The dynamics of sounds and contrasts on reality television

Morgan Sonderegger

McGill University

Apr 22, 2015

Introduction

- What does phonetic/phonological variation in individuals look like over time?
  - dynamics

- Causes of dynamics?

- Relationship to community-level sound change?

Variation in individuals over time

- Short term: phonetic imitation/convergence/ accommodation (Giles et al., 1991; Goldinger,1998; Pardo, 2006; Babel, 2009...)
  - Widespread, robust
    - Most variables (VOT, vowels, …) , most speakers
  - Mediated by social, linguistic factors
  - Minutes-days

- Hypothesis: Short-term accommodation/ imitation a major source of language change (Neogrammarians; Pardo, 2006; Delvaux & Soquet, 2007)

Variation in individuals over time

- Long term
  (Munro et al., 1999; Harrington et al., 2000; Evans & Iverson, 2007; Siegel, 2010)
  - Panel studies (Sankoff, 2005, 2012)
    - Individuals stay in same speech community
  - Dialect change/acquisition/shift (Siegel, 2010)
    - Individuals move
  - Measure at a few time points years apart

- Huge variation among speakers, variables
  - Adults: Stability the norm, some change significantly

What is the relationship between the different patterns seen in short-term and long-term dynamics?

A “medium term” experiment

- Months
- Trajectories of
  - Phonetic & phonological variables
  - (Social dynamics)

- Track how variables change between endpoints
  - Longitudinal variation

- Link between short and long term.

Big Brother

- Reality TV program from the Netherlands
- Exported to UK, US, Germany…

Big Brother UK: Season 9

- Contestants spend 3 months in BB house
- Each week one is voted off
  (+ sporadic additions)
- Last remaining wins £100,000

- No outside contact: closed system
- Continuous surveillance
  - Cameras in every room
  - Wearable microphones

11 native speaker contestants on for >50 days:
≈ 80% of data

Data

- Live 24-hour feed (!)

- Daily produced episodes (1 hour)
  - Easier to obtain

- Speech data from diary room clips
  - Talk to Big Brother, semi-spontaneous (c.f. Buckeye)
  - Constant recording environment, social context.
  - ≈10.5 hours

Speaker origin

- England: 3 northern, 3 southern, 1 W midlands

- Scotland: 1

- Wales: 1

- US/UK: 1

- Australia: 1

Analysis

- High level:
  for each variable

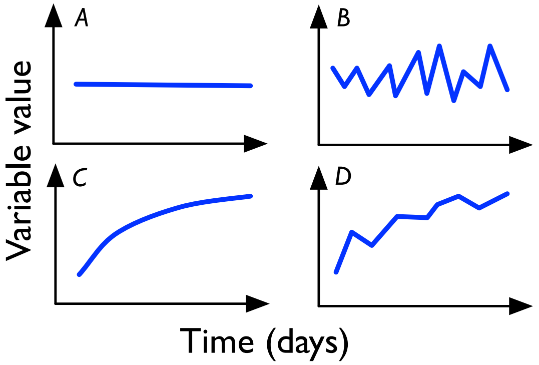
VOT

Coronal stop deletion

Vowel formants

  – Determine
    time dependence
    within individual
    speakers

=



  – Controlling for
    static factors

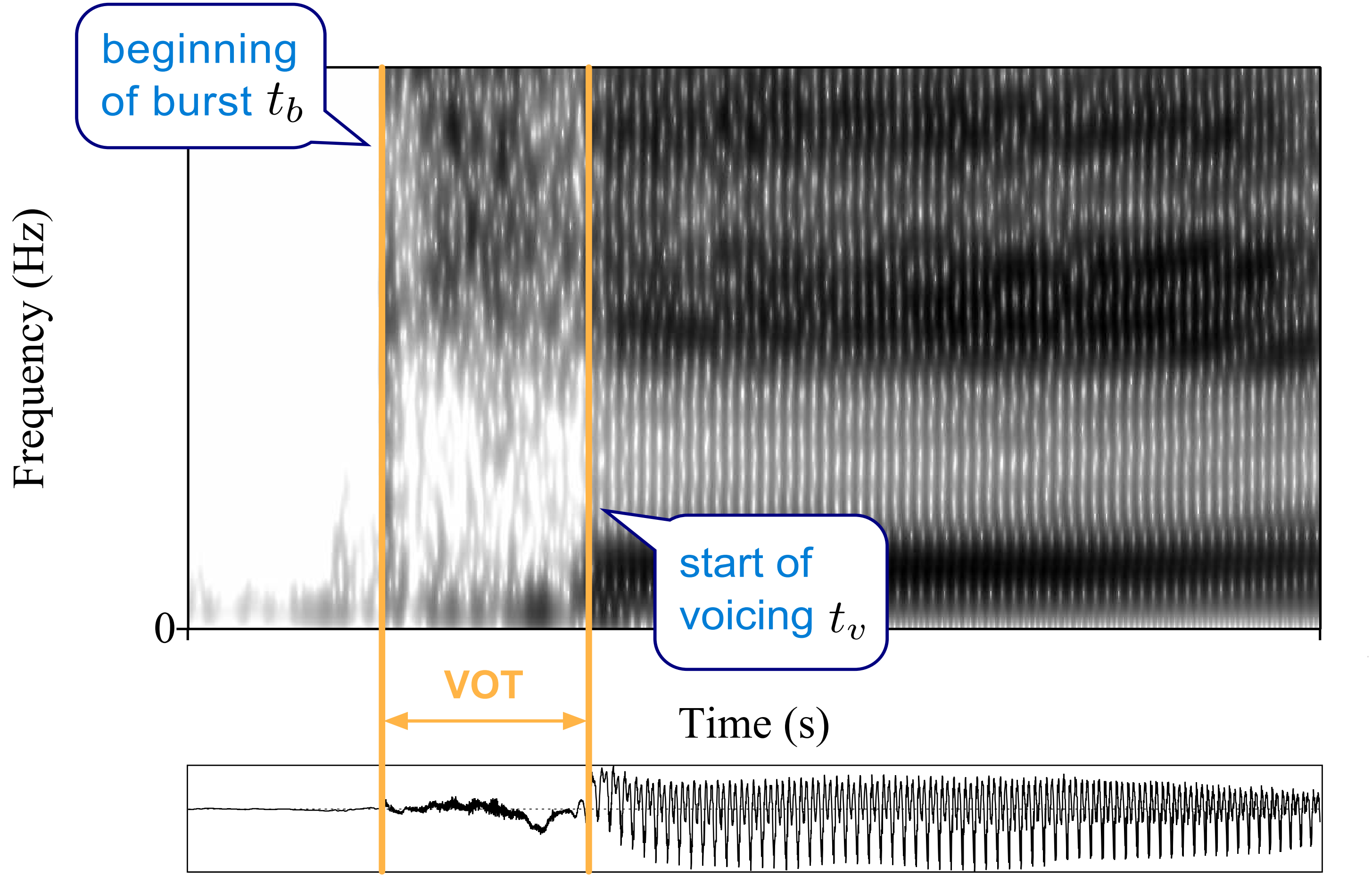
–

Speaking rate

coarticulation

syllable structure

Variable 1:VOT



- Primary cue to voicing contrast, for stop consonants

Data

- Procedure:
  - Semi-automatic measurement
    1. Automatic: AutoVOT (Keshet et al. 2014; Sonderegger & Keshet 2012)
    2. Manual correction
       - Including exclusions (fricatives, deleted, …)
  - vs. fully manual measurement:
    - 20-30x faster
    - very similar measurements*

https://github.com/mlml/autovot

* Auto/manual reliability same order as intertranscriber reliability

Data

- Which stops?
  - “VOT” complex in spontaneous speech
    - Strict definition: lose >50% possible tokens
    - Loose definition: include tokens w/o closure, etc.
  - Our choice: loose
    - positive VOT, ≈ any stop with a burst
    - ⇒ VOT ≈ burst duration
    - (voicing duration, neg. VOT not examined)
- All word-initial stops
  - *can, burning, today, *today*

Data

- Dataset:
  - Voiced: 10.6k tokens (709 words)
  - Voiceless: 10.1k tokens (893 words)

(phonologically)

  - 11 speakers (>50 days, native)
    - 800-3300 tokens/speaker
    - 32-80 clips/speaker

0+ clips per speaker per day

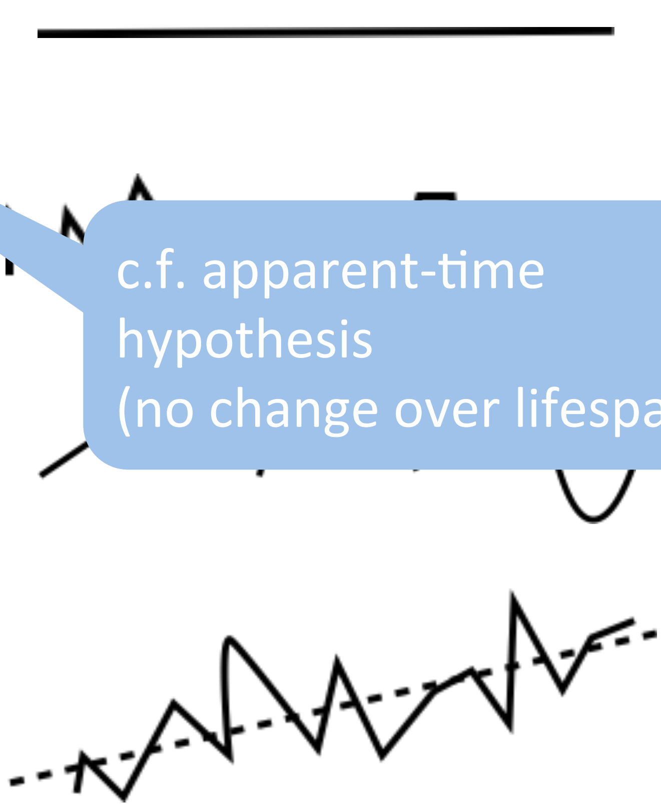
Analysis

- Many static factors affecting VOT:
  - Speaking rate (slower > faster)
  - Place of articulation (p ≤ t ≤ k)
  - Following segment (C > V)
  - Following V height (high > non-high)
  - Stress (stressed > unstressed)
  - Word frequency (higher > lower)

(Allen et al., 2003; Baran et al., 1977; Crystal & House, 1988; Klatt, 1973,1975; Lisker & Abramson, 1965; Miller, 1986; Miller et al., 1986; McCrea & Morris, 2005; Morris et al, 2008; Nearey & Rochet, 1994; Ohala, 1981; Port & Rotunno, 1979; Randolph, 1989; Schertz 2013; Stuart-Smith et al., in press; Summerfield, 1975; VanDam and Port, 2005; Volaitis and Miller, 1992; Yao, 2009; Zue, 1976...)

