MODELING SOCIAL MEANINGS OF PHONETIC VARIATION AMID VARIABLE CO-OCCURRENCE: A MACHINE-LEARNING APPROACH

Dan Villarreal & James Grama

University of Pittsburgh; Sociolinguistics Lab, University of Duisburg-Essen
d.vill@pitt.edu, james.grama@uni-due.de

ABSTRACT

The social meaning of phonetic variation is central to sociophonetics. Despite the co-occurrence of multiple potentially meaningful variables in speech, previous experimental approaches generally focus on single features. This analysis investigated whether machine-learning methods can uncover vowel features that impact sociolinguistic perceptions. Ninety-seven listeners rated short, spontaneous California English stimuli on 12 affective scales; /æ/ and /u/ tokens were manipulated to create “California-shifted” and “unshifted” guises, with other vowels left to vary naturally. Mirroring “bag-of-words” approaches to text corpora, we treated stimuli as “bags of features” based on 11 vowel phonemes. We used the Boruta feature-selection algorithm to assess the importance of these features, plus guise, on affective scale ratings. The most frequently selected variables included both predictable (/æ/) and less well-attested variables (/ʊ/). However, guise was never selected, suggesting it was an overly-coarse axis of relevant variation. We argue that “bottom-up” approaches can model social meanings amid variable co-occurrence.

Keywords: vowel variation, computational sociolinguistics, social meaning, methods, California English

1. INTRODUCTION

Social meanings are the glue that connects the social and linguistic dimensions of language variation and change.[1] More formally, a social meaning is an ideologically mediated relationship between a linguistic structure and some stance, identity, or personal characteristic; these meanings emerge (and may be observed) in both production and perception. For example, a New York City English speaker may delete postvocalic /r/ to emphasize that they are an “authentic” New Yorker [1]; listeners may likewise perceive a New York City English speaker who deletes postvocalic /r/ as better suited for working-class than middle-class jobs [2]. These meanings are a central concern of the “third wave” of language variation research, which views social meaning as an essential (rather than incidental) part of language [3].

One strand of research attempts to investigate social meanings in perception, most often via matched-guise tasks (MGTs), which compare listeners’ reactions to pairs of stimuli (i.e., guises) that differ only in their use of a particular variable (e.g., [ɪn] vs. [ɪŋ] variants of English *-ing* [4]). This method necessarily entails hypothesizing that a particular variable is socially meaningful, but this researcher choice means we may fail to account for other meaningful variables that are present in the stimulus. In other words, variables do not occur in isolation, including in chain shifts where multiple changes are occurring at once. This phenomenon of *variable co-occurrence* thus creates a dilemma for operationalizing and analyzing social meanings.

Given this quandary, the present analysis investigated whether machine-learning methods can uncover vowel variables that impact sociolinguistic perceptions, as these methods are particularly good at sorting through complex data. To test this methodology, we undertake a re-analysis of Villarreal [5], which investigated the social meanings of California English vowels (TRAP backing and GOOSE fronting) using an MGT. (In this paper, we use Wells’ [6] notation for English vowel phonemes.) We use the Boruta algorithm [7], which selects all predictors that are important to some dependent variable (in this case, attributes like *familiar* and *feminine*). Our findings suggest that “top-down” approaches to sociolinguistic perception can miss crucial details about which co-occurring variables influence social meanings.

1.1. California English

The last 30 years has seen a groundswell of work on the vowel system of California English as spoken by white, coastal speakers (but see [8–11]). The bulk of this work has focused on two phenomena [10–15]:

1. The *Low-Back-Merger Shift* (LBMS) [16], consisting of:
   a. the merger (or near-merger) of LOT and THOUGHT in the low-back corner of the vowel space; this is hypothesized to trigger
   b. lowering and/or retraction of the short front vowels TRAP, DRESS, and KIT;
2. *Back Vowel Fronting* (BVF), involving long back vowels GOOSE and GOAT, particularly following coronal consonants.

While neither phenomenon is restricted to California or the western US states (e.g., [17–20]), perceptual research has suggested that Californians and non-Californians alike associate TRAP-backing with a Californian identity [5, 21, 22]. Proportionally less attention has been paid to the putative fronting of short back vowels FOOT and STRUT despite structural parallels to BVF [23]. The work that directly investigates these vowels rarely identifies compelling evidence of STRUT fronting over time [11], and identifies FOOT centralization as having reached completion [24].

