0001_	M	aa	laan	la	5	76	1594.77
R1	1001	R1_0001_	M	aa	73	l	n	laan	163	la	75	R1_0001_	M	aa	laan	la	5	76	1594.77
R1	1001	R1_0001_	M	aa	73	l	n	laan	163	la	75	R1_0001_	M	aa	laan	la	5	76	1594.77
R1	1001	R1_0001_	M	aa	73	l	n	laan	163	la	75	R1_0001_	M	aa	laan	la	5	76	1594.77
R1	1001	R1_0001_	M	aa	73	l	n	laan	163	la	75	R1_0001_	M	aa	laan	la	6	85	1018.34
R1	1001	R1_0001_	M	aa	73	l	n	laan	163	la	75	R1_0001_	M	aa	laan	la	6	85	1018.34
R1	1001	R1_0001_	M	aa	73	l	n	laan	163	la	75	R1_0001_	M	aa	laan	la	6	85	1018.34
R1	1001	R1_0001_	M	aa	73	l	n	laan	163	la	75	R1_0001_	M	aa	laan	la	6	85	1018.34

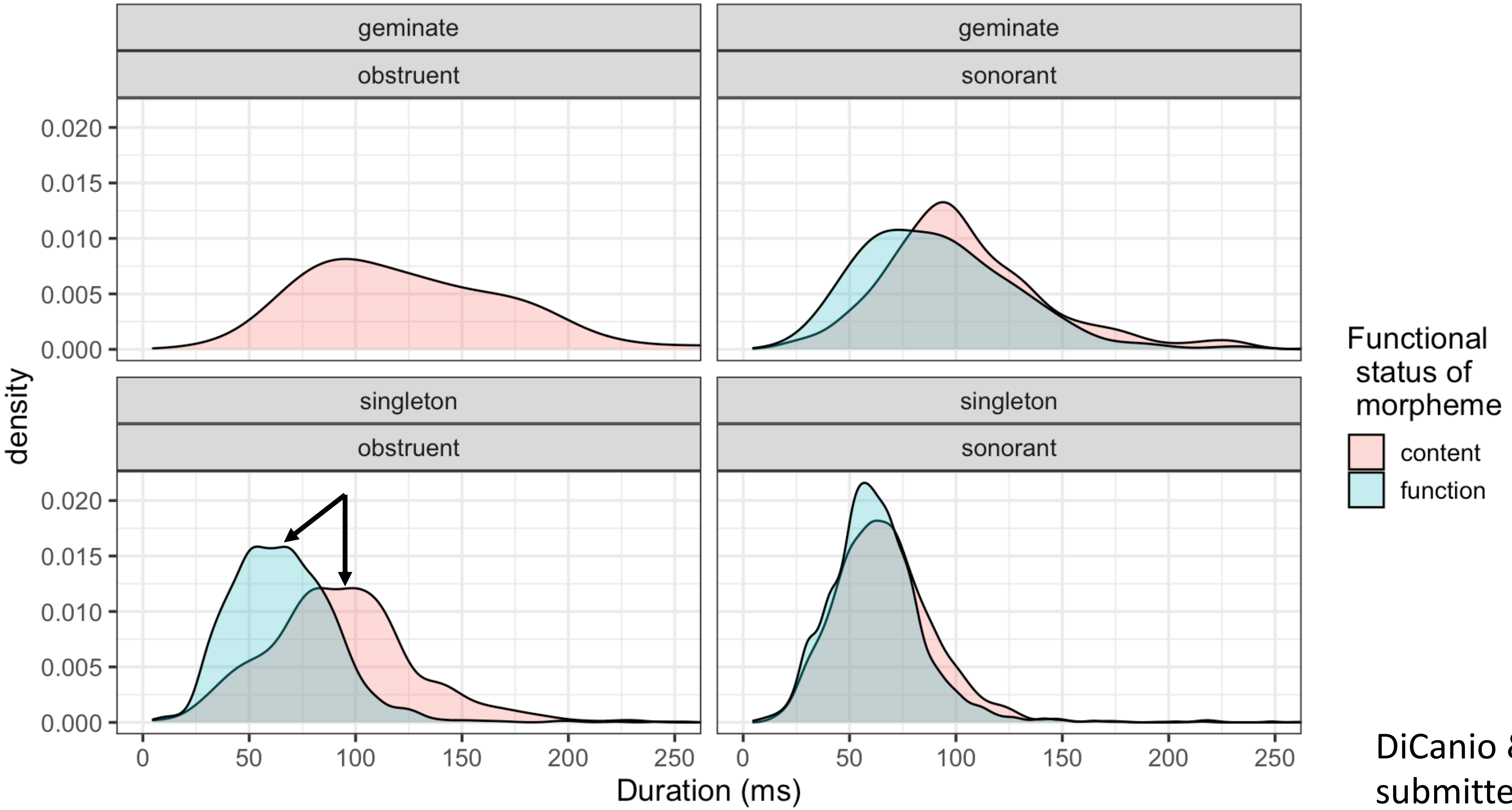
Utilize rich dictionaries

Flag words for content/function distinction as it tends to correlate with observed patterns of speech reduction (Gahl et al 2012).

Lenition_data_allcons

	A	B	C	D	E	F	G	H	I	J	K	L
1	file	speaker	labelNumbe	Word	**functional**	**Size**	consonant	contrast	length	manner	**before**	**post-pausal**
2	CCR_WCM_	CCR	3161	be4	function	monosyllabi	B	B	singleton	approximan	i~	no
3	CCR_WCM_	CCR	2037	yu3be32	function	polysyllabic	B	B	singleton	approximan	u	no
4	CMC_NLG_c	CMC	193	ka3bin3	content	polysyllabic	B	B	singleton	approximan	a	no
5	BMC_histori	BMC	309	chu4ba43	function	polysyllabic	B	B	singleton	approximan	u	no
6	BMC_histori	BMC	426	chu4ba43	function	polysyllabic	B	B	singleton	approximan	u	no
7	MJM_La_mu	MJM	792	be4	function	monosyllabi	B	B	singleton	approximan	sp	yes
8	BMC_histori	BMC	1624	be4	function	monosyllabi	B	B	singleton	approximan	e	no
9	BMC_histori	BMC	1354	chu3ba32	content	polysyllabic	B	B	singleton	approximan	u	no
10	ALM_Rana_	ALM	1681	be4	function	monosyllabi	B	B	singleton	approximan	sp	yes
11	CMC_NLG_c	NLG	90	stu3bej3	content	polysyllabic	B	B	singleton	approximan	u	no
12	ALM_Rana_	ALM	1302	chu4ba43	function	polysyllabic	B	B	singleton	approximan	u	no
13	MRO_Histori	MRO	2024	kbin3	content	polysyllabic	B	B	singleton	approximan	k	no

Functional factors influence durational variation in Triqui singleton obstruents, but not in singleton sonorants

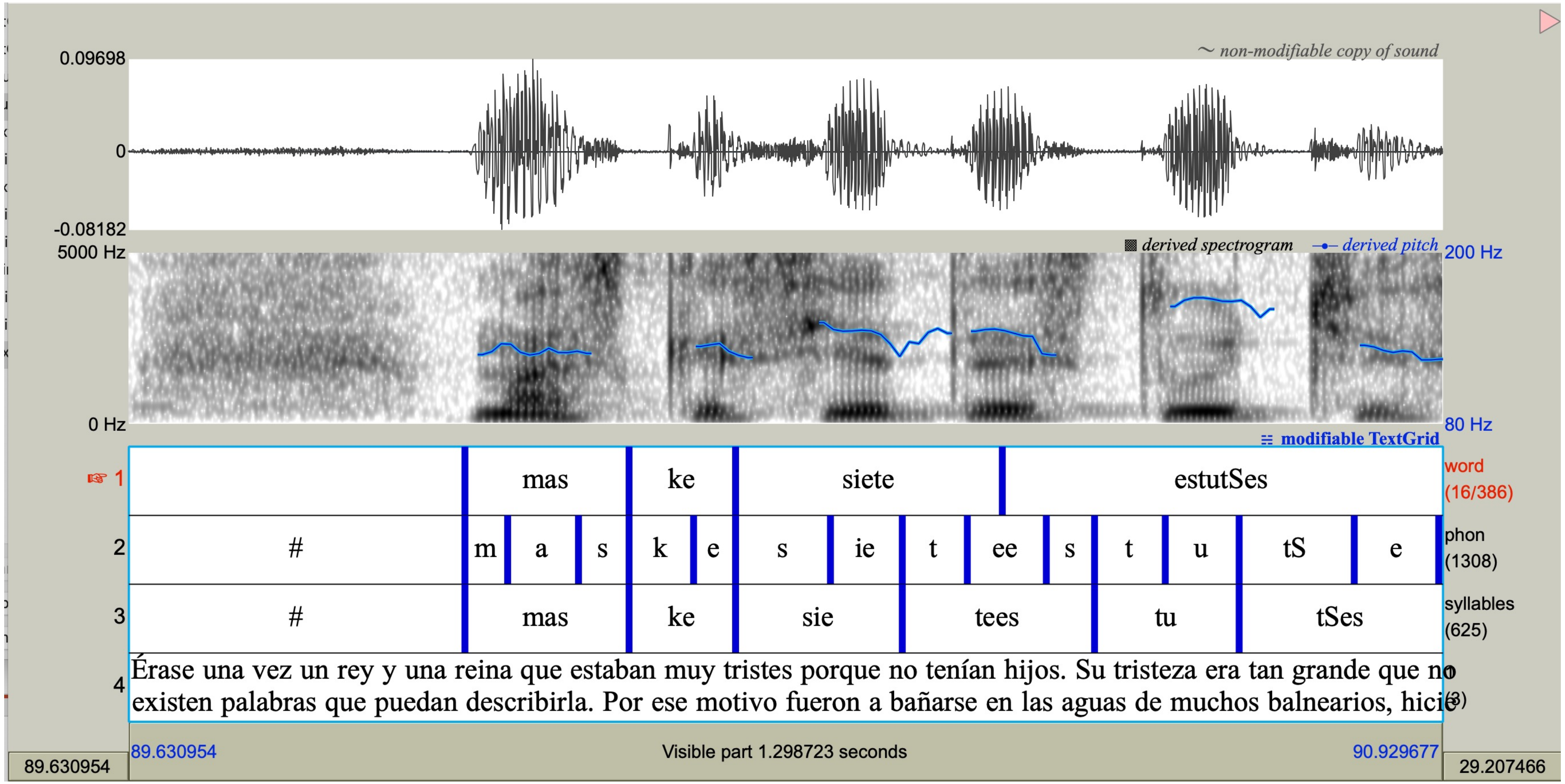


Approximately 7,000 obstruents were analyzed from 90 minutes of corpus data.

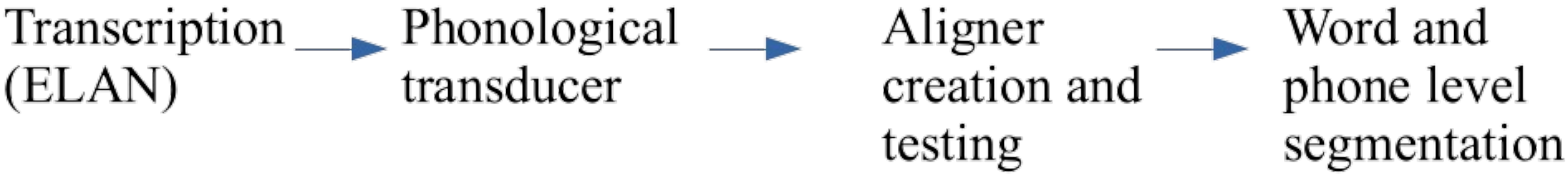
DiCanio & Sharp, submitted

Syllabification of the speech signal

- Accomplished either via Python script or Praat script.
- Syllabifier_for_Spanish_full.zip (demo)



Conclusions for Part I
(processing documentatary recordings)



1. Creation of segmentable corpus for alignment.
2. Alignment of corpus
3. Additional annotation of corpus

But what about the scientific questions? (Part II)

Question period for Part I.