Analysis

- Time dependence: no a priori hypothesis!
- Possibilities:
  - None (null hypothesis)
  - By-day variability
  - Time trend
  - Time trend and by-day variability



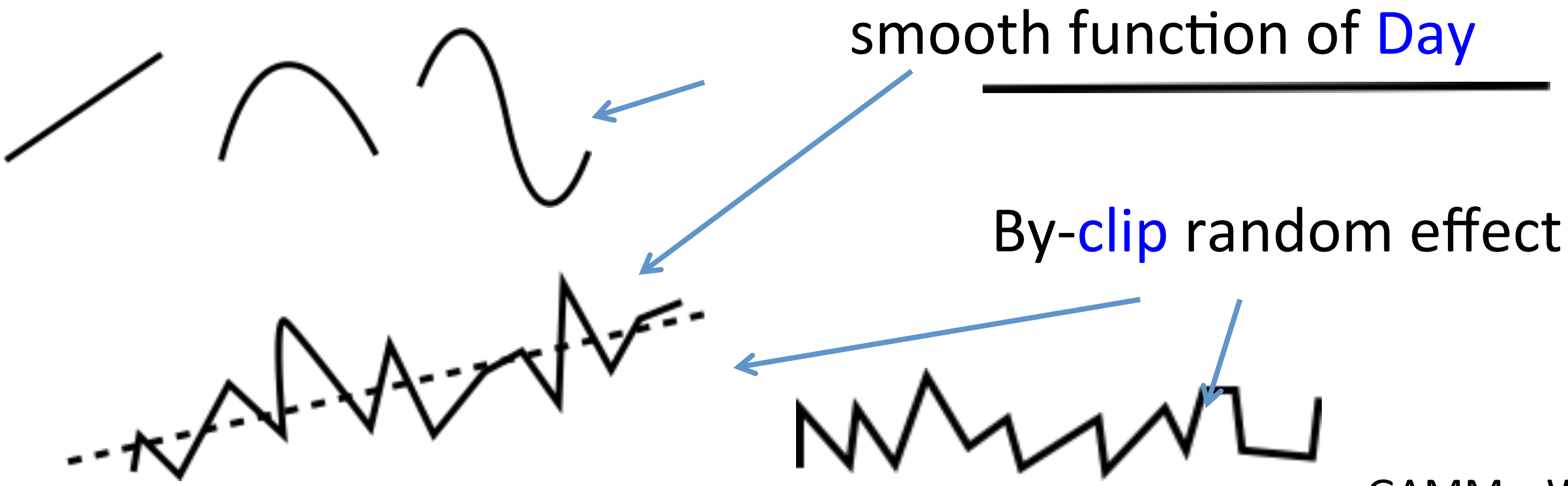
c.f. apparent-time hypothesis
(no change over lifespan)

Analysis: models

1. Build 2 linear mixed-effect models (voiced, voiceless) of static factors, across all speakers
   - Response: log(VOT)
   - Fixed effects: static factors (+ interactions)
   - Random effects: (speaker, word) x (intercept, slopes)

   - Residuals of these models : normalized VOT for speaking rate, context, etc.

2. For each speaker, for voiced/voiceless subset, four models of time dependence
   - Response: normalized VOT
   - Generalized additive mixed model
   - By-word random effect
   - Time dependence: one of



smooth function of Day

By-clip random effect

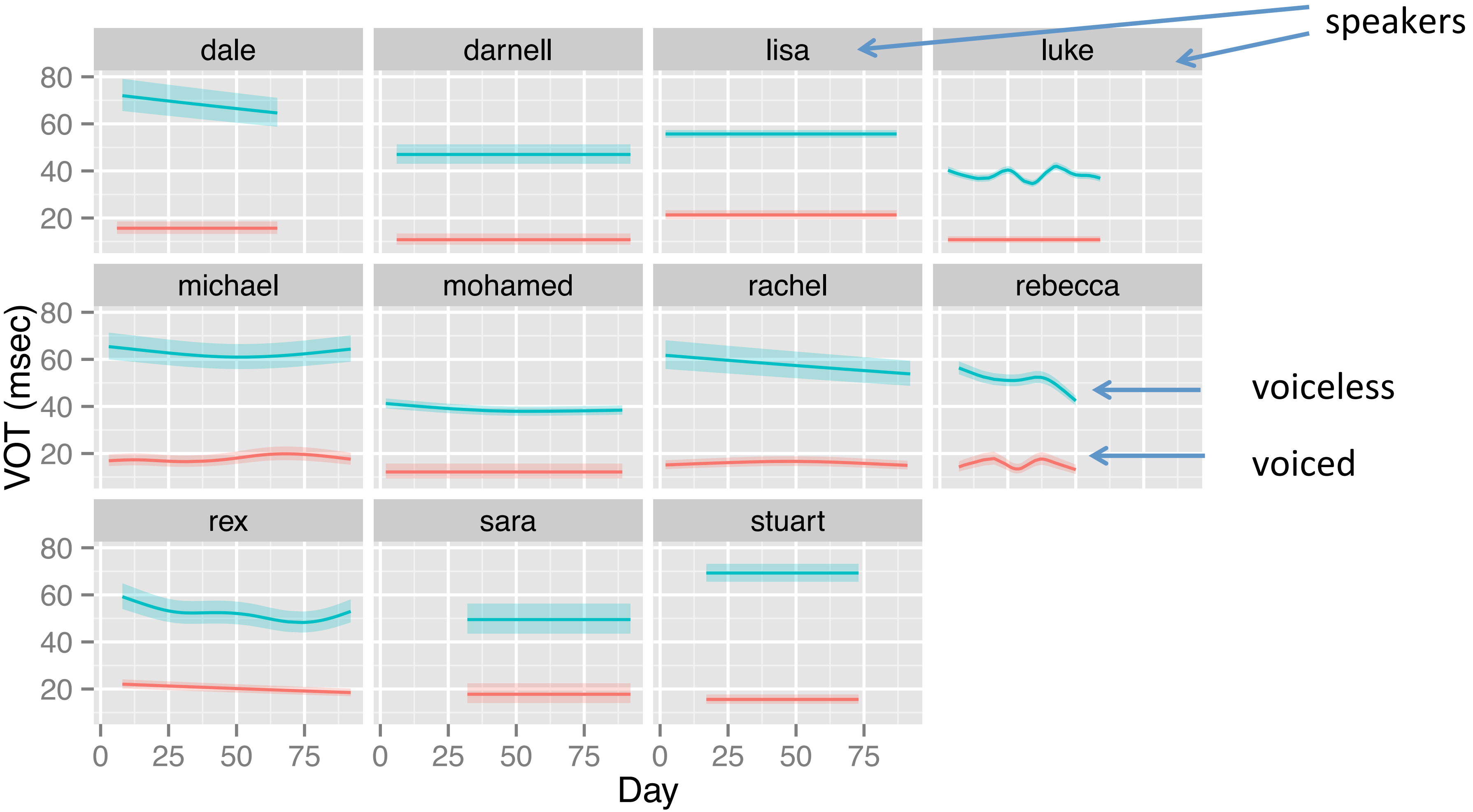
GAMMs: Wood 2006

Analysis: models

- Choose best of four models using Akaike Information Criterion (AIC)

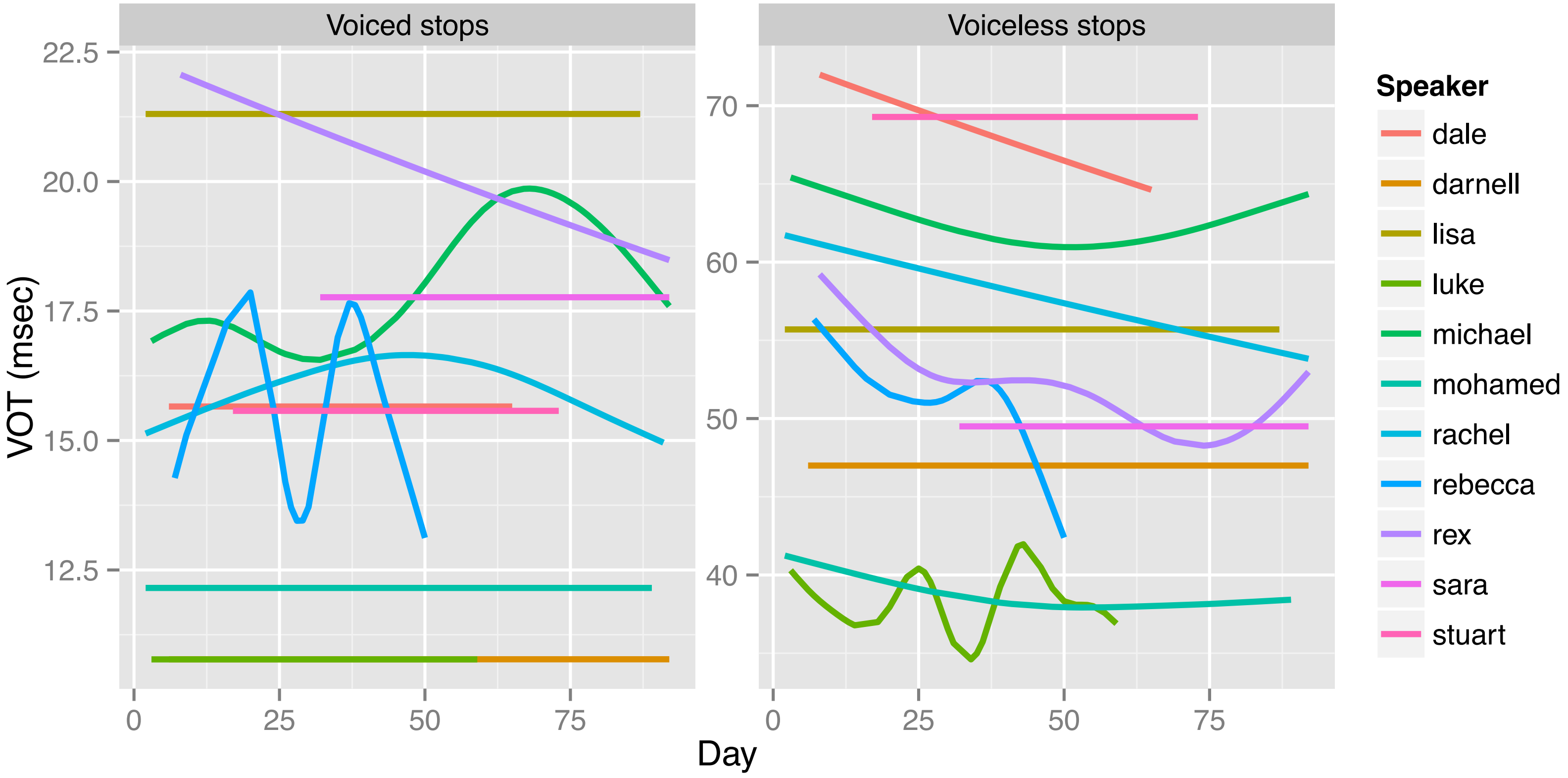
- ⇒ <u>one</u> model of time dependence for
  - Speaker 1, voiceless stops
  - Speaker 1, voiced stops
  - (etc.)

