Corpus studies of variation in obstruent ‘voicing’ across languages and speakers:

phonetic variation and implications for phonology

Morgan Sonderegger

AMP 2021

3 Oct., 2021

Introduction

- Phonological features and phonetic realization
  - There is some link
- Debated:
  - how direct, by what criteria?

(e.g. Jakobson et al., 1952; Clements, 1985; Stevens, 1989; Flemming 1995; Hall, 2001; Avery & Idsardi, 2001; …)

- Especially for laryngeal contrasts
- Highly variable phonetics, across many acoustic cues
- By position (*<u>b</u>at*, *ra<u>b</u>id*, *ta<u>b</u>*)
  - Initial: VOT, closure voicing, F0, …
  - Final: VDur, burst, F0, …
- By language
  - 'True voicing' (French, Turkish)
  - 'Aspirating' (English, German)

Questions

1. What is the relationship between phonological representation and phonetic realization?
2. What is the typology of phonetic 'laryngeal' contrasts?
   – They lie in some space: what are the dimensions?
   – Discrete or continuous?

- Approach: cross-language/dialect corpus studies, (mostly) large-scale

Questions

- Related to questions in:
- Phonetics
  - Automatic vs. controlled (Kingston & Diehl, 1994; Solé, 2007)
  - Individual differences (Yu & Zellou, 2019)
- Sound change
  - What phonetic precursors can seed change? (Ohala)
- Sociolinguistics
  - What defines a speech community? (Labov, 1972)
  - Does 'English' phonetics exist (across dialects)?

Outline

- Study 1: Laryngeal timing
- Study 2: Intrinsic F0 effects
- Study 3: Vowel dur. effects

Q1

Q2

- Study 1: Laryngeal timing across 7 languages

Collaborators:

Michael McAullife

Jurij Bozić

Amy Bruno

September Cowley

- Study ... nsic ... ts ... 14 ...

Bing’er Jiang

Jeff Lamontagne

Martha Schwarz

Jiajia Su

- Study ... du ... 34

Laryngeal features-phonetics theories

- How to capture 'voicing' (etc.) contrasts, x-ling?
- Traditional: [±voice]
  - Indirect phonetics-feature link
  - Broad similarities in phonetic cue patterns x-ling
    - Ex: 'voiced'/'voiceless' differences in VOT , F0 (initial)
  - 'phonetic implementation' minimally predictable from features
    (L&Abramson, 1964; Keating 1984; Kingston & Diehl 1994; Kohler 1984; Lombardi 1991)

- Laryngeal realism: [voi], [sg] (+ [cg]) ← spread, constricted glottis
  - More direct phonetics/feature link
  - Ex: German: [sg] contrast, French [voi] contrast
  - Differences in `phonetic implementation' x-ling: predictable

(Jakobson, 1949; Iverson & Salmons, 1995 et seq.; Beckman et al., 2011, 2013; Avery & Idsardi 2001)

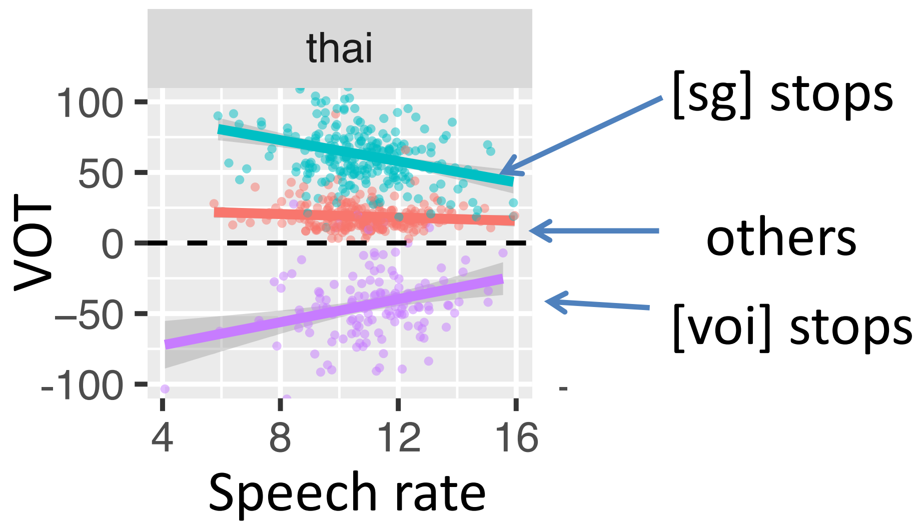
	Traditional	Laryngeal realism
German, English	/p/ = [-voice] /b/ = [+voice]	/p/ = [sg], /b/ = [ ]
Turkish, French		/p/ = [ ], /b/ = [voi]
Thai	/p/ = [-voice], [-spread] /b/ = [+voice], [-spread /p^h/ = [-voice], [+spread]	/p/ = [ ] /b/ = [voi] /p^h/ = [sg]

- Phonetically similar

Criteria

- LR criteria linking features & phonetic realization

1. Prevoicing
   - [voi] stops vs. others
2. Speech rate ~ VOT
3. Voicing during closure
   - [voi] stops (near 100%) vs. others



(e.g. Jakobson, 1949; Beckman et al., 2011, 2013; Jessen, 2001)

Research questions: Study 1

- Criteria (1)-(3) often tested in isolation or in 1-2 languages
  (e.g. Beckman et al., 2011, 2013; Helgason & Ringen 2008; Jessen, 1998; Lisker & Abramson, 1964; Kessinger & Blumstein, 1997; Ringen & Kulikov 2012; M. Schwartz et al., 2019)

- Do they hold in a wider sample of languages?
- Give convergent evidence?

- Today: 7 languages, comparable data

Data

- 7 languages:

	Croatian, French, Turkish		Swedish		Thai			German		Korean		
IPA	b	p	b	p^h	b	p	p^h	p	p^h	p^*	p	p^h
Features	[voi]	[ ]	[voi]	[sg]	[voi]	[ ]	[sg]					g]

"voiced" "voiceless unaspirated" "voic ”

21 language corpora
~100 speakers/language,
~100 sentences/speaker

- Read sentences from GlobalPhone, force-aligned (Schultz et al. 2013; MFA – McAuliffe et al. 2017)
- n ~100-300 per laryngeal class/position/language

Data

Data from two positions:

Utterance-initial (##C)

- Sentence-initial or post-pause

Intervocalic (VCV)

- Word-medial
- #*V**C**V*# words

Examine criteria:

1. Prevoicing

2. Speech rate

3. Voicing during closure

Data: manual annotations

- ##C: presence + duration of

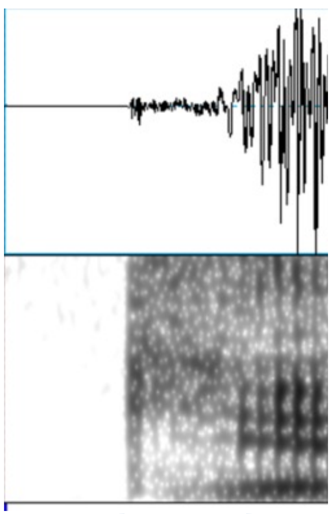
positive VOT + negative VOT ⇒ "VOT"

