Affect Structures Variation in Vowel Quality

The Influence of Smiling on the Front Lax Vowels in California

Robert J. Podesva
Stanford University

Introduction

Variation as Stancetaking

Attention to "interactionally specific" (Bucholtz and Hall 2015: 592) dimensions of identity construction like stance (Kiesling 2009, Freeman 2014) is recent.

Vowels and Affective Stance

- fronter vowels, positive affect (Johnson 2006, Podesva et al. 2015)
- backer vowels, negative affect (Eckert 2010, Eckert 2011)

The Challenge of Affect

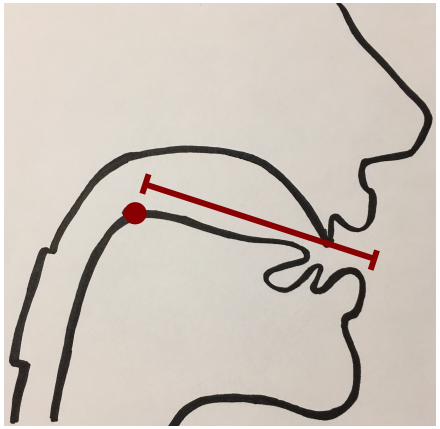
"In probably all speech communities, emotions can be described (e.g. *I hate him*), although such overt avowals in the first person are likely to be associated with rather marked situations. More commonly, emotions are alluded to, and the decoding task is a process of 'reading off' complex covert messages." (Besnier 1990: 428)

Formant Frequency and Smiling

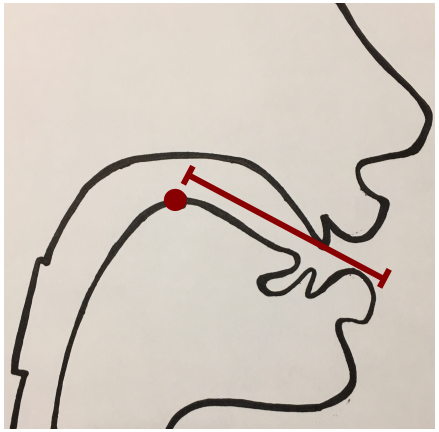
Affective stancetaking is accomplished multimodally (C. Goodwin 2000, 2007; M. Goodwin, Cekaite & C. Goodwin 2012; M. Goodwin 2016).

Embodied Affect: Smiling

Lip retraction



Tongue fronting

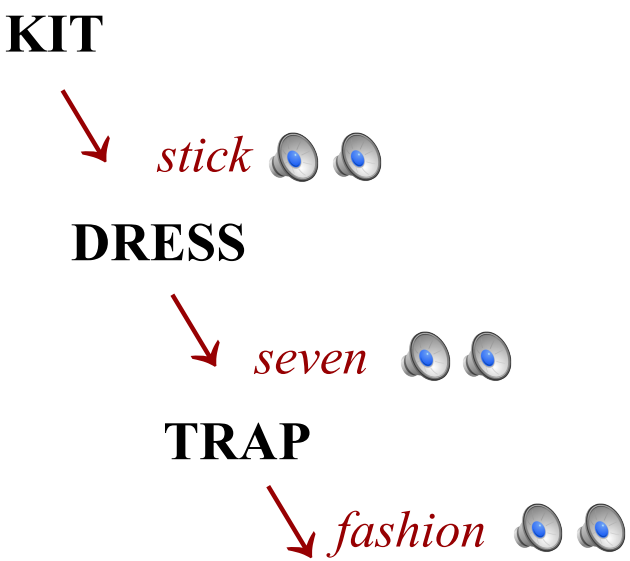


shorter vocal tract → higher F2

Is the connection between affect and formant frequency reducible to smiling, which frequently accompanies emotions with positive valence?

Smiling and F1

If there is a fundamental connection between affect and vowel quality, we should still observe a correlation between smiling and F1 (which is not directly influenced by the physical act of smiling).



Previous Studies on California

Hinton et al. 1987
Eckert 2008
Kennedy and Grama 2012
Holland 2014
Hall-Lew et al. 2015
King 2015
King and Calder 2016
Van Hofwegen, Pratt & D'Onofrio 2016

Body Movement and F1

If there is a fundamental connection between affect and vowel quality, we should observe a correlation between other forms of embodied affect (e.g., body movement) and F1.