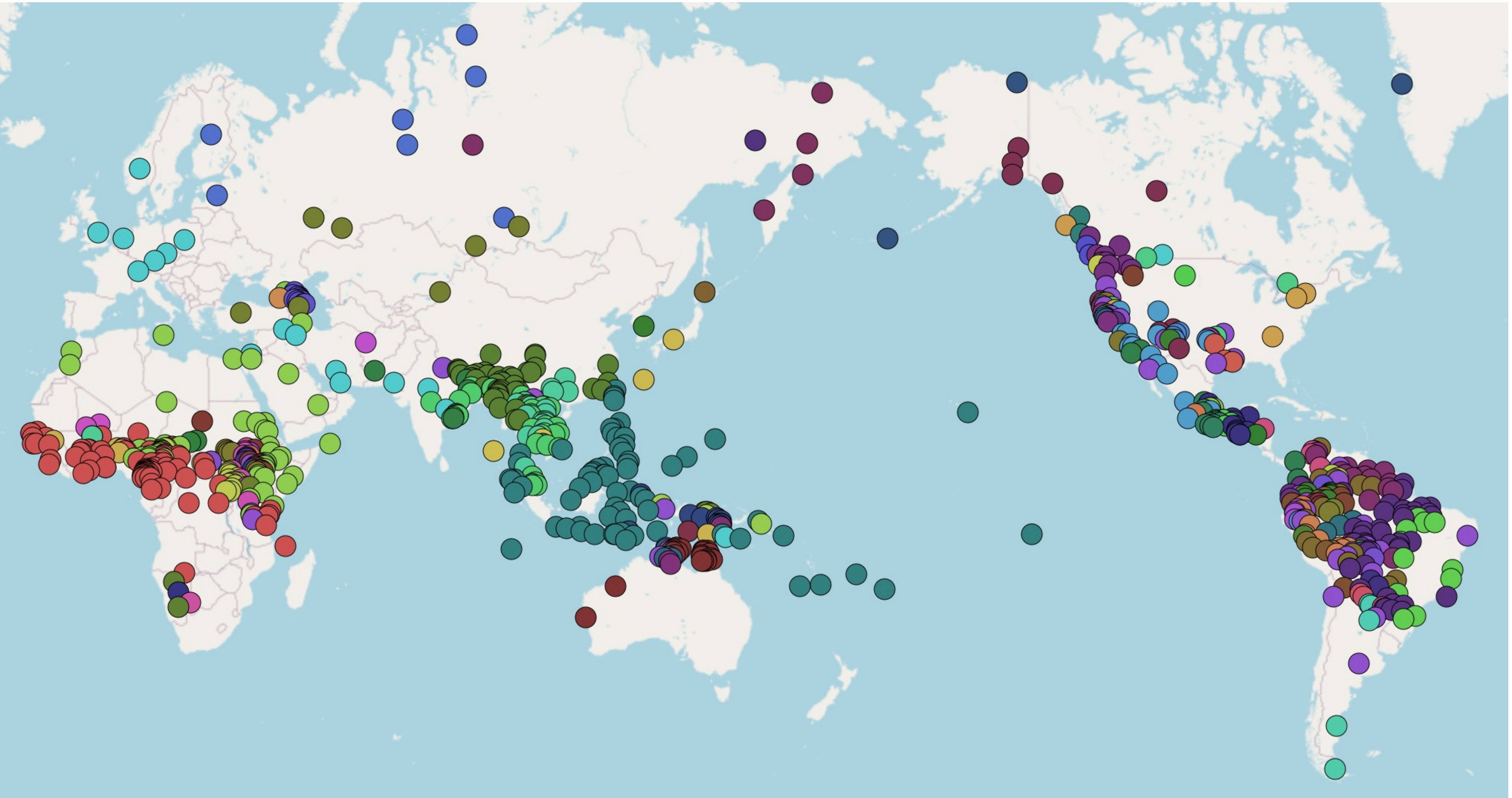
And a 15 minute break after questions

Part II: Addressing phonetic questions with language documentation corpora: case studies

1. Theoretically/typologically-motivated research
   Glottalization in spontaneous speech in Itunyoso Triqui (in progress)

2. Exploratory research
   Consonant reduction in Yoloxóchitl Mixtec (DiCanio et al 2022)

2.1 Glottal stop variation



Glottal stops are frequent in the languages of the world.

They are listed as allophones in 1131/3020 (37.5%) of the consonant inventories in PHOIBLE (Moran & McCloy 2019).

What qualifies as a glottal stop?

- In **isolated words** produced during **careful speech**, a glottal stop is often produced with full glottal closure.

- What counts as full glottal closure though? 20 ms silence during closure? 40 ms?

- Phonological descriptions of languages of the world frequently focus on carefully-produced allophones produced in elicitation contexts. This obscures observed variation.

  (DiCanio et al 2015, 2022; DiCanio & Whalen 2015, Keating & Huffman 1984, Koopmans van Beinum 1980)

Variation in speech production is the norm *regardless of phonological status*

- Three major phonetic variants of glottal stops:
  - **full stops** [aʔa]
  - **non-modal phonation** – either creak or vocal fry or diplophonia [aa̰a]
  - **omitted** [aː] (but do subtle intensity perturbations count?)

- In languages where glottal stops are unequivocally phonemes, they are usually just produced as non-modal phonation and no closure, e.g.
  - Arapaho /ʔ/: full glottal closure 25-27% of the time, omitted 13-18% of the time, non-modal phonation 55-62% of the time (Whalen et al 2016).
  - Hawaiian /ʔ/: full glottal closure just 7% of the time, omitted 20% of the time, non-modal phonation 73% of the time (Davidson 2021).

Variation, duration, and prosodic boundaries

- Full glottal closure occurs more often in contexts of increased duration whereas non-modal phonation occurs more often in shorter duration contexts.
    - Ixcatec glottalized stops (DiCanio 2011)
    - Arapaho glottal stops (Whalen et al 2016)
    - Hawaiian glottal stops (Davidson 2021)

- Full glottal closure is more common at prosodic boundaries in Itunyoso Triqui (DiCanio 2012).

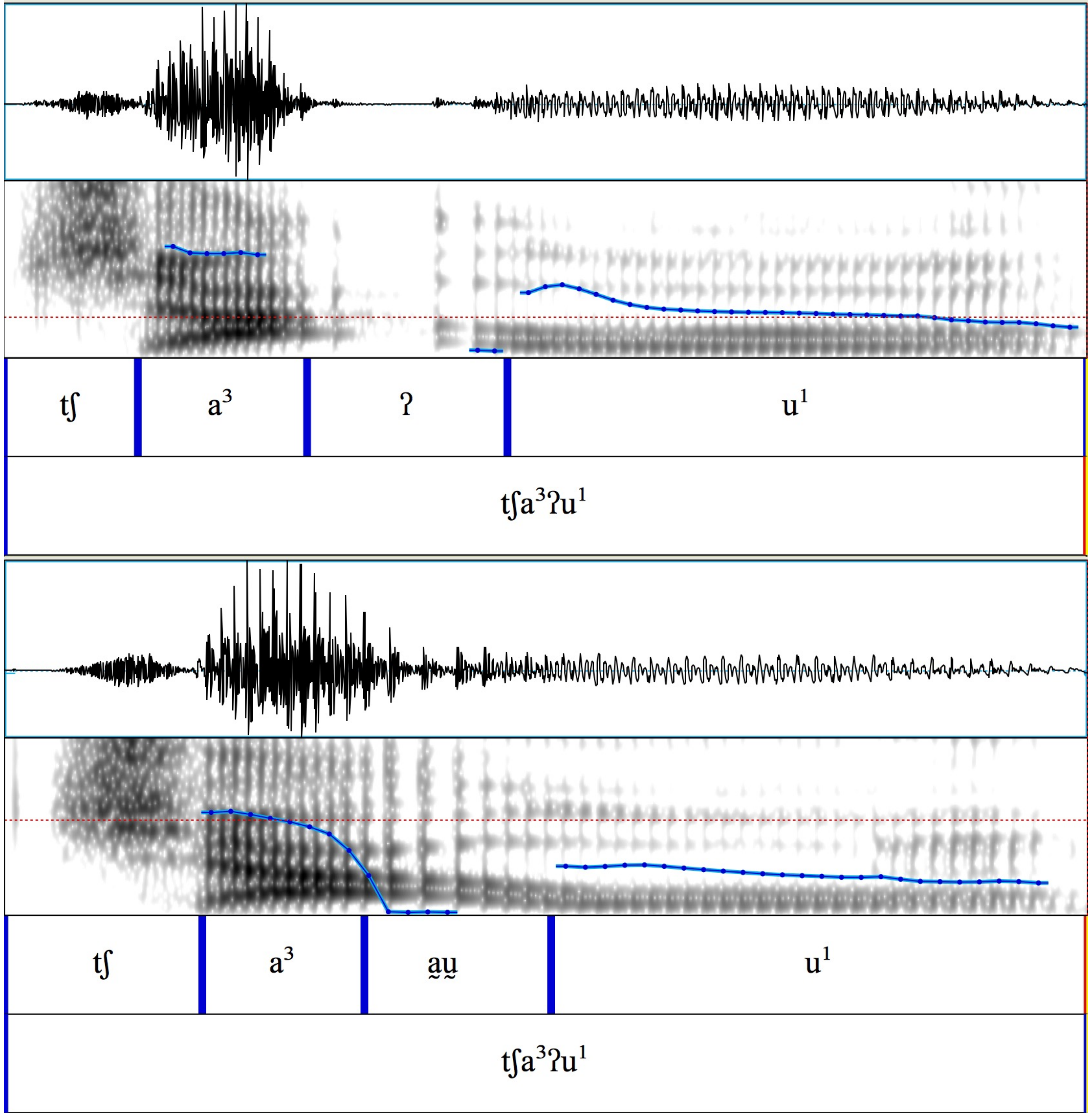
- Incomplete closure is a type of *speech reduction*, but it is far more common with glottal stops than with oral stops.

Variation I

In Itunyoso Triqui (Otomanguean), medial glottal stops can be realized with complete closure (above)

or without complete closure (below)

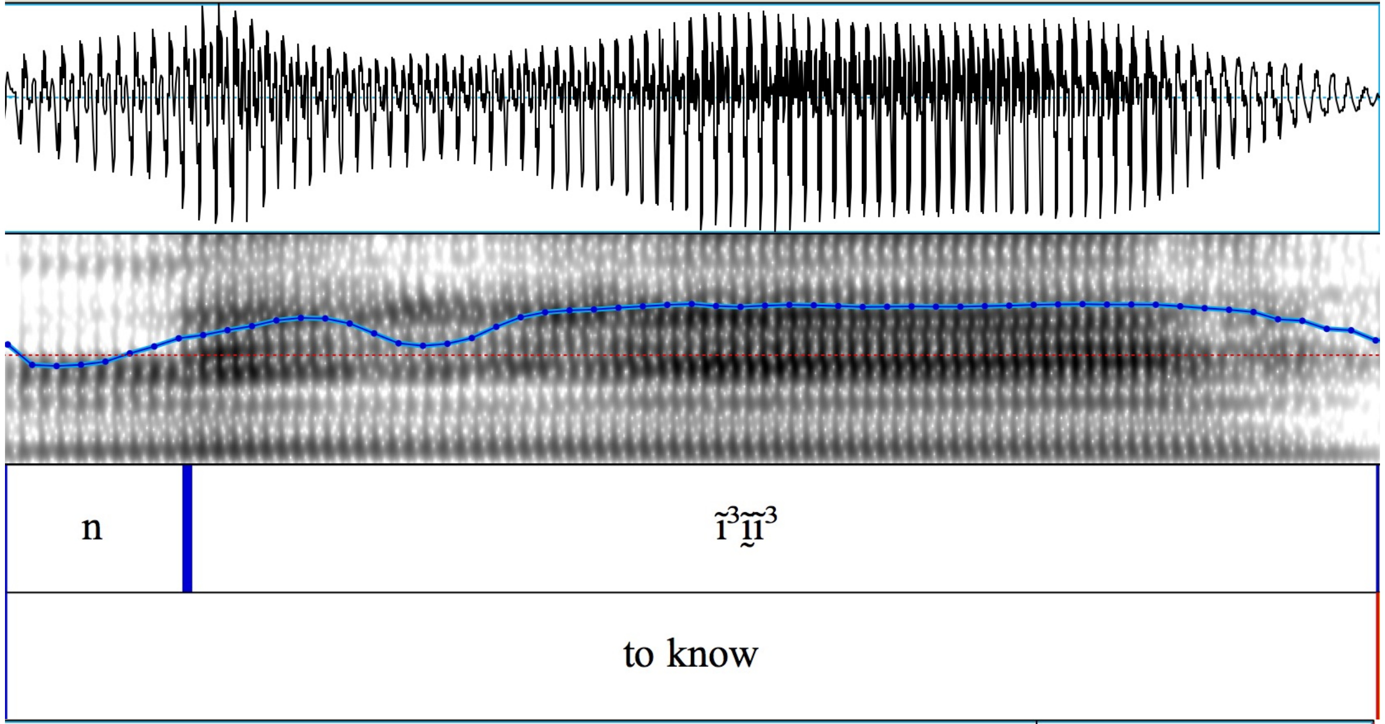
/tʃa^{3}ʔu^1/
'barn owl / tecolote'



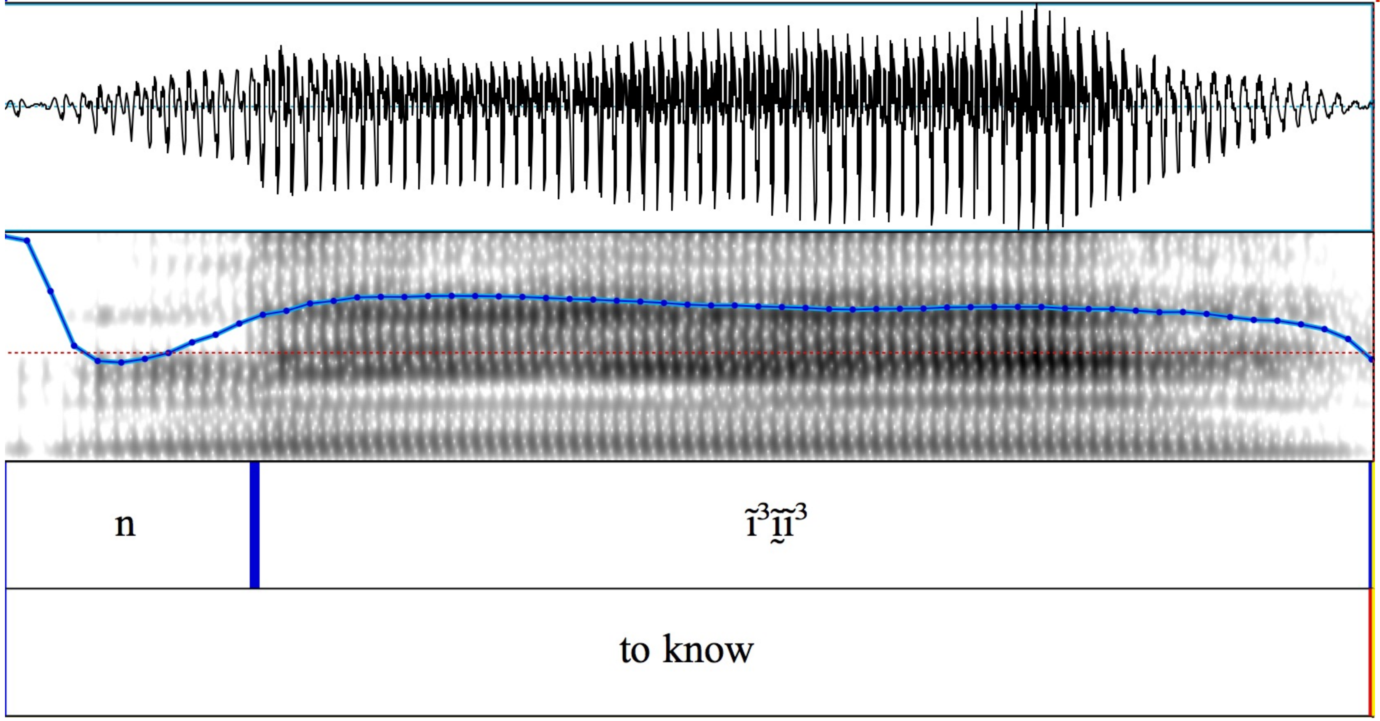
Both sound like they have closure, right?

