Vowel Quality and Affective Stance

Robert J. Podesva

Stanford University

Sociolinguistic, Psycholinguistic, and Formal Perspectives on Meaning

July 2, 2018

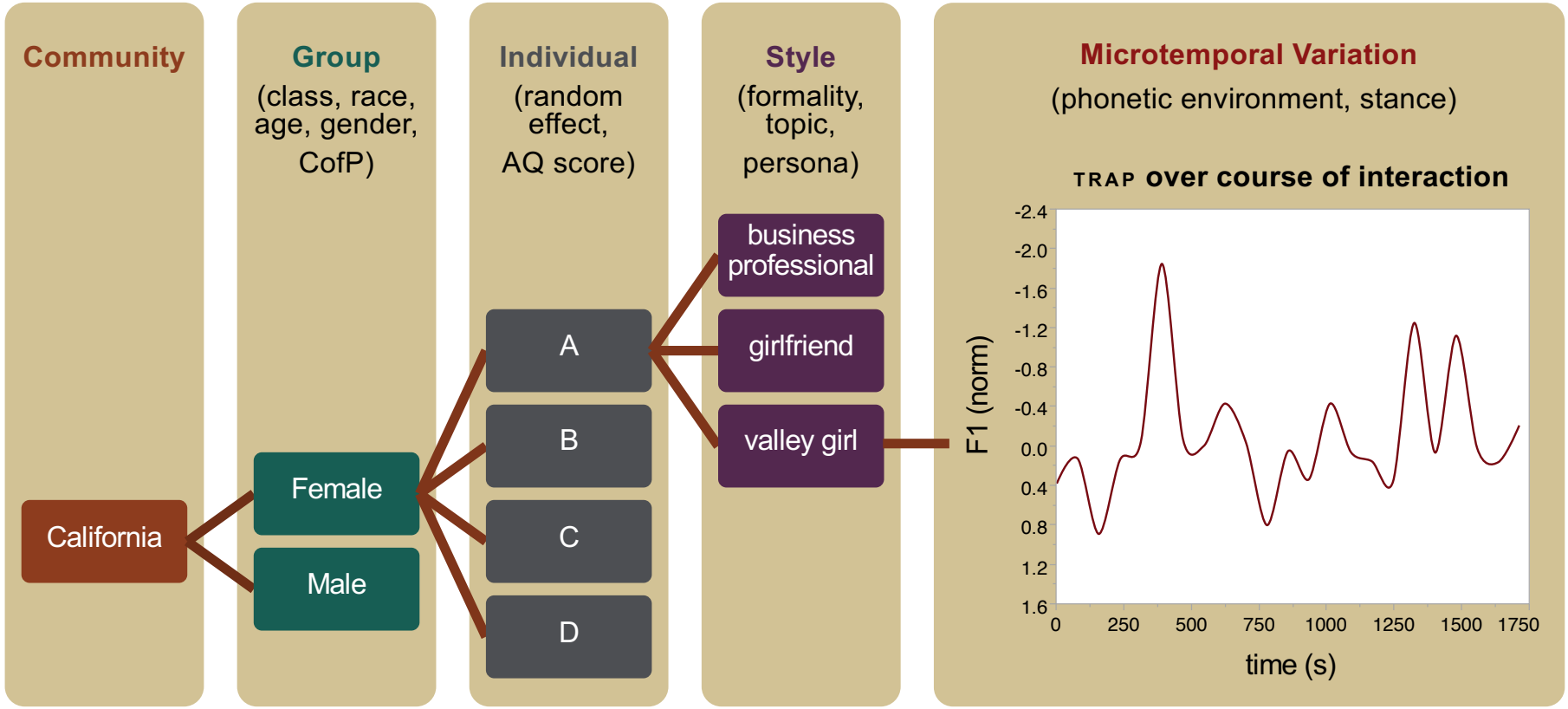
A Goal of Sociolinguistics

Understanding and modeling the linguistic and social factors that structure variation patterns

Variation is introduced at different levels of granularity:

- Across groups
- Across individuals
- Within individuals

Capturing Variation at Different Levels



Stance in Variationist Sociolinguistics

The term *stance* has been used to refer to wide variety of distinct interactional practices (Kiesling 2009, Freeman 2014, Podesva 2016, Levon 2016)

- Evaluation
- Expression of affect
- Epistemic authority
- Relationships among people

Levon (2016) illustrates that any given stancetaking act can serve multiple functions simultaneously.

Affective Stance

- Plays a fundamental role in structuring vocalic variation patterns
- Attending to stance must be part of the variationist enterprise (and it is empirically possible)