Affect and Body Movement

Direct correlation between how much people move and emotional arousal (Pollick et al. 2001, Pollick et al. 2002, Camurri et al. 2003, Atkinson et al. 2007, Crane and Gross 2007)

Body Movement and Prosody

Speakers move more in phrases with higher and more variable pitch and intensity (Voigt, Podesva & Jurafsky 2013)

Current Study

Acoustic and visual analysis of the front lax vowels in California

Visual analysis

- Whether speakers are smiling
- How much speakers are moving

Findings

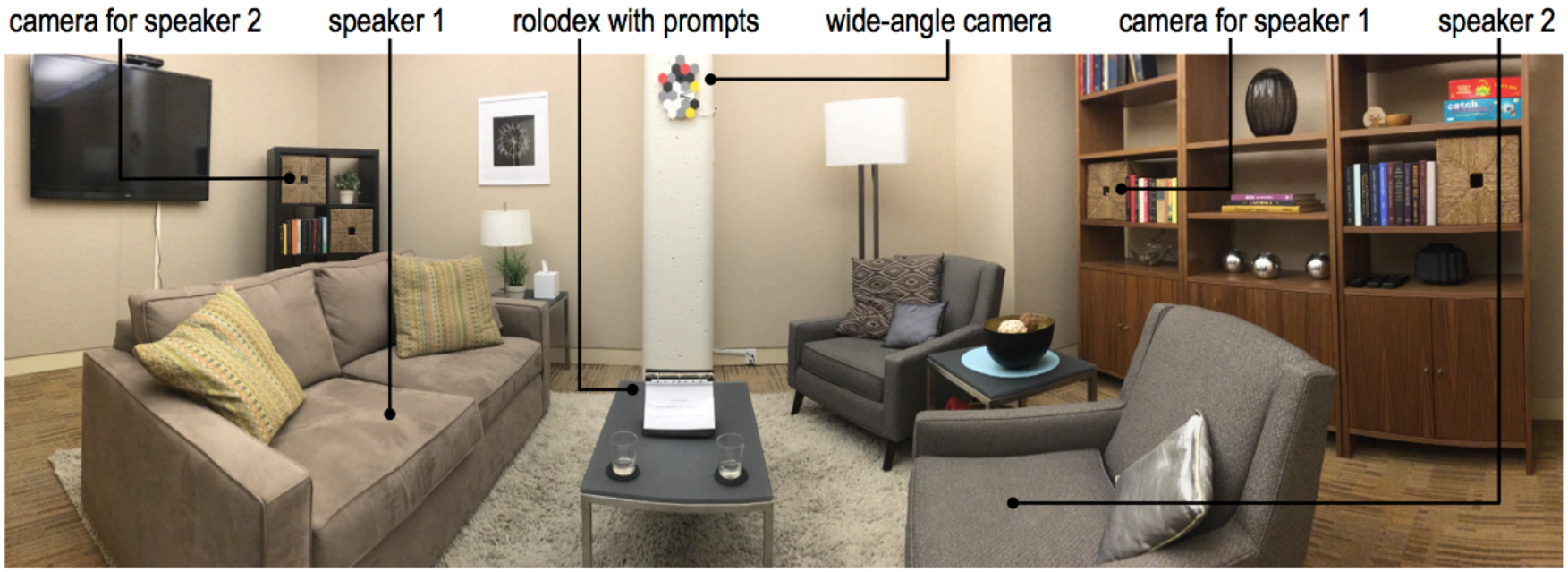
Speakers produce lower (higher F1) front lax vowels when…

- …they are smiling.
- …they are moving more.

Claim

The connection between affect and vowel quality is fundamental, not a mere consequence of smiling.