Results: predicted time dependence



- By-day variability (ribbons): all cases
- Time trends (non-horizontal lines): 50% of cases

Results: time trends



- No clear convergence

Results: by-day variability

- Time dependence is ubiquitous
  - Is it important?

Predicted diff between +-1σ days

- By-day variability effect size :
  - Voiced: 43-180% / 8-13 ms
  - Voiceless: 13-48% / 7-26 ms
- Compare: place of articulation (strongest static factor)
  - Voiced: 9 ms
  - Voiceless: 27 ms

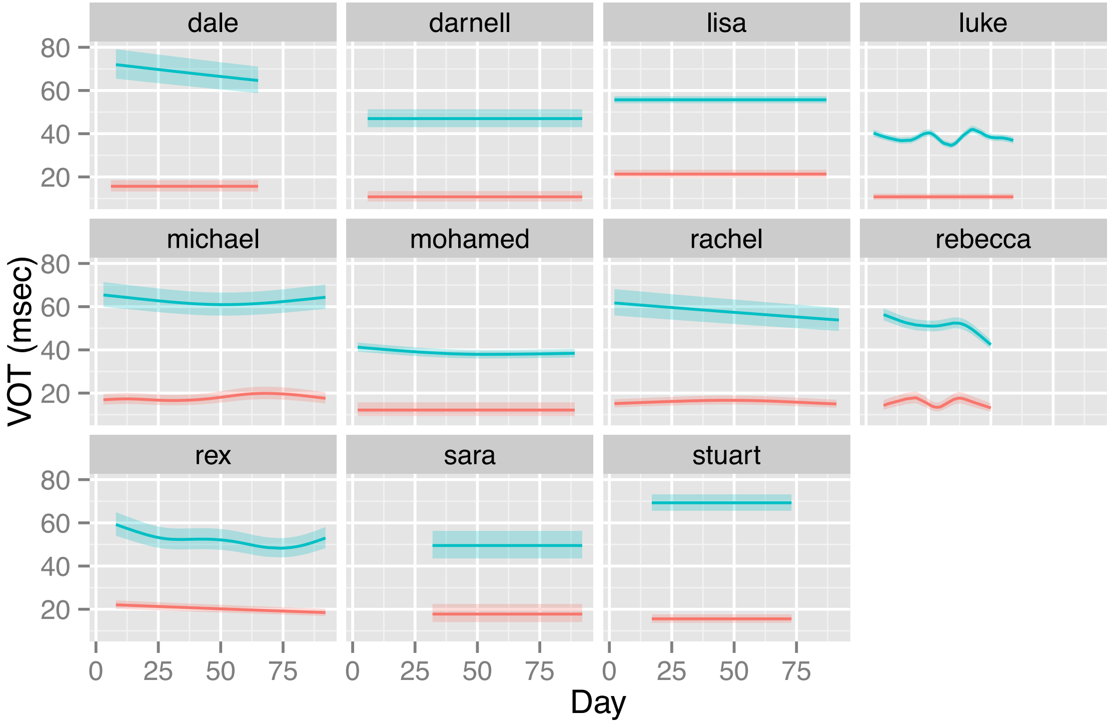
By-day fluctuations are of similar magnitude to contextual effects

- Compare: short-term voiceless VOT shifts (Nielsen, 2011; Shockley et al., 2004)
  - Shadowing: 12 msec (avg)
  - Imitation: 0-30 msec

By-day fluctuations are of similar magnitude to accommodation effects

Results: voiced and voiceless

- Compare: magnitude of voiced/voiceless VOT difference (primary cue to contrast)



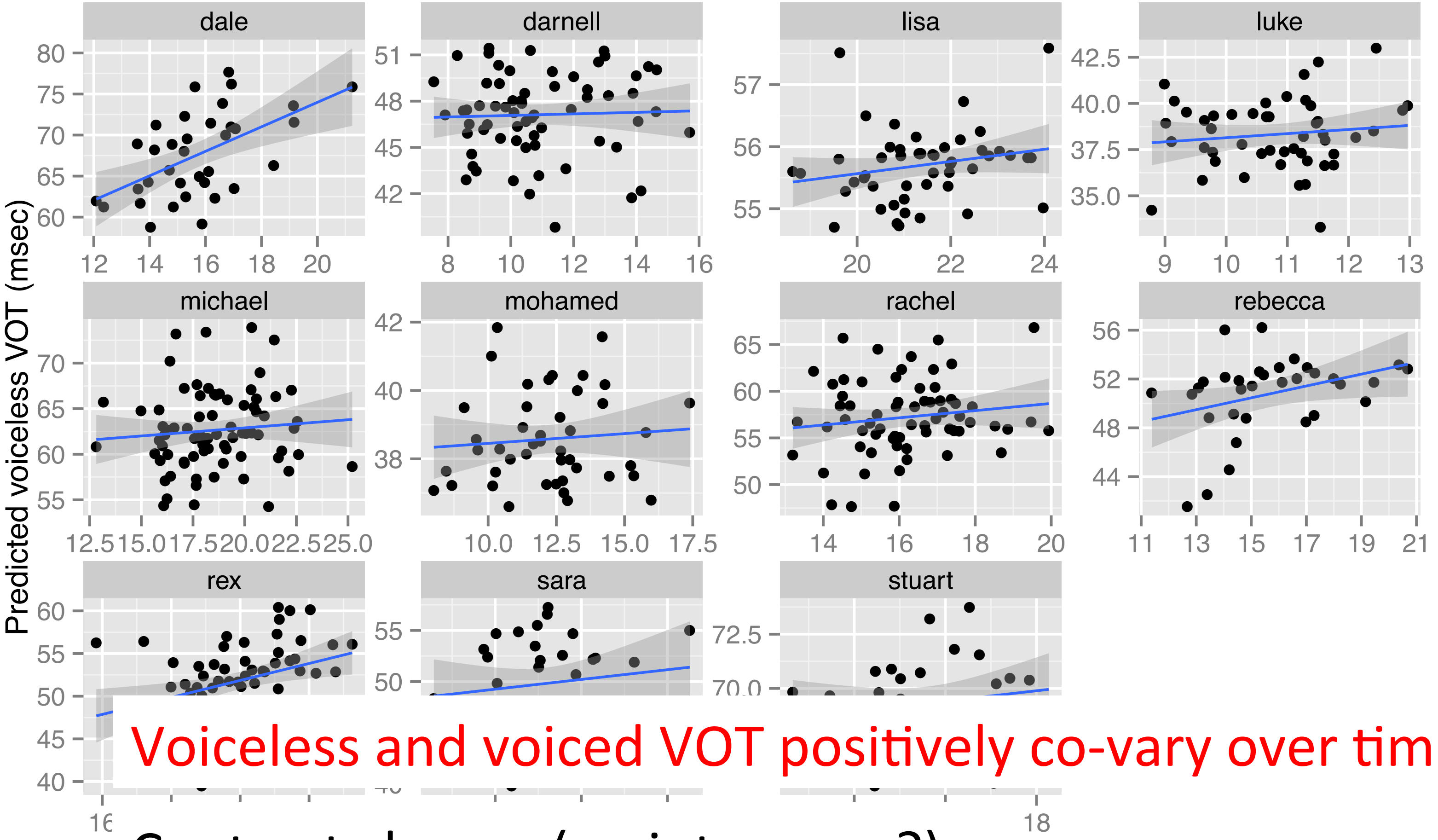
Magnitude of time dependence never sufficient to endanger contrast

Results: voiced and voiceless

- Change in sounds, or voicing contrast?
  - Do voiced, voiceless change together?

Results: voiced and voiceless predictions

- (1 point = 1 clip)



Voiceless and voiced VOT positively co-vary over time

Contrast change (maintenance?)