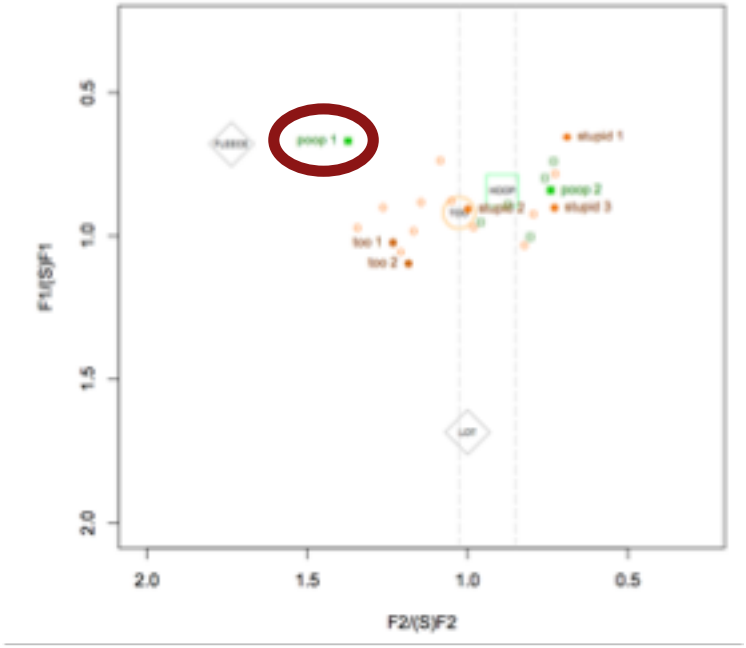
Complementary Approaches

- Tamminga (forthcoming) centers vowel quality, and looks for systematicity in stance-taking practices that occur with particular vowel qualities.
- This work centers affective stance, and looks for systematicity in the kinds of vowel qualities produced when such stances are taken.

Vowels and Affective Stance

fronter vowels → positive affect

(Johnson 2006, Wong 2014, Podesva, Callier, Voigt, and Jurafsky 2015)



Irene

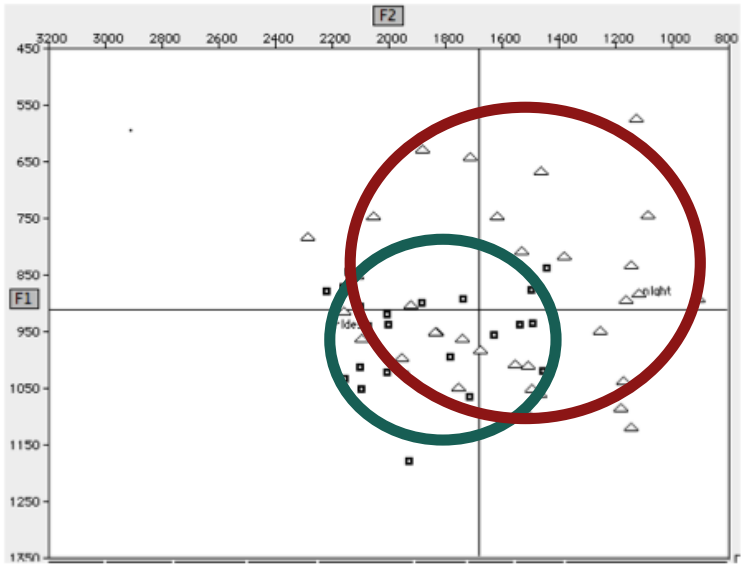
fronting of GOOSE

(Wong 2014: 214)

Vowels and Affective Stance

backer vowels → negative affect

(Eckert 2010, Eckert 2011, Wong 2014, Pratt 2018)



Collette

nucleus of PRICE

- nice
- negative

(Eckert 2010: 75)

The Challenge of Affect

Previous approaches rely on detailed analysis of discourse and/or ethnographic insight.

"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)

Claim

Affective stancetaking practices structure vowel quality variation at the microtemporal level.

Hypothesis

Vowel quality correlates with one or more modality through which affective stance is dynamically expressed:

- Lexical (valence/arousal, sentiment)
- Vocal (voice quality, prosody)
- Embodied (smiling, body movement)