Interactional Sociophonetics Laboratory



Acoustical specifications of sound booth, staged as living room

METHODS

Data Collection

Separate audio and video recordings for each speaker

- Dyadic interactions, video and audio (wireless microphones) recorded
- Part 1: “would you rather…” questions, recording levels checked
- Part 2: ~30 min of conversation, with aid of prompts
- Part 3: survey (demographic information, assessments)

Sample: 42 speakers from California

Sex:	26 female
	16 male
Age:	25 undergraduates (18-22 years old)
	17 older adults (23 years old and up)

Race:	21 white
	6 African American/white
	5 Asian/Pacific Islander
	3 Asian/white
	2 Native American
	2 other multiracial
	1 each of African American, Latinx, Middle Eastern
Sexual Orientation:	32 straight
	7 LGBTQ
	3 unspecified

Acoustic Analysis

- Approximately 21 hours of speech
- Transcriptions in ELAN (Lausberg and Sloetjes 2009)
- Forced alignments using FAVE (Rosenfelder et al. 2011)
- For every vowel interval, a number of acoustic measures were taken every 10 ms via Praat (Boersma and Weenink 2015) script
  - F1-F3
  - Spectral tilt
  - F0 and periodicity measurements
- Acoustic measures reduced to median value/vowel
- Each segment classified as ±creaky using Kane et al. (2013) method
- All stressed vowels > 75 ms normalized using Lobanov (1971)
  - Excluding preceding vowels, glides, /r/
  - Excluding following vowels, glides, liquids
- N = 23,311

METHODS

Smiling Annotation

Haar cascade classifier trained on open source corpus of photographs hand-annotated for ±smiling (http://github.com/hromi/SMILEsmileD).

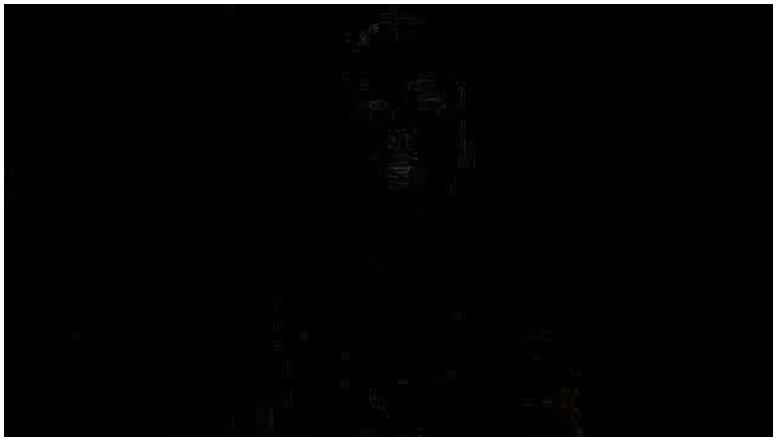
Each frame of video run through classifier.

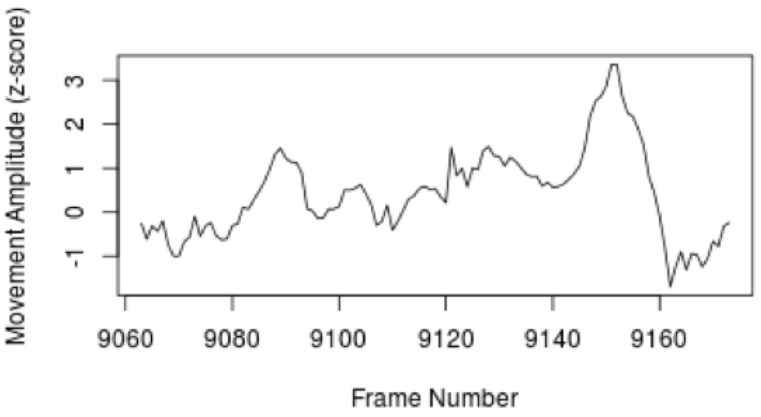
smiling = TRUE FALSE

Quantifying Movement

Movement Amplitude

Voigt, Podesva & Jurafsky (2014)





METHODS

Linear Mixed-Effects Regression Analysis

Observations:	stressed KIT, DRESS, TRAP exclusions: __ {*nasals, velars*} N = 5,255
Responses:	F1, F2
Random:	speaker, word, pre and fol segment
Linguistic predictors:	duration (log), phrase position, ±creak
Social predictors:	sex, age
Embodied predictors:	±smiling, movement amplitude (each at the segmental and phrase levels)

Smiling During Speech

We have *so* many *freshman* with like eager *attitudes* and *perceptions* about Stanford, and you're like, "That's just *not* how it works." Like, "You're not- I can't tell you *that*. It's something you gotta experience, and that sucks."

