SPADE

SPeech Across Dialects of English

Multidimensional acoustic variation in vowels across English dialects

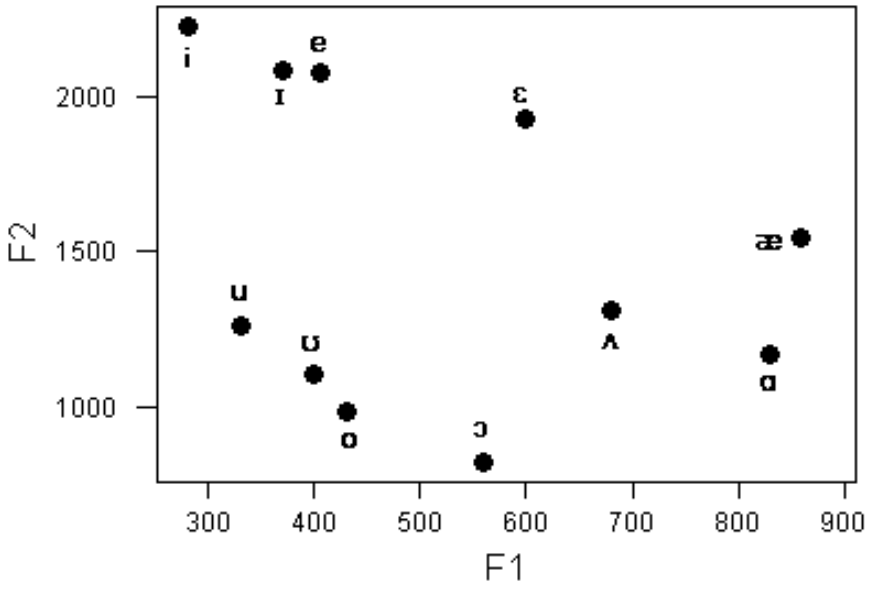
James Tanner[a], Morgan Sonderegger[a], Jane Stuart-Smith[b], The SPADE Consortium

[a]McGill University

[b]University of Glasgow

Introduction

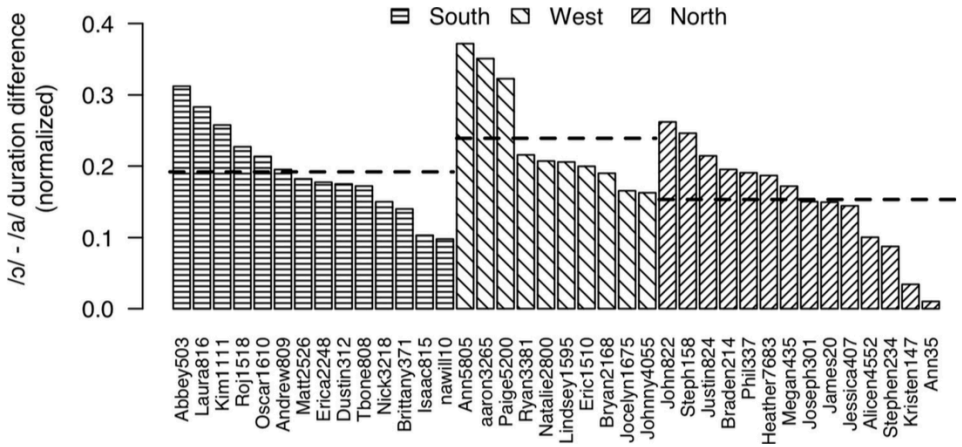
- Vowels vary across dialects in a number of dimensions
  - F1 x F2



Peterson & Barney (1952), Fant (1960), House (1961), Watson & Harrington (1994), Jacewicz et al. (2007), Morrison (2013)