Variation II

But medial glottal stops can *also* be realized with non-modal phonation (creak, vocal fry, diplophonia)



or only very subtle visible changes in voice quality



/ni³ʔiː³/

'to know'

'saber, conocer'

Previous accounts of production variation

In the great majority of languages we have heard, glottal stops are apt to fall short of complete closure, especially in intervocalic positions.

p.75, Ladefoged and Maddieson (1996) (cited in Davidson, 2021)

Also, it is not uncommon to find segments described as glottal stops surfacing phonetically in one of the following articulatory forms: (1) glottal stops or creaky transitions; (2) glottal approximants; (3) voiced laryngealized transitions between vowels. Some of the questions raised by these facts are: if glottal stops are not always phonemes or stops, what are they? Features? Suprasegmentals? Are they linked to stress or syllable boundaries? Do they operate at larger prosodic levels, for instance, the prosodic word or phonological phrase?

p.352, Storto & Demolin (2012) (on South American languages)

Small perturbations = "glottalization"

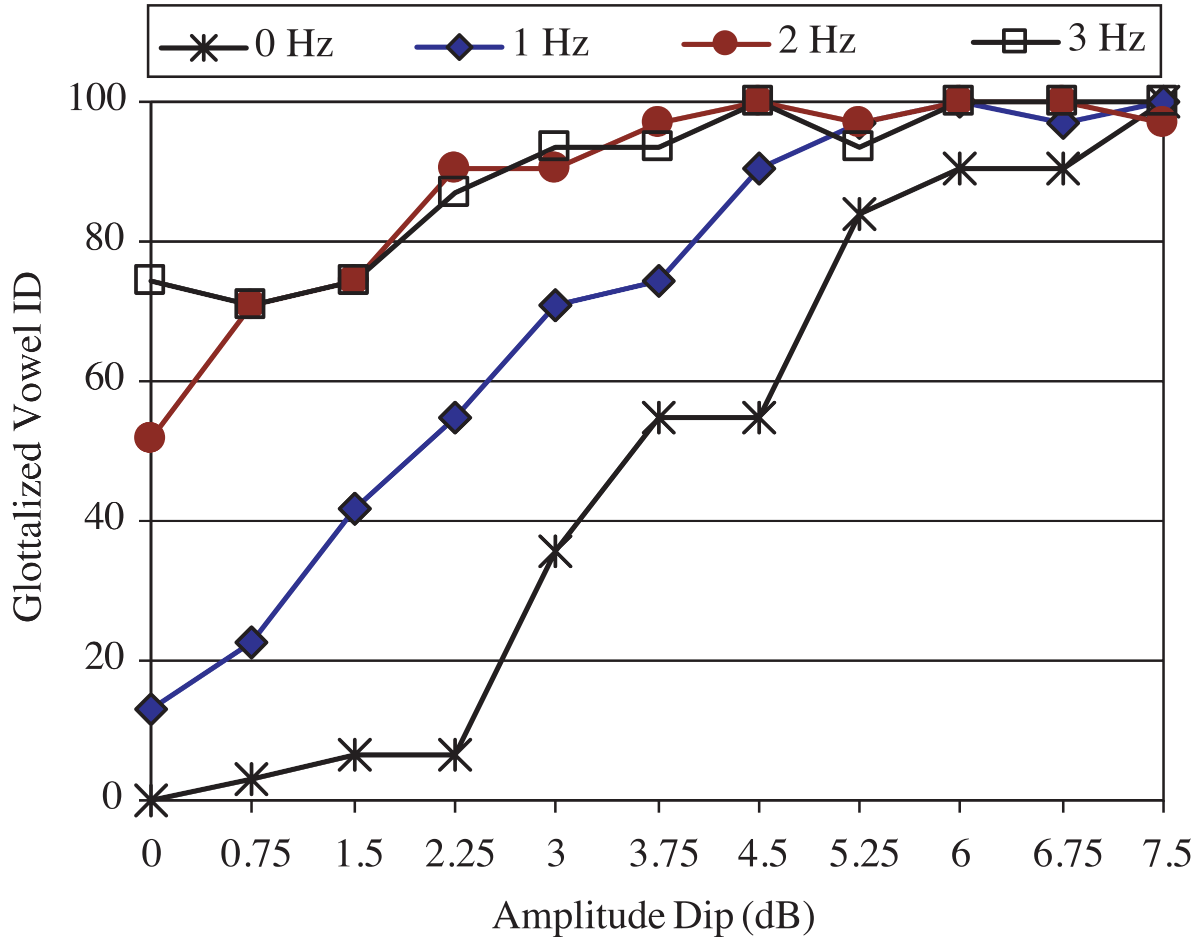


Figure from Gerfen & Baker (2005) with Coatzospan Mixtec listeners

Speakers use small, local f0, intensity and H1-H2 perturbations as cues to the presence of glottalization in intervocalic position (Gerfen & Baker 2005, Hillenbrand & Houde 1996).

Listeners can use even more subtle cues – just a 2Hz, 3dB perturbation induces the perception of glottalization at >90% in Coatzospan Mixtec listeners.

Why does this phonetic variation matter?

- Non-modal phonation as *the primary realization* of a glottal stop involves overlap with adjacent speech sounds, making segmentation/parsing particularly challenging.

- In many languages, glottalization and glottal stops have a high functional load, meaning that it is important for morphological identification.

- Acoustic landmark detection of non-modal phonation often requires accurate f0 estimation (e.g. spectral tilt), but it is precisely during non-modal phonation that *one can not reliably estimate f0*.

It is important for speech but a very difficult thing to recognize/detect.

(see also – nasalization)

Approach – human categorization

- Several recent studies have attempted to categorize variation in the production of glottal stops (Davidson 2021, Whalen et al 2016).

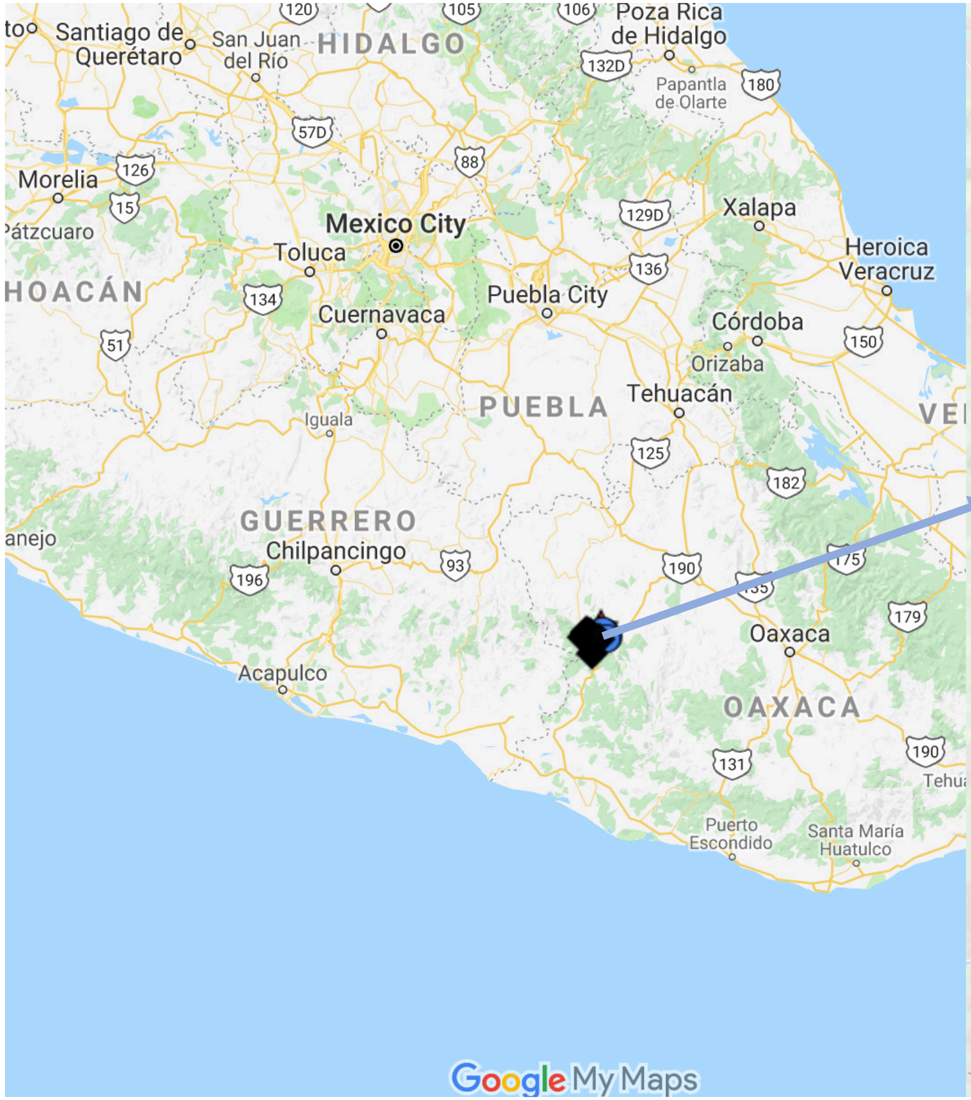
- Coming up with reliable inter-transcriber categories is a challenge, but it is also rewarding since computational models can be trained on them.

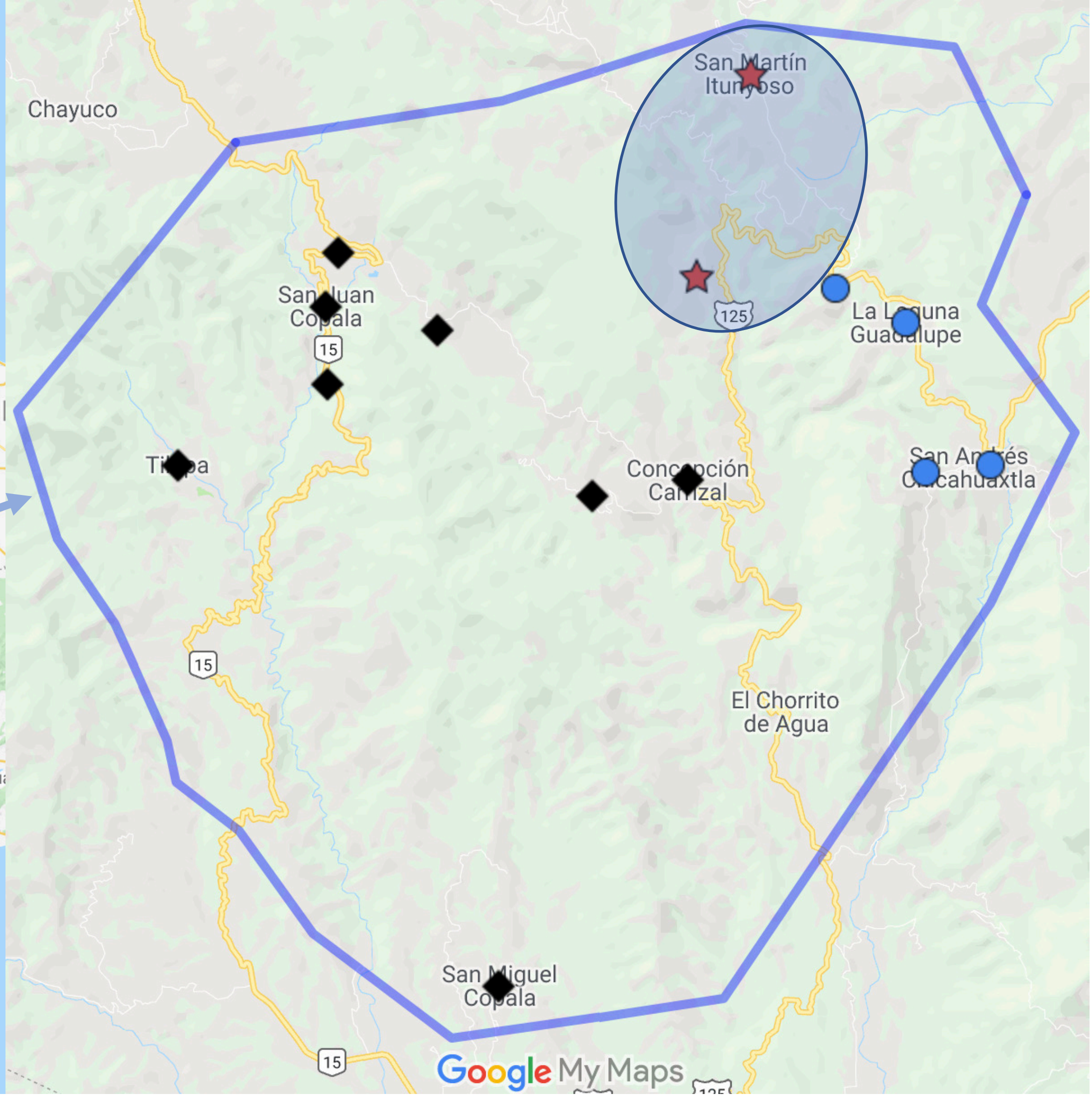
- A DNN trained on minimal human categorization of stop allophones is able to accurately categorize lenited stop allophones at over 90% accuracy (DiCanio, Chen, Benn, Amith, and Castillo García, 2022).