$p < 0.01$

Variable 2: coronal stop deletion

- Word final t/d variably deleted in consonant clusters
  - *wan'~want* , *slep'~slept*
  - *bes'~best*

BEST BUY

(Labov et al., 1968; Wolfram, 1969; Fasold, 1972; Labov, 1975; Wolfram & Christian, 1976; Guy, 1980, 1991; Neu, 1980; Labov, 1989; Guy & Boyd, 1990; Santa Ana, 1992, 1996; Bayley, 1994; Reynolds, 1994; Roberts, 1995, 1997; Patrick, 1999; Schreier, 2005; Tagliamonte & Temple, 2005; Hazen, 2011 …)

Data

- Annotation
  - Spectral cues + auditory
  - 9 labels (burst, glottal stop…) collapsed to present/absent
    (c.f. Temple, 2014)

- Dataset
  - 11.6k tokens, 538 types
  - 11 speakers
    - 551-1174 tokens/speaker

Analysis

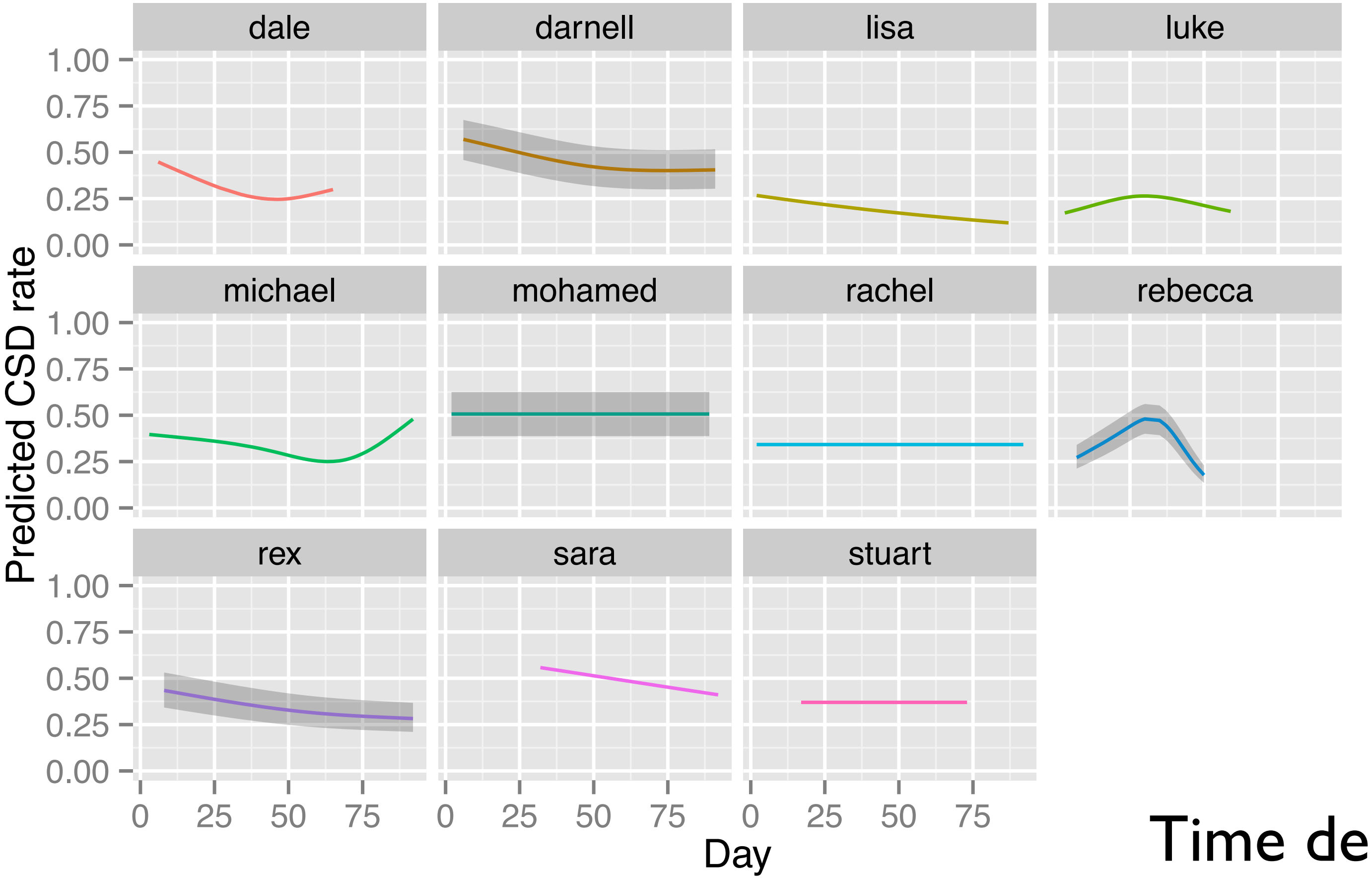
- Static factors affecting CSD rate:
  - Following context (t/d > consonants > vowels ~ pauses)
  - Preceding context (Tagliamonte & Temple, 2005)
    - /s/ > liquids > nasals > stops > sibilants
  - Frequency (higher > lower)
  - Speaking rate (higher > lower)
  - Voicing (*bust* > *want*)
  - Morphological class (*mist* > *missed)*

Analysis: models

- For each speaker, build mixed effects logistic regression models
  - Response: t/d realization
  - Accounting for static factors
  - Different types of time dependence
- Choose best one (AIC)
- (similar procedure to VOT)

- ⇒ one model of time dependence per speaker

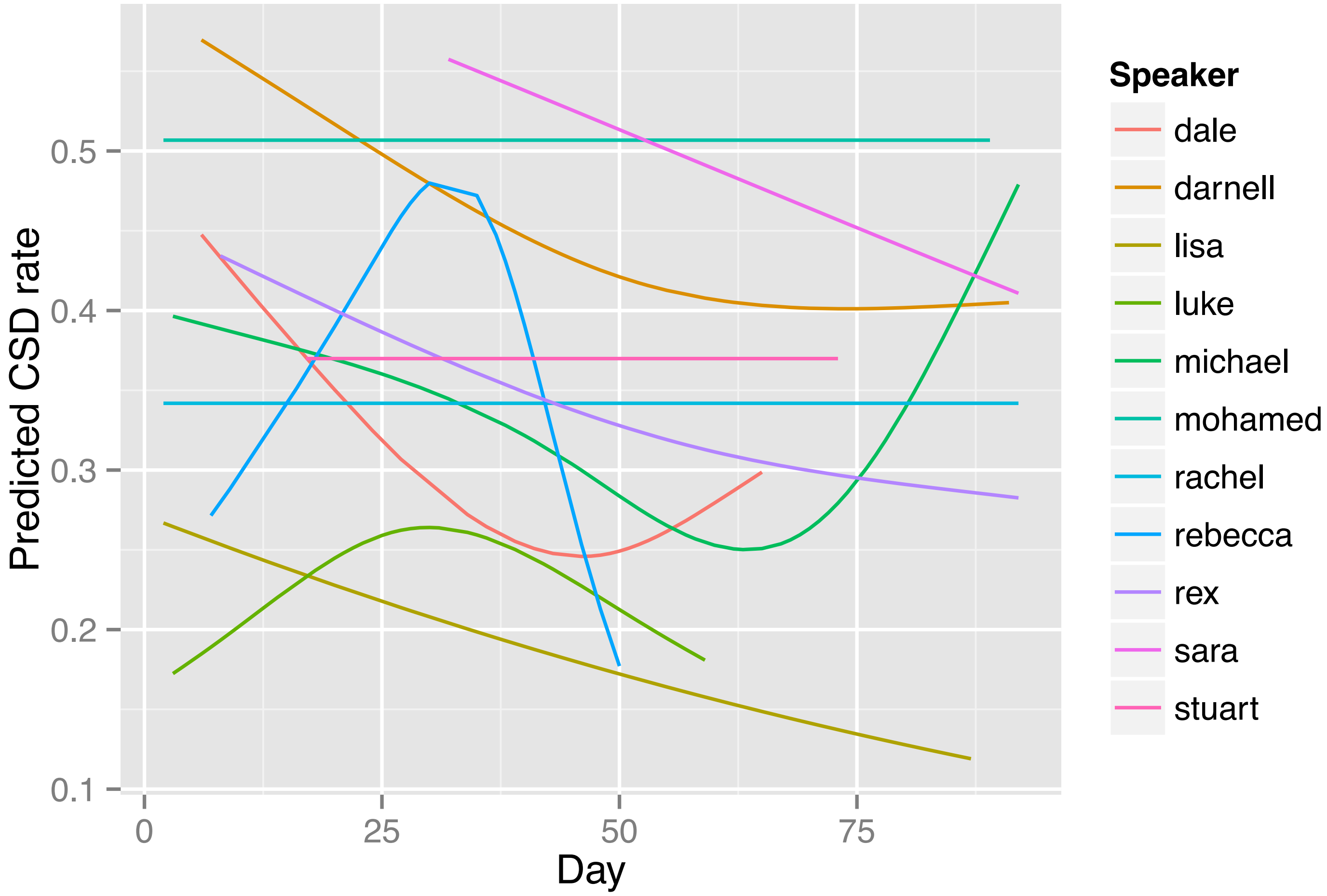
Results: predicted time dependence



Time dependence: 82%

- By-day variability (ribbons): 36% of cases
- Time trends (non-horizontal lines): 73% of cases

Results: time trends



- Downward trend (more casual)?
- No clear overall convergence

Results: by-day variability

- Effect size:
  - 8/12 speakers: 0
  - Rest: 1.9-2.6x increase in CSD odds
    - ≈ 16-24% `` `` CSD rate

- Compare: strongest static factors
  - Speaking rate: 5.0
  - Following context: 2.9

By-day fluctuations smaller than contextual effects

- Compare: short-term shifts
  - No imitation studies to compare to, but..
  - by-day fluctuations similar magnitude to style-shifting effects
    (Hazen, 2011)

Variable 3: vowel formants

1. GOOSE

[u]	[ʉ]~[u]	[ʉ]
🏴󠁧󠁢󠁷󠁬󠁳󠁿 🇺🇸	🏴󠁧󠁢󠁥󠁮󠁧󠁿N 🏴󠁧󠁢󠁥󠁮󠁧󠁿S	🏴󠁧󠁢󠁳󠁣󠁴󠁿 🇦🇺

2. TRAP′

[a]	[a]~[æ]	[æ]
🏴󠁧󠁢󠁥󠁮󠁧󠁿N 🏴󠁧󠁢󠁳󠁣󠁴󠁿 🏴󠁧󠁢󠁷󠁬󠁳󠁿	🏴󠁧󠁢󠁥󠁮󠁧󠁿S	🇺🇸 🇦🇺

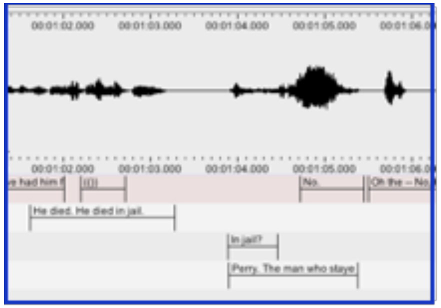
3. STRUT

[ʌ] ~ [ɐ]	[ʊ] (=FOOT)
🏴󠁧󠁢󠁷󠁬󠁳󠁿 🏴󠁧󠁢󠁳󠁣󠁴󠁿 🇦🇺 🇺🇸 🏴󠁧󠁢󠁥󠁮󠁧󠁿S	🏴󠁧󠁢󠁥󠁮󠁧󠁿N

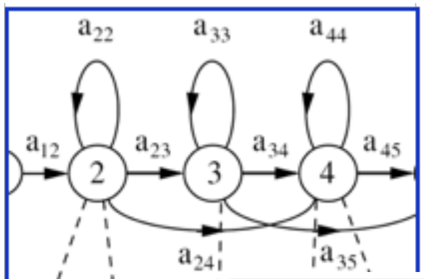
(Ferragne & Pellegrino, 2010; Wells, 1982)