Outline

Introduction

The Study: Linguistic Variable, Community, Speakers, Data

Part 1. Lexical Affective Stance: Valence/Arousal, Sentiment

Part 2. Vocal Affective Stance: Voice Quality, Prosody

Part 3. Embodied Affective Stance: Smiling, Movement

Conclusion

Collaborators

Patrick Callier

data collection and
acoustic analysis

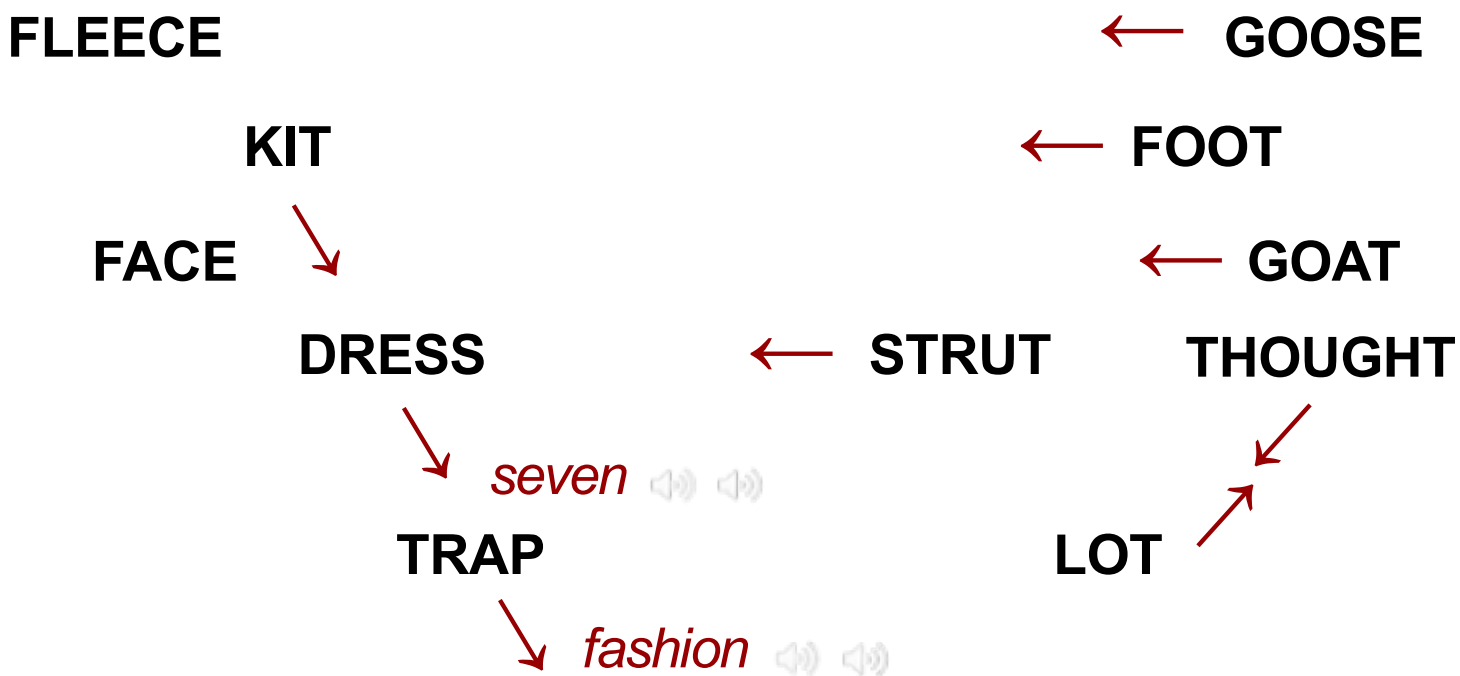
Rob Voigt

data collection and
computer vision analysis

Katherine Hilton

design of data
collection procedures

Western Vowel Shift



Interactional Sociophonetics Laboratory

Acoustical specifications of sound booth, staged as living room

Data Collection

Separate audio and video recordings for each speaker

Dyadic interactions, video and audio (wireless microphones) recorded

~30 min of conversation, with aid of prompts

Post-recording survey (demographic information, assessments)

Example Recording

Sample: 42 speakers from the Western United States

Gender	26 women
	16 men
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

- ~21 hours of speech
- Transcribed in ELAN (Lausberg and Sloetjes 2009), aligned with 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
  - Segments classified as ±creaky (Kane et al. 2013)
- Acoustic measures reduced to median value/vowel
- All stressed vowels > 75 ms normalized using Lobanov (1971)
  - Excluded preceding vowels, glides, /r/; following vowels, glides, liquids
- N = 23,311

Statistical Models: Linear Mixed-Effects Regression

Observations	stressed KIT, DRESS, TRAP (N = 5,255) exclusions: __ {*nasals, /g/*}
Responses	F1, F2, F2-F1
Random	speaker, word, pre and fol segment
Linguistic predictors	vowel class, duration (log), phrase position
Social predictors	gender, age

Separate models created for lexically-expressed, vocal, and embodied affect.

Outline

Introduction

The Study: Linguistic Variable, Community, Speakers, Data

Part 1. Lexical Affective Stance: Valence/Arousal, Sentiment

Part 2. Vocal Affective Stance: Voice Quality, Prosody

Part 3. Embodied Affective Stance: Smiling, Movement

Conclusion

Examining Lexical Affective Stance

Two Challenges

- Coding (previous work relies on discourse analysis and ethnography)
- Quantification (previous word is based on small datasets)

Computational Approaches Taken Here

- Valence and Arousal Lexicons (Mohammad 2018)
- Sentiment Analysis (Socher, Perelygin, Wu, Chuang, Manning, Ng, and Potts 2013)

Valence and Arousal Lexicons

word	valence	arousal
emptiness	0.177	0.173
menace	0.125	0.894
floral	0.833	0.235
champion	0.89	0.88

(Mohammad 2018)

Operationalization of Valence/Arousal

For each vowel, the valence and arousal for both the word and phrase that the vowel occurred in was calculated.

utterance	"the	most	fashionable	person"
valence	X	X	0.75	0.741
arousal	X	X	0.51	0.363