Introduction

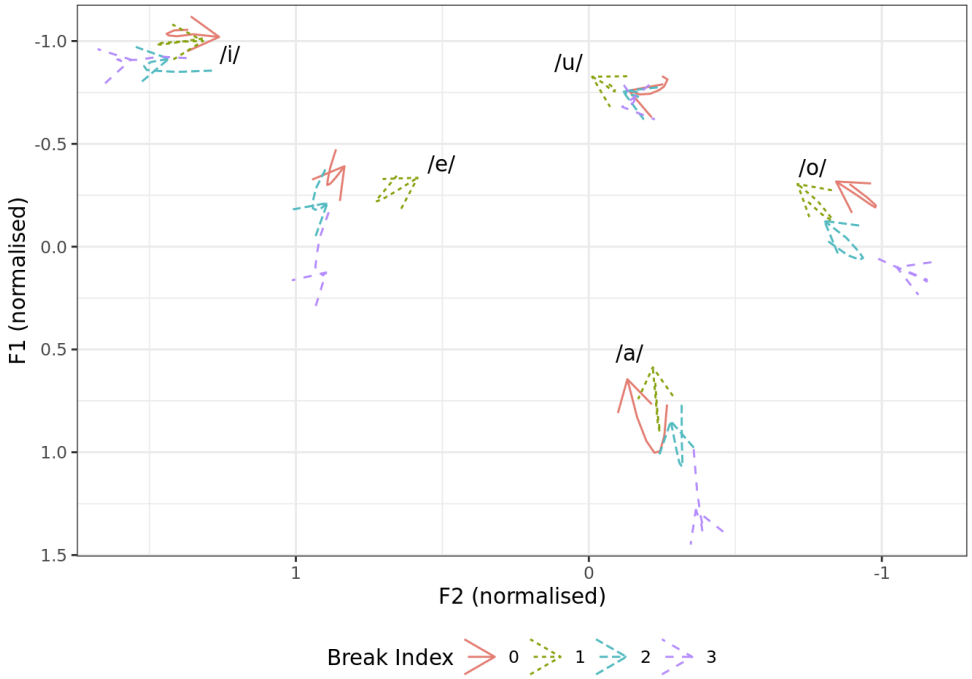
- Vowels vary across dialects in a number of dimensions
  - F1 x F2
  - Duration



Peterson & Barney (1952), Fant (1960), House (1961), Watson & Harrington (1994), Jacewicz et al. (2007), Morrison (2013)

Introduction

- Vowels vary across dialects in a number of dimensions
  - F1 x F2
  - Duration
  - Dynamic change



Peterson & Barney (1952), Fant (1960), House (1961), Watson & Harrington (1994), Jacewicz et al. (2007), Morrison (2013)

Introduction

- *How do these dimensions capture systematic vowel variation across dialects?*
  - *In what ways can a single sound vary within a language?*
- *Do patterns of dynamic variation correspond to traditional categories e.g. monophthong vs. diphthong?*

Introduction

- **This study:** explore variation in vowel dimensions across large cross-dialectal English dataset
  - Exploratory analysis
  - Dialect classification experiment
- **Findings:** Vowels show *structured variation* across dialects
  - Corresponding to monophthong/diphthong distinction
- First *large-scale* analysis of formant dynamics across a single language