II. Test case – Itunyoso Triqui (Otomanguean)

Otomanguean languages, spoken Southern Mexico, have contrastive glottalization and/or glottal stops.

Itunyoso Triqui – about 2,500 speakers





Tone and syllable structure

All syllables are open in non-final syllables, but two glottal consonant codas are possible in final syllables. There are nine lexical tones.

	Open syllable		Coda /h/		Coda /ʔ/	
Tone	Word	Gloss	Word	Gloss	Word	Gloss
/4/	$yũ^{4}$	‘earthquake’	$yãh^{4}$	‘dirt’	$niʔ^{4}$	‘see.1DU’
/3/	$yũ^{3}$	‘palm leaf’	$yãh^{3}$	‘paper’	$tsiʔ^{3}$	‘pulque’
/2/	$ũ^{2}$	‘nine’	tah^{2}	‘delicious’	$ttʃiʔ^{2}$	‘ten’
/1/	$yũ^{1}$	‘loose’	$kãh^{1}$	‘naked’	$tsiʔ^{1}$	‘sweet’
/45/			toh^{45}	‘forehead’		
/13/	yo^{13}	‘fast (adj.)’	toh^{13}	‘a little’		
/43/	$ɾa^{43}$	‘want’	$nnãh^{43}$	‘mother!’		
/32/	$ɾã^{32}$	‘durable’	$nnãh^{32}$	‘cigarette’		
/31/	$ɾã^{31}$	‘lightning’				

Contour tones do not occur before a coda /ʔ/, but otherwise, tone and glottal consonants are orthogonal.

Glottal features in Itunyoso Triqui

Phonemic glottal stops occur

- intervocalically as onsets of root-final syllables, e.g. [ɾã^{3}ʔãː^{3}] ‘mushroom’
- in codas of root-final syllables, e.g. [ɾãʔ^{3}] ‘to happen’

Pre-glottalized sonorants occur

- in word-initial position of monosyllables, e.g. [ˀnĩɦ^{32}] ‘to hold up’
  [nĩɦ^{32}] ‘to fall down’
- intervocalically as onsets of root-final syllables, e.g. [ba^{1}ˀnĩ^{3}] ‘three of them’

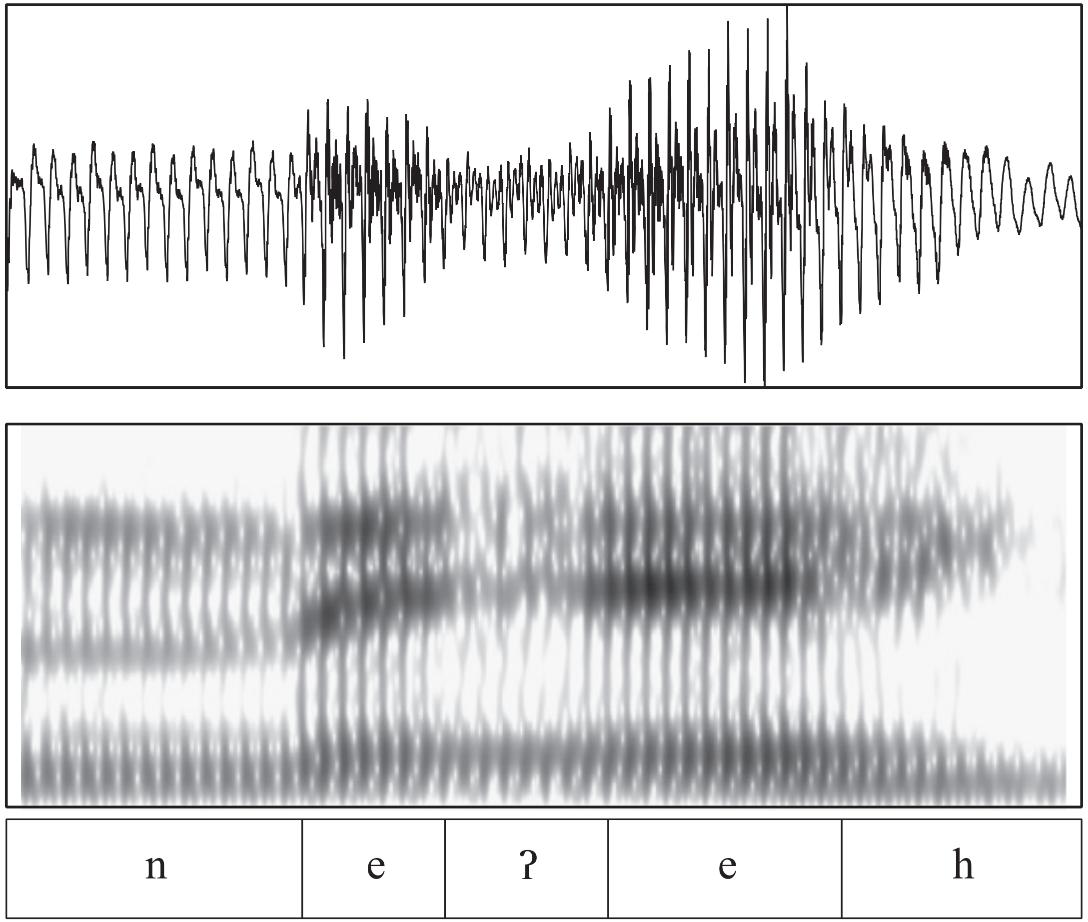
Previous phonetic research

- Intervocalic glottal stops are more rarely produced with full stop closure than coda glottal stops (DiCanio 2012).

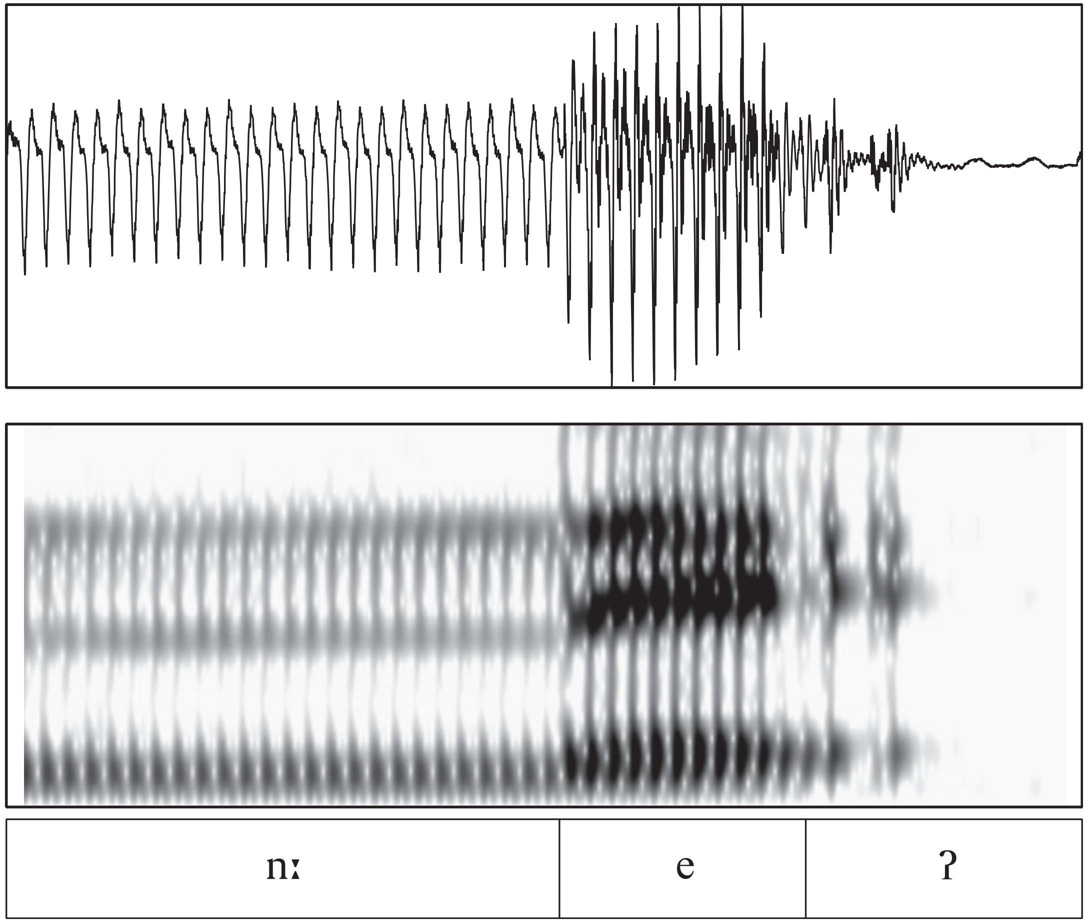
- Vowels preceding a coda glottal consonant are much shorter than vowels in open syllables and listeners rely on vowel duration to as a cue for coda glottalization (DiCanio 2012, 2014).

- Additional acoustic cues to glottalization include corrected spectral slope measures like H1-H2, H1-A3 (DiCanio 2012).

Lenited vs. non-lenited glottal stops



ne³ʔeɦ³ ‘child’



nneʔ³ ‘straw rope’

Figures from DiCanio (2012)

Glottalization and tone

- Glottalization perturbs f0 targets. This predicts that for languages where f0 is also contrastive, tone and glottal targets may be timed so as to prevent acoustic overlap and maintain tonal recoverability (Silverman 1997).
- Caveat – linguists are likely to transcribe glottalization as [ʔ] ~ [ˀ]. Transcriptional practice biases phonological accounts to assume separability, i.e. if it's transcribed as [ʔ] instead of [a̰], it must always be a stop, right?
- For Triqui, prosodic position appears to play a more important role in glottal stop realization than tone.

III. Categorization

- Human categorization of allophony in running speech is useful not only for descriptive/empirical purposes but also for the construction of predictive models trained on the acoustics.

- Arapaho classification of glottal stop realizations (Whalen et al 2016)
  Full glottal stop
  Creaky phonation.
  Creaky phonation with local glottalization.
  No glottalization.
  "Problematic" (hard to classify)

Categorization of Triqui glottal stops

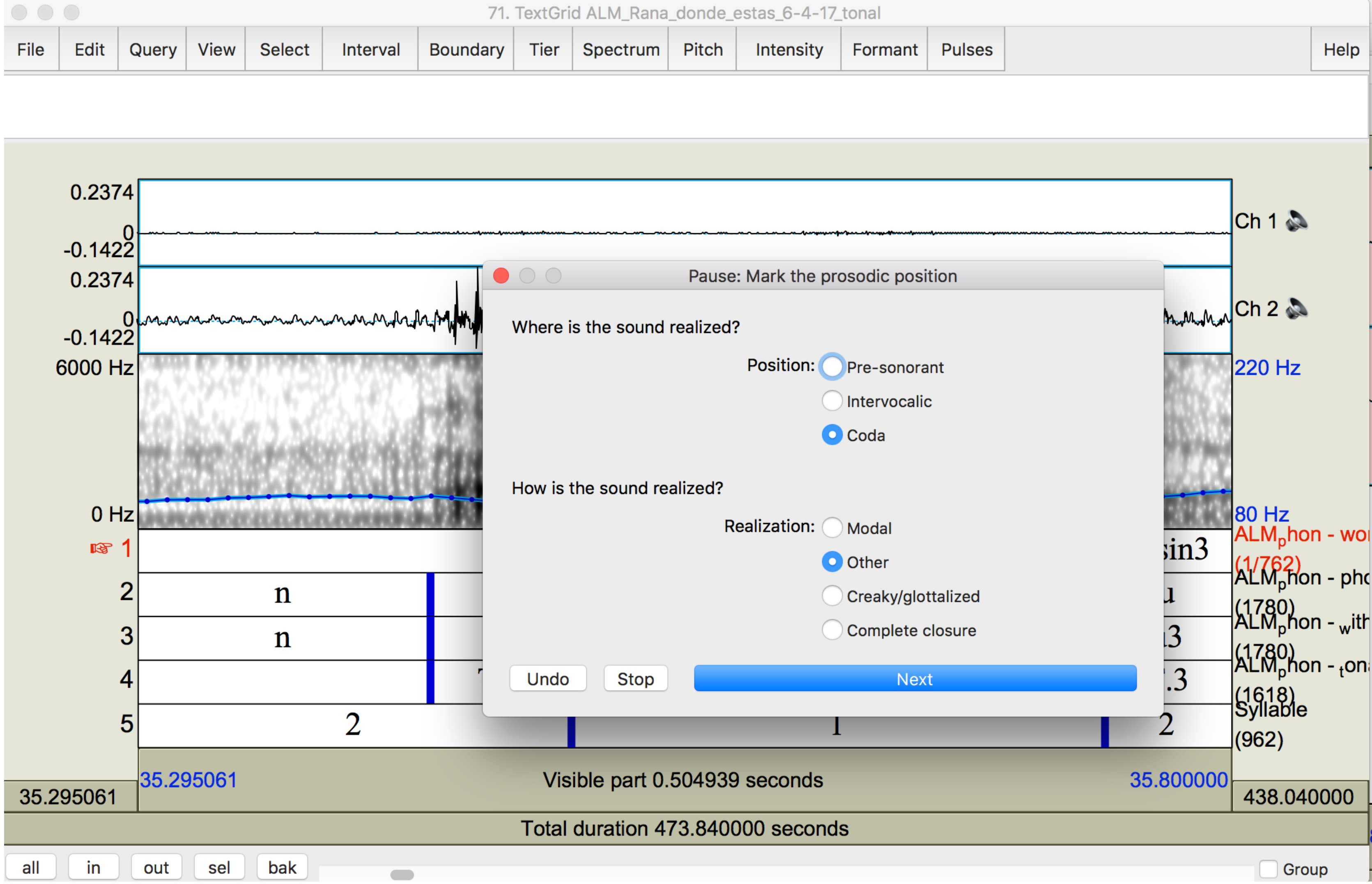
- What does the variation in Triqui spontaneous speech look like?