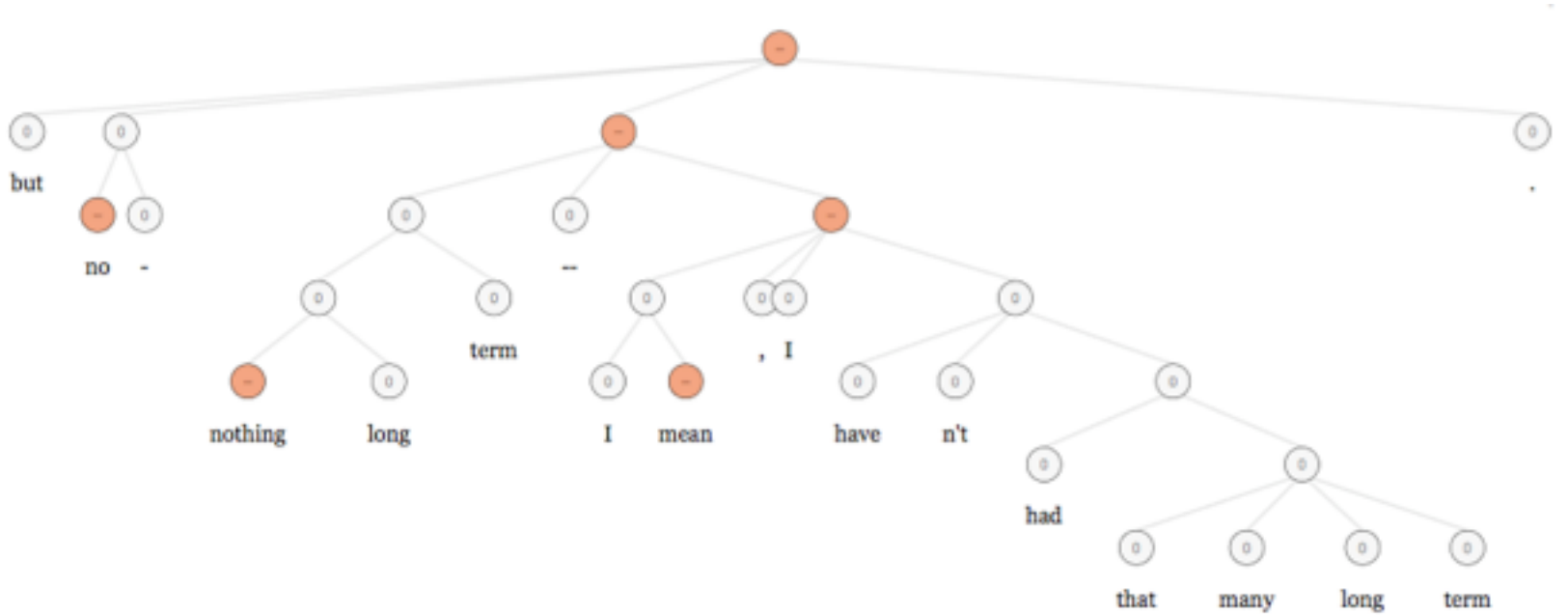
TRAP in fashionable

word valence = 0.75
word arousal = 0.51

mean phrase valence = 0.7455
mean phrase arousal = 0.4365

Sentiment Analysis

"but no- nothing long term -- I mean, I haven't had that many long term"



(Socher, Perelygin, Wu, Chuang, Manning, Ng, and Potts 2013)

Sentiment Analysis

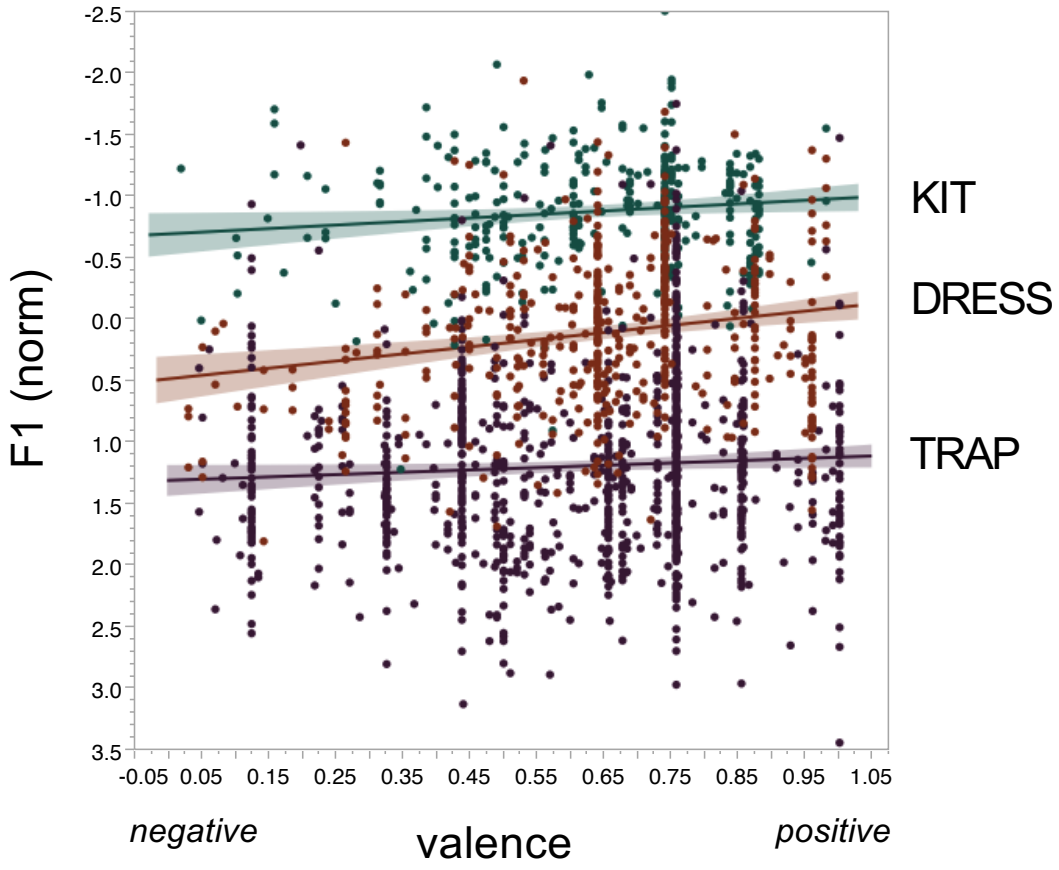
Sentiment Analyzer Output (for Each Phrase)	Converted to Score
Very positive	+2
Positive	+1
Neutral	0
Negative	-1
Very negative	-2