SPADE
SPeech Across Dialects of English

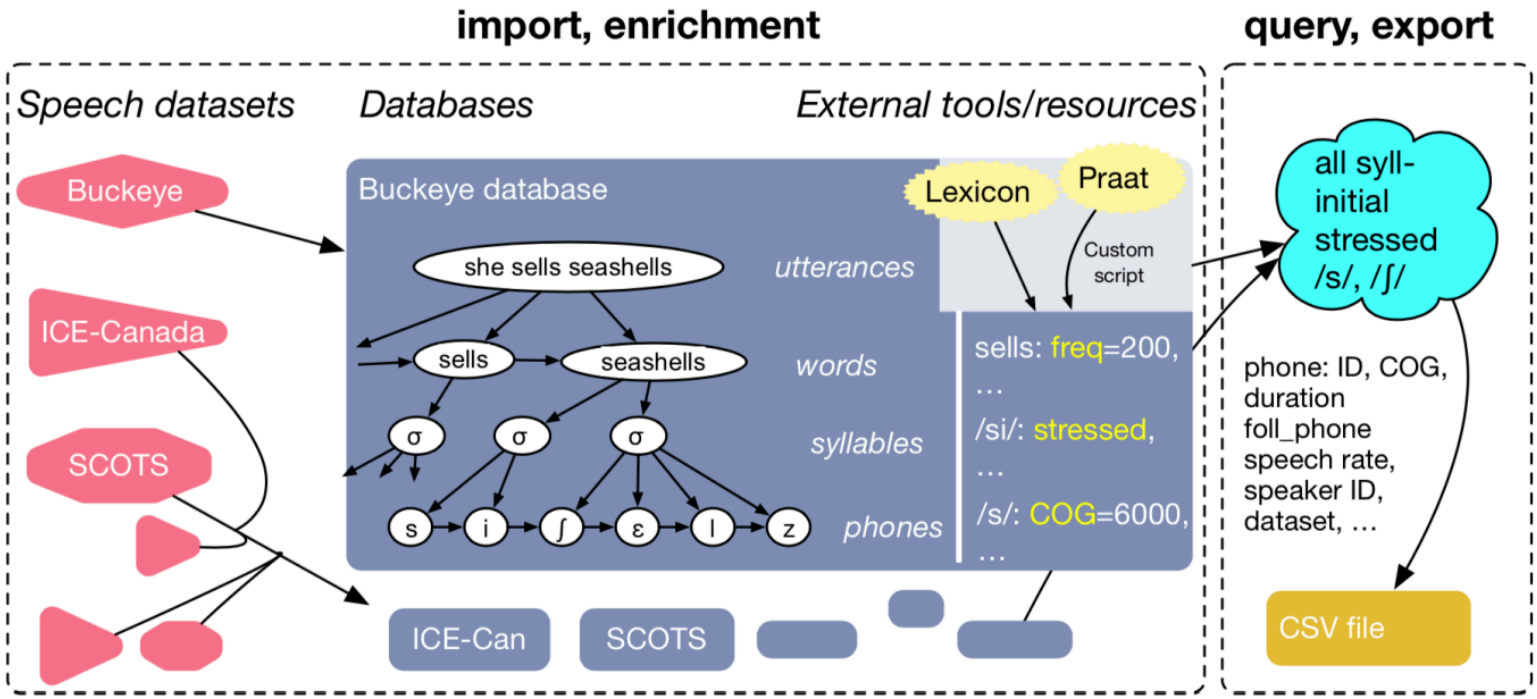
https://spade.glasgow.ac.uk/the-spade-consortium/

Data

- **44** public/private datasets from 4 countries
- **This study: 21** English dialects from 11 speech corpora

https://github.com/MontrealCorpusTools/PolyglotDB

Data: Processing



- Forced alignment
- Store in graph database

McAuliffe et al. (2017, 2019), Rosenfelder et al. (2014), Schiel (1999)