fronted GOAT	*so*
lowered DRESS	*freshmen*
lowered TRAP	*attitudes*
lowered DRESS	*perceptions*
backed LOT	*not*
lowered TRAP	*that*
fronted STRUT	*sucks*

RESULTS

F2 Model

Term	Estimate	Std Error	DFDen	t Ratio	Prob>|t|
Intercept	0.182	0.076	105.0	2.38	0.0191*
vowel [KIT] – vs. DRESS	0.298	0.019	723.9	15.69	<0.0001*
vowel [TRAP] vs. DRESS	-0.258	0.018	594.5	-14.18	<0.0001*
duration (log)	0.090	0.015	5175.0	5.89	<0.0001*
phrase position	-0.071	0.022	5118.6	-3.21	<0.0013*
F1 (normalized)	-0.119	0.009	5185.1	-13.27	<0.0001*
creak [TRUE]	-0.059	0.009	5129.2	-6.77	<0.0001*
smiles during vowel [TRUE]	0.026	0.008	4928.1	3.18	<0.0015*
sex [female]	-0.062	0.019	40.1	-3.26	<0.0023*
age	0.006	0.002	40.5	3.77	<0.0005*

KIT F2 > DRESS F2 > TRAP F2

Longer vowels have higher F2 (are fronter).

Vowels have lower F2 (are backer) as phrases progress.

RESULTS

F2 Model

Term	Estimate	Std Error	DFDen	t Ratio	Prob>|t|
Intercept	0.182	0.076	105.0	2.38	0.0191*
vowel [KIT] – vs. DRESS	0.298	0.019	723.9	15.69	<0.0001*
vowel [TRAP] vs. DRESS	-0.258	0.018	594.5	-14.18	<0.0001*
duration (log)	0.090	0.015	5175.0	5.89	<0.0001*
phrase position	-0.071	0.022	5118.6	-3.21	<0.0013*
F1 (normalized)	-0.119	0.009	5185.1	-13.27	<0.0001*
creak [TRUE]	-0.059	0.009	5129.2	-6.77	<0.0001*
smiles during vowel [TRUE]	0.026	0.008	4928.1	3.18	<0.0015*
sex [female]	-0.062	0.019	40.1	-3.26	<0.0023*
age	0.006	0.002	40.5	3.77	<0.0005*

Lowering predicts backing.
Creaky vowels are more shifted (i.e., backed).

F2 Model

Term	Estimate	Std Error	DFDen	t Ratio	Prob>|t|
Intercept	0.182	0.076	105.0	2.38	0.0191*
vowel [KIT] – vs. DRESS	0.298	0.019	723.9	15.69	<0.0001*
vowel [TRAP] vs. DRESS	-0.258	0.018	594.5	-14.18	<0.0001*
duration (log)	0.090	0.015	5175.0	5.89	<0.0001*
phrase position	-0.071	0.022	5118.6	-3.21	<0.0013*
F1 (normalized)	-0.119	0.009	5185.1	-13.27	<0.0001*
creak [TRUE]	-0.059	0.009	5129.2	-6.77	<0.0001*
smiles during vowel [TRUE]	0.026	0.008	4928.1	3.18	<0.0015*
sex [female]	-0.062	0.019	40.1	-3.26	<0.0023*
age	0.006	0.002	40.5	3.77	<0.0005*

Smiled vowels are fronter (no observed effect of smiling at phrase level).
No observed effect of movement amplitude.

F2 Model

Term	Estimate	Std Error	DFDen	t Ratio	Prob>|t|
Intercept	0.182	0.076	105.0	2.38	0.0191*
vowel [KIT] – vs. DRESS	0.298	0.019	723.9	15.69	<0.0001*
vowel [TRAP] vs. DRESS	-0.258	0.018	594.5	-14.18	<0.0001*
duration (log)	0.090	0.015	5175.0	5.89	<0.0001*
phrase position	-0.071	0.022	5118.6	-3.21	<0.0013*
F1 (normalized)	-0.119	0.009	5185.1	-13.27	<0.0001*
creak [TRUE]	-0.059	0.009	5129.2	-6.77	<0.0001*
smiles during vowel [TRUE]	0.026	0.008	4928.1	3.18	<0.0015*
sex [female]	-0.062	0.019	40.1	-3.26	<0.0023*
age	0.006	0.002	40.5	3.77	<0.0005*

Younger speakers produce lower F2 (backer vowels), more shifted.
Female speakers produce lower F2 (backer vowels), more shifted.