≈ burst duration = prevoicing

Criterion 2

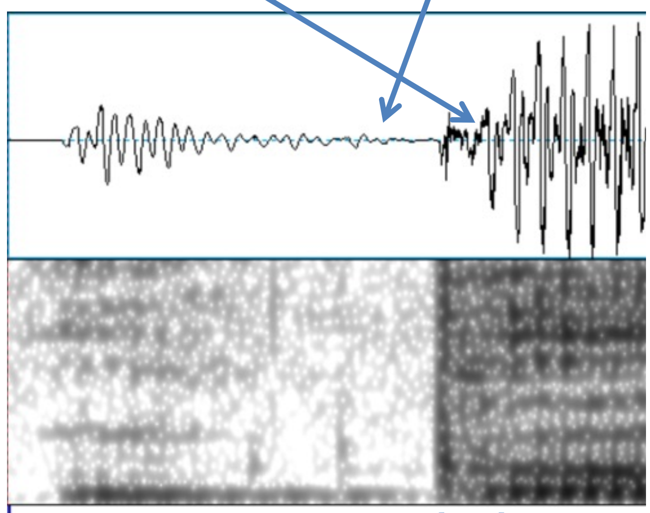
Criterion 1

Turk. ***ge**nsoru*



+ VOT, VOT

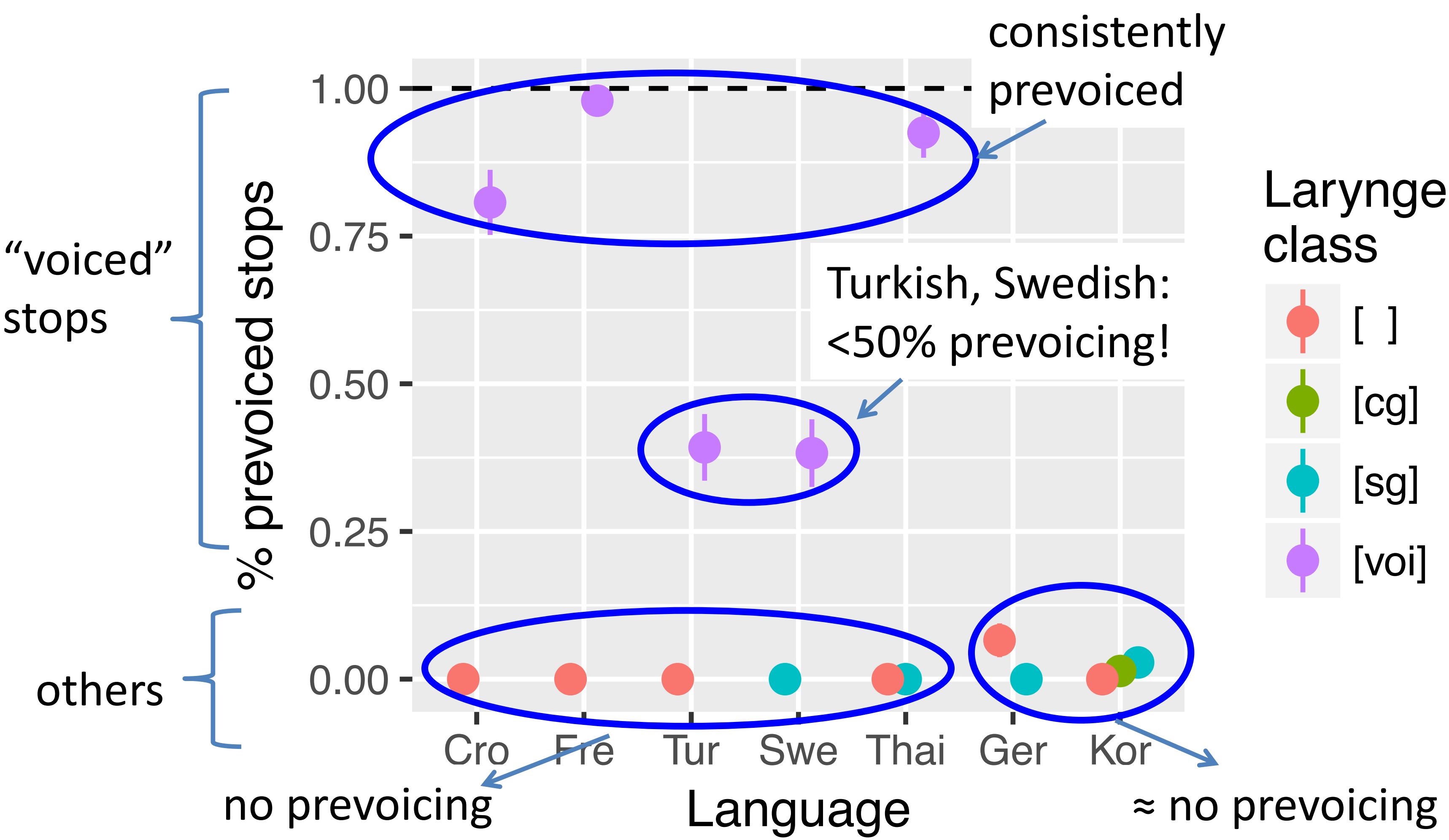
French ***ba**nques*



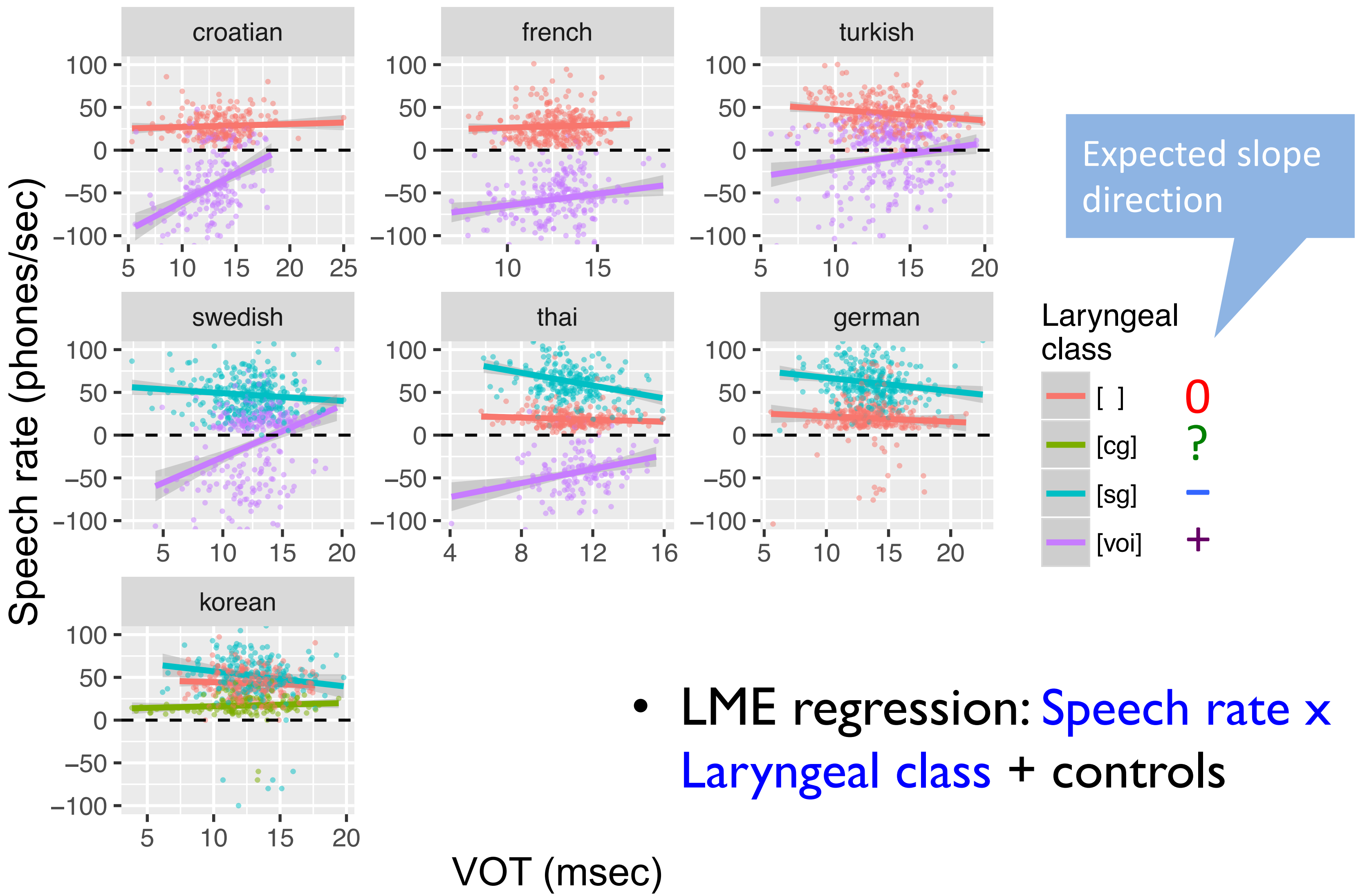
− VOT, VOT   + VOT

VCV: annotated percent **voicing during closure**
(0-100%)
-- Criterion 3

Results: ##C prevoicing

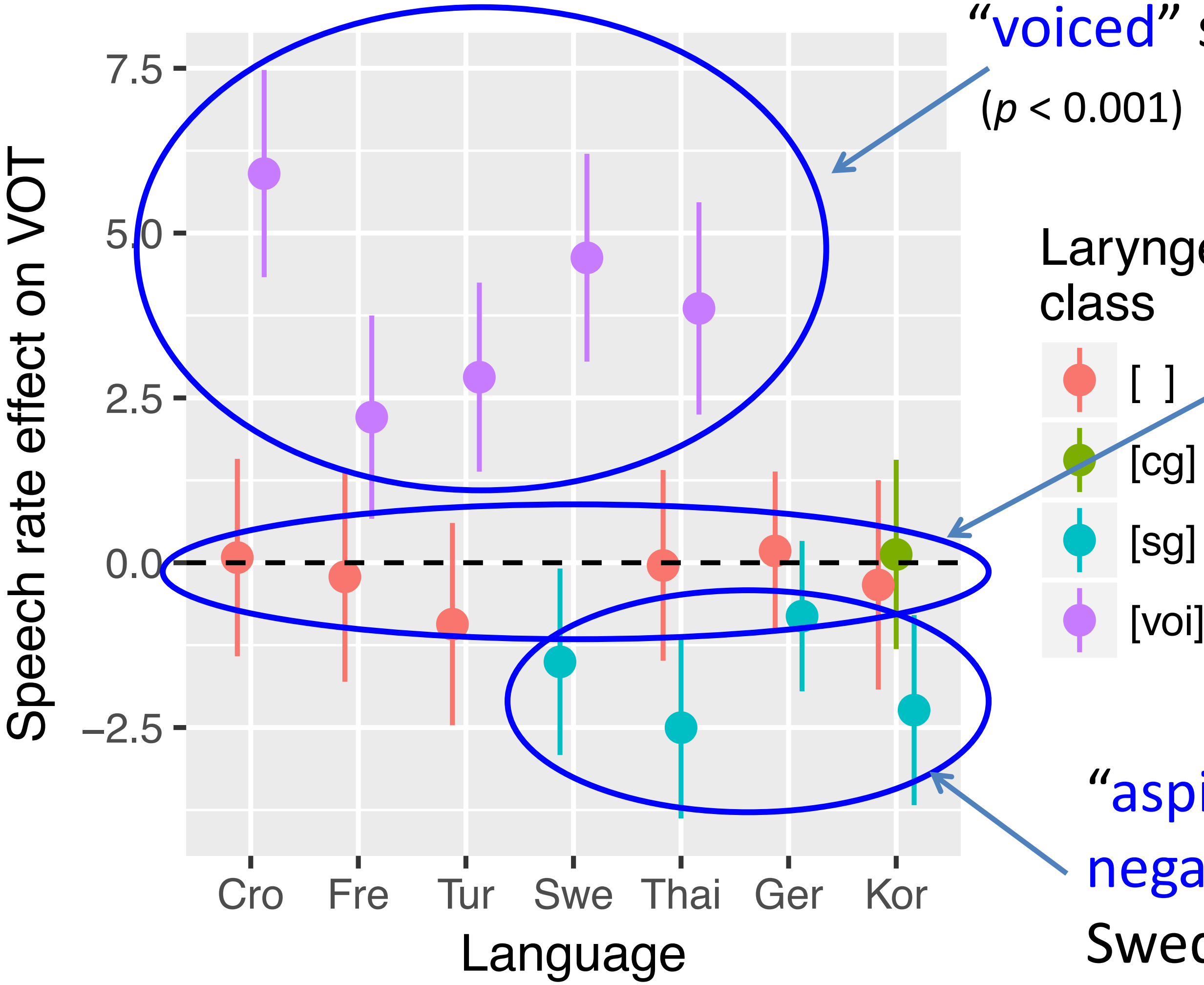


Results: ##C speech rate vs. VOT



- LME regression: Speech rate x Laryngeal class + controls

Results: ##C speech rate vs. VOT



"voiced" stops: positive effect

($p < 0.001$)

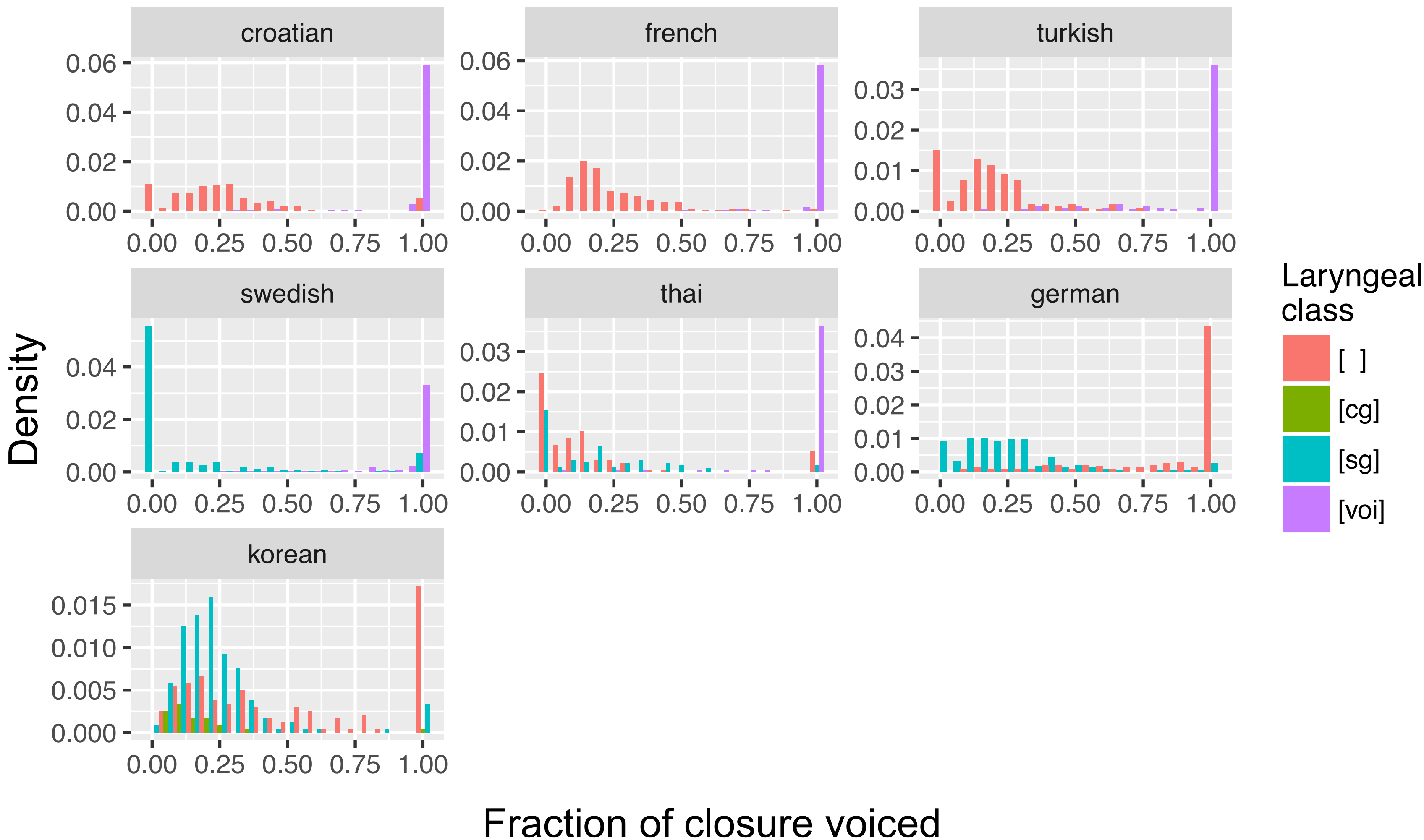
"voiceless unasp" and "tense" stops: n.s.