Data

- Select vowels across monophthong/diphthong continuum

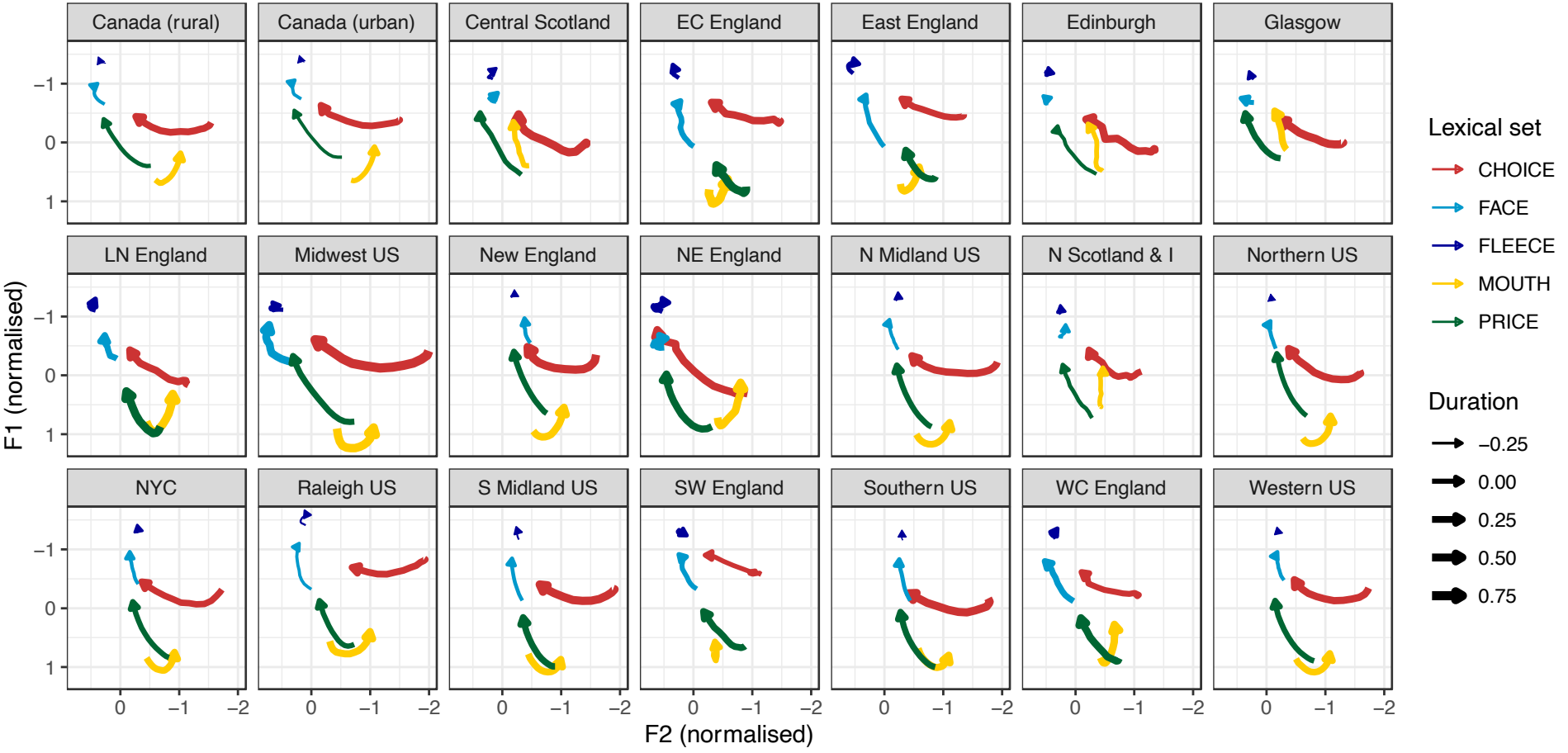
FLEECE FACE MOUTH PRICE CHOICE

More monophthongal More diphthongal

- ~1200 speakers, ~320k tokens
- Formants measured at 5% timepoints; middle 60% used

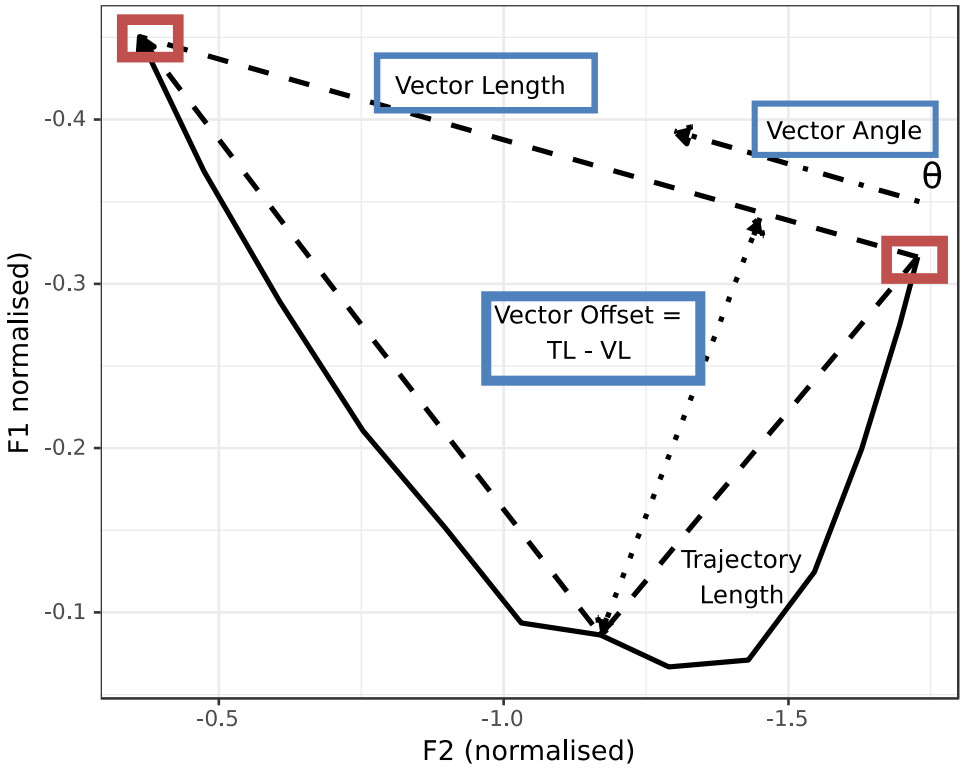
Fruewald (2013), McAuliffe et al. (2019), Mielke et al. (2019), Williams et al. (2019)