F1 Model

Term	Estimate	Std Error	DFDen	t Ratio	Prob>|t|
Intercept	0.692	0.092	89.8	7.54	<0.0001*
vowel [KIT] – vs. DRESS	-0.863	0.027	529.5	-31.79	<0.0001*
vowel [TRAP] vs. DRESS	0.838	0.026	419.2	32.54	<0.0001*
duration (log)	0.244	0.023	5157.6	10.58	<0.0001*
phrase position	-0.108	0.033	5123.5	-3.27	<0.0011*
F2 (normalized)	-0.276	0.021	4968.9	-13.43	<0.0001*
smiles during phrase [TRUE]	0.028	0.01	1618.2	2.8	<0.0051*
movement amplitude during vowel	0.041	0.014	5175.9	2.89	<0.0038*
movement amp * vowel [KIT]	-0.483	0.021	5159.4	-2.3	<0.0217*
movement amp * vowel [TRAP]	0.021	0.017	5153.4	1.24	0.2154
sex [female]	0.007	0.017	35.6	0.39	0.6958
age	0.001	0.001	31.0	0.83	0.4116

KIT F1 < DRESS F1 < TRAP F1
Longer vowels have higher F1 (are lower).
Vowels have lower F1 (are higher) as phrases progress.
Backing predicts lowering.

RESULTS

RESULTS

F1 Model

Term	Estimate	Std Error	DFDen	t Ratio	Prob>|t|
Intercept	0.692	0.092	89.8	7.54	<0.0001*
vowel [KIT] – vs. DRESS	-0.863	0.027	529.5	-31.79	<0.0001*
vowel [TRAP] vs. DRESS	0.838	0.026	419.2	32.54	<0.0001*
duration (log)	0.244	0.023	5157.6	10.58	<0.0001*
phrase position	-0.108	0.033	5123.5	-3.27	<0.0011*
F2 (normalized)	-0.276	0.021	4968.9	-13.43	<0.0001*
smiles during phrase [TRUE]	0.028	0.01	1618.2	2.8	<0.0051*
movement amplitude during vowel	0.041	0.014	5175.9	2.89	<0.0038*
movement amp * vowel [KIT]	-0.483	0.021	5159.4	-2.3	<0.0217*
movement amp * vowel [TRAP]	0.021	0.017	5153.4	1.24	0.2154
sex [female]	0.007	0.017	35.6	0.39	0.6958
age	0.001	0.001	31.0	0.83	0.4116