- The Itunyoso Triqui documentation corpus contains about 30 hours of transcribed recordings.

- Results from the initial categorization of some of the hand-corrected portion of the aligned corpus, done by three categorizers: Christian DiCanio (UB), Richard Hatcher (Hanyang U.), and Lisa Davidson (NYU)

Methods

We used a categorization script written for Praat that allows users to visually examine an annotated portion of the speech signal, click the qualitative judgment, and then proceed to the next annotation.

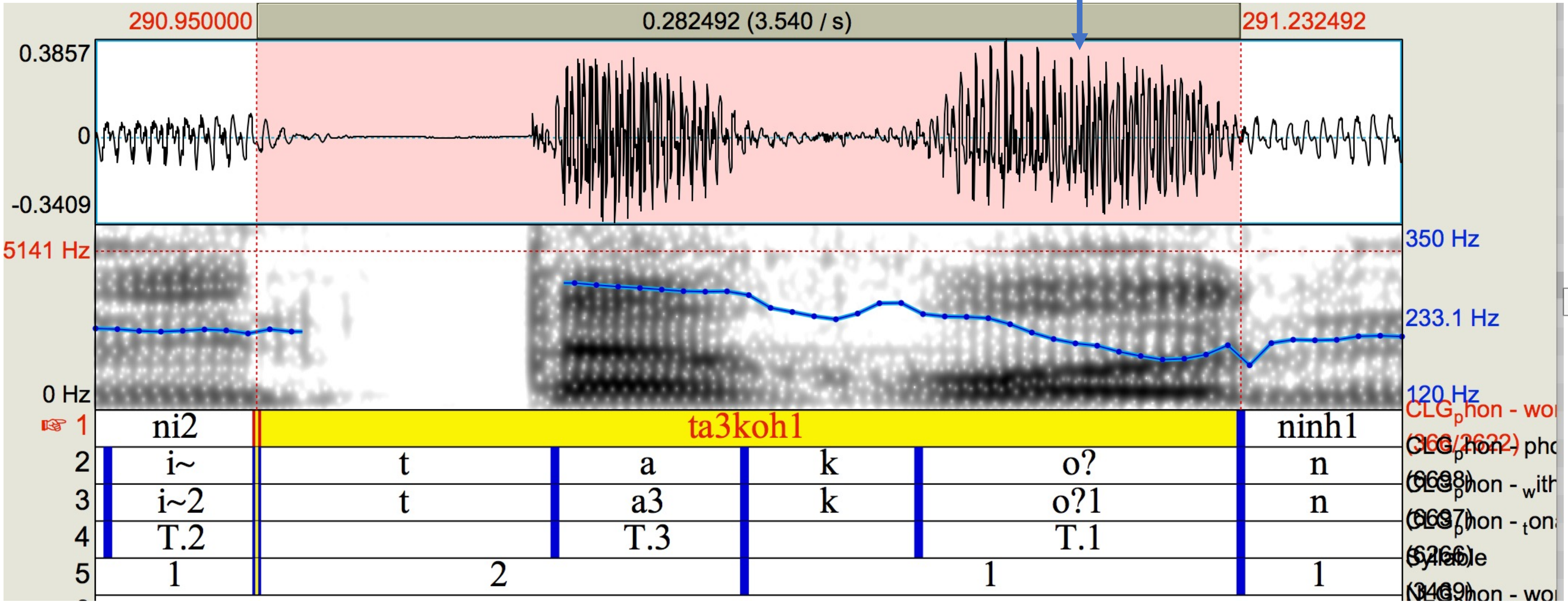


Spontaneous speech sample

- We examined the text “Derechos de Mujeres Triquis” - a 14 minute 15 second conversation between two women (who speak somewhat quickly).
- There were **604 glottal stops** in the recording.
- Five categories were used:
  1. Glottal stop (closure at least 20 ms)
  2. Aperiodic – some clear loss of periodicity
  3. Perturbation – periodic but with some intensity/f0 dip
  4. Modal – periodic with no apparent change in intensity/f0
  5. Unsure

Perturbation

Increased shimmer in production



ta^{3}koʔ^{1} ‘hangs’

Inter-categorizer reliability

Richard and Christian – 70.8% agreement

Christian and Lisa – 69.4% agreement

Lisa and Richard – 67.5% agreement

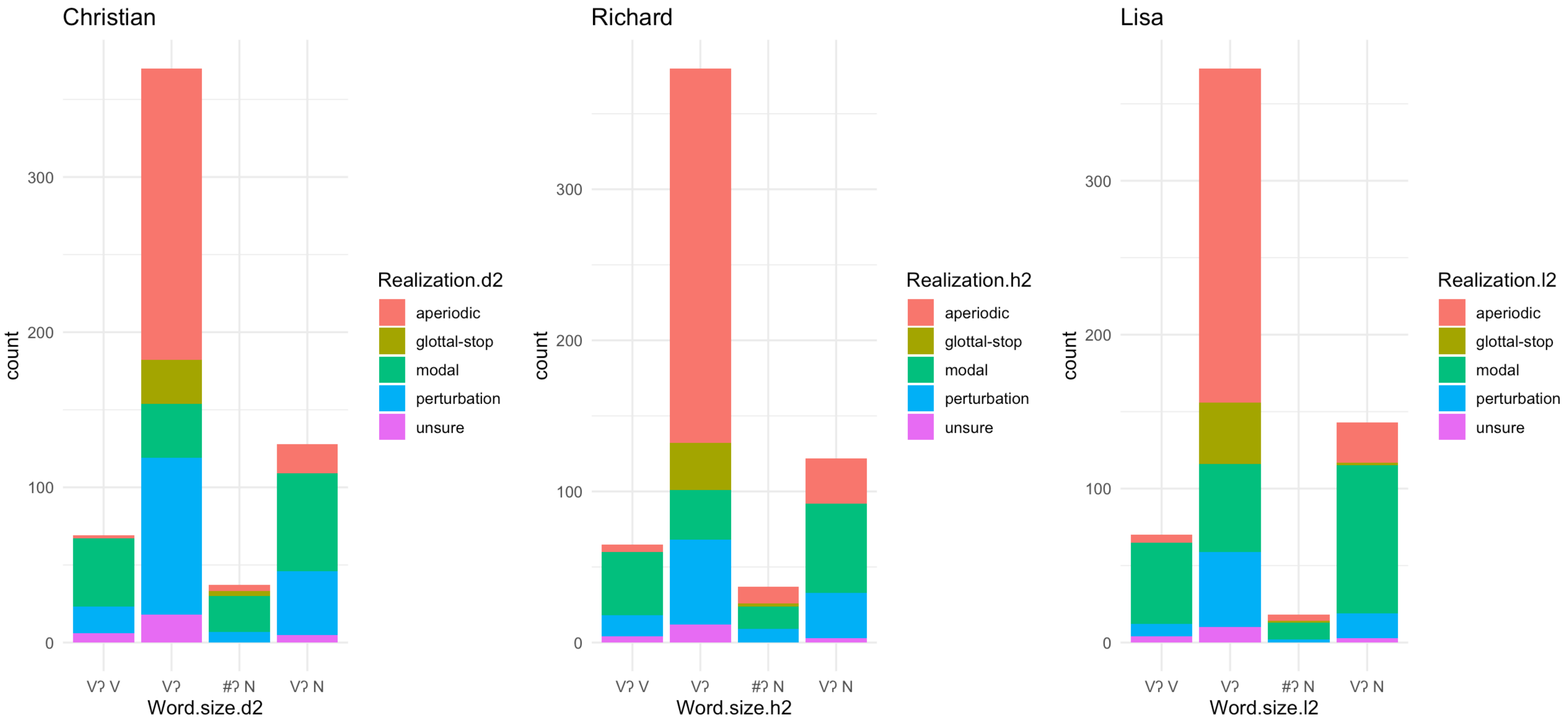
Most of the disagreement was between the categories “aperiodic” and “perturbation.” If we merge these categories, we agree more often.

Richard and Christian – 84.3% agreement

Christian and Lisa – 77.0% agreement

Lisa and Richard – 76.3% agreement

Our categorizations



How often is a glottal stop a stop? Rarely.

Realization	**Frequency across categorizers**
Glottal stop	5.1 – 7.1%
Aperiodic	35.2 – 48.7%
Perturbation	12.4 – 27.5%
Modal	24.6 – 35.9%
Unsure	2.8 – 4.8%

Prosodic influence on production

- 28-40 realizations of the complete glottal stops occurred in root-final position, but <u>none</u> occurred in intervocalic position.

- This finding mirrors past work suggesting that prosodic position plays a role in the realization of glottal stops (cf. Davidson 2021).

Does the presence of tonal contrast matter?

- The proportion of glottal stop realizations (5-7%) is very close to values reported for Hawaiian (Davidson 2021).
- Hawaiian is non-tonal and Itunyoso Triqui is heavily tonal.
- The degree to which glottal stops are realized as non-modal phonation seems to independent from whether they may influence tone, contra Silverman (1997).

2.1 Findings – corpus phonetics

- Utilized human categorization of groupings via a Praat script.

- Phonetic variation does not always have to be captured via continuous phonetic measurements; it is possible to quickly mark variants.

- Typologically-motivated research question addressed with a documentation corpus.

Questions?

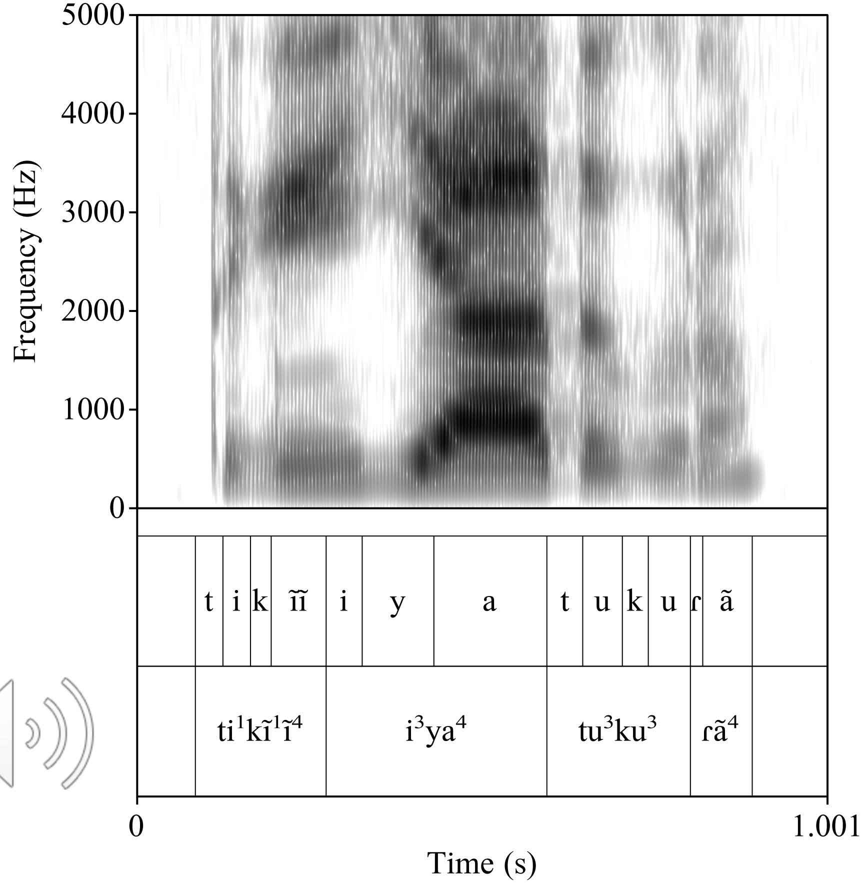
2.2 Exploratory phonetic research with documentation corpora (DiCanio et al 2022)

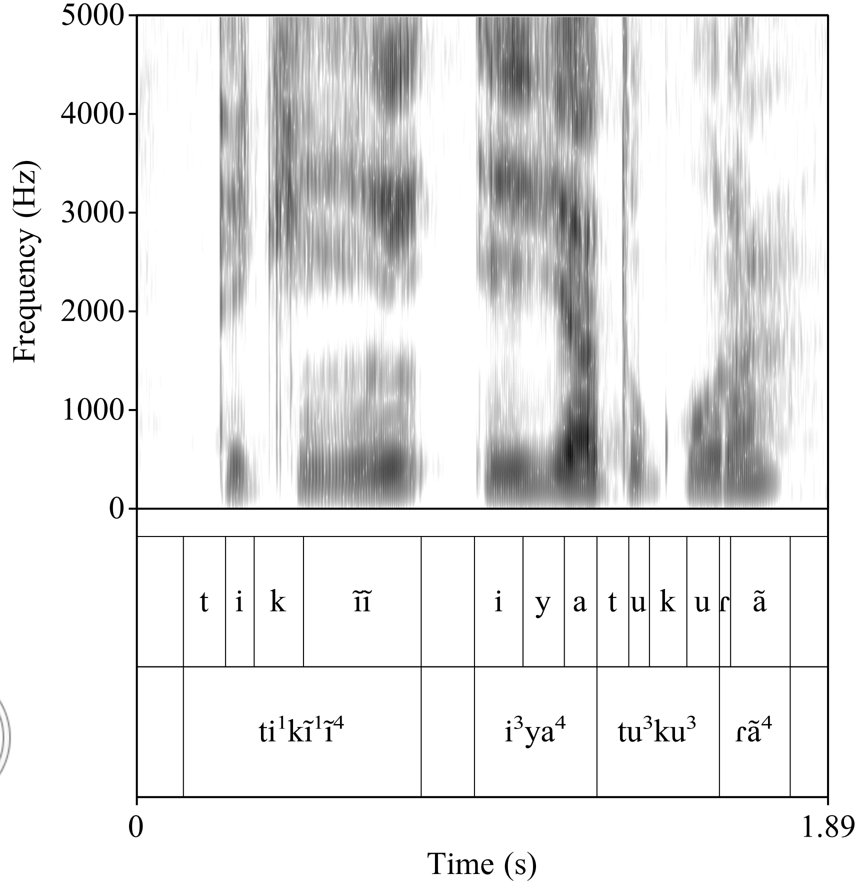
- Yoloxóchitl Mixtec (Otomanguean: Mixtecan) is an endangered language spoken in Guerrero, Mexico.
- Large-scale language documentation project (Amith and Castillo García, no date).
- Roughly 100+ hours of recordings that have been transcribed, transduced, and force-aligned.
- Complex tone language with a noticeable pattern of obstruent reduction.