2. METHODS

This study represents a re-analysis of Villarreal's matched-guise perception task, which used an MGT to investigate which social meanings Californian listeners perceive in California English vowels [5].

In this study, TRAP and GOOSE were chosen to represent California English vowels more broadly; each is implicated in a major vocalic subsystem of California English (TRAP in LBMS, GOOSE in BVF), and there is evidence linking each to Californian social meanings [8, 22, 25, 26]. Listeners (*n*=97), all from California, heard stimuli created from 12 voices (also from California) producing spontaneous speech retelling a short cartoon. In each trial (*n*=580), the listener identified the speaker's region and rated the speaker on 12 attribute scales. Scales were chosen based on a pretesting task and previous research on California English social meanings. Each stimulus contained either a "California-shifted" guise (where TRAP and GOOSE were acoustically manipulated to be backer and fronter, respectively), or a "conservative" guise (where TRAP and GOOSE were fronter and backer, respectively). Other vowel phonemes were left to vary naturally. For further details on these methods, including acoustic manipulation, see [5] and https://github.com/djvill/Vowel-Manipulation.

Villarreal [5] found that guise significantly affected perceived attributes for *Californian*, *sounds like a Valley girl*, and (for male speakers) *confident*; however, there was substantial variance in attribute ratings that guise failed to capture. In other words, while TRAP and GOOSE have a clear impact on listener assessments of Californian-ness, listeners were clearly also attending to cues beyond these two vowels in the signal. Furthermore, because TRAP and GOOSE were shifted in tandem in the stimuli, it was not possible to disentangle the effects of either vowel individually. Thus, we turn here to machine-learning methods to see how multiple vowel variables may have influenced listener behavior.

2.1. Feature set

A conceptual challenge to modeling the effects of multiple cues is sparsity. English has a large number of vowel phonemes unevenly distributed through the lexicon, meaning short stretches of speech are unlikely to include all possible vowels. Because stimuli were spontaneously produced (albeit all on the same topic), they all contain slightly different content and thus different vowel variables. For example, 13 of the 24 stimuli in the perception task include zero FLEECE tokens. To deal with sparsity, we treated each stimulus as a "bag of features" (based on the computational approach to text corpora as "bags of words" [27]).

These features sorted vowel variation into discrete bins, in order to assess the effect of (e.g.) high FOOT or front TRAP on listeners' perceptions. Our procedure for creating features is schematized in Fig. 1. Vowels' midpoint F1 and F2 were measured via Praat [28], hand-checked, and normalized using the *Atlas of North American English* (ANAE) procedure (based on 466–712 tokens per speaker) [13]. In order to contextualize vowel variation in our stimuli against the broader range of North American vowels, we used ANAE's "natural breaks", which define F1 and F2 bins for North American English vowel variation. This procedure yielded features across 11 vowel phonemes: FLEECE, KIT, FACE, DRESS, TRAP, PRICE, LOT, STRUT, GOAT, FOOT, and GOOSE. Four vowels (CHOICE, MOUTH, THOUGHT, and pre-nasal TRAP-N) were discarded from the final analysis as they were not adequately represented across all stimuli. The final dataset comprised 379 tokens across 24 stimuli.

2.2. Analysis

To determine which vowel variables were important for sociolinguistic perception, our analysis used the Boruta algorithm [7]. This algorithm's primary use case is feature selection, answering the question "which predictor(s) appear to impact the dependent variable?" For example, Dickson & Durantin used Boruta to determine important predictors of reflexive pronoun choice in Australian Kriol [29]. Briefly, Boruta identifies whether features increase the prediction accuracy of the dependent variable, by comparing each feature's prediction accuracy to that of randomly shuffled "shadow" features.

In this study, separate Boruta models were run with each attribute as the dependent variable, via the `Boruta` package in R [30, 31]. This procedure returned a list of selected features for each attribute scale. We interpret a feature's selection as indicating that the feature impacted listeners' perceptions vis-à-vis the attribute scale in question.