"aspirated" stops: negative effect

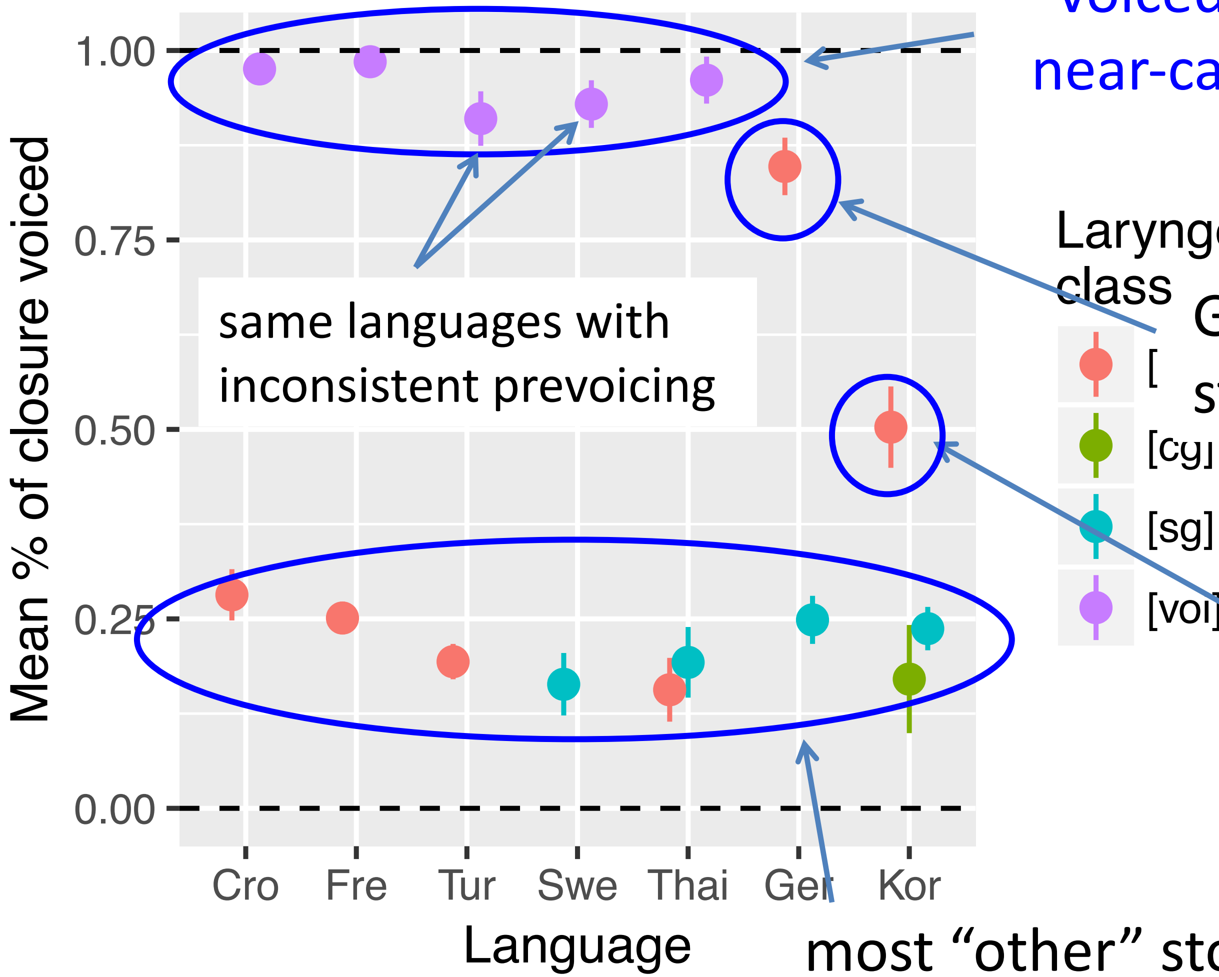
Swedish: $p = 0.04$

German: $p = 0.16$

Results: voicing during closure (VCV)



Results: voicing during closure



"voiced" stops:
near-categorical VDC

Germ
stops: ??

Expectations

"passive voicing"
or none*

near-100% voicing

Korean "lax" stops:
allophonically
voiced
(Jun ,1994)

most "other" stops:
passive voicing (15-25%)

* LR theories differ

Summary of results

	Croatian	French	Turkish	Swedish	Thai	German	Korean
Prevoicing	?	✓	X	X	✓	✓	✓
Rate ~ VOT	✓	✓	✓	✓	✓	?	✓
Closure voicing	✓	✓	✓	✓	✓	?	?

- Q1: do criteria hold up across 7 languages?
- Q2: do criteria give convergent evidence `` ``?

Discussion: Study 1

As in other corpus studies
(van Alphen & Smits, 2004; Davidson, 2015
Sonderegger et al., 2020)

- Criterion 1: prevoicing
  - [voi] stops: ✗
  - non-[voi] stops: ✓
  - ##C "Voiced" stops
    not consistently prevoiced in read sentences
- Criterion 2: speech rate effects on VOT
  - ✓ (mostly)
- Criterion 3: voicing during closure
  - [voi] stops: ✓
  - non-[voi] stops: ?
    - Mostly: low/inconsistent VDC.
    - No evidence that having active [voi] matters

(Beckman et al., 2013; c.f. Kirby & Ladd 2019)

Discussion: Study 1

- 2/3 criteria (voicing during closure, speaking rate) give largely convergent evidence across 7 languages
  - … with some gaps
  - Assuming "laryngeal realism" (privative) features + diagnostics

'Laryngeal realism' predictions (partially) borne out across 7 languages

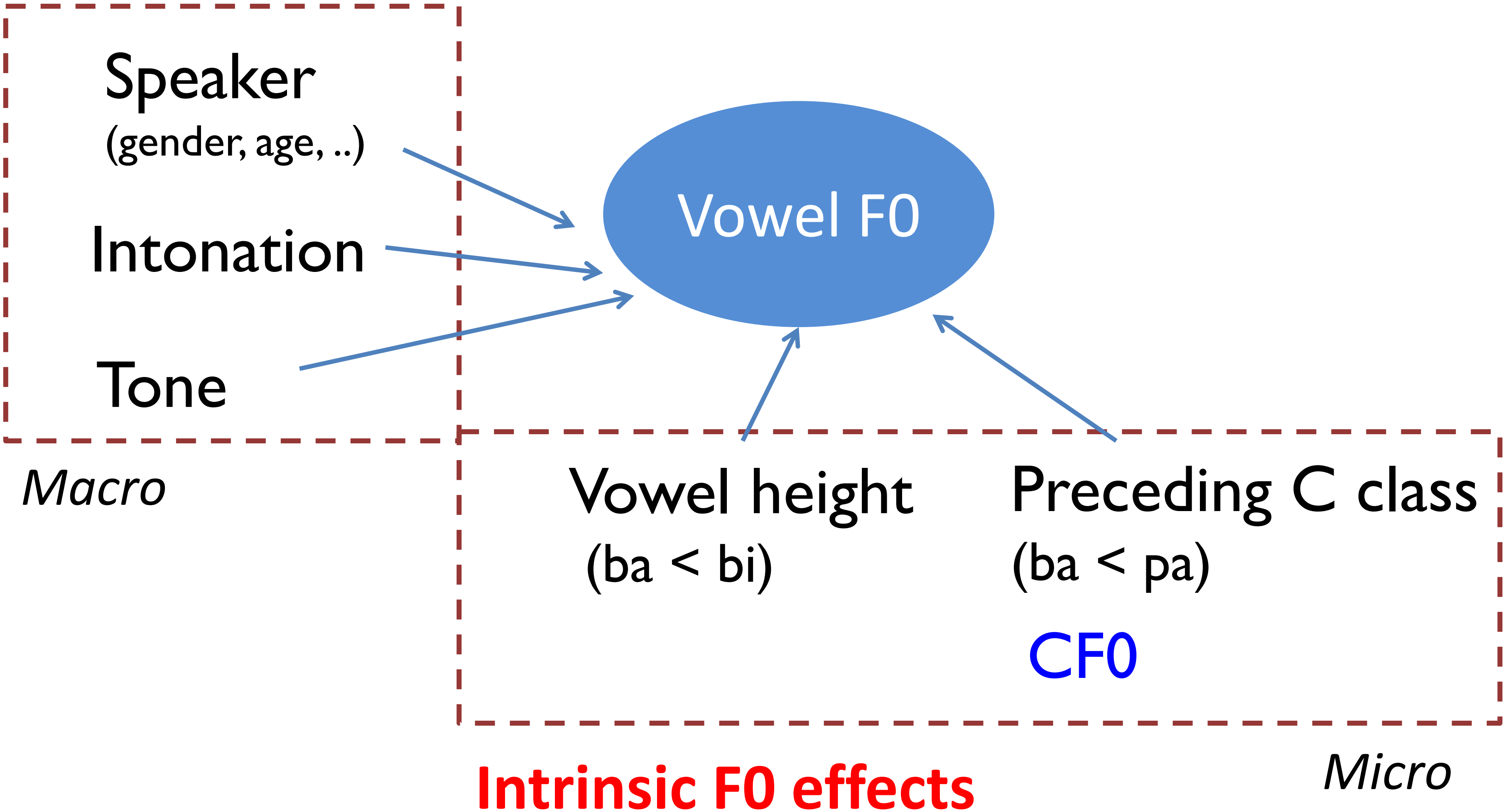
- Study 1: Laryngeal timing across 7 languages

- Study 2: Intrinsic F0 effects across 14 languages

Collaborators: Michael McAuliffe, Hye-Young Bang

- Study 3: Vowel du acr ects

Introduction: Study 2



(e.g. Babinski, 2021; Chen, 2011; Connell 2002; Fischer-Jørgenson, 1990; Hanson, 2009; Hoole & Honda, 2011; House & Fairbanks, 1953; Kingston & Diehl, 1994; Kirby & Ladd, 2016, XX; Kingston, 2007; Ladd & Schmid, 2018; Ladd & Silverman, 1994; Meyer, 1896; Whalen & Levitt, 1995)

Introduction: Study 2

- Today: CF0
- Many languages: 'voiced' < 'voiceless' F0
  - Evidence for [±voice] (Kingston & Diehl, 1994)
  - Some conflicting or null effects (e.g. Mandarin)
  - Effect size: variable
    - Tonal ⇒ smaller effect?
- Need: comparable data, from many languages

- Previous work
  - Primarily 1-3 languages, lab speech
  - Speakers differ, how much unclear
  - Focus: mechanism (automatic vs. controlled)

RQ1: How much variability in CF0 across 14 languages?