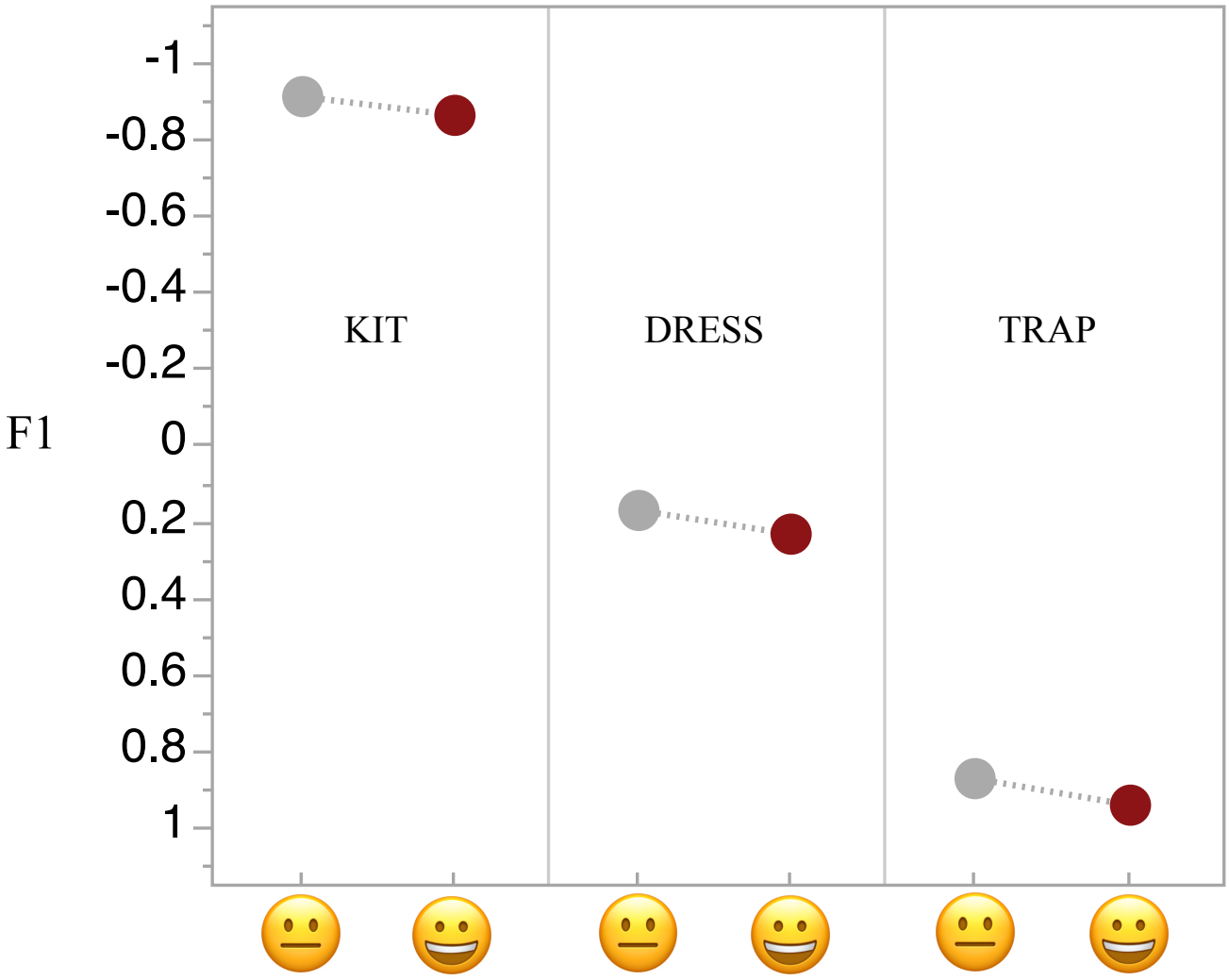
Vowels in smiled phrases are lower (no observed effect at segmental level).
Vowels characterized by more movement are lower.

F1 Model

Term	Estimate	Std Error	DFDen	t Ratio	Prob>|t|
Intercept	0.692	0.092	89.8	7.54	<0.0001*
vowel [KIT] – vs. DRESS	-0.863	0.027	529.5	-31.79	<0.0001*
vowel [TRAP] vs. DRESS	0.838	0.026	419.2	32.54	<0.0001*
duration (log)	0.244	0.023	5157.6	10.58	<0.0001*
phrase position	-0.108	0.033	5123.5	-3.27	<0.0011*
F2 (normalized)	-0.276	0.021	4968.9	-13.43	<0.0001*
smiles during phrase [TRUE]	0.028	0.01	1618.2	2.8	<0.0051*
movement amplitude during vowel	0.041	0.014	5175.9	2.89	<0.0038*
movement amp * vowel [KIT]	-0.483	0.021	5159.4	-2.3	<0.0217*
movement amp * vowel [TRAP]	0.021	0.017	5153.4	1.24	0.2154
sex [female]	0.007	0.017	35.6	0.39	0.6958
age	0.001	0.001	31.0	0.83	0.4116

No observed effect of sex or age.