Data

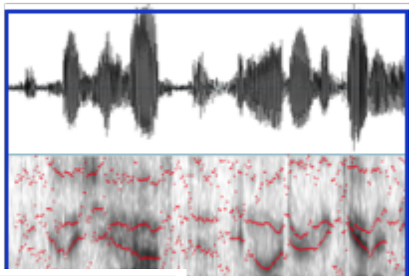
- Semi-automatic F1, F2 measurement
  1. FAVE suite (Rosenfelder et al, 2011)



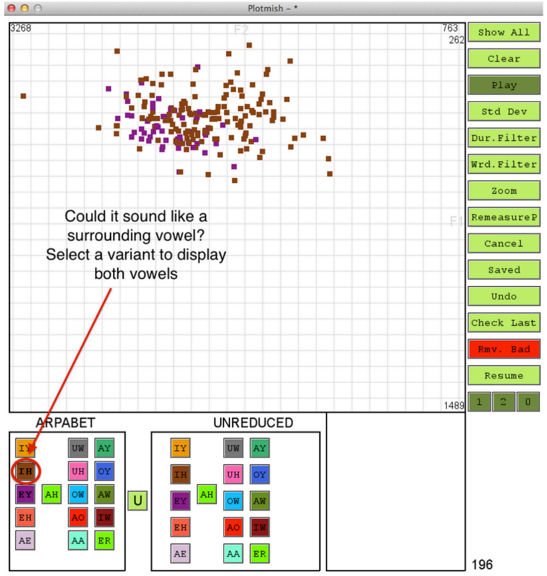
Transcription



Forced al



measurement



Could it sound like a surrounding vowel? Select a variant to display both vowels

  2. Manual correction: Plotmish (github.com/mlml/plotmish)

Data

- Dataset:
  - GOOSE: 2.9k tokens
  - TRAP' : 2.3k tokens
  - STRUT: 4.9k tokens
- Exclusions:
  - Reduced
  - Highest-freq words (e.g. *and*)
  - (etc.)

Analysis

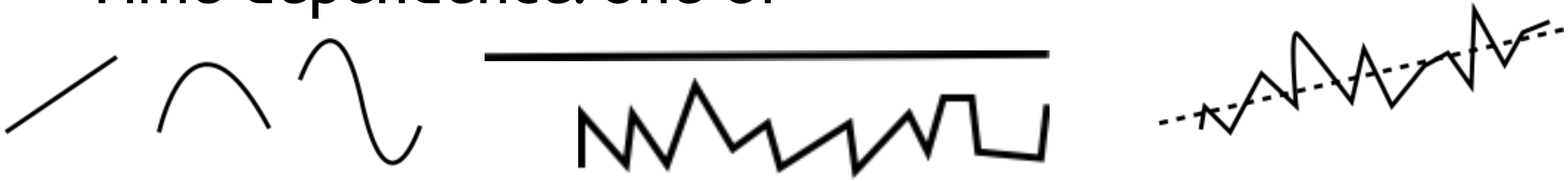
- Static factors affecting F1, F2:
  - Preceding consonant
  - Following C
    - Manner, place, voicing

(e.g. Stevens & House, 1963; Hillenbrand et al., 2001)

  - Others:
    - Can’t model due to <u>sparse data</u>

Analysis: models

- Similar to VOT
- For each vowel/formant/speaker, build linear mixed-effect models:
  - Response: normalized F1 or F2
  - Static factors
  - Time dependence: one of



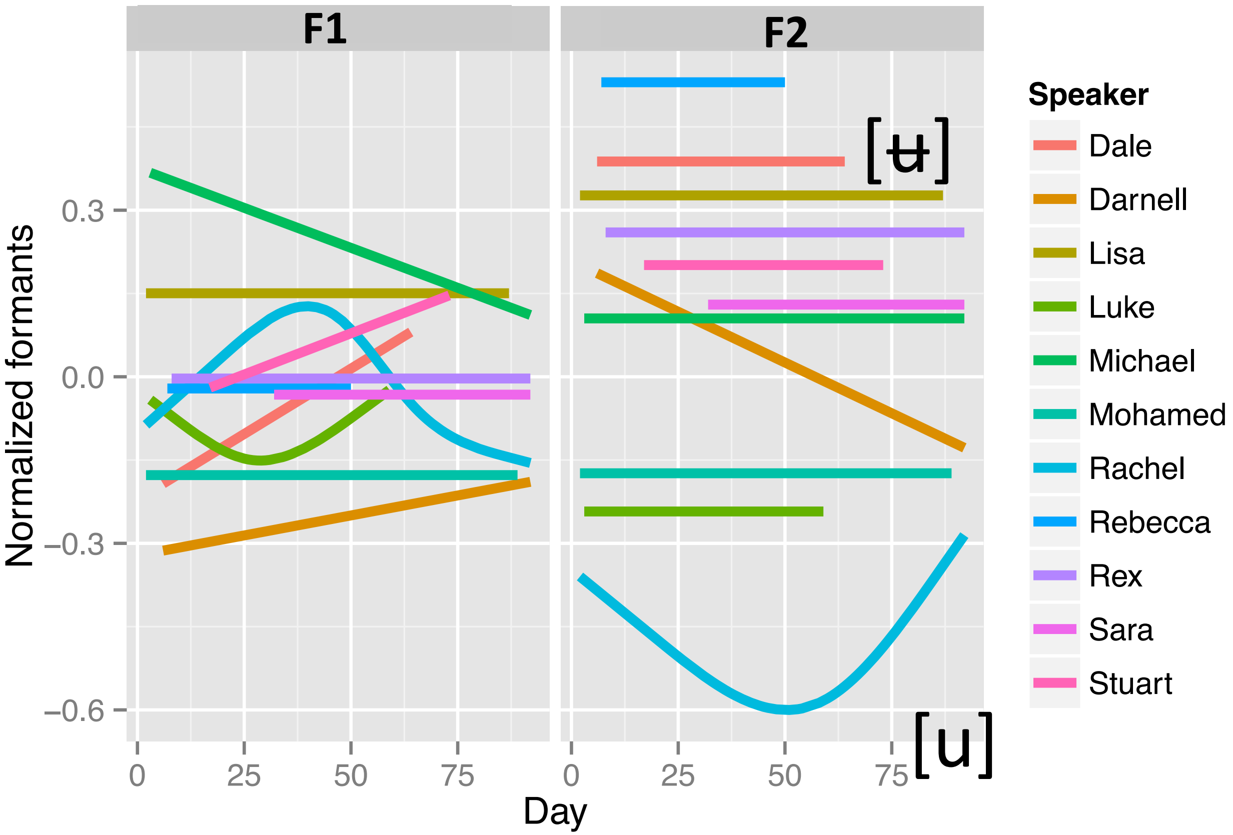
- Pick best model (AIC): one model of time dependence for
  - Speaker 1 GOOSE F1, GOOSE F2, …
  - (etc.)

Results: predicted time dependence

	GOOSE	TRAP	STRUT
Any time dependence	91%	91%	100%
By-day variability	91%	73%	91%
Time trend	55%	73%	64%

Results: time trends

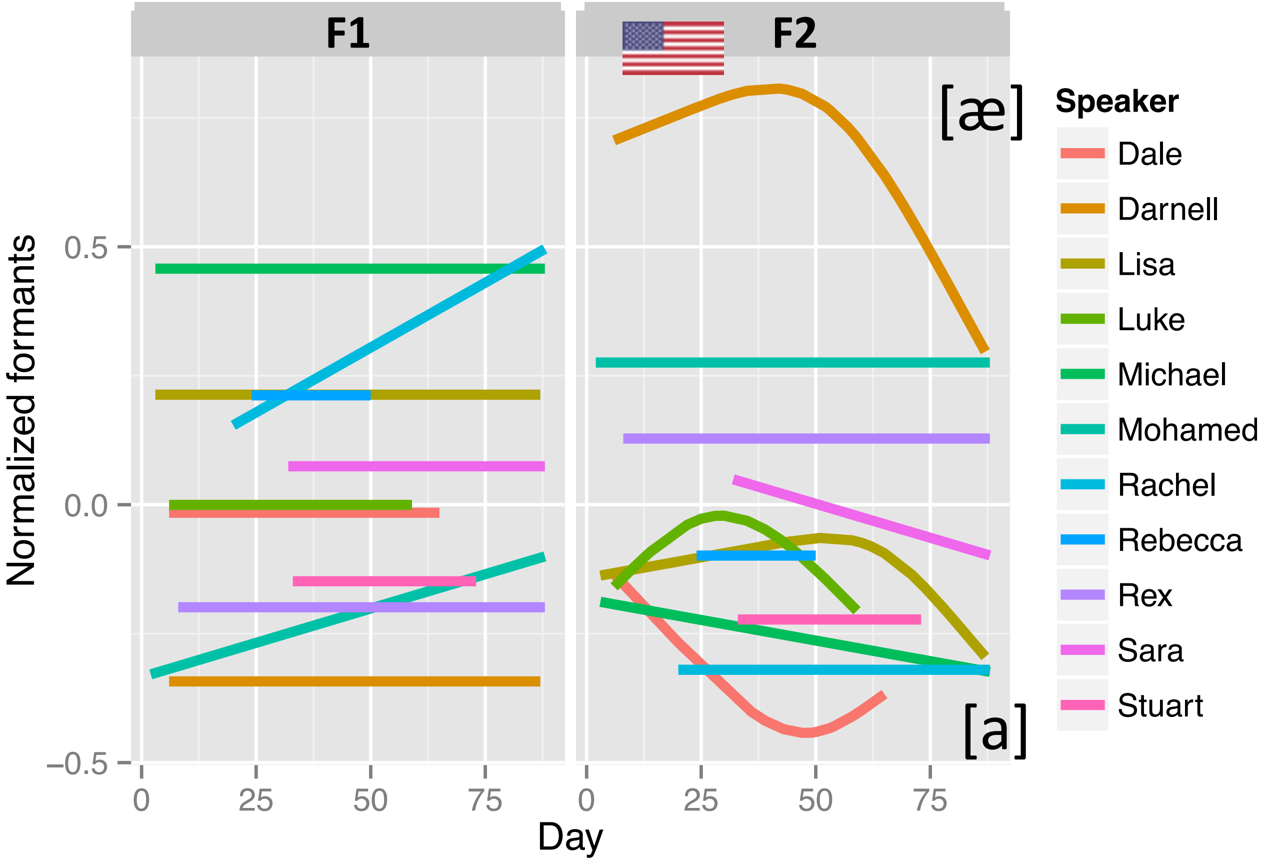
- GOOSE



- Convergence in F1?