- Important for sound change, as example of how changes originate:

  phonetic effect → phonological pattern

"phonetic precursors"

F0 perturbations around p/b → lexical tone

What kind of precursor can be a source of change?

- robust
  - Across speakers, languages (e.g. Hombert et al., 1979, Ohala)
- … but variable
  - Individual differences, language-specific phonetics

    (e.g. Baker et al., 2011; Labov, 1967; Kingston, 2007; Yu, 2013)

tension

RQ 2: How robust/variable is CF0, across languages and individuals?

Datasets

English	Russian
French	Polish
German	Spanish
Korean	Turkish

Hausa
Mandarin
Thai
Vietnamese

- Globalphone
  - Read sentences
  - ~20 hours each
  - Force-aligned (MFA)

Croatian
Swedish

GlobalPhone (Schulz et al., 2013), Librispeech (Panyatov et al., 2015)

Datasets

- “Utterance-initial”

  > 150 ms pause or file-initial

C V

obstruent /a/, /i/, /u/

1.9-9.5k tokens (~2000/language)
76-132 speakers (~100/language)

- vowel F0 (Praat)
  - F0 histogram → speaker min, max → re-extract F0
- Other info:
  - Speaker: ID, gender, mean F0
  - Utterance: length (syllables)
  - Surrounding segments
  - Word

Polyglot/ISCAN

https://iscan.readthedocs.io/
https://polyglotdb.readthedocs.io/
McAulffe et al. (2017, 2019)

Analysis

- One linear mixed effects model / language
- Main terms:

Response: mean F0 in first 50 ms

C_1 V X

consonant "voicing"*

overall effect + interspeaker variability

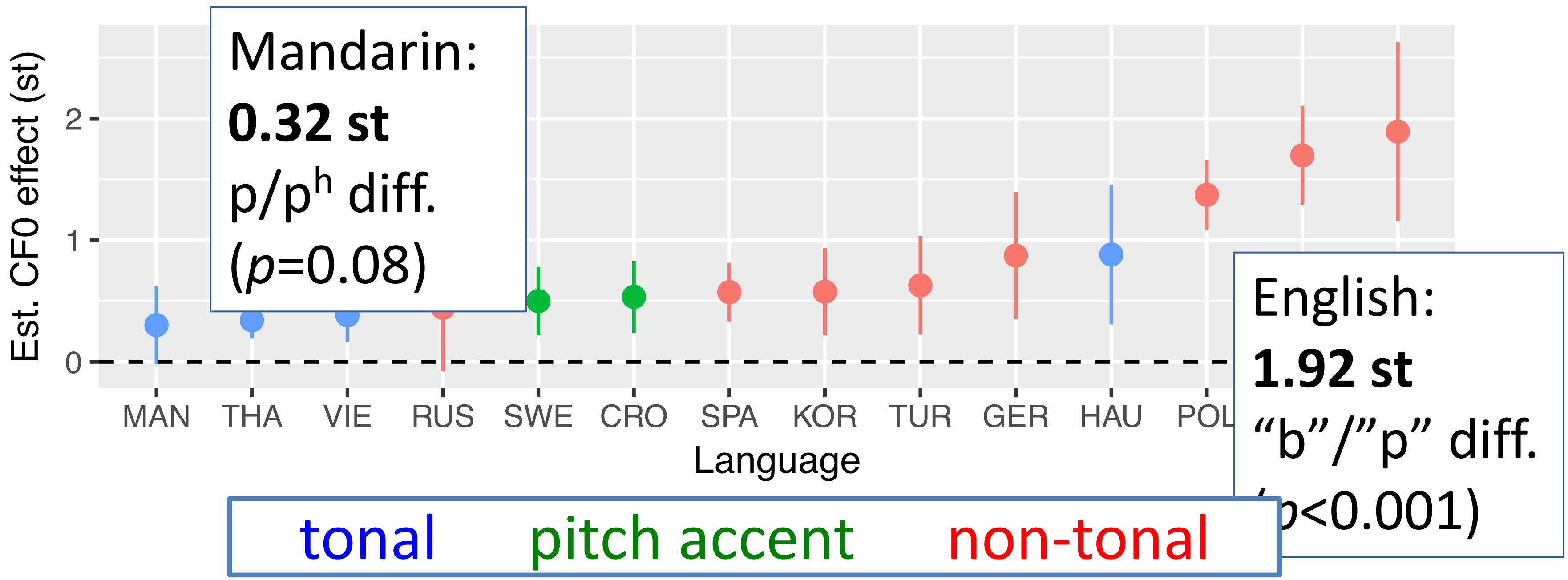
fixed effect by-speaker random slope

- + controls

* Ex: French p/b, Mandarin p/p^h

CF0 across languages

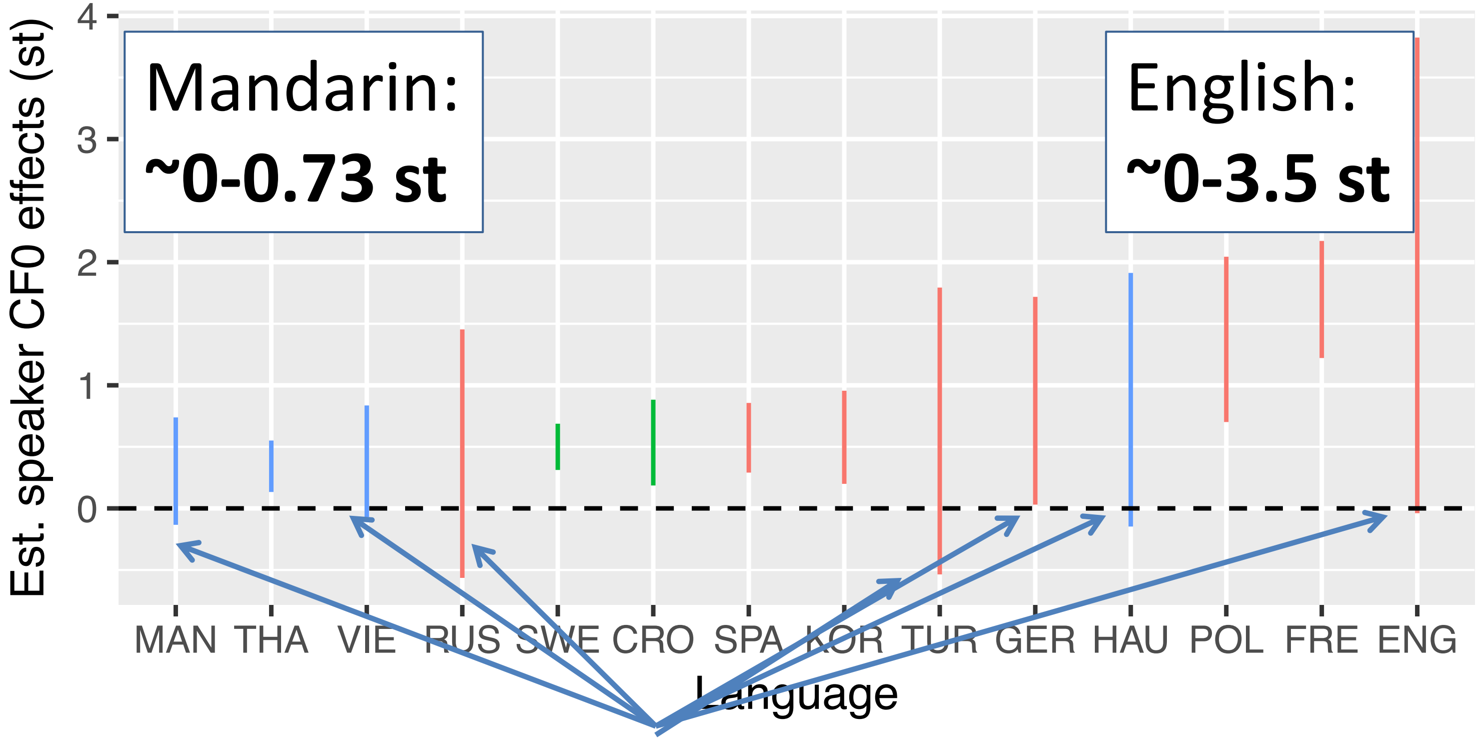
- "most voiceless" – "most voiced" effect:



- Robust across languages
- Variable effect size
  - Non-tonal ⇒ larger effect

CF0 across speakers

- Predicted effects for 95% of speakers:



- Common: large interspeaker variability

Discussion: Study 2

- Robust group-level CF0 effects across languages
  - same direction
  - In line with "universality"
    (Kingston & Diehl, 1994; c.f. Whalen & Levitt, 1995 for VF0)

- Very different effect sizes
  - One reason: tonal/pitch accent language
    ⇒ smaller IF0 more likely
    (hypothesized: Connell 2002)

- Fits with automatic + controlled mechanism
  - Strong <u>automatic</u> basis (vocal fold tension)
  - Controlled: individual languages
    (many 'knobs' to turn in larynx)

(e.g. Hoole & Honda, 2011; Kirby & Ladd, 2018)

Discussion: Study 2

- Large interspeaker variability in IF0 magnitude common, within language
  - $\Rightarrow$ there are some speakers with null/large effects
  - Still, most speakers show effect in same direction

- Overall: IF0 effects
  - robust across languages
  - variable across speakers
- Both important for sound change

- May be related to actuation: why are sound changes from IF0 possible, but rare? (Kingston, 2007)

Outline

- Study 1: Laryngeal timing

Collaborators: James Tanner, Jane Stuart-Smith, Joe Fruehwald

- Study 2: Intrinsic F0 effects

- Study 3: Vowel dur. effects

Tanner et al. (2020), *Frontiers in AI*

Introduction: Study 3

- *bead* > *beat* vowel dur.

- the voicing effect
- Received wisdom
  - Near-universal cue to 'voicing' word-finally, x-ling
  - Very large effect in English ← lab speech, ~2 US dialects
- RQ1: robust effect?
  - Across dialects, speakers
  - Spontaneous speech

James Tanner

Tanner et al. (2020), *Frontiers in AI*

Introduction: Study 3

- Textbook allophonic rule 'of English'
  - Is 'English' defined in part by phonetics, across dialects?
  - RQ2: is there an 'English' voicing effect?