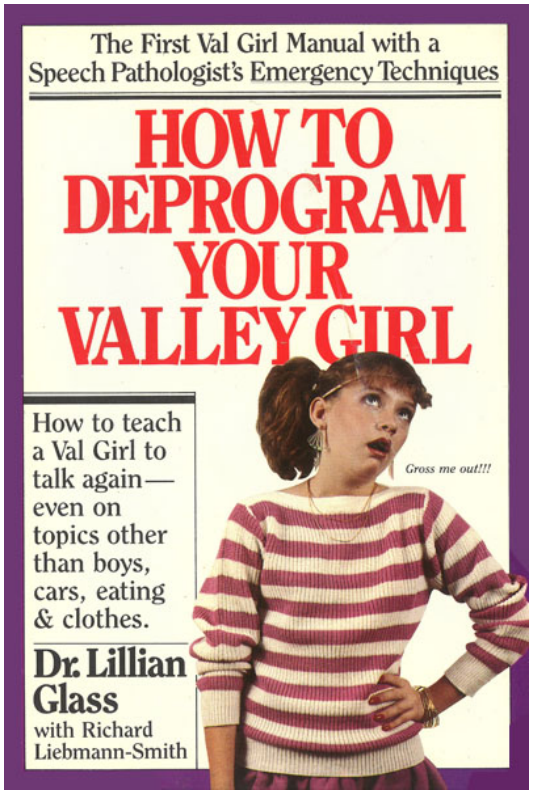
Utterance sentiment = mean score for phrases in utterance

Summary of Results

	F1 Model	F2 Model	F2-F1 Model
word valence	✓	X	X
word arousal	X	X	X
phrase valence	X	X	X
phrase arousal	X	X	X
sentiment	X	X	X

Effect of Word Valence on F1





Pratt and D'Onofrio 2017

Interim Discussion

Word valence predicts vowel quality (F1) for KIT, DRESS, and TRAP, suggesting that vowel quality might serve as a resource for conveying affect.

Correlation between lexically-conveyed affect and vowel quality obtains only at the level of the word.

Outline

Introduction

The Study: Linguistic Variable, Community, Speakers, Data

Part 1. Lexical Affective Stance: Valence/Arousal, Sentiment

Part 2. Vocal Affective Stance: Voice Quality, Prosody

Part 3. Embodied Affective Stance: Smiling, Movement

Conclusion

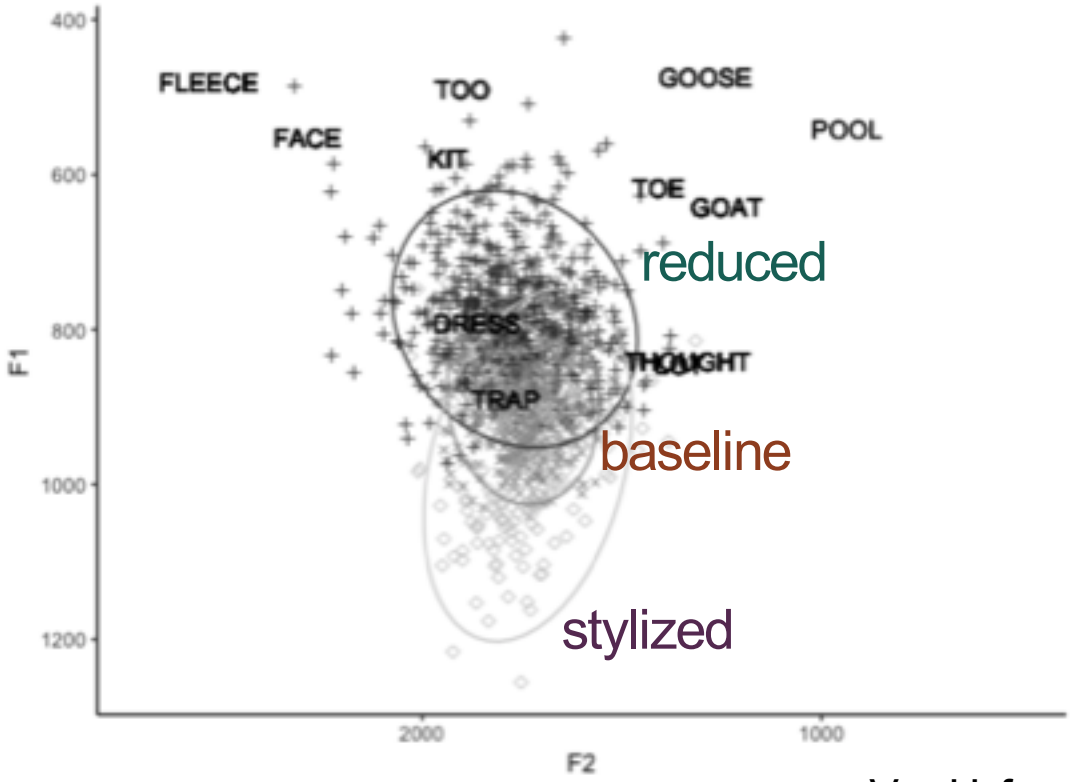
Voice Quality/Prosody and Affect

Voice quality as the most “interior” of linguistic variables (Eckert 2018, Starr 2015).

Iconic readings of affect through voice quality enable listeners to correctly recognize affect (in the absence of lexical material) across a variety of languages (Pell, Paulmann, Dara, Alasseri, and Kotz 2009).

Affective prosody on affectively neutral words primes recognition of words affectively valenced words (Kim 2017).