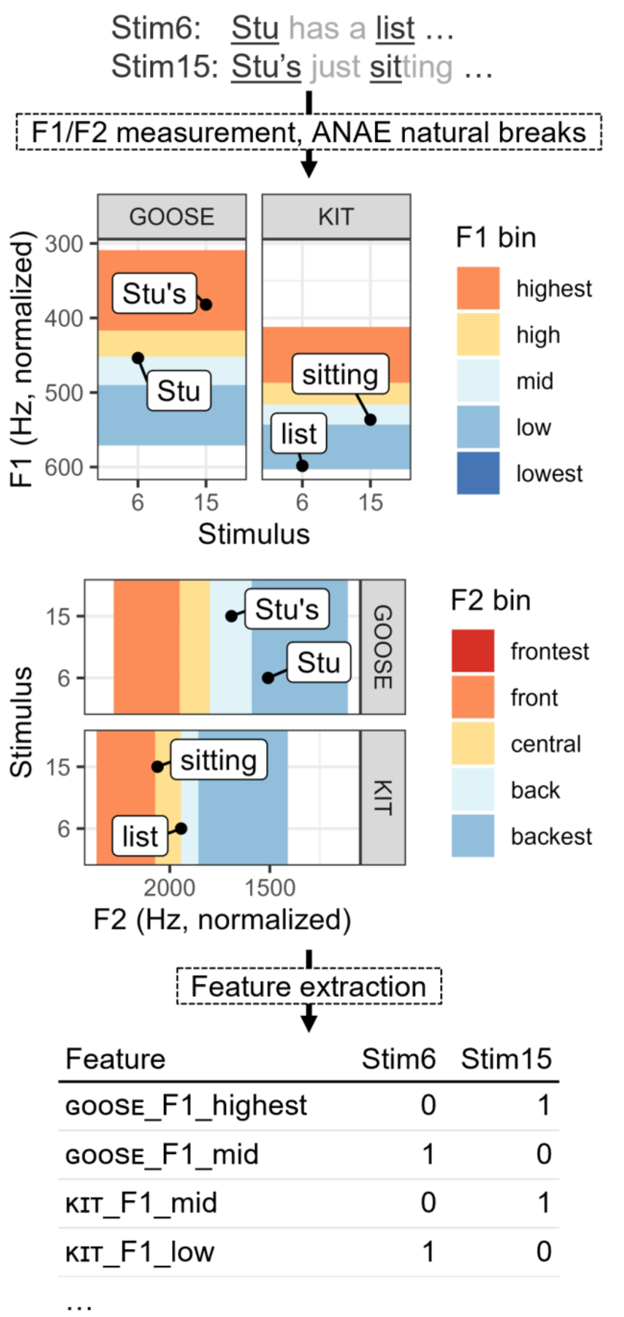

Figure 1: Schematic for creation of feature set from stimuli, based on ANAE [13] natural breaks.

3. RESULTS

Here, we primarily focus on which variables were most frequently selected by the Boruta algorithm. These results are presented in Table 1. We identify three key findings:

1. Vowels that are changing (or have changed) in California English are not necessarily more impactful to listeners' perceptions.
2. The most impactful vowel variables were FOOT, TRAP, and GOAT.
3. Guise was never selected as important to attribute scale ratings.

Variable	F1	F2
FOOT	10	8
TRAP	8	6
GOAT	4	10
FLEECE	7	6
KIT	6	6
FACE	4	6

Variable	F1	F2
LOT	5	5
GOOSE	5	5
PRICE	5	4
STRUT	5	3
DRESS	3	4

Table 1: For each variable and formant, the number of times at least one feature was selected for an attribute scale (out of 12); bold = change in progress; italics = marginal evidence of change in progress.

As Table 1 indicates, different vowel variables did not all influence attribute scale ratings to the same degree. For example, at least one FOOT F1 feature was selected as important for 10 of 12 scales, but only 3 of 12 scales for DRESS F1. Pooling these results across F1 and F2 presents a picture of each variable's impact on listeners' perceptions overall, and which variables were more *impactful* than others.

These results indicate that oft-studied variables in California English are not necessarily more impactful to listeners' perceptions, at least with regard to this particular set of attributes. The most impactful variable identified here, FOOT, is rarely investigated in California despite being structurally related to well-attested GOOSE and GOAT fronting; FLEECE, which is almost completely unattested as undergoing change (but see [11]), outranks several well-attested California English variables. Moreover, the historical ordering of LBMS sound changes (TRAP then DRESS then KIT) is not reflected in these variables' selection frequencies relative to one another. This evidence bolsters Eckert and Labov's claim that only individual variables, not the structural phenomena that they comprise (e.g., chain shifts), are available for social meaning [32].