Results: time trends

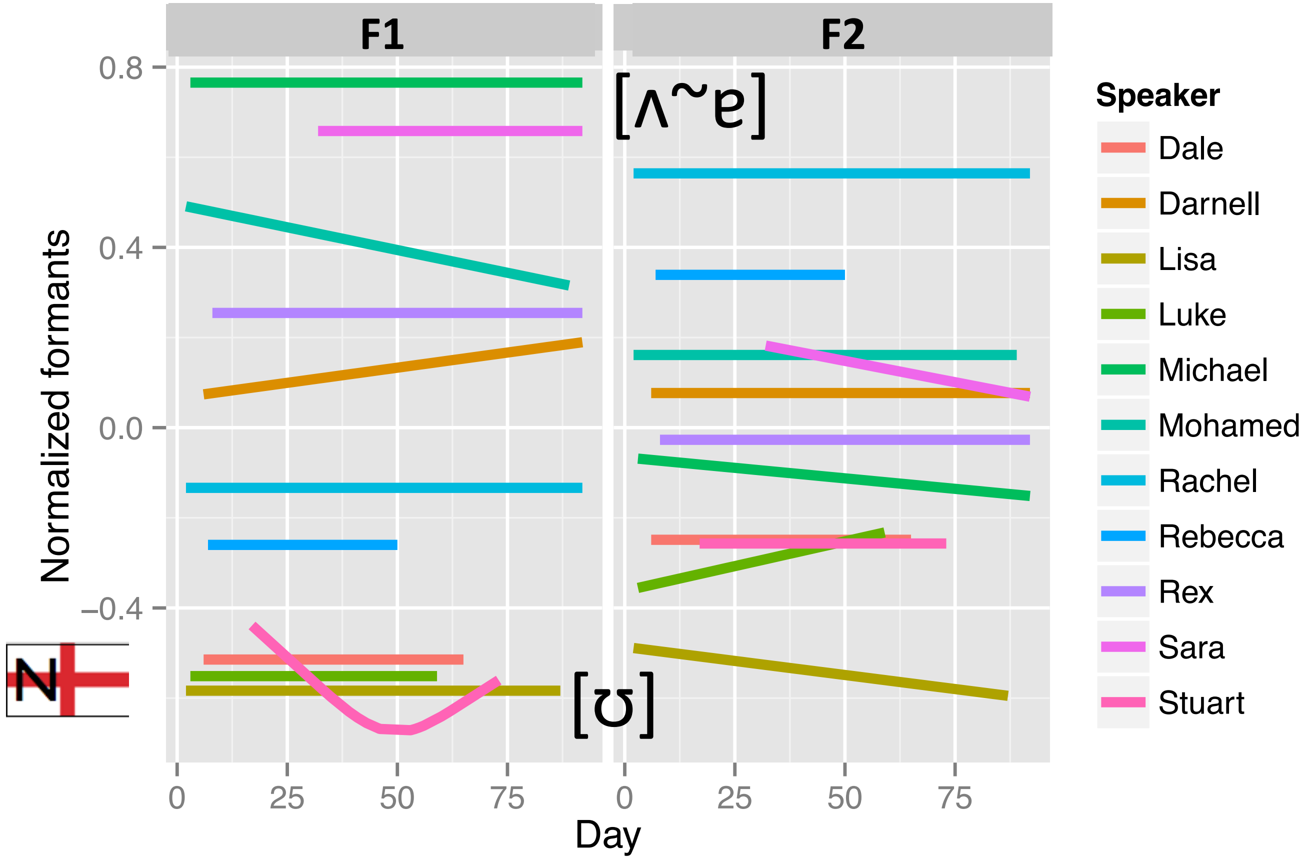
- TRAP



- No overall convergence

Results: time trends

- STRUT



- No overall convergence

Results: by-day variability

- Effect size:
  - F1: 0.13-0.94
  - F2: 0.11-0.72

±1σ normalized formant

- Compare: strongest static factors
  - F1: 0.26
  - F2: 1.04

By-day fluctuations similar magnitude to contextual effects

- Compare: short-term shifts
  - Babel (2011) vowel imitation: most subjects < 0.15

By-day fluctuations similar magnitude to accommodation effects

Discussion

- What *are* medium-term phonetic/phonological dynamics?
- Relationship to short-term, long-term dynamics?
  - Including community-level change
- Causes?
  - Convergence?

Medium-term dynamics

- Variability over time of sounds in individuals is the norm
  - 82-100% of speakers
  - Reject null hypothesis

- More variability detected for larger dataset
  - 2x larger than Sonderegger (2012): greater power
  - ⇒ we're likely underestimating

Medium-term dynamics

- By-day variability is very common
  - Vowels, VOT: 70-100%
  - CSD: 35%

Discrepancy makes sense if BDV due to accumulated accommodation effects

- Longer-term time trends less common
  - Vowels, VOT: < BDV
  - CSD: > BDV

- Hypothesis: by-day variability in phonetic parameters is the norm

Medium-term dynamics

- Overall: pronunciation of sounds fluctuates on timescale of days-months
  - VOT: also contrasts

- Important?
  - Effect size comparable to:
    - Coarticulation, speaking rate
  - But: not large enough to endanger contrasts
    - VOT
    - More generally: hypothesis for future work

Short, medium, long

- Medium-term change
  - Qualitatively different types of dynamics
  - High inter-speaker, variable variation
  - Robust: some time dependence

- Previous work:
  - Short-term: accommodation robust, widespread
  - Long-term: highly variable, majority don't change

- Medium-term is in between

- Mismatch between short and long-term dynamics

- Proposal:
  - Speakers robustly vary on timescale of days
    - (In part) due to accommodation effects persisting: c.f. similar effect sizes
  - But these fluctuations often don't accumulate into longer-term trends
    - Fits with relatively rarity of change over lifespan

Sources of dynamics

- Why these dynamics?
  - Huge intervariable, -speaker differences

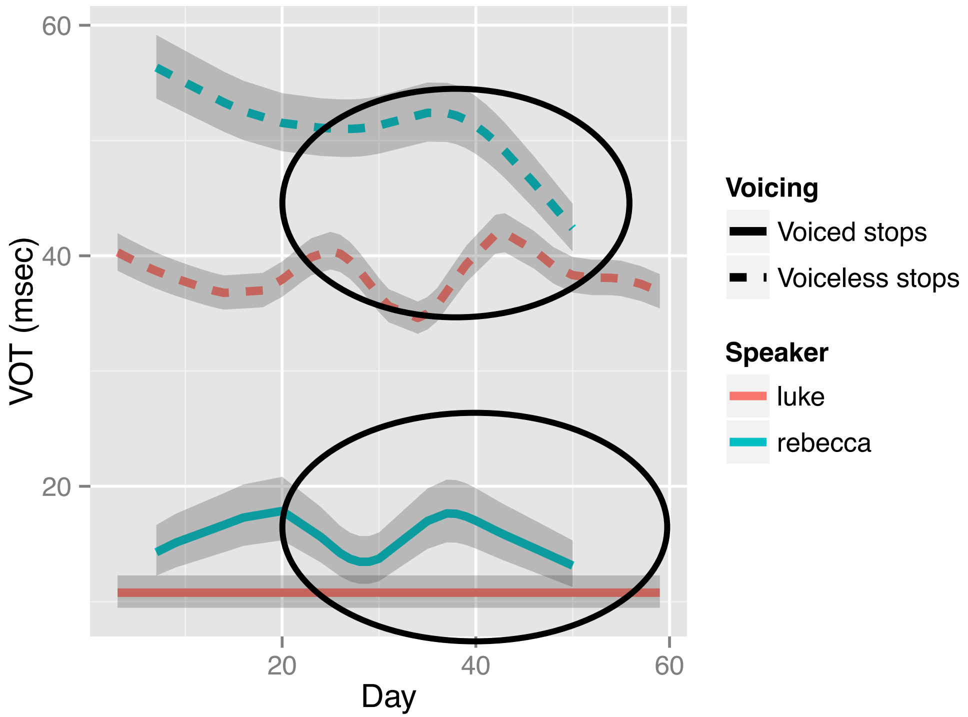
- Mostly still unknown

- Across variables: no clear overall convergence!
  - But…

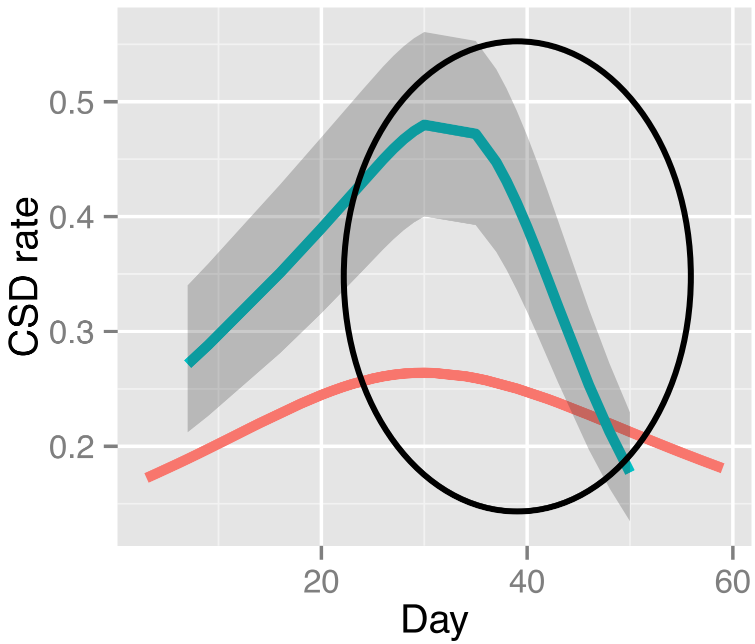
Luke and Rebecca

- Enemies → couple (≈ day 30)