Vowel Quality and Voice Quality/Prosody



Stylized clusters for this and other vowels characterized by longer duration, greater intensity, and creakier phonation.

Van Hofwegen (2017: 68)

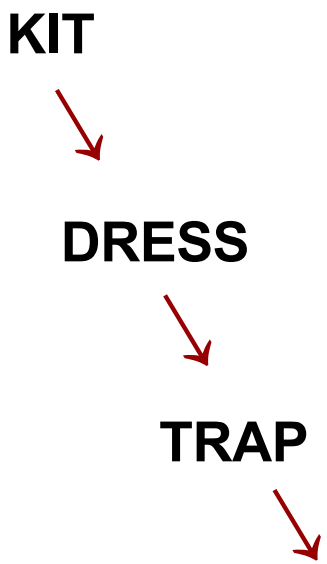
Quantifying Voice Quality and Prosody

Phonetic Property	Acoustic Measure
length	log duration
pitch	f0
loudness	intensity
periodicity	CPPS
voice quality	±creak
nasality	A1-P0

Summary of Results

	F1 Model	F2 Model	F2-F1 Model
log duration	✓	✓	✓
f0	X	✓	✓
intensity	✓	X	✓
CPPS	✓	✓	X
creak	✓	✓	✓
A1-P0	✓	✓	✓

Effects of Voice Quality/Prosody on Vowel Quality



Lower and backer (more shifted) vowels if

- creaky (therefore lower f0, CPPS)
- more oral/less nasal

- longer in duration
- higher in intensity

Interim Discussion

Lowered and backed vowels predicted by creaky phonation, which conveys low arousal, emotional detachment; vowel quality may carry similar affective potential.

Lowered and backed vowels also predicted by greater orality, the social meaning of which is unclear given its unmarked status (see Acton 2014, Beltrama and Staum Casasanto 2017).