- Known extreme cases:
  - Scottish English

    'beat', 'bead' (V)

    'bee', 'bees' (V:)

  - African American Vernacular English (some speakers)

    'bag'

    'back'

Source: CORAAL, Rachel MacDonald

SPADE

SPeech Across Dialects of English

UK PI: Jane Stuart-Smith

US PI: Jeff Mielke

2017-2021

http://spade.glasgow.ac.uk/

Trans-Atlantic Platform Social Sciences and Humanities

DIGGING INTO DATA CHALLENGE

THE UNIVERSITY of EDINBURGH

University of Glasgow

McGill

NC STATE UNIVERSITY

UNIVERSITY OF OREGON

Project goals

SPADE
SPeech Across Dialects of English

Software large-scale speech analysis

Data from ~40 datasets (socio)linguistic surveys

Research 'English' sounds over time and space

Sonderegger et al. (2021), *Open Handbook Ling. Data Management*

Data: The SPADE Consortium

- ~40 corpora: public/private, 6 countries, 115 years
- <u>Processing</u>: cleaning, (forced) alignment, acoustic measurement

https://spade.glasgow.ac.uk/the-spade-consortium/

Voicing effect: data & analysis

- Utterance-final CVC words
  - n = 229k tokens
  - 1964 speakers, 30 'dialects'
- Bayesian linear mixed-effects model
  - `stan/brms` (Carpenter et al., 2017; Bürkner 2018)
- Effect of interest: following C voicing

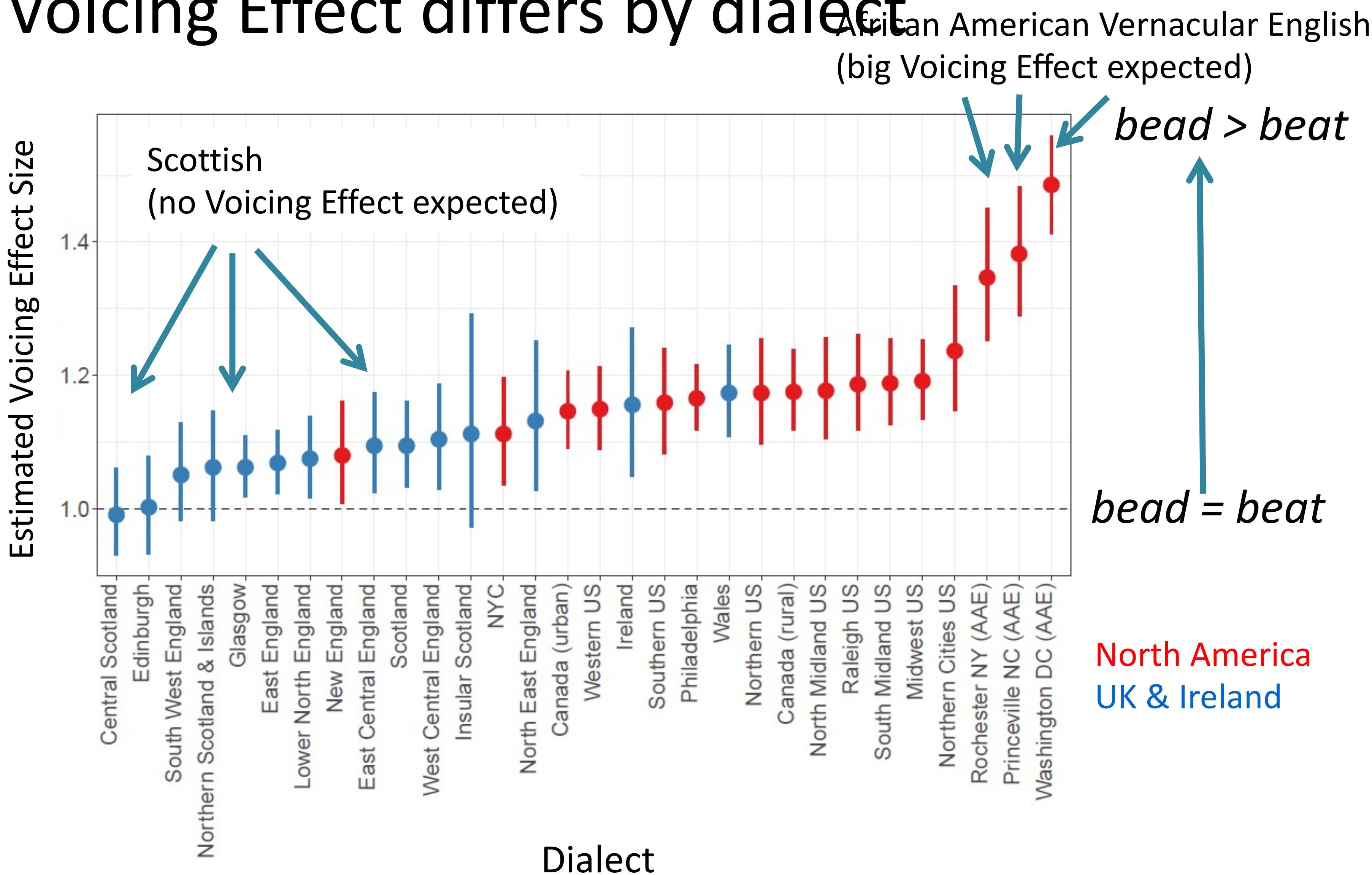
+ controls (speech rate, word freq., vowel height..)

- Random effects:
  - Dialect
  - Speaker, <u>within dialect</u>
  - Word

Dialect mean

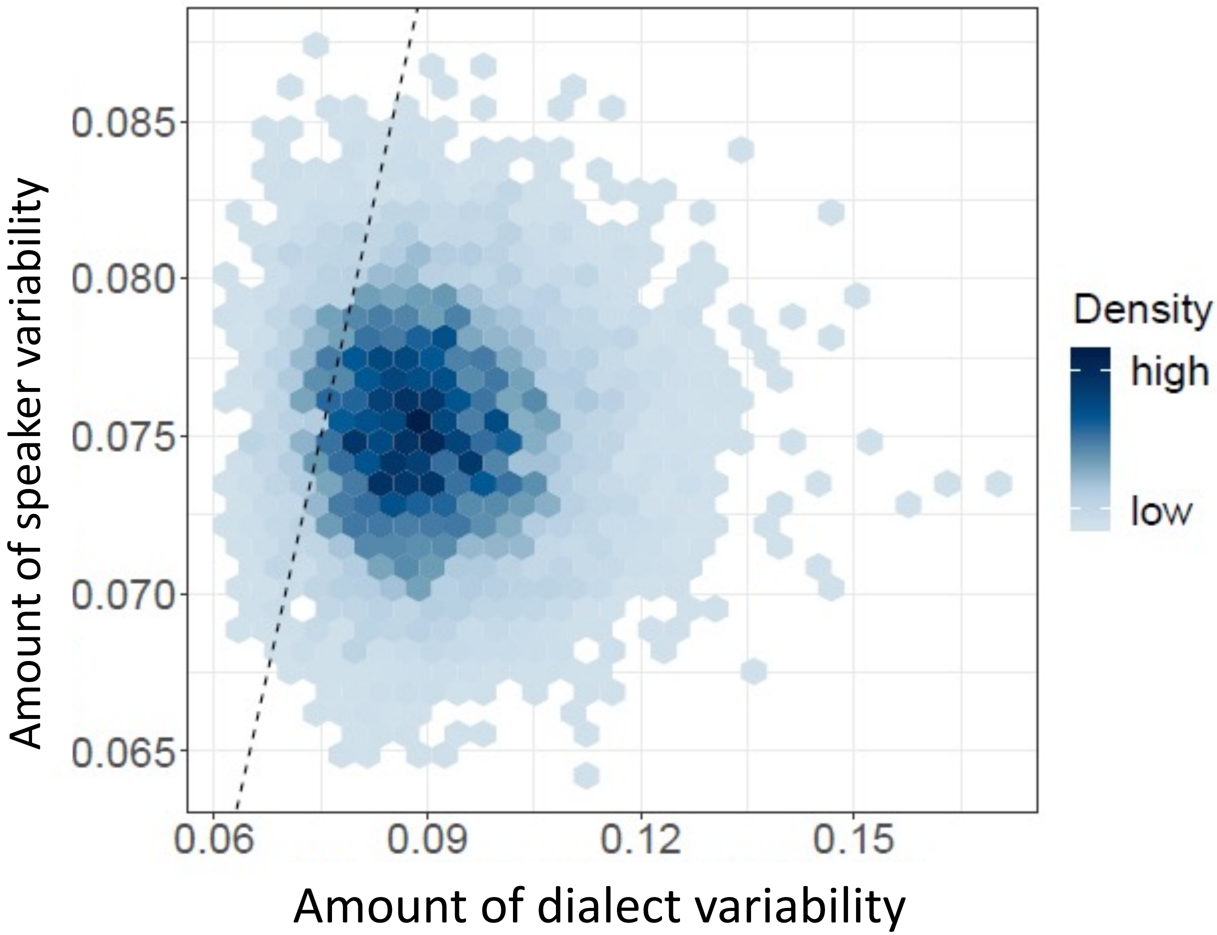
Speaker <u>offset</u> from dialect mean

Voicing Effect differs by dialect



1964 speakers
~230k tokens

Voicing Effect differs more by dialect than by speakers



Discussion: Study 3

- Voicing effect:
  - … but <u>some</u> effect always there
  - Smaller effect size in spontaneous speech
  - <u>Not robust</u> to context, speech rate (not shown)

- High dialect variability
  - Partially due to dialect-specific [voice] processes

- Scaling up analysis allows new perspective

Discussion: Study 3

- Speaker < dialect variability:
  - Kleinschmidt (2019): similar finding for Am. English vowel formants

- Why? Speculation:
  - What defines a 'speech community'? (Labov, 1972)
  - Perhaps patterns like VE (dialect > speaker variability)

- Other SPADE work: sibilants show speaker > dialect variability
  (Stuart-Smith et al. 2020 *Labphon*)
  - Such sounds may signal social-indexical informativity better within community
  - Compared to group-level attributes (dialect)
- More cross-dialect studies needed!

Summary

- Study 1
  - ‘Laryngeal realism’ phonetic diagnostics partially hold up across 7 languages
- Study 2
  - Consonant F0 effects exist across 14 languages
  - Effect size varies greatly, but not direction
  - Phonology may matter (tone languages)
- Study 3
  - Voicing effect exists across 30 English dialects
  - Effect size varies greatly, but not direction
  - Phonology matters (Scottish Eng./AAVE)

Discussion

Q1: relationship between phonological representation and phonetic realization

- Study 1: partially supports laryngeal realism (data: laryngeal timing)
- Study 2 (&3?): broadly supports 'traditional' view (data: F0)
- Conflict

- Simple versions of both must be wrong

- Across languages:
  - Too much predictable variation in phonetic realization for the relationship to be arbitrary, or restricted to single cues (e.g. VOT) (LR motivation)
  - Too much unpredictable variation for a tight link to features (LR critics, e.g. Kirby & Ladd 2018; Ladd & Shmid 2019)

- Solution: I don't have one
  - but Q2 suggests a way forward

Discussion

Q2 .What is the typology of phonetic 'laryngeal' contrasts?

- What is clear is: there is structure here
- Possible dimensions (for initial position)
  - 'fortis'/'voicing': shown by F0 effects (just magnitude varies)
  - 'Slack' dimension (is /b/ "voiced"?)
  - 'Aspiration' dimension (is /p/ "aspirated"?)

VOT: first principal component

- Much recent work suggests this view:
  - Laryngeal contrasts lie in a space, but it is much more complex than traditional VOT-based (Abramson & Whalen, 2017)

2017 J. Phon special issue; *AMP* 2021 Burroni et al.; Indic languages (e.g. Schertz & Kahn, 2020), Swiss German (Ladd & Schmid, 2018)

Interim Discussion

- Whatever the dimensions are, they are constrained by
  - Articulation & perception
  - Larger system, e.g.
    - F0 use for lexical contrast
    - Phonological rules involving ‘voicing’
- We need much more phonetic data on laryngeal contrasts, cross linguistically, to map out the space they lie in!