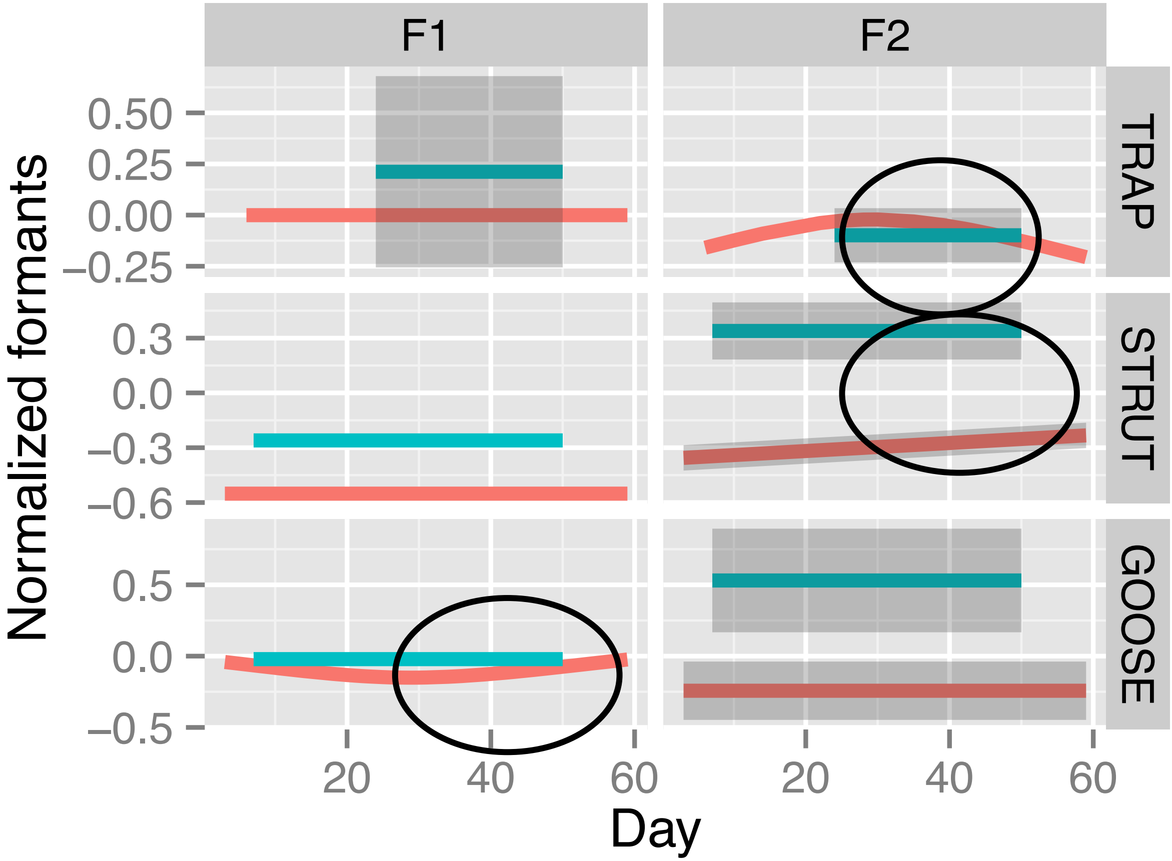
VOT



Coronal stop deletion



Vowels

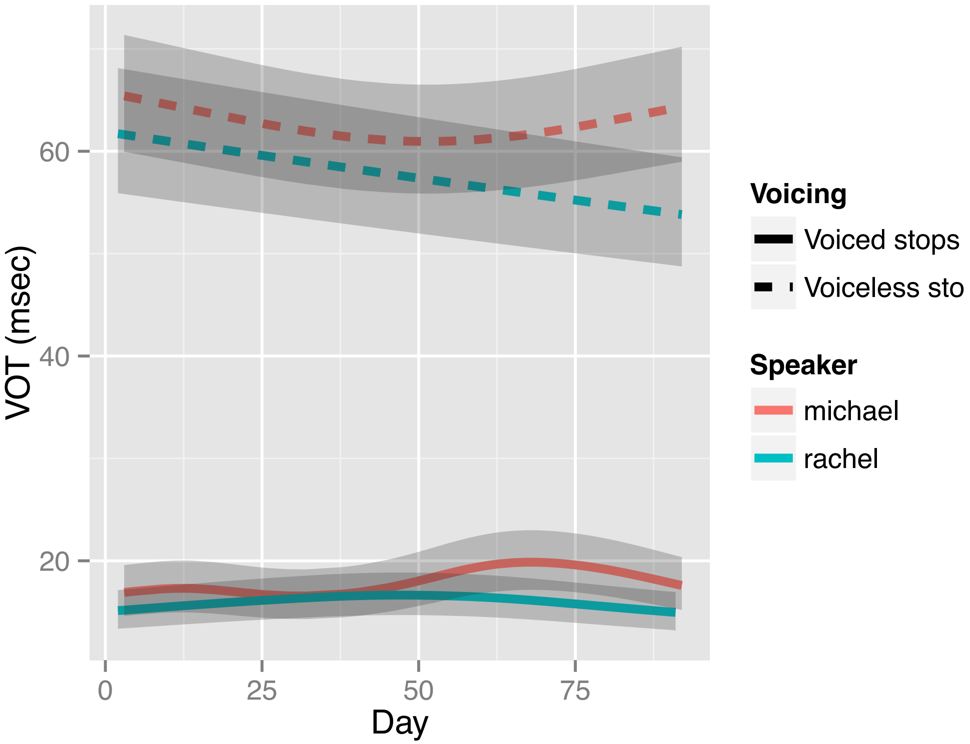


- Convergence, across variables
  - (?)

Michael and Rebecca

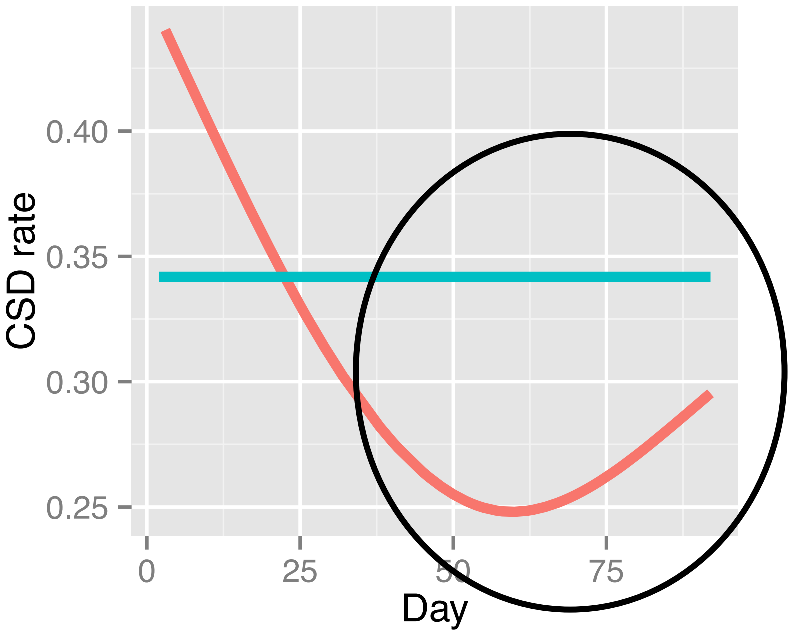
- Best friends in house, from early on

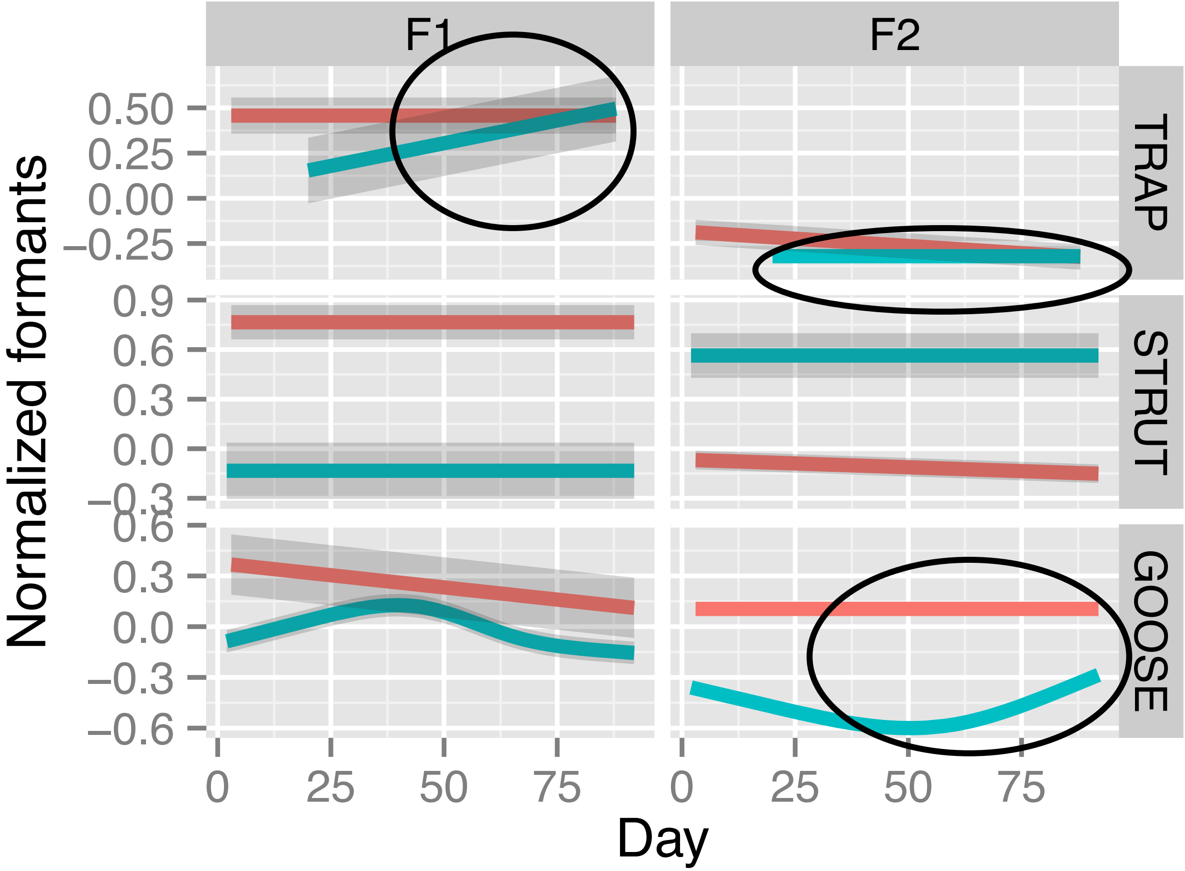
VOT



(very similar throughout show)