Underneath the results in Table 1 were different distributions of features selected as important. This is exemplified by the F2 features from the highest-ranking variables: FOOT, TRAP, and GOAT (see Fig. 2). In some cases, a single attribute scale was influenced by multiple features from the same variable (e.g., both front GOAT and central GOAT influenced *friendly* perceptions); in others, a single feature acted alone. Pooling across scales, it is clear that some features stood out to listeners more than others. For FOOT, fronter features had a greater impact on listeners' perceptions than backer features; for GOAT, however, impact was assigned relatively evenly across features.

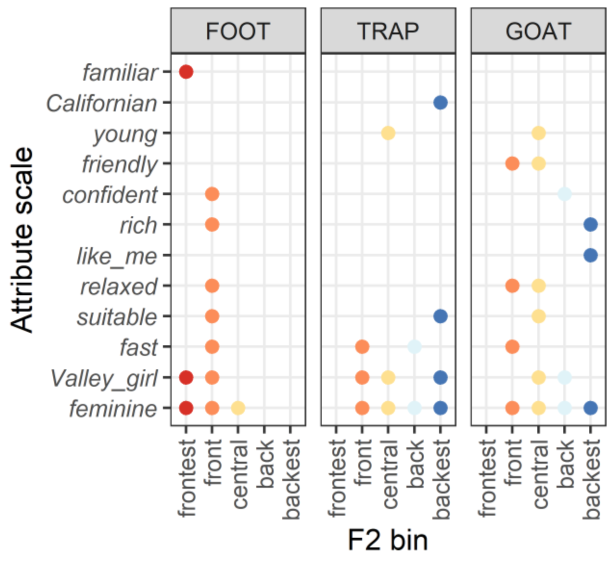

Figure 2: F2 features selected by Boruta models for FOOT, TRAP, and GOAT.

None of the Boruta models selected guise as an important predictor, despite the frequent selection of TRAP and GOOSE, the variables implicated in the guise manipulation. This result is surprising given Villarreal's [5] original finding that guise affected *Californian*, *sounds like a Valley girl*, and *confident*. In addition, our finding that TRAP was more impactful than GOOSE (Table 1) suggests that this original guise finding was driven by more by TRAP than GOOSE. This re-analysis suggests that as operationalized in [5], guise was an overly-coarse method of modeling socially meaningful vowel variation.

4. DISCUSSION

Social meanings are crucial to understanding language variation and change, and sociolinguists have an array of methodological tools at their disposal for investigating them, both in production and perception. Our findings raise critical questions about what gets overlooked by existing methods such as MGTs. However, the feature-selection method we have presented here is not without its limitations. We thus suggest that, rather than undermining the established methodological toolkit for investigating social meanings, the feature-selection method we have presented complements this toolkit.

The status of FOOT as the most frequently selected variable suggests that it may be a more prominent indicator of Californian-ness than previous research on California English identifies. Despite being part of early descriptions of California vowels [23], FOOT is often left out of discussions of California English vowel variation (but see [24]), and almost completely absent from research on California vowels' social meanings (but see [25] for an example of fronted FOOT in Kristen Wiig's parodic performance and [33] for a putative KIT–FOOT merger among South Asian and Korean-identifying Californians). This overlooking of FOOT is likely a matter of methodological expedience—compared to better-studied long back vowels GOOSE and GOAT, FOOT tokens are shorter (and thus more prone to measurement error), and less frequent in content words. Still, our findings imply that FOOT may be more relevant to Californian social meanings than previously assumed (see also [34]).

A further strength of this feature-selection method is that it accounts for variable co-occurrence when measuring social meanings in perception. Variable co-occurrence is a "known problem" in sociolinguistics (e.g., [35]), but actually accounting for co-occurrence in perception is challenging. Production research on co-occurrence typically appeals to analytic constructs such as stances and styles [36], indexical fields [37], and sociolects [38]. Studies of real-time listener reactions [39] represent a significant step forward, but can only measure one perceived attribute at a time.