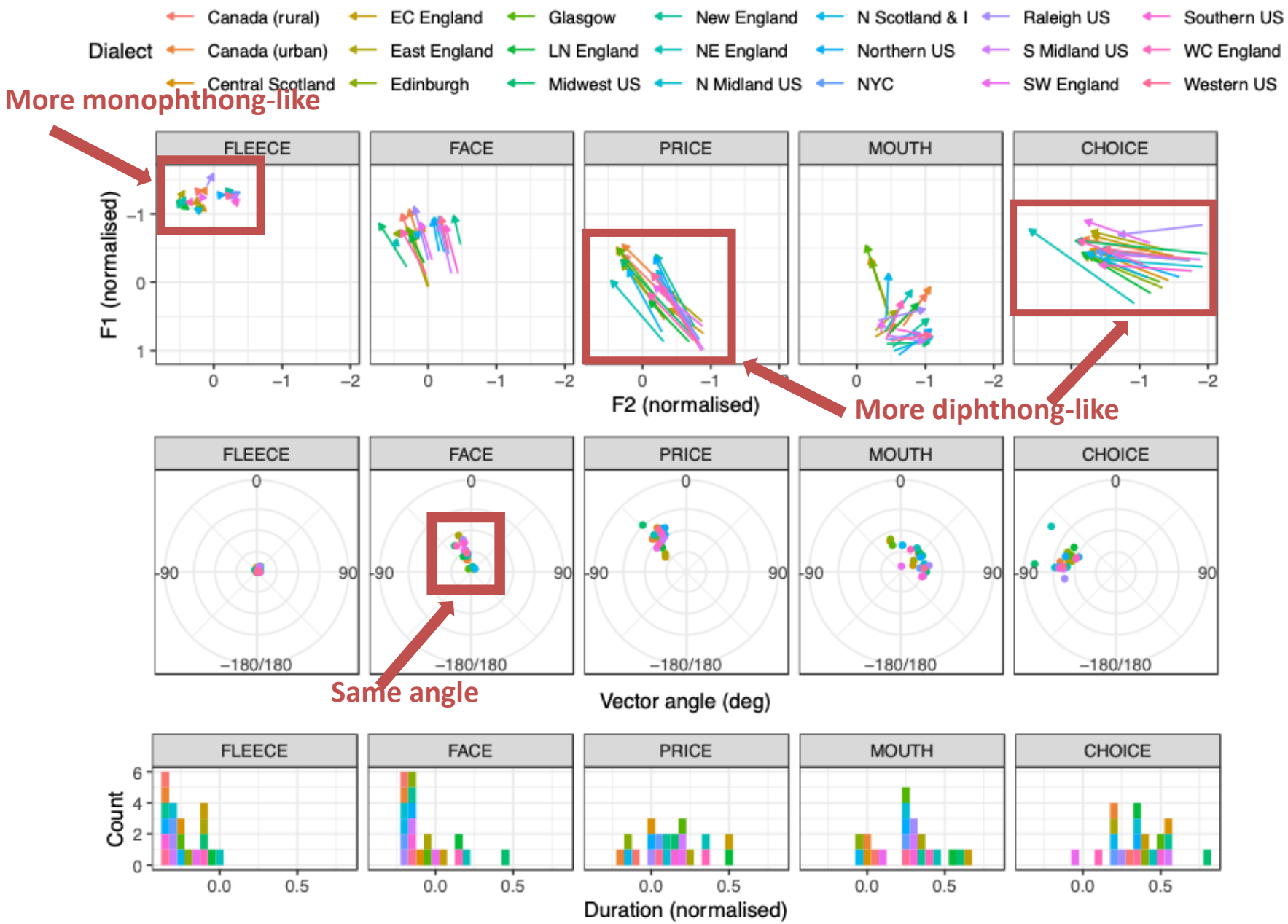
Data


Canada (rural)
Canada (urban)
Central Scotland
EC England
East England
Edinburgh
Glasgow
LN England
Midwest US
New England
NE England
N Midland US
N Scotland & I
Northern US
NYC
Raleigh US
S Midland US
SW England
Southern US
WC England
Western US
F1 (normalised)
F2 (normalised)
−1
0
1
0
−1
−2
Lexical set
CHOICE
FACE
FLEECE
MOUTH
PRICE
Duration
−0.25
0.00
0.25
0.50
0.75

Vowel measurements

- *F1 & F2 start/end*
- *Vector Length*
  - Diff in F1/F2 space between start and end point
- *Vector Offset*
  - Amount of *non-linear change*
- *Vector Angle*
  - Direction of change
- *Duration*





Variation ***constrained*** across dialects along monophthong-diphthong continuum

Dialect Classification

Measures	FLEECE	FACE	PRICE	MOUTH	CHOICE
Baseline (most common dialect label)	50	50	50	50	50
Formants (F1, F2 onset + offset)	58	61.3	62.2	61.4	56.7
Trajectory (Vector Length, Offset, Angle)	54.5	62.1	56	63.6	56
Duration	55	52.9	57.4	52.7	51.9
{Formants, duration}	62.5	**65.3**	66.2	66.4	**60.3**
{Trajectory, duration}	56.7	65.1	60.6	65.4	55.9
{Formants, trajectory}	60.8	62.7	65	67.4	57.6
{Formants, trajectory, duration}	**63.4**	**64.2**	**69.4**	**70**	**59.2**

Table 2: Balanced accuracy (%) for each SVM, trained with different configurations of formant position, trajectory shape, and duration measures.

- Trajectory measures more informative where monophthong/ diphthong varies across dialect (FACE, MOUTH)

Dialect Classification

Measures	FLEECE	FACE	PRICE	MOUTH	CHOICE
Baseline (most common dialect label)	50	50	50	50	50