Coronal stop deletion



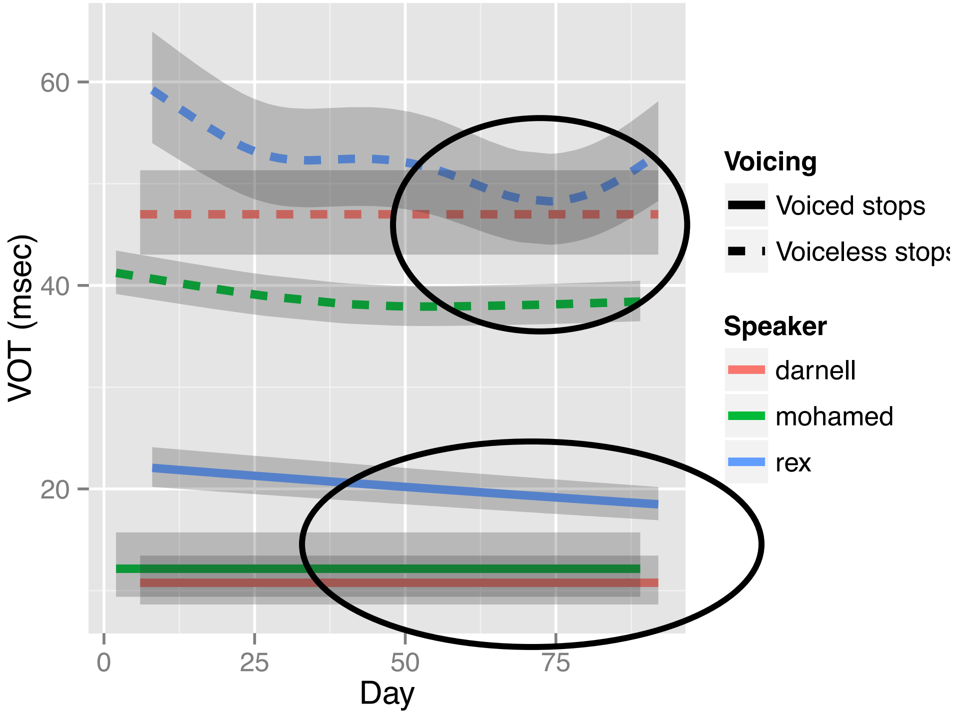


- Convergence across variables (?)

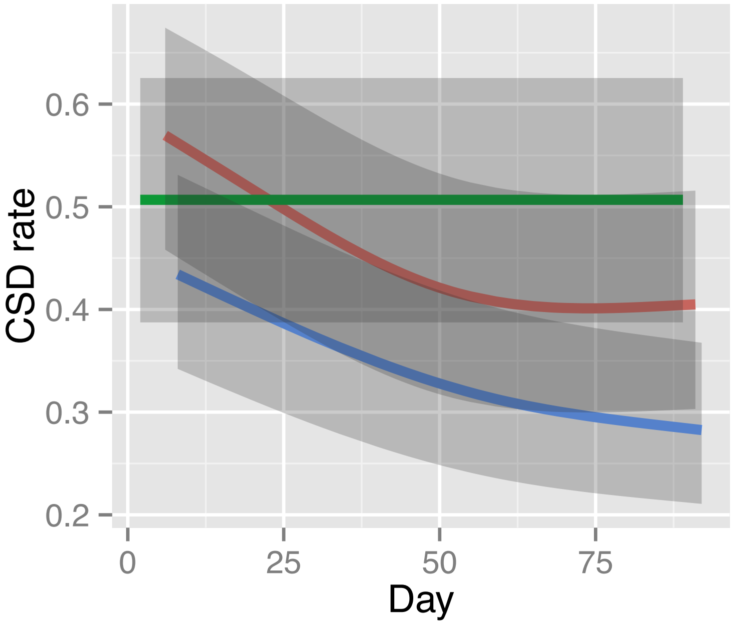
Darnell, Mohamed, Rex

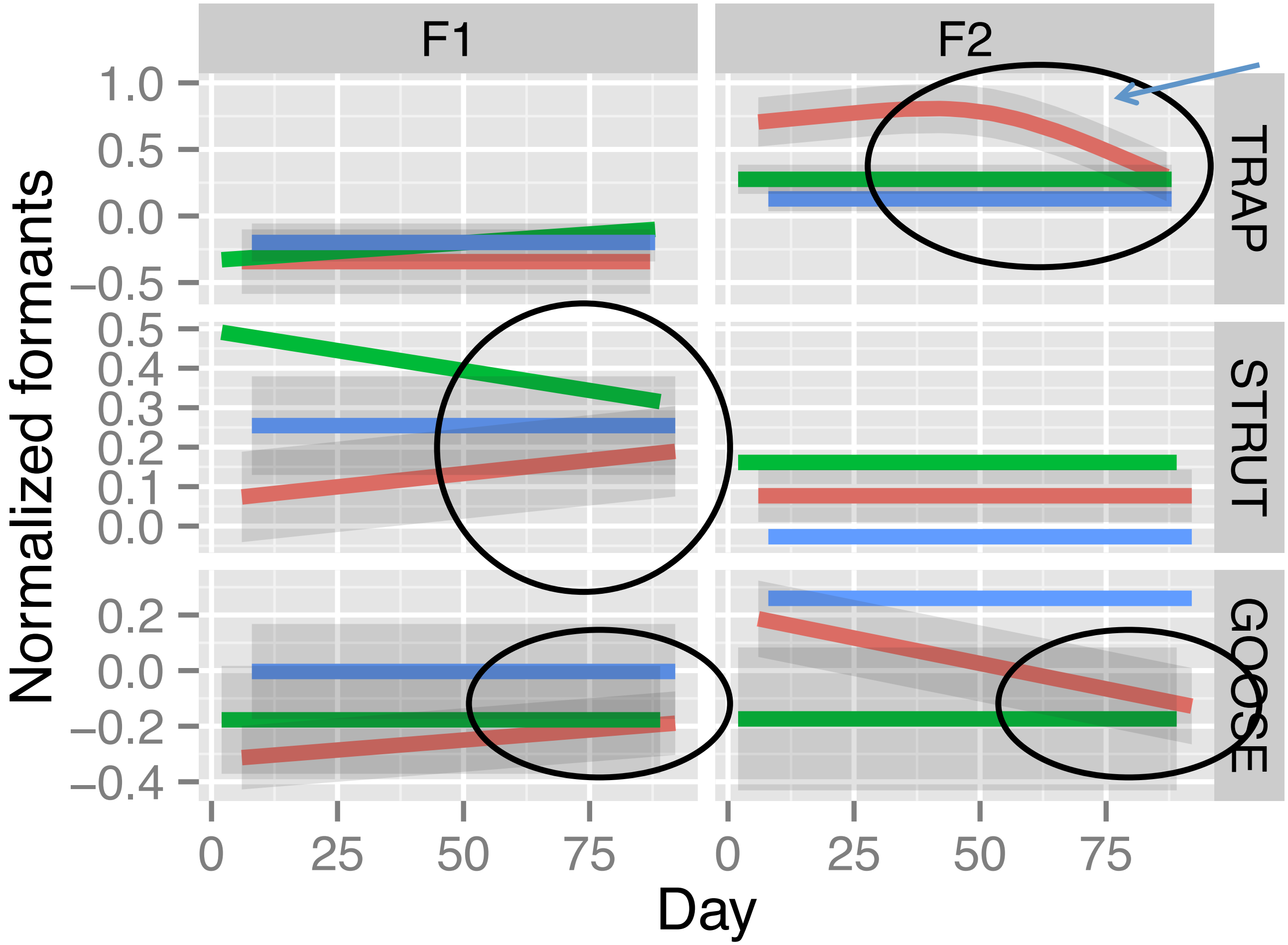
- Form an "outsiders" group from early on

VOT



Coronal stop deletion





Darnell's (American) TRAP → much closer to UK norm

- Convergence across variables except CSD
  - (?)

Sources of dynamics

- Big Q: what explains observed dynamics?

- Little-no evidence for convergence across speakers

- But: suggestive evidence for convergence within socially-meaningful subsets of speakers!
  - Especially during last part of show
    ⇒ fewer people, more concentrated interactions

- Consistent with a role for accommodation effects in language change (Neogrammarians on)
  - But, socially-mediated (Babel, 2011)

- For now, post-hoc/qualitative!
  - Ongoing work: hypotheses based on social interaction data (20k obs)

- Other future work:
  - Is "grammar" changing, or just phonetic parameters?

- Other future work:
  - High variability ⇒ much more study needed of dynamics in individuals
  - Many variables
  - Trajectories!

Thanks

- Max Bane, Peter Graff, Tyler Schnoebelen
- Montreal Language Modeling Lab RAs :
  - Thea Knowles, Liam Bassford, Hannah Cohen, Maggie Labelle, Misha Schwartz
- Permission: Channel 4/Endemol
- Funding:

Fonds de recherche
Société et culture
Québec

INNOVATION.CA
CANADA FOUNDATION FOR INNOVATION | FONDATION CANADIENNE POUR L'INNOVATION

SSHRC CRSH

McGill
UNIVERSITY

Questions