Thanks

- RAs: Arlie Coles, Michael Goodale, Elias Stengel-Eskin, many more
- The SPADE team
- Comments: Hye-Young Bang, Pat Keating, Heather Goad, James Kirby, Simon King, MCQLL members
- Funding:

Fonds de recherche Société et culture Québec

SSHRC CRSH

INNOVATION.CA

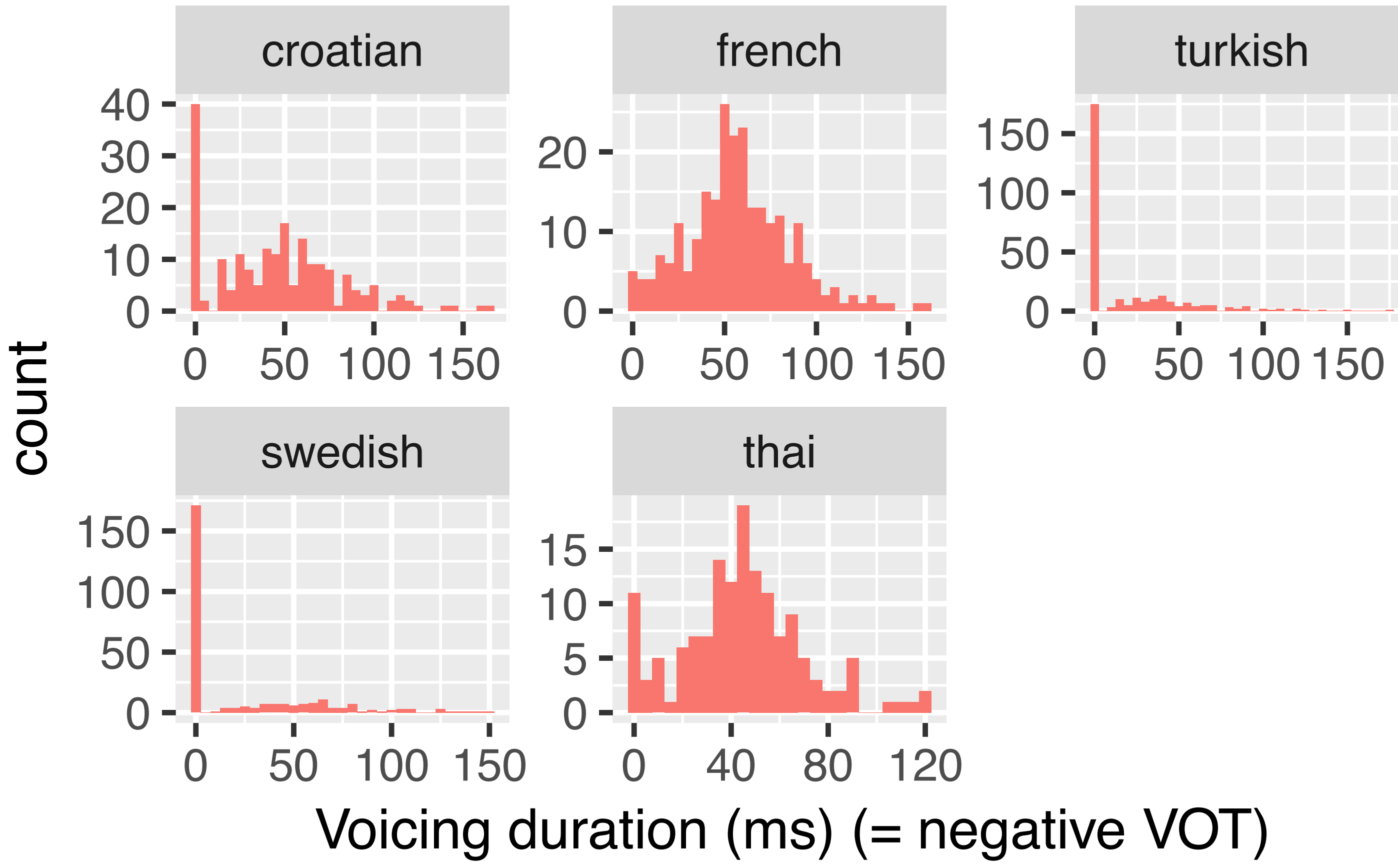
CANADA FOUNDATION FOR INNOVATION | FONDATION CANADIENNE POUR L'INNOVATION

Questions

Extra slides

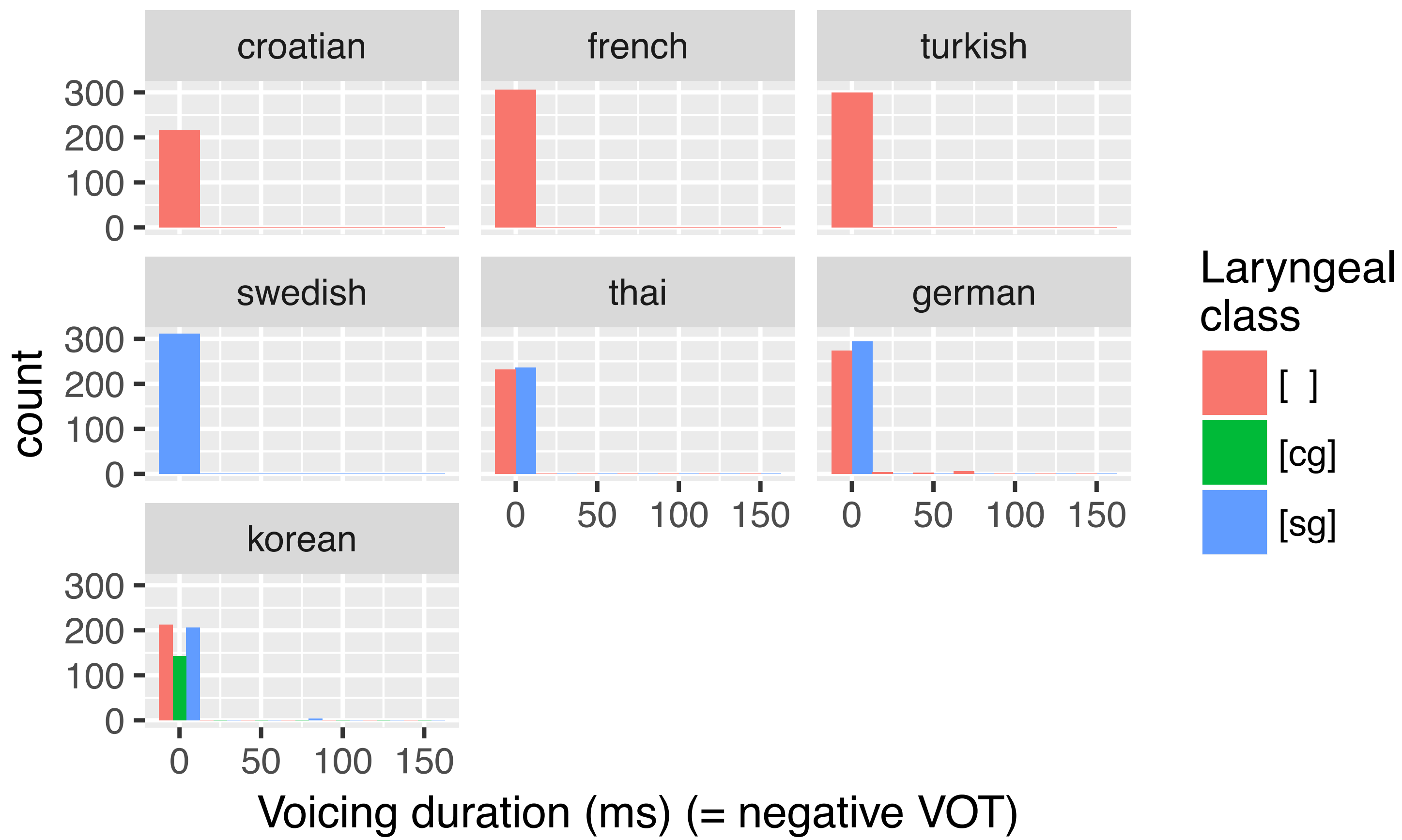
Negative VOT: "voiced" stops

- Amount of prevoicing for ##C [voi] stops :



Negative VOT: other stops

- Amount of prevoicing for ##C non-[voi] stops:



Data

: ##C position

	Cro	Fre	Tur	Swe	Thai	Ger	Kor
n	415	549	588	588	616	583	569

: VCV position

	Cro	Fre	Tur	Swe	Thai	Ger	Kor
n	349	367	293	310	344	344	369

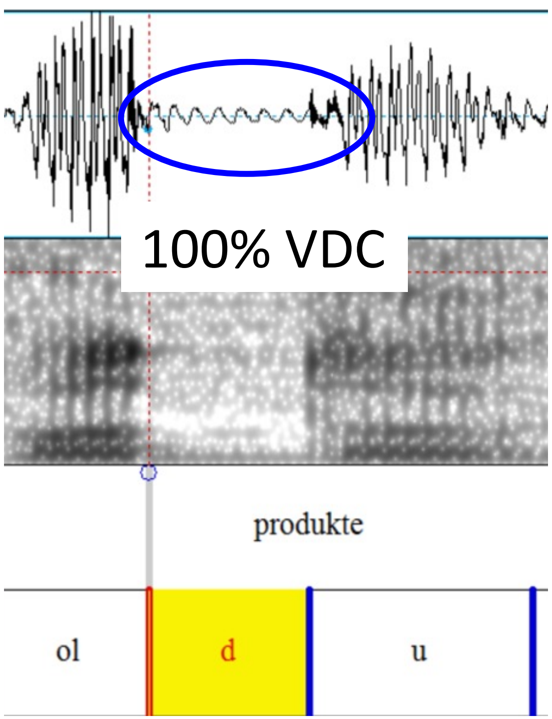
- *n* ~100-300 per laryngeal class/position/language
  - ≈ balanced by place of articulation
- Speech rate:
  - Phones/second, from forced alignment
  - Montreal Forced Aligner (McAuliffe et al., 2017)