Formants (F1, F2 onset + offset)	58	61.3	62.2	61.4	56.7
Trajectory (Vector Length, Offset, Angle)	54.5	62.1	56	63.6	56
Duration	55	52.9	57.4	52.7	51.9
{Formants, duration}	62.5	**65.3**	66.2	66.4	**60.3**
{Trajectory, duration}	56.7	65.1	60.6	65.4	55.9
{Formants, trajectory}	60.8	62.7	65	67.4	57.6
{Formants, trajectory, duration}	**63.4**	**64.2**	**69.4**	**70**	**59.2**

Table 2: Balanced accuracy (%) for each SVM, trained with different configurations of formant position, trajectory shape, and duration measures.

- Multiple measures > single measures
- Adding trajectory to static measures (generally) improves most for more-diphthongal vowels

Summary

- While vowels vary in all measures, that variability is *constrained* along monopthong-diphthong continuum
- Trajectory provides ‘additional resolution’ over static formants, esp for diphthongs

Discussion

- Looking across dialects provides a window into how the *same* speech sound can vary
- An example of *large-scale analysis*:
  - Access to a large number of high-quality speech corpora
  - Tools for automatic measurement (forced alignment, formant tracking, etc)

Fromont & Hay (2012), McAuliffe et al. (2017, 2019),

Discussion: large-scale analysis

- Take a ‘high-level’ approach to study of phonetic variation
- ‘Check’ our current knowledge built from previous smaller-scale research
- Limitations
  - Automation: simply too much data to check by hand; instead apply reasonable filtering criteria
  - *How comparable are corpus data from different sources?*