Why are the obstruents pronounced so differently in these sentences?

[ti^{1}kĩ^{1}ĩ^4 i^{3}ja^4 tu^{3}ku^4 rã^4] (left) vs. [ti^{1}kĩ^{1}ĩ^4 i^{3}ja^4 tu^{3}ku^4 rã^4] (right)

‘...the sour tamale again, then.’





Obstruent reduction

Relative to carefully-produced speech, an obstruent is produced with **reduced spatial excursion** of articulators and reduced constriction degree.

Reduced obstruents are shorter than carefully-produced ones (Lavoie, 2001; Parrell and Narayanan, 2018)

Reduced voiceless obstruents may undergo a process of **passive voicing** (Beckman et al., 2013; Davidson, 2018; DiCanio, 2012; Schwarz et al., 2019, Stevens, 2000; Westbury and Keating, 1986).

But no reduction in elicited speech!

- DiCanio et al. (2020) did not find much lenition of obstruents in careful speech. Stop closure was found for all speakers.

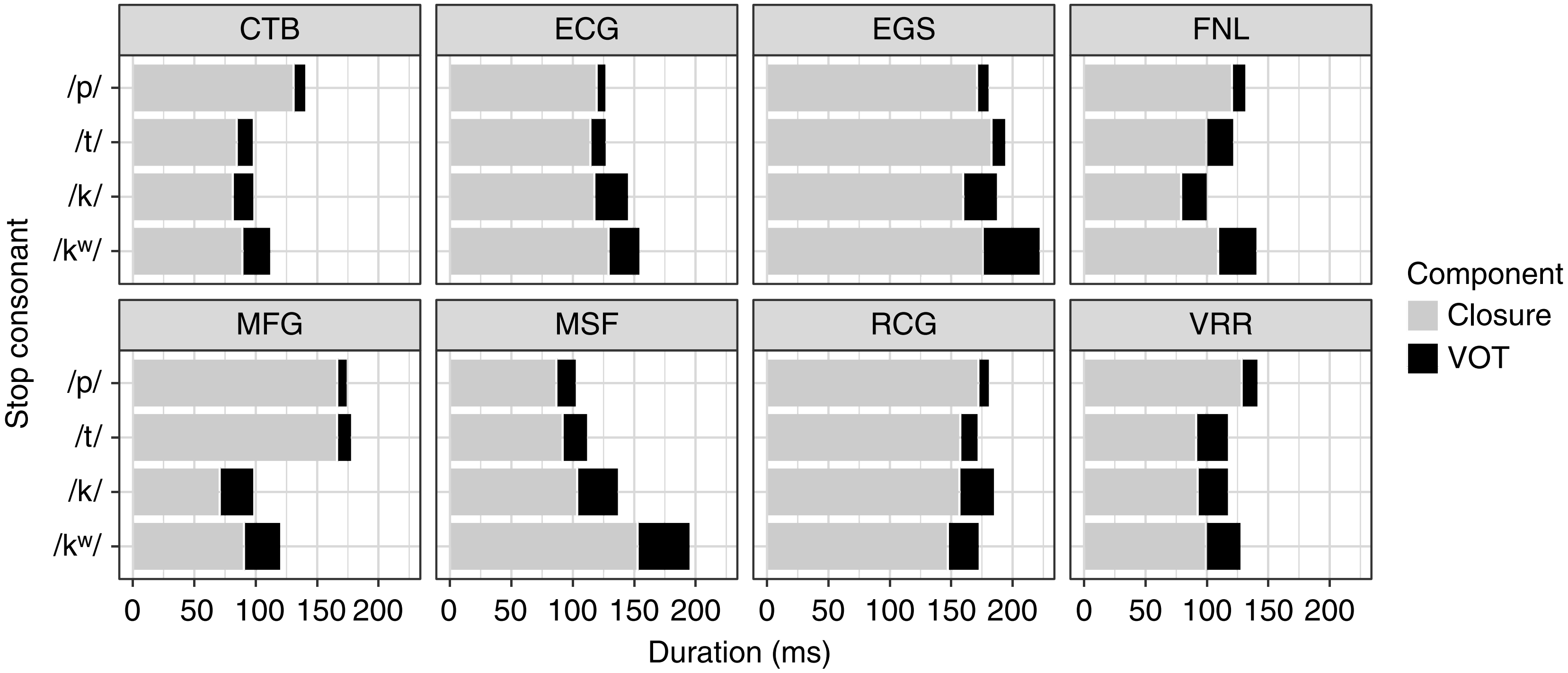


Figure 7 Stop component duration across speakers.

What is behind this lenition?

Possible factors:

- prosody/stress – perhaps the onset of the final (stressed) syllable is not reduced but all other obstruents are reduced
- stop type – perhaps some obstruents are more likely to lenite

But this would be rather odd – because of final stress there would be more reduction in *word-initial* position than in word-medial position.

Exploratory study using (a) categorical measures of lenition and (b) continuous measures of lenition/reduction

Consonant inventory – no voicing contrast

Table 1 Yoloxóchitl Mixtec consonant inventory.

	Bilabial	Dental	Alveolar	Post-alveolar	Palatal	Velar	Labialized velar
Plosive	p	t̪				k	kʷ
Pre-nasalized plosive	(ᵐb)		(ⁿd)			ᵑg	
Affricate				tʃ			
Nasal	m		n				
Tap			(ɾ)				
Fricative		s̪		ʃ		(x)	
Approximant	β				j		
Lateral approximant			l				

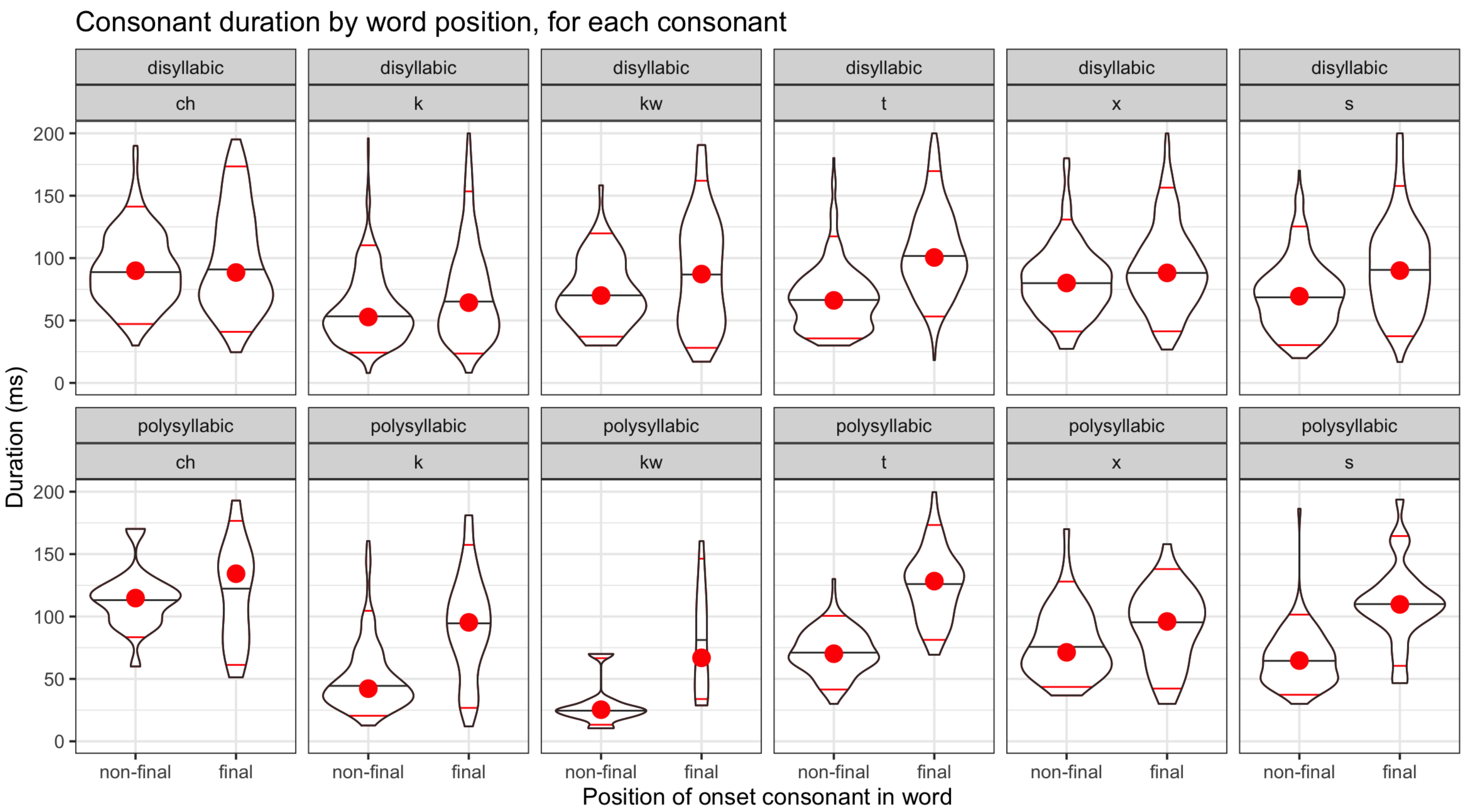
Most roots are composed of bimoraic feet (CVV or CVCV) with optional aspectual prefixes on verbs.

The final syllable of roots is stressed, realized via distributional asymmetries in the phonology and consonant lengthening (DiCanio et al., 2020).