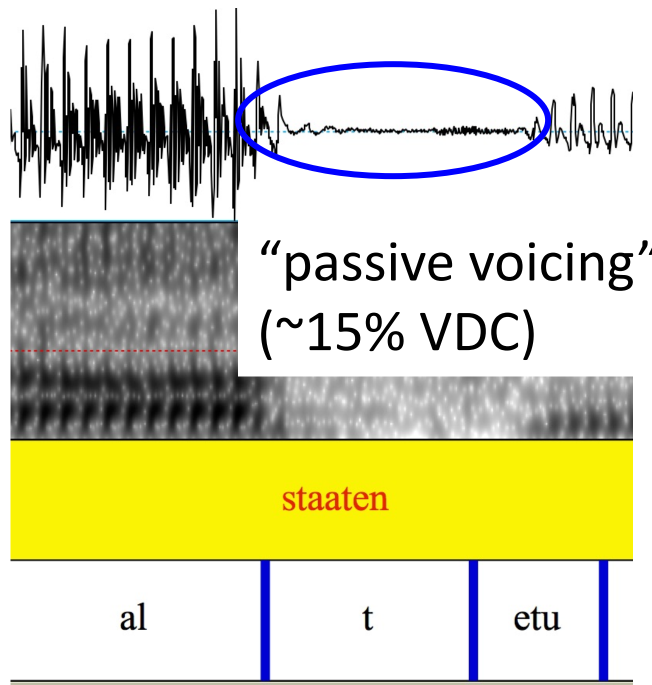
Data

Criterion 3

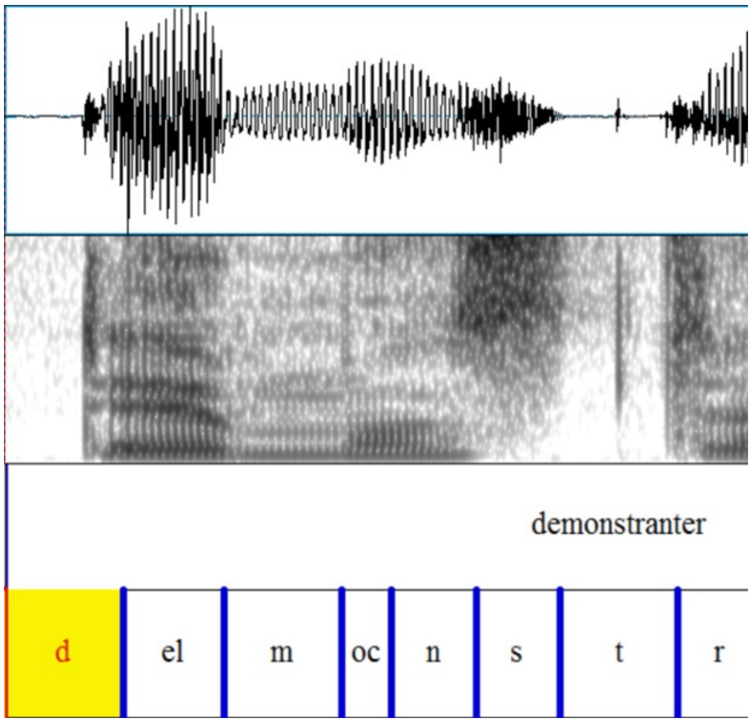
- Hand annotated: percent voicing during closure

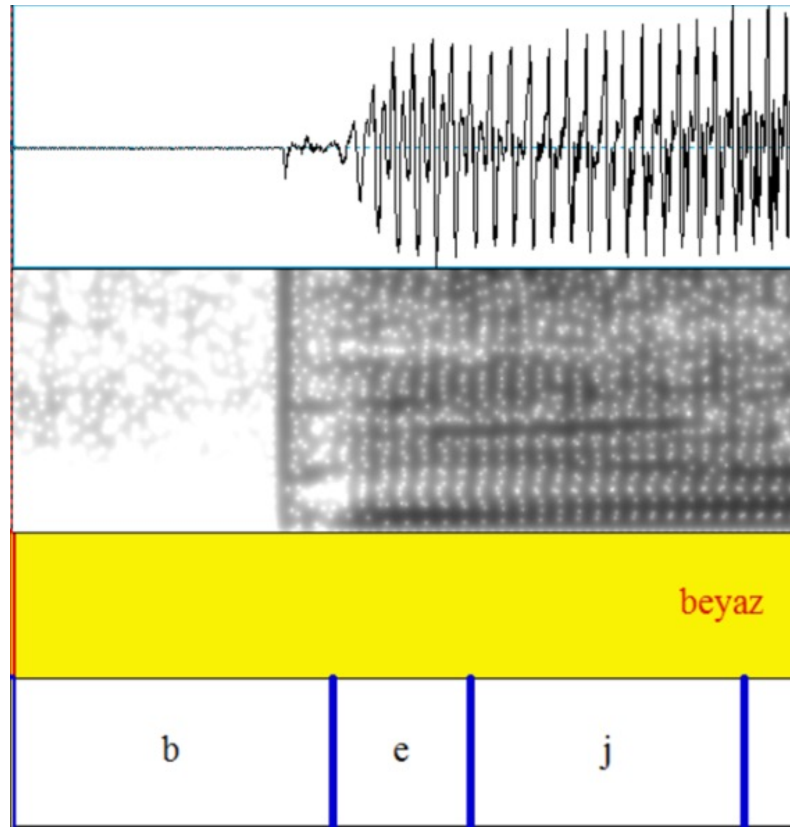
Examples: German





Swedish & Turkish examples





VCV position: Discussion

- Criterion 3: voicing during closure
  - ✓ (mostly)
- [voi] stops: near-full VDC
- non-[voi] stops
  - Mostly: low/inconsistent VDC
  - Exceptions:
    - Korean (due to phonology)
    - German
  - No evidence for distinction between languages with/without active [voi]
    - important LR prediction (Beckman et al., 2013; c.f. Kirby & Ladd 2019)

Discussion

[voi], [+voice], [stiff], etc.

- ##C prevoicing: standardly used to diagnose "voiced" stops cross-linguistically (Lisker & Abramson, 1964; etc.)
  - in lab speech / isolated words

artefact of hyperarticulation?

- Our data:
  - ##C "Voiced" stops not consistently prevoiced in read sentences: Turkish, Swedish, Croatian
- Also:
  - Dutch, Am. English (van Alphen & Smits, 2004; Davidson, 2015)
  - Glasgow spontaneous speech (Stuart-Smith et al. 2015)

Discussion

- Relationship between how languages realize laryngeal contrasts <u>across positions</u>
  - Novel
  - Could account for via features, or "controlled" phonetics (Solé, 2007)?
  - More data needed to test

- Future work:
  - – Codas
  - – More languages
  - – More cues (e.g. F0)

TODO: cut/combine

Datasets

- Data cleaning: minimize F0 errors, reduced vowels
- Exclusions:
  - Speakers: multimodal F0 distribution (non-tonal langs)
  - Vowel tokens:

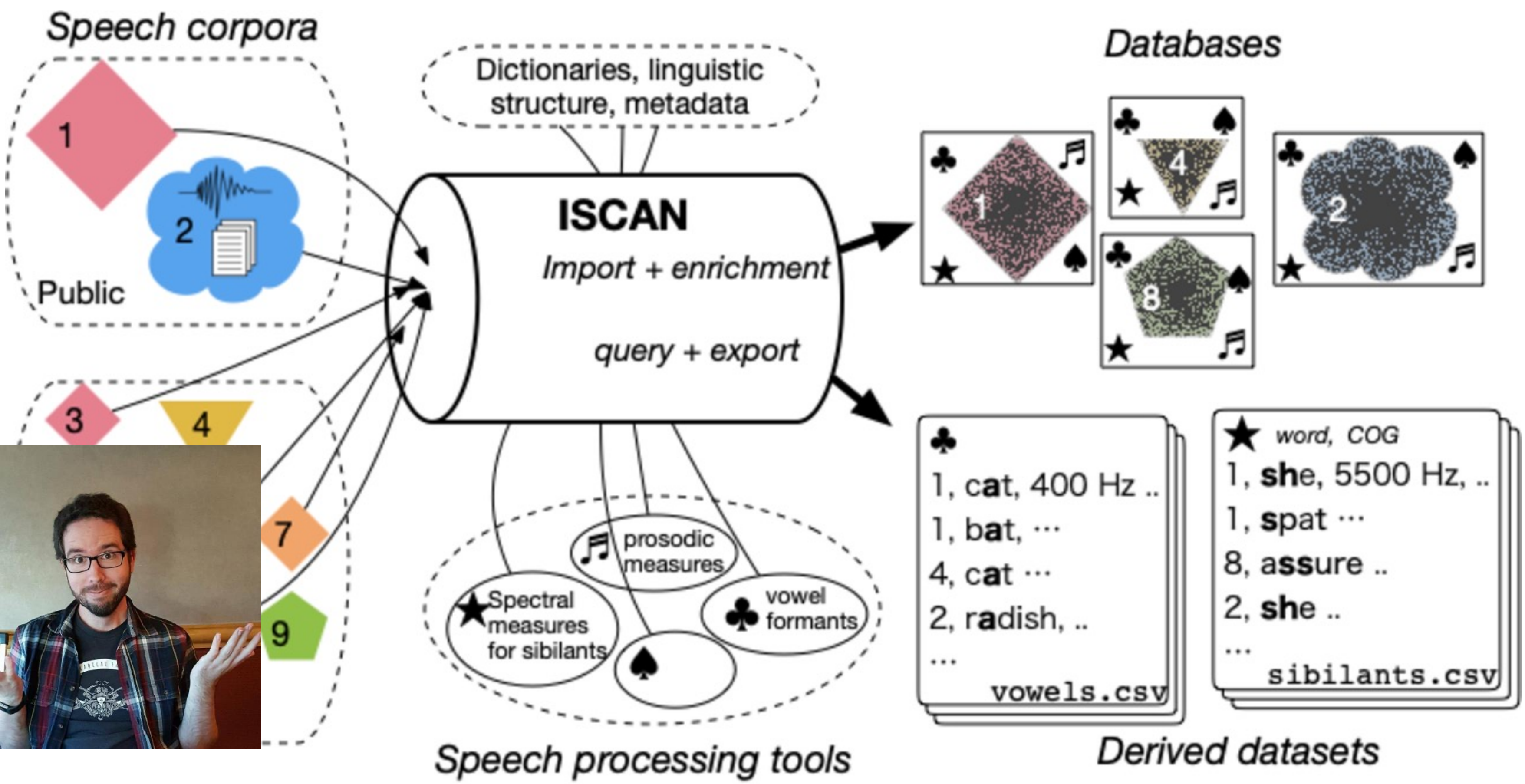
< 50 msec

< 50% voiced

Extreme values of DV, within-speaker

- Data per language:
  - 1.9-9.5k tokens (~2000)
  - 76-132 speakers (~100)