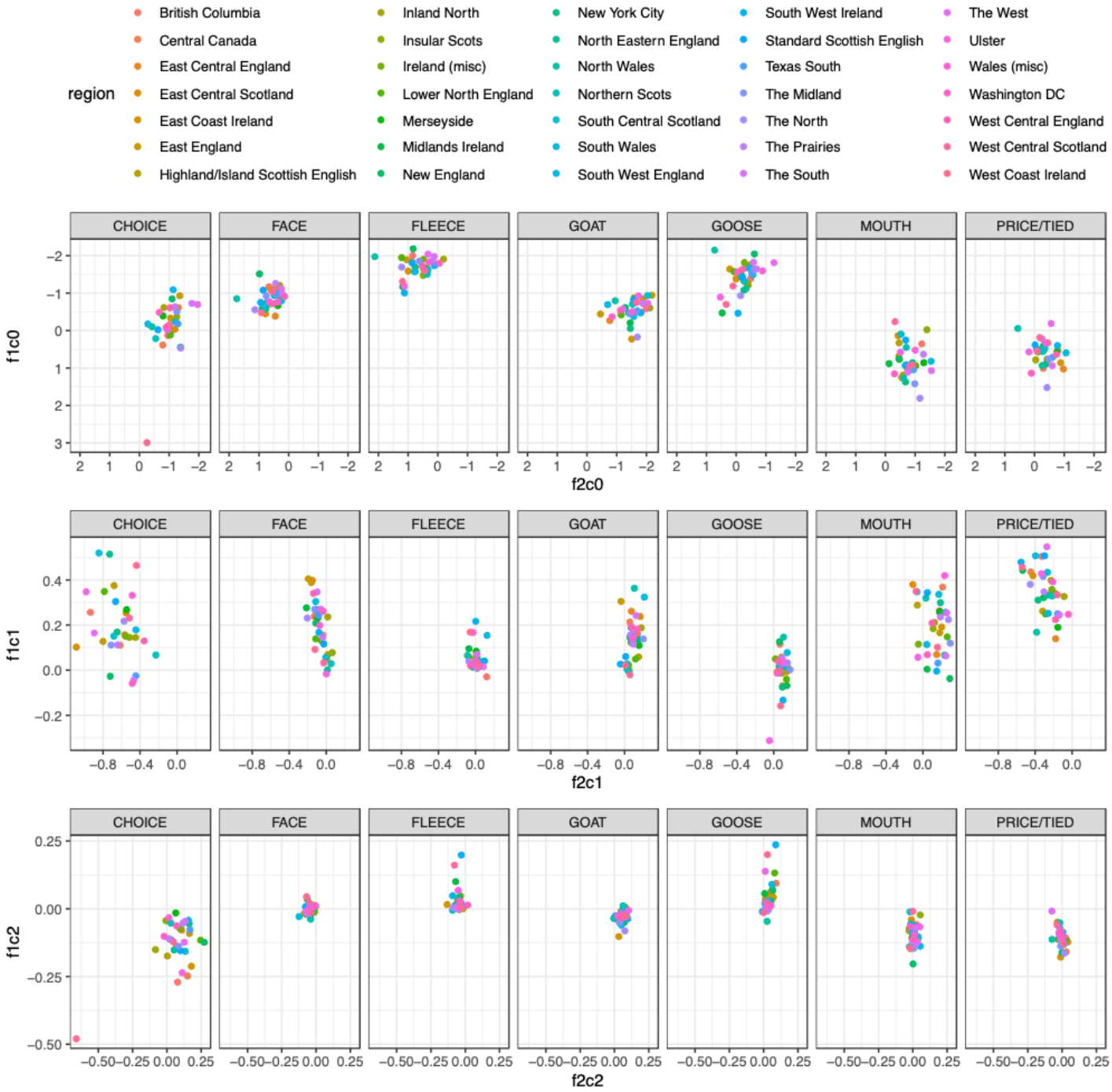
Liberman (2018), Salesky (2020)

Thank you!