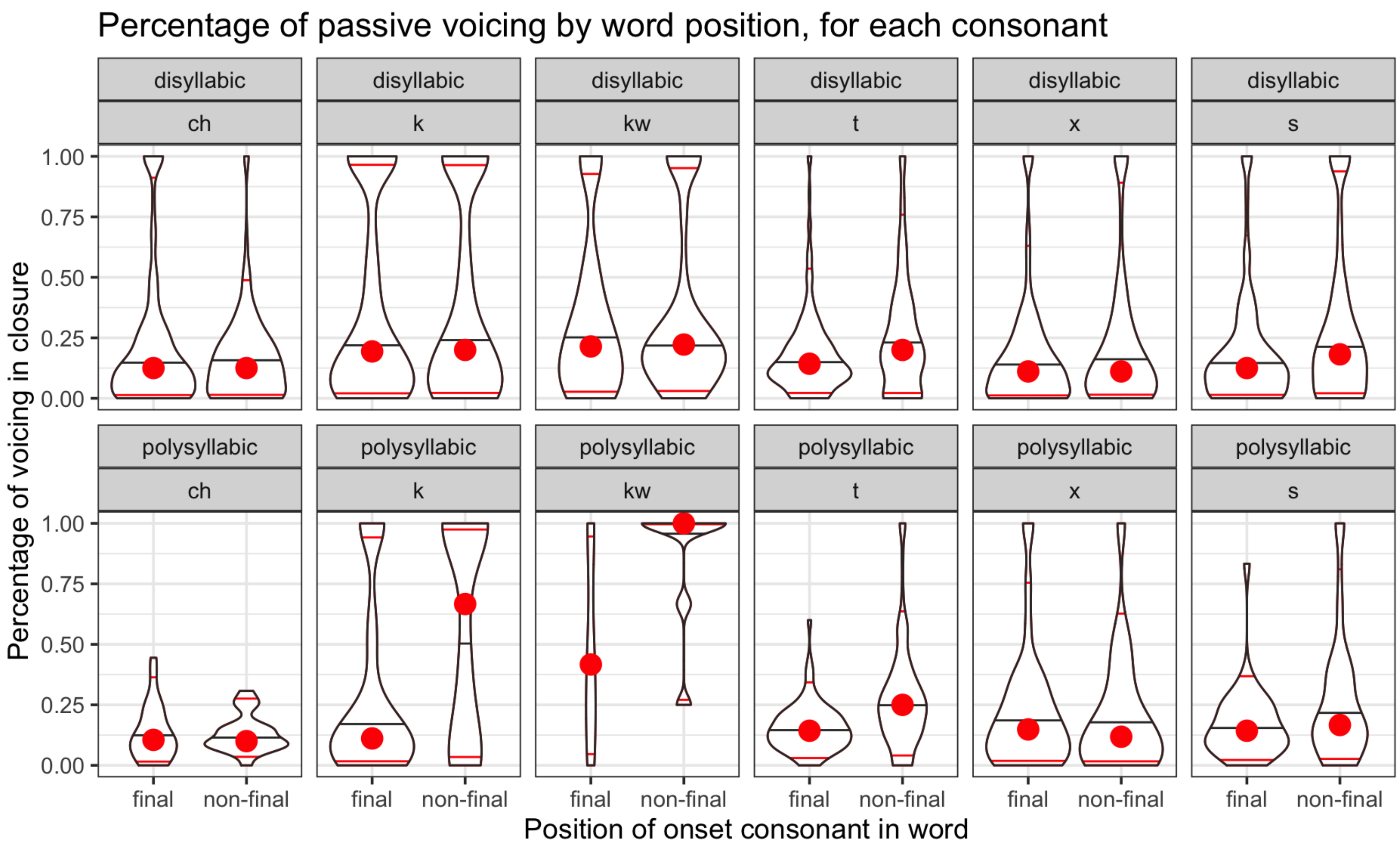
Corpus and methods

- Corpus of 6 speakers (3 male, 3 female) producing spontaneous narratives in YM, totalling 86 minutes.
- Analysis of duration, spirantization, and percentage of voicing during constriction/closure for /t, k, k^w, s, ʃ, tʃ/.
- A total of 7923 segments were analyzed.
- Hand-labelling of corpus was done in a previous study (DiCanio et al., 2015), but words here were coded by stem position (initial, medial, final syllable), and word size (monosyllabic, disyllabic, polysyllabic).

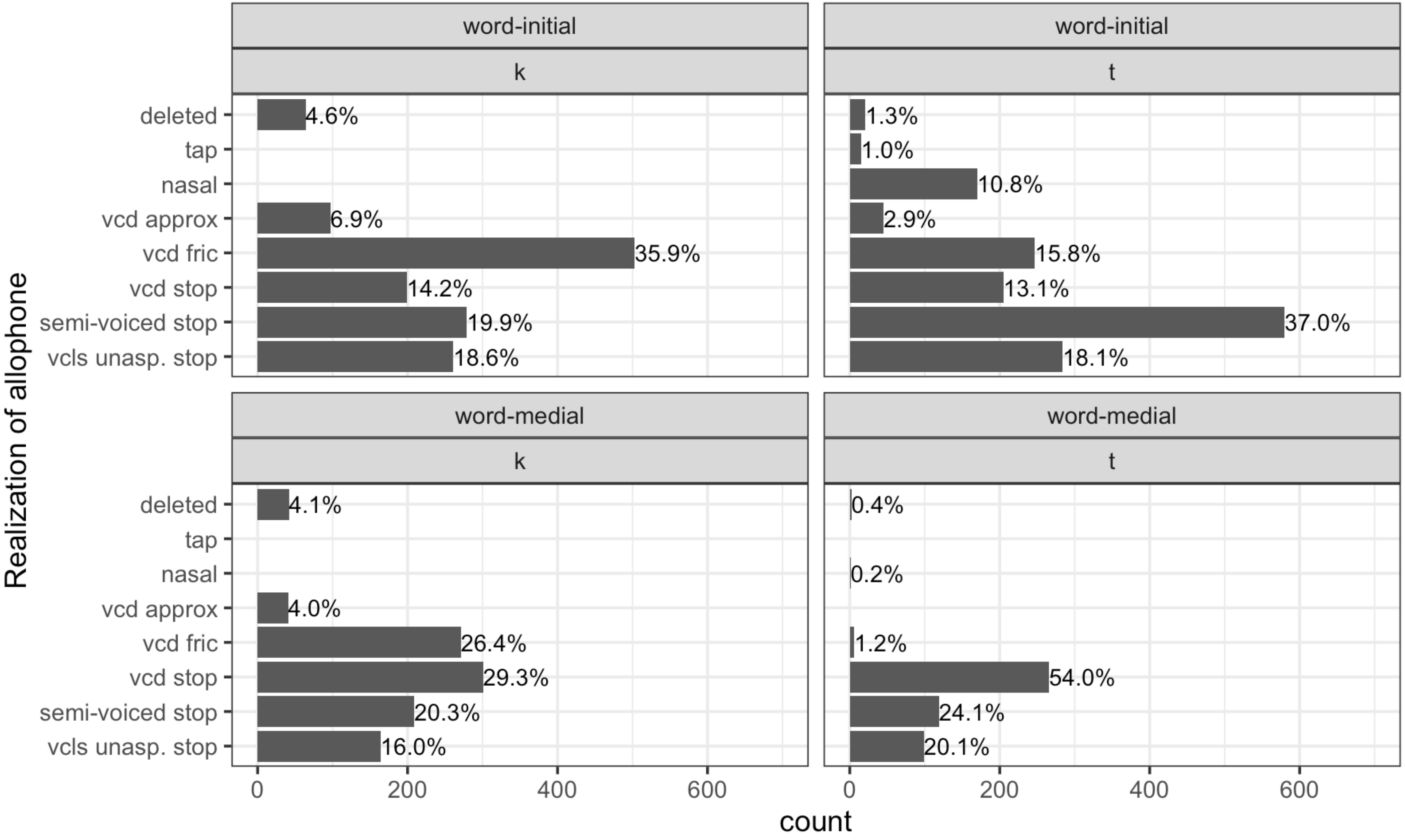
A general pattern of lengthening in stressed onsets (final σ) in polysyllabic words

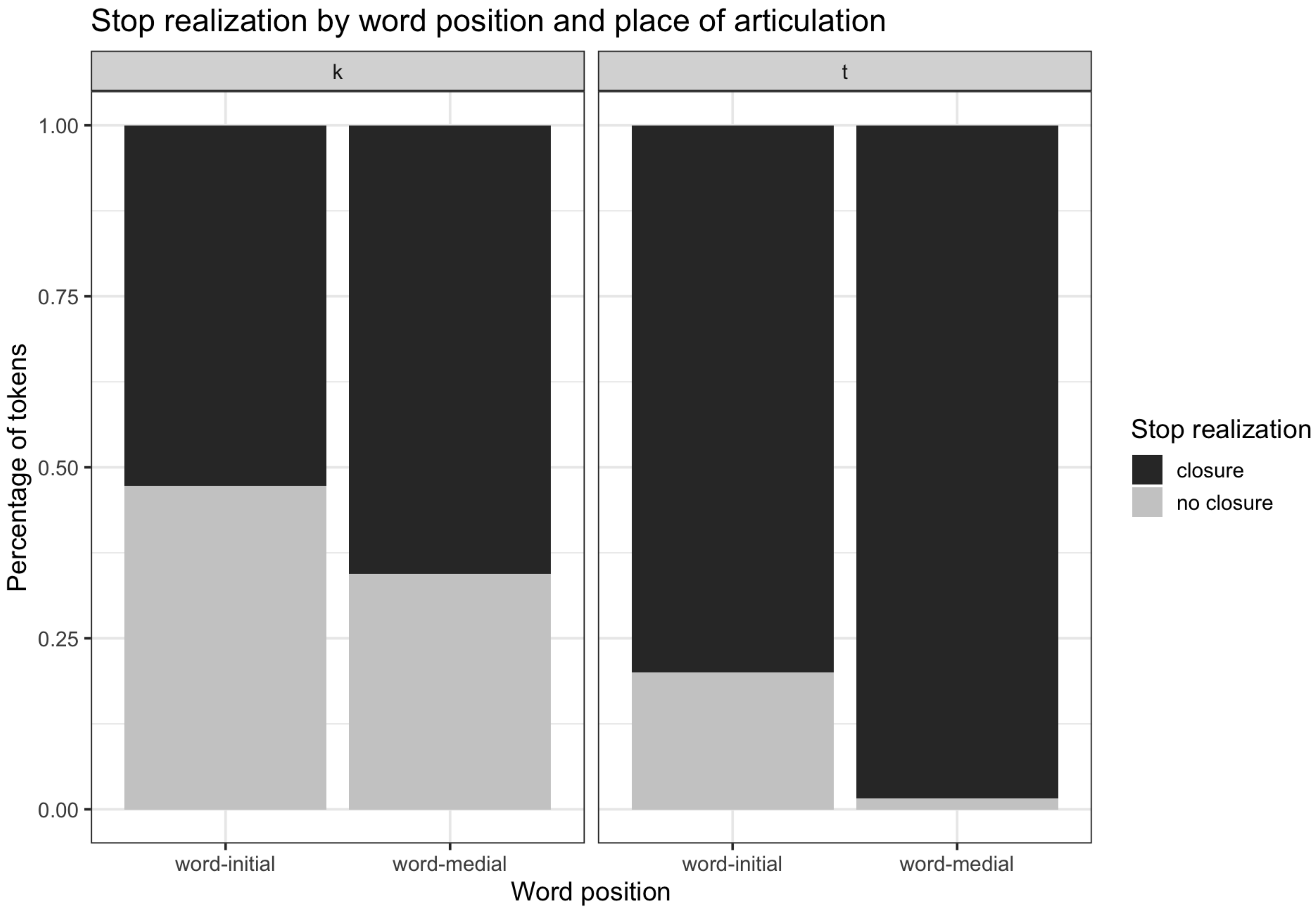


This corresponds with more passive voicing in unstressed syllables.



And there is more reduction in word-initial position of disyllabic words than in word-medial position





Lenition is more common in word-initial (unstressed) position than in word-medial (stressed) position.

Discussion

- Obstruents in YM are frequently lenited.
- Voicing lenition and spirantization realizations were more frequent in word-medial (stressed syllables) than in non-final syllables (regardless of word position).
- Patterns of lenition largely follow durational differences within the word.

- An exception to intervocalic spirantization?

- Word prosodic patterns on consonant production may be stronger than patterns related just to word position.

- Stem-final stress is fairly common in the world's languages. Among languages with phonologically-predictable stress, between 18% - 31% have stem-final stress (depending on the typological survey) (Gordon, 2016).

- No work discussing initial strengthening has examined languages with final stress.

Might morphology matter?

- Yoloxóchitl Mixtec has only prefixal inflectional morphology on verbs, historical derivational prefixes on nouns, and pronominal enclitics which can freely apply to most parts of speech (Castillo García, 2007).

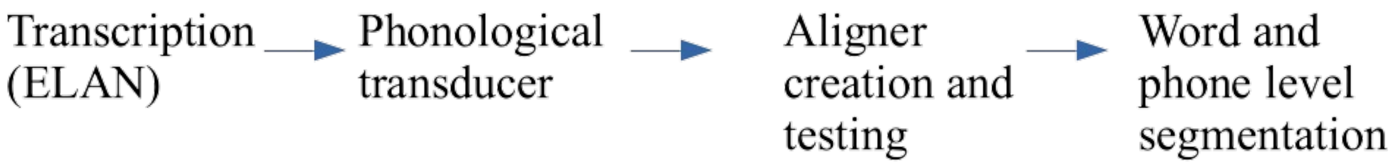
- This patterning is common among Otomanguean languages (Beam de Azcona, 2004; Campbell, 2014; DiCanio, 2016; Macaulay, 1996).

- If the goal of word-initial strengthening is to ensure reliable cues to word segmentation (Katz and Fricke, 2018; White et al., 2020), ideally speakers should ensure clear acoustic or articulatory cues in the initial portion of words which happens to be co-extensive with a lexical stem.

- The initial portion of Yoloxóchitl Mixtec words is *not* coextensive with the stem though, which might explain why we find reduction here.

- More to explore!