Outline

Introduction

The Study: Linguistic Variable, Community, Speakers, Data

Part 1. Lexical Affective Stance: Valence/Arousal, Sentiment

Part 2. Vocal Affective Stance: Voice Quality, Other Suprasegmentals

Part 3. Embodied Affective Stance: Smiling, Movement

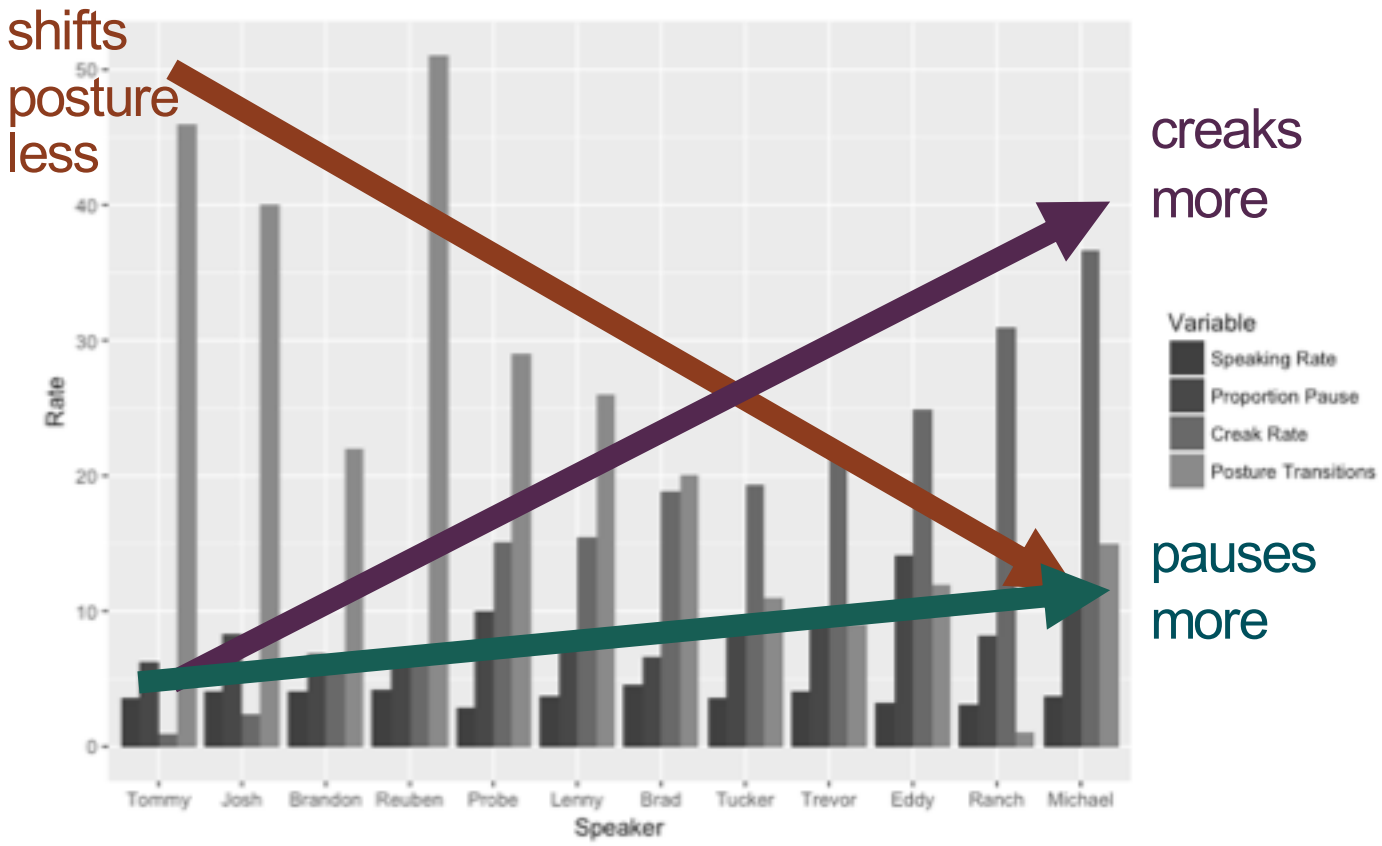
Conclusion

Embodiment, Affect, and Variation

- Embodiment: semiotically meaningful use of the body, constrained by bodily form, the physical environment, and discourse, which encompasses a wide range of phenomena (gesture, hexis, posture, physical stance, gaze, actions, adornment).

- Recent studies have begun to attend to embodied affect and how it structures variation patterns (Calder 2017, Levon 2017, Pratt 2018).

Embodied and Linguistic Performances of Affect



Pratt (2018: 86)

Smiling

Smiling ⟶ Affect

- Subjects who smiled more likely to have reported being happy (Ekman, Friesen, and Ancoli 1980).

[Smiling ⟶ Affect] ⟶ Phonetics

- Subjects can discern whether (and how) speakers are smiling on the basis of audio (Drahota, Costall, and Reddy 2008).

Smiling Annotation

smiling = TRUE FALSE

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.

Body Movement

Body Movement ⟶ Affect

- Observers (dancers, psychotherapists) able to infer affective information from the way subjects move (Dittman, Parloff, and Boomer 1965)
- Direct correlation between how much people move and emotional arousal (Ekman and Friesen 1967, Pollick et al. 2001, Pollick et al. 2002, Camurri et al. 2003, Atkinson et al. 2007, Crane and Gross 2007)

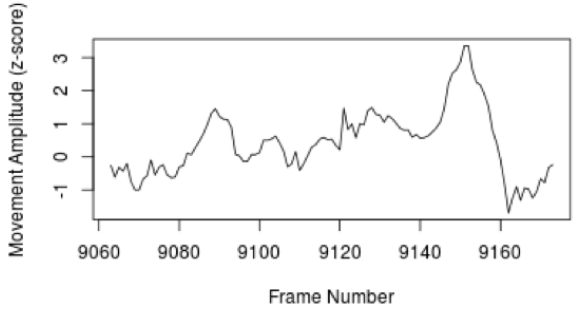
[Body Movement ⟶ Affect] ⟶ Phonetics

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

Quantifying Movement

Movement Amplitude

Voigt, Podesva & Jurafsky (2014)



Smiling and Movement During Speech

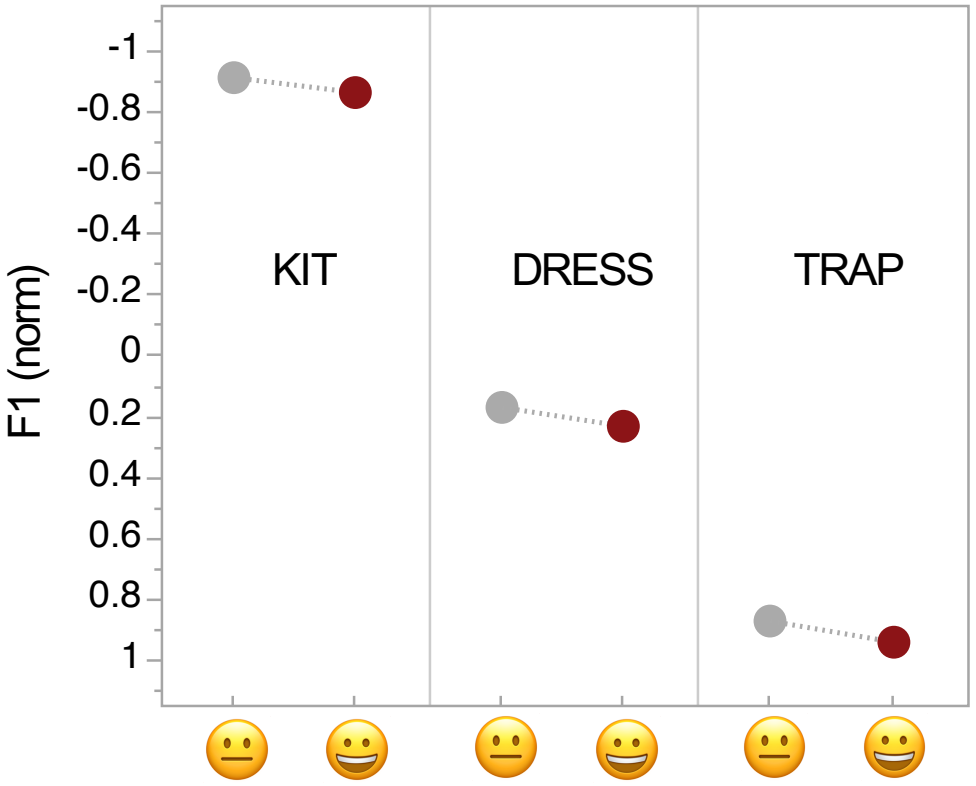
We have *so* many *freshmen* with like eager *attitudes* and *perceptions* about Stanford, and you're like, "That's just *not* how it works." Like, "You're not- I can't tell you *that*. It's something you gotta experience, and that sucks."