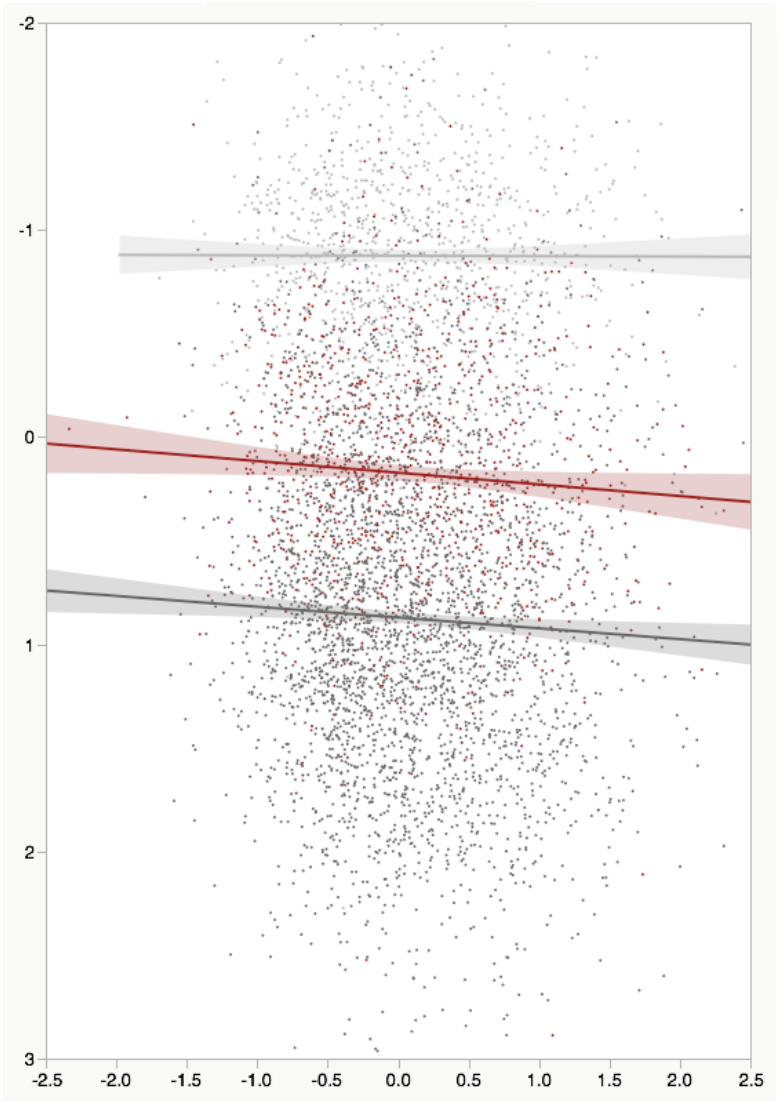
Effect of Smiling on F1



Across classes, vowels in smiled phrases are lower/more shifted (no observed effect at segment-al level).

RESULTS

Effect of Movement on F1



Vowels are lower/more shifted for tokens in which speakers are moving their bodies more (no observed effect at phrase level).

Interaction with vowel class indicates stronger effect for DRESS and TRAP.

Discussion

Affect structures vocalic variation patterns.

Lip configuration does not wholly explain observed patterns (so tongue position, larynx height likely play a role).

Innovative variants coincide with embodied displays of affect.

- Affect imbues vowel quality with meaning.
- Vowel qualities become resources for signaling affect.

Multimodal affective stancetaking using embodied and vocalic resources

Implications

Affect and Region

- How are regional accents ideologically tied to particular affective valences?

Significance of Affect

- Embodied affect structures variation at least as strongly as age and sex.
- Methodological challenges are surmountable.
- Speakers express affect and move bodies in most of our data.

Embodiment and Variation

- How does affect endow linguistic forms with meaning?
- Pratt (yesterday): creaky voice and posture
- Calder (yesterday): fronted /s/ and the gendered body
- The body as a context of variation, a constraint on variation, and a resource for variation.

Thank You!

This research was supported by a grant from the Roberta Bowman Denning Initiative in the Digital Humanities at Stanford University.

Data collection and analysis has been collaborative, with **Patrick Callier**, **Katherine Hilton**, and **Rob Voigt** playing central roles. Thanks also to Annette D'Onofrio, Penny Eckert, Sunwoo Jeong, Teresa Pratt, Janneke Van Hofwegen, Rob Xu, and Lal Zimman for discussions at various points of development.

Thanks to audiences at Brown University, The Ohio State University, The University of California (Merced), and the University of Nevada (Reno) for feedback on earlier versions of this research.

Thanks to Quan Nguyen for figure drawing.

Questions? podesva@stanford.edu