- SPADE Data Guardians
- Jeff Mielke
- Rachel Macdonald
- Vanna Willerton
- Michael McAuliffe
- SPADE team

Continent	Dialect	Corpus	Speakers	Tokens
North America	Canada (rural)	Canadian-Prairies	44	20042
	Canada (rural)	ICE-Canada	8	2764
	Canada (urban)	Canadian-Prairies	67	38021
	Canada (urban)	ICE-Canada	8	877
	Midwest US	Buckeye	40	17669
	New England	Switchboard	18	2868
	North Midland US	Switchboard	44	7126
	Northern US	Switchboard	53	7494
	NYC	Switchboard	19	3183
	Raleigh US	Raleigh	100	64659
	South Midland US	Switchboard	106	20327
	Southern US	Switchboard	37	5595
	Western US	Switchboard	45	6376
United Kingdom	Central Scotland	SCOTS	23	5237
	East Central England	Audio BNC	30	3877
	East England	Audio BNC	100	13429
	East England	Hastings	49	25477
	East England	IViE	12	972
	East England	IViE	11	992
	East England	ModernRP	48	2811
	Edinburgh	SCOTS	18	2361
	Glasgow	SCOTS	26	4432
	Glasgow	SOTC	155	45487
	Lower North England	Audio BNC	41	5445
	Lower North England	IViE	11	891
	Lower North England	IViE	10	760
	North East England	Audio BNC	10	917
	North East England	IViE	12	1018
	Northern Scotland & Islands	SCOTS	31	3998
	South West England	Audio BNC	37	3458
	West Central England	Audio BNC	32	4497
Total	**21**	**11**	**1245**	**323060**

Table 1: Speaker and token count for each dialect used in this study, separated by the corpus from which the data was originally sourced.


region
British Columbia
Central Canada
East Central England
East Central Scotland
East Coast Ireland
East England
Highland/Island Scottish English
Inland North
Insular Scots
Ireland (misc)
Lower North England
Merseyside
Midlands Ireland
New England
New York City
North Eastern England
North Wales
Northern Scots
South Central Scotland
South Wales
South West England
South West Ireland
Standard Scottish English
Texas South
The Midland
The North
The Prairies
The South
The West
Ulster
Wales (misc)
Washington DC
West Central England
West Central Scotland
West Coast Ireland
CHOICE
FACE
FLEECE
GOAT
GOOSE
MOUTH
PRICE/TIED
f1c0
f1c1
f1c2
f2c0
f2c1
f2c2

Trajectory shape

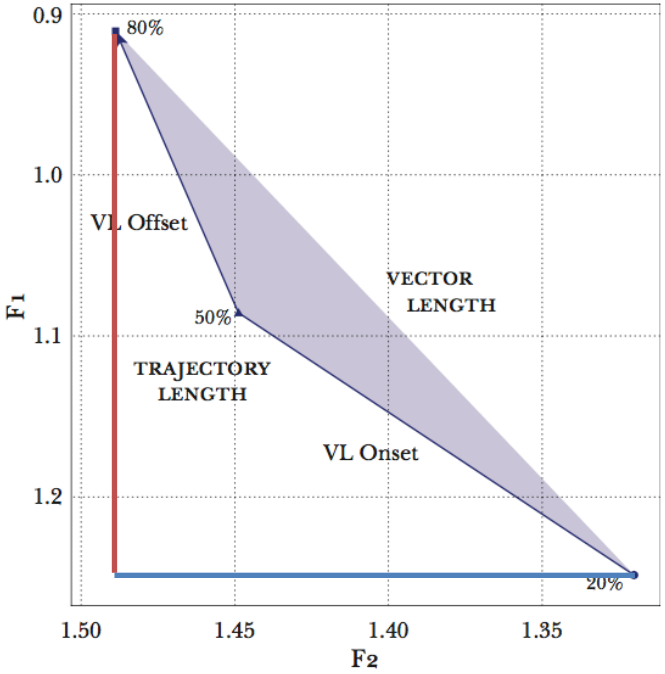
$$VSL_{n,m} = \sqrt{(F1_n - F1_m)^2 + (F2_n - F2_m)^2}$$

Vector length = VSL(start, end)

Trajectory length = VSL(start, mid) + VSL(mid,end)

Vector offset = Vector Length - Trajectory Length

Vector Angle = arctan(F1diff/F2diff)



Balanced accuracy

$$BA = \frac{Sens + Spec}{2}$$

- Sens = correctly-identified positive
- Spec = correctly-identified negative
- Accounts for imbalance between positive & negative cases