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

Summary of Results

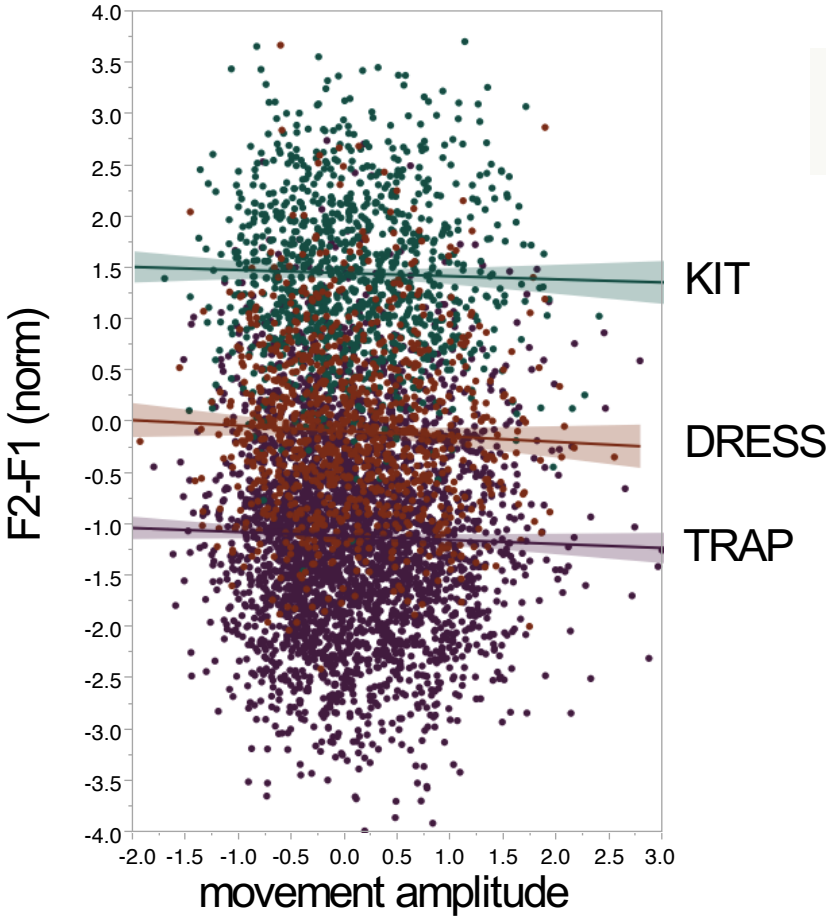
	F1 Model	F2 Model	F2-F1 Model
smiling (during vowel)	X	✓	X
smiling (during phrase)	✓	X	X
movement (during vowel)	✓	X	✓
movement (during phrase)	X	X	X

Effect of Smiling During Phrase on F1



Lower (more shifted) when smiling

Effect of Movement During Vowel on F2-F1



Lower and backer (more shifted) when speakers moving more

Same pattern evident for F1: lower vowels when speakers moving more; interaction indicates stronger effect for DRESS and TRAP

Outline

Introduction

The Study: Linguistic Variable, Community, Speakers, Data

Part 1. Lexical Affective Stance: Valence/Arousal, Sentiment

Part 2. Vocal Affective Stance: Voice Quality, Other Suprasegmentals

Part 3. Embodied Affective Stance: Smiling, Movement

Conclusion

Summary

More shifted vowels are observed when speakers are expressing affect...

	vowel quality measure	valence	arousal	level
lexically	F1	negative		word
vocally (creak)	F1, F2, F2-F1		low	vowel
bodily (smiling)	F2	negative		vowel
bodily (smiling)	F1	positive		phrase
bodily (movement)	F1, F2-F1		high	vowel

But is this “contradiction” really a problem?

1. Affect is more complicated than a set of two binaries: negative vs. positive valence; low vs. high arousal

2. Western-shifted vowels can index two distinct social types, each with a different affective profile:
   - Negative valence, low arousal – valley girl, slacker
   - Positive valence, high arousal – enthusiastic airhead, surfer dude

Compatible with current theories of social meaning, none of which assume a one-to-one mapping between form and meaning. Indexical fields can consist of contradictory meanings (Maegaard and Pharao forthcoming).

Conclusion

Lexical, vocal, and embodied expressions of affect structure vocalic variation patterns.

Attending to them helps capture the social motivations for microvariation.

It's also empirically possible, even from a big data, first wave perspective.

Speakers are expressing affect in all the data (socio)linguists analyze.

- Quantitative models could be improved
- We are missing a fundamental insight about what motivates speakers to choose one linguistic form over another

Thank You!

The collection of audiovisual data was supported by a grant from the Roberta Bowman Denning Initiative in the Digital Humanities. Data collection for the Voices of California Project was supported by Stanford University and the Richard A. Karp Foundation. Many thanks to audiences at Cornell University, Sheffield University, University of Nevada, Reno, University of California, Merced, Brown University, the Ohio State University, Michigan State University, and the University of California, Berkeley for feedback on segments of the work presented here. Thanks also to participants in a seminar on Language and Embodiment, and Annette D'Onofrio, Penny Eckert, Lauren Hall-Lew, Katherine Hilton, Sunwoo Jeong, Emma Moore, Teresa Pratt, Janneke Van Hofwegen, Rob Voigt, and Rob Xu for discussions of this material.

Questions?

podesva@stanford.edu