General discussion: corpus phonetics and language documentation



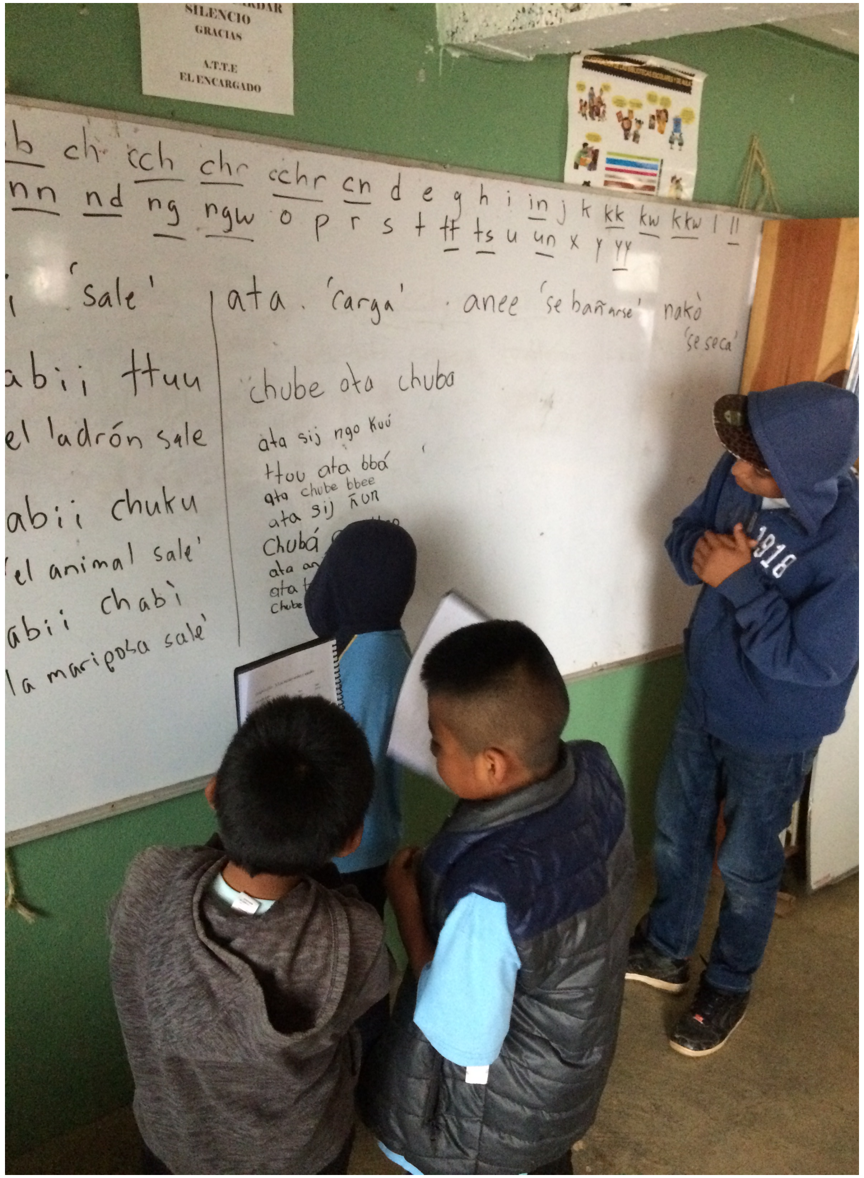
The development of a corpus for analytical purposes is not equivalent to just taking the documentation corpus "off the shelf.”

Very few archives are ready like this.

Pre-processing the documentation corpus

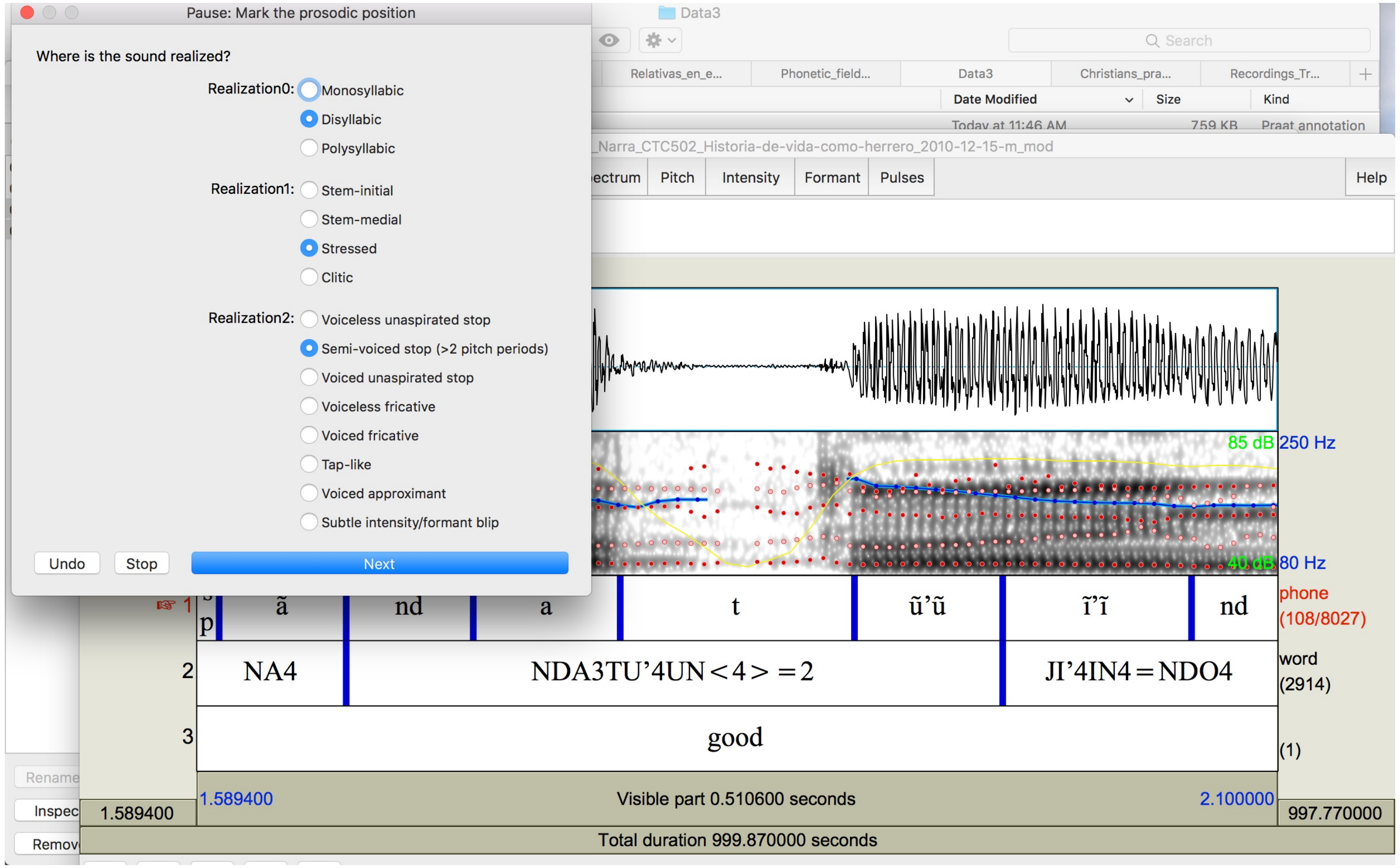
Pre-processing of the corpus is not insurmountable, but knowledge of both the corpus and the language is usually also necessary.

Collaborate with fieldworkers and communities if you wish to use archival materials!



Consider multiple approaches to data analysis!

- Phonological variation coding (categorical)
- Utilize scripts which provide you with contextual information from the corpus.
- Utilize additional scripts which enhance the richness of the corpus data (because it is already quite rich)



References

Babinski, S., Dockum, R., Goldenberg, D., Craft, J. H., Fergus, A., and Bowern, C. (2019). A Robin Hood approach to Forced Alignment: English-trained algorithms and their use on Australian languages. In *Proceedings of the 93rd Annual Meeting of the Linguistic Society of America. Linguistic Society of America*.

Beam de Azcona, R. (2004). *A Coatlán-Loxicha Zapotec Grammar.* PhD thesis, University of California, Berkeley.

Beckman, J., Jessen, M., and Ringen, C. (2013). Empirical evidence for laryngeal features: Aspirating vs. true voice languages. *Journal of Linguistics*, 49(2):259–284.

Campbell, E. (2013). The internal diversification and subgrouping of Chatino. *International Journal of American Linguistics*, 79(3):395–400.

Castillo García, R. (2007). *Descripción fonológica, segmental, y tonal del Mixteco de Yoloxóchitl*, Guerrero. Master's thesis, Centro de Investigaciones y Estudios Superiores en Antropología Social (CIESAS), Mexico City.

Davidson, L. (2018). Phonation and laryngeal specification in American English voiceless obstruents. *Journal of the International Phonetic Association*, 48(3):331–356.

Davidson, L. (2021). Effects of word position and flanking vowel on the implementation of glottal stop: Evidence from Hawaiian. *Journal of Phonetics,* 88:1–14.

DiCanio, C. (2010). Illustrations of the IPA: San Martín Itunyoso Trique. *Journal of the International Phonetic Association*, 40(2):227–238.

DiCanio, C. (2011). The Phonetics of Glottalized Consonants in Ixcatec. *SSILA Annual Meeting.*

DiCanio, C. (2012). Coarticulation between tone and glottal consonants in Itunyoso Trique. *Journal of Phonetics*, 40(1):162–176.

DiCanio, C. (2012). The Phonetics of Fortis and Lenis Consonants in Itunyoso Trique*. International Journal of American Linguistics*, 78(2):239–272.

DiCanio, C. (2014). Cue weight in the perception of Trique glottal consonants. *Journal of the Acoustical Society of America*, 135(2):884–895.

DiCanio, C. (2016). Abstract and concrete tonal classes in Itunyoso Trique person morphology. In Palancar, E. and Léonard, J.-L., editors, *Tone and Inflection: New Facts and New Perspectives*, pages 225–266. Berlin: De Gruyter Mouton.

DiCanio, C., Chen, W.-R., Benn, J., Amith, J. D., and Castillo Garćıa, R. (2022). Extreme stop allophony in Mixtec spontaneous speech: Data, word prosody, and modelling. *Journal of Phonetics*, 92:1–18.

DiCanio, C., Martínez Cruz, B., Cruz Martínez, B., and Martínez Cruz, W. (2020). Glottal toggling in Itunyoso Triqui. *Phonological Data & Analysis*, 2(4):1–28.

DiCanio, C., Nam, H., Whalen, D. H., Bunnell, H. T., Amith, J. D., and Castillo García, R. (2013). Using automatic alignment to analyze endangered language data: Testing the viability of untrained alignment. *Journal of the Acoustical Society of America*, 134(3):2235–2246.

DiCanio, C., Nam, H., Amith, J., Castillo García, R., and Whalen, D. (2015). Vowel variability in elicited versus spontaneous speech: Evidence from Mixtec. *Journal of Phonetics*, 48:45 – 59. The Impact of Stylistic Diversity on Phonetic and Phonological Evidence and Modeling.

DiCanio, C. and Whalen, D. H. (2015). The interaction of vowel length and speech style in an arapaho speech corpus. In *Proceedings of the 18th International Congress of the Phonetic Sciences*, Glasgow, Scotland, pages 513–517.

DiCanio, C., Zhang, C., Whalen, D. H., and Castillo García, R. (2020). Phonetic structure in Yoloxóchitl Mixtec consonants. *Journal of the International Phonetic Association,* 50(3):333– 365.

Gahl, S., Yao, Y., and Johnson, K. (2012). Why reduce? Phonological neighborhood density and phonetic reduction in spontaneous speech. *Journal of Memory and Language*, 66:789–806.

Gerfen, C. and Baker, K. (2005). The production and perception of laryngealized vowels in Coatzospan Mixtec. *Journal of Phonetics*, 33(3):311–334.

Gordon, M. K. (2016). *Phonological Typology.* Oxford University Press.

Hillenbrand, J. M. and Houde, R. A. (1996). Role of F0 and Amplitude in the Perception of Intervocalic Glottal Stops. *Journal of Speech and Hearing Research*, 39:1182–1190.

Johnson, L. M., DiPaolo, M., and Bell, A. (2018). Forced Alignment for Understudied Language Varieties: Testing ProsodyLab-Aligner with Tongan Data. *Language Documentation and Conservation*, 12:80–123.

Katz, J. and Fricke, M. (2018). Auditory disruption improves word segmentation: A functional basis for lenition phenomena. *Glossa,* 3(1):1–25.

Keating, P. A. and Huffman, M. K. (1984). Vowel variation in Japanese. *Phonetica*, 41:191–207.

Koopmans van Beinum, F. (1980). *Vowel contrast reduction, an acoustic and perceptual study of Dutch vowels in various speech conditions.* PhD thesis, University of Amsterdam, The Netherlands., Academische Pers B. V., Amsterdam.

Lavoie, L. M. (2001). *Consonant Strength: Phonological Patterns and Phonetic Manifestations.* New York: Garland.

Macaulay, M. (1996). *A Grammar of Chalcatongo Mixtec*, volume 127 of University of California Publications in Linguistics. University of California Press.

McAuliffe, M., Socolof, M., Mihuc, S., Wagner, M., and Sonderegger, M. (2017). Montreal Forced Aligner: trainable text-speech alignment using Kaldi. In *Proceedings from Interspeech 2017.*

Parrell, B. and Narayanan, S. (2018). Explaining coronal reduction: Prosodic structure and articulatory posture. *Phonetica*, 75:151–181.

Roettger, T. B. (2019). Researcher degrees of freedom in phonetic research. *Laboratory Phonology*, 10(1):1–27.

Schwarz, M., Sonderegger, M., and Goad, H. (2019). Realization and representation of Nepali laryngeal contrasts: Voiced aspirates and laryngeal realism. *Journal of Phonetics,* 73:113–127.

Silverman, D. (1997). *Phasing and Recoverability.* Outstanding Dissertations in Linguistics. Routledge.

Stevens, K. N. (2000). *Acoustic Phonetics*. MIT Press, first edition.

Storto, L. R. and Demolin, D. (2012). The phonetics and phonology of South American languages. In Campbell, L. and Grondona, V., editors, *The indigenous languages of South America: A comprehensive guide*, pages 331–390. De Gruyter.

Tang, K. and Bennett, R. (2019). Unite and conquer: bootstrapping forced alignment tools for closely-related minority languages Mayan. In Calhoun, S., Escudero, P., Tabain, M., and Warren, P., editors, *Proceedings of the International Congress of Phonetic Sciences (ICPhS) 2019*, pages 3584–3552. Canberra, Australia: ASSTA.

Westbury, J. R. and Keating, P. A. (1986). On the naturalness of stop consonant voicing. *Journal of Linguistics,* 22:145–166.

Whalen, D. H., DiCanio, C., Geissler, C., and King, H. (2016). Acoustic realization of a distinctive, frequent glottal stop: The Arapaho example. *Journal of the Acoustical Society of America*, 139(4):2212.

White, L., Benavides-Varela, S., and Mády, K. (2020). Are initial-consonant lengthening and final-vowel lengthening both universal word segmentation cues? *Journal of Phonetics*, 81:1–14.