Software: Integrated Speech Corpus ANalysis



Michael McAuliffe
Software development

McAuliffe et al. (2019) *Proc. ICPhS*

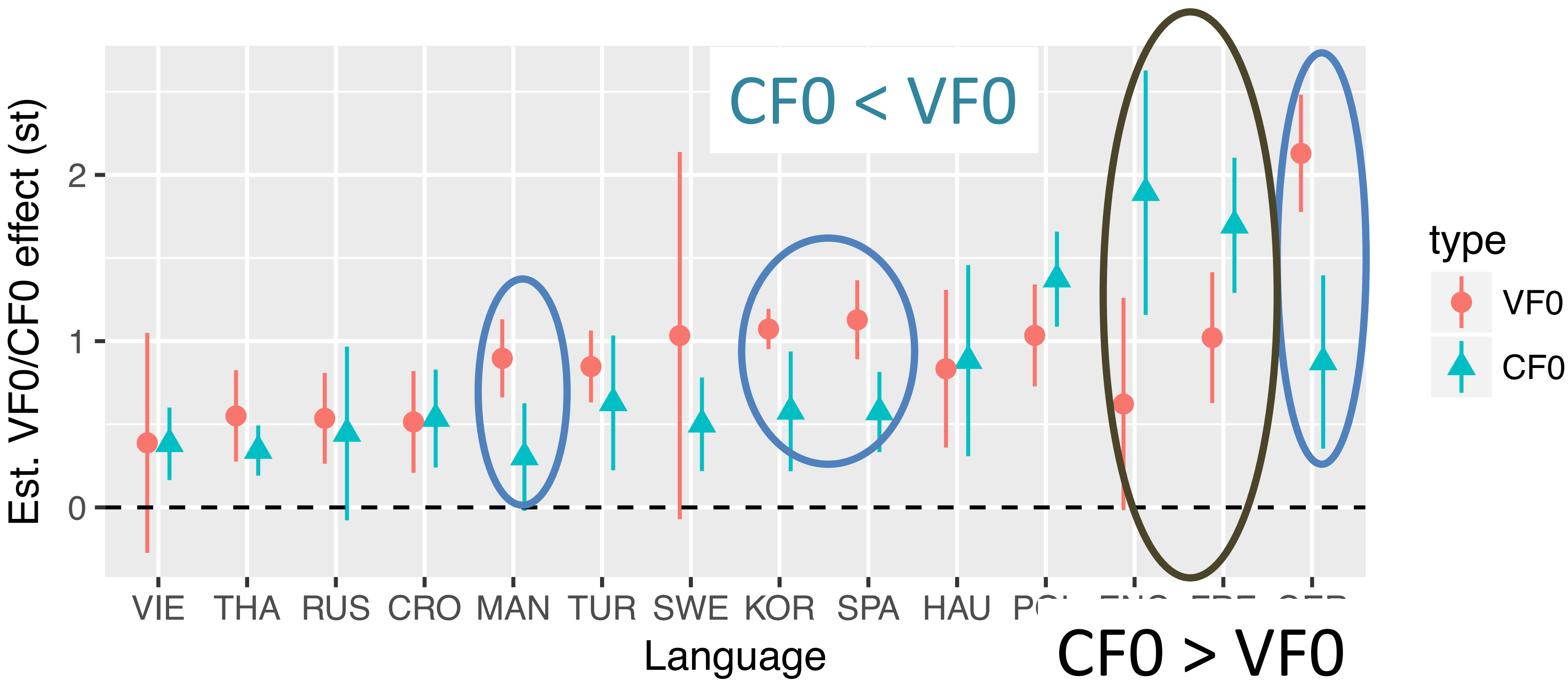
Extra: VF0 vs. CF0

- Asymmetry between IF0 effects w.r.t. sound change:
  - CF0: many attested changes
  - VF0: ~none

- Why?
  - VF0/CF0 magnitude roughly similar? (Hombert et al., 1979)
  - Perhaps perception is different (Hombert, 1979)
  - VF0 effects show more variability? (Kingston, 2011)

- Q4: Relative magnitude, variability of CF0 & VF0 across languages?

VF0 vs. CF0: effect size

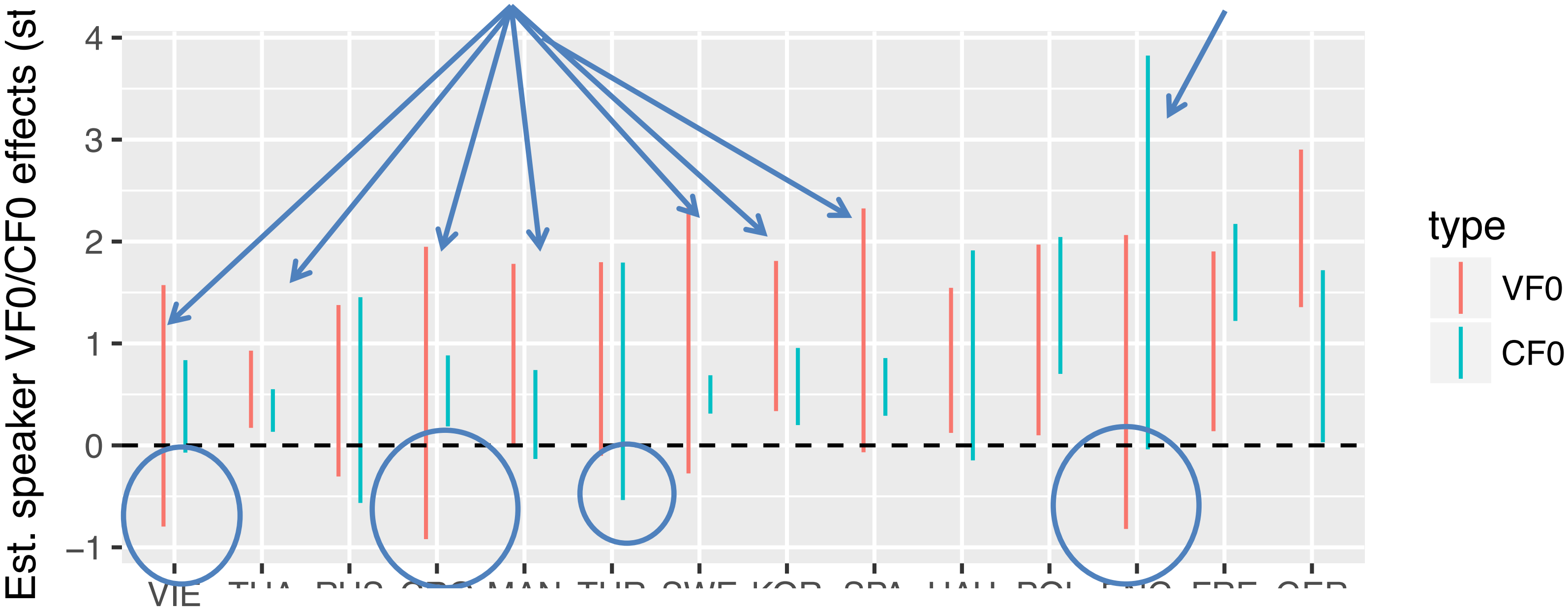


- No clear pattern
- CF0, VF0 of ~comparable size

VF0 vs. CF0: speaker variability

VF0 >> CF0

CF0 >> VF0



Minority of speakers show reverse effects

- Overall: no obvious pattern
- But: some evidence that VF0 “more variable” than CF0

- VE very sensitive to context and style:

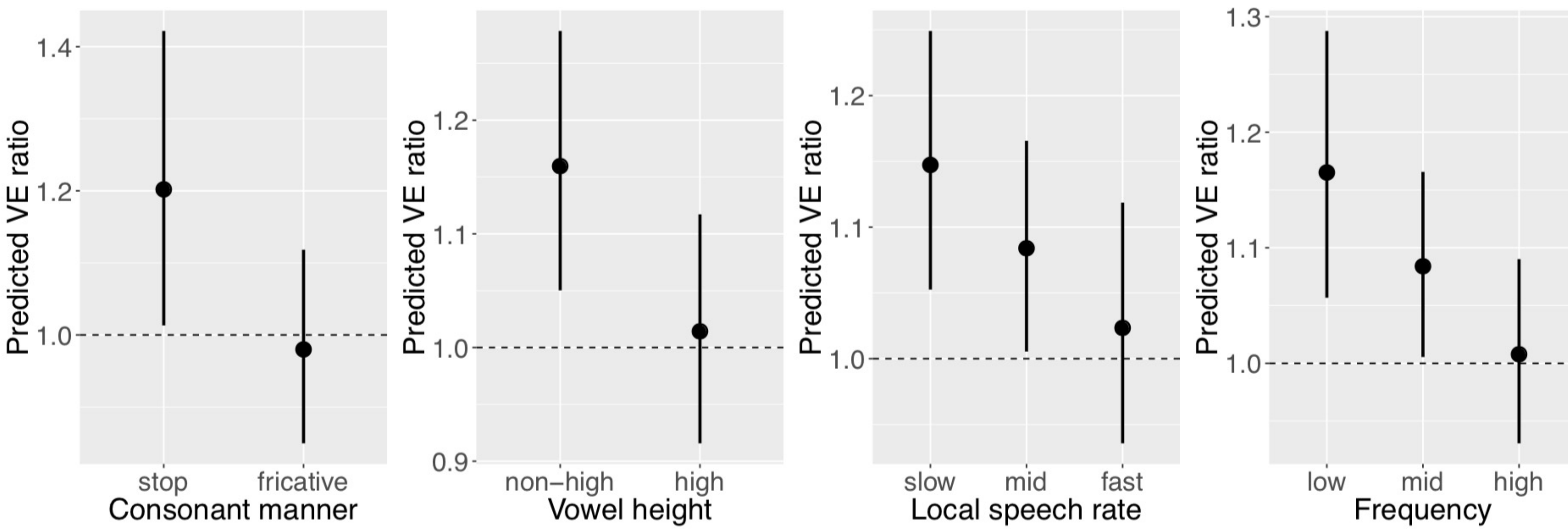


Figure 1: Predicted voicing effect (median and 95% CrI) as a function of consonant manner, vowel height, local speech rate (at -1, 0, 1), and word frequency (at -1, 0, 1). Predictions based on regression lines computed from the model's posterior, marginalising over all other covariates.

Discussion

- IF0 effects can be detected using
  - Corpus data
  - Fully automatic analysis
  - Basic statistical controls
  - *n* =~2-4k
- Not obvious!

- Demonstrates feasibility of large-scale studies of phonetic precursors (involving F0)

- TODO: More extra for vowel duration

Table 3.1: Number of speakers and tokens per dialect (left), and by corpora from which each dialect was derived.

Region	Dialect	n speakers	n tokens	Corpus	n speakers	n tokens
North America	Canada (rural)	52	9313	Canadian Prairies	44	8316
				ICE-Canada	8	997
	Canada (urban)	64	12124	Canadian Prairies	56	11939
				ICE-Canada	8	185
	Midwest US	40	5567	Buckeye	40	5567
	New England	24	1336	Santa Barbara	7	174
				Switchboard	17	1162
	North Midland US	46	3084	Switchboard	46	3084
	Northern Cities US	21	1377	Santa Barbara	21	1377
	Northern US	58	3086	Switchboard	58	3086
	NYC	25	1477	Santa Barbara	6	158
				Switchboard	19	1319
	Philadelphia	371	59581	PNC	371	59581
	Princeville NC (AAE)	71	6759	CORAAL	17	6759
	Raleigh US	92	3282	Raleigh	92	3282
	Rochester NY (AAE)	14	6308	CORAAL	14	6308
	South Midland US	108	8188	Switchboard	108	8188
	Southern US	44	2738	Santa Barbara	6	345
				Switchboard	38	2393
	Washington DC (AAE)	50	21205	CORAAL	50	21205
	Western US	100	5456	Santa Barbara	50	2900
				Switchboard	50	2556
United Kingdom & Ireland	Central Scotland	24	2426	SCOTS	24	2426
	East Central England	51	2544	Audio BNC	51	2544
	East England	229	20727	Audio BNC	132	6622
				Doubletalk	5	726
				Hastings	44	12642
				ModernRP	48	737
	Edinburgh	18	1148	SCOTS	18	1148
	Glasgow	177	33938	Brains in Dialogue	23	9210
				SCOTS	27	2294
				SOTC	127	22434
	Insular Scotland	8	351	SCOTS	8	351
	Ireland	19	624	Audio BNC	19	624
	Lower North England	60	3325	Audio BNC	60	3325
	North East England	17	488	Audio BNC	17	488
	Northern Scotland & Islands	33	2280	SCOTS	33	2280
	Scotland	70	3468	Audio BNC	65	2633
				Doubletalk	5	835
	South West England	50	2067	Audio BNC	50	2067
	Wales	41	2524	Audio BNC	41	2524
	West Central England	41	2615	Audio BNC	41	2615
Total		1964	229406			