This method is not without its limitations. While it mitigates the potential bias MGTs introduce by choosing particular variables to manipulate, there is still an element of researcher choice with respect to which features are in the "bag". Our feature set included a broad range of vowel categories, but we lack features (e.g., consonants, prosody) that may shape social meanings. Next, the particular selection of attribute scales in any perception task should not be taken to represent all possible social meanings. We would not expect the selection frequencies in Table 1 to be exactly replicated under a different set of attribute scales. Finally, this study benefited from benchmarks for discretizing vowel variation based on a large reference corpus, which is unavailable in under-resourced language contexts.

Perhaps most importantly, this method (and indeed large-scale methods more broadly) can only give a "bird's-eye" view of social meanings. Feature-selection algorithms like Boruta cannot, on their own, identify the direction or magnitude of any feature's effect, only that it has been assessed as important. MGTs offer unparalleled experimental control for zeroing in on particular variables. Like MGTs themselves, this method does not account for listener expectations that are known to affect sociolinguistic perceptions (e.g., [40]). The true value of this feature-selection method is guiding researchers' choice of variables for further study using tried-and-true methods for investigating social meaning.

5. REFERENCES

[1] Becker, K. 2009. /r/ and the construction of place identity on New York City's Lower East Side. *J. Socioling.* 13, 634–658.

[2] Labov, W. 2006. *The social stratification of English in New York City*, 2nd ed. Cambridge U. Press.

[3] Eckert, P. 2012. Three waves of variation study: The emergence of meaning in the study of sociolinguistic variation. *Ann. Rev. Anthro.* 41, 87–100.

[4] Campbell-Kibler, K. 2007. Accent, (ING), and the social logic of listener perceptions. *Am. Speech* 82, 32–64.

[5] Villarreal, D. 2018. The construction of social meaning: A matched-guise investigation of the California Vowel Shift. *J. Eng. Ling.* 46, 52–78.

[6] Wells, J. C. 1982. *Accents of English*. Cambridge U. Press.

[7] Kursa, M. B., Jankowski, A., Rudnicki, W. R. 2010. Boruta: A system for feature selection. *Fundamenta Informaticae* 101, 271–285.

[8] Eckert, P. 2008. Where do ethnolects stop? *Intl. J. Biling.* 12, 25–42.

[9] Podesva, R. J., D'Onofrio, A., Van Hofwegen, J., Kim, S. K. 2015. Country ideology and the California Vowel Shift. *Lang. Var. and Change* 27, 157–186.

[10] D'Onofrio, A., Eckert, P., Podesva, R. J., Pratt, T., Van Hofwegen, J. 2016. The low vowels in California's Central Valley. In: Fridland, V., Evans, B., Kendall, T., Wassink, A. (eds), *Speech in the Western states, volume 1*. Duke U. Press, 11–32.

[11] D'Onofrio, A., Pratt, T., Hofwegen, J. V. 2019. Compression in the California Vowel Shift: Tracking generational sound change in California's Central Valley. *Lang. Var. and Change* 31, 193–217.

[12] Hagiwara, R. 1997. Dialect variation and formant frequency: The American English vowels revisited. *The Journal of the Acoustical Society of America* 102, 655–658.

[13] Labov, W., Ash, S., Boberg, C. 2006. *The atlas of North American English*. Mouton.

[14] Kennedy, R., Grama, J. 2012. Chain shifting and centralization in California vowels: An acoustic analysis. *Am. Speech* 87, 39–56.

[15] Hall-Lew, L., Cardoso, A., Kemenchedjieva, Y., Wilson, K., Purse, R., Saigusa, J. 2015. San Francisco English and the California Vowel Shift, in *Proc. 18th ICPhS*.

[16] Becker, K. 2019. *The Low-Back-Merger Shift*. Duke U. Press.

[17] Boberg, C. 2019. A closer look at the short front vowel shift in Canada. *J. Eng. Ling.* 47, 91–119.

[18] Fridland, V., Kendall, T., Evans, B., Wassink, A. B. 2016. *Speech in the Western states, volume 1*. Duke U. Press.

[19] Fridland, V., Kendall, T., Evans, B., Wassink, A. B. 2017. *Speech in the Western states, volume 2*. Duke U. Press.

[20] Nesbitt, M., Stanford, J. N. 2021. Structure, chronology, and local social meaning of a supra-local vowel shift: Emergence of the Low-Back-Merger Shift in New England. *Lang. Var. and Change* 33, 269–295.

[21] Villarreal, D., Kohn, M. 2021. Local meanings for supra-local change: Perceptions of TRAP backing in Kansas. *Am. Speech* 96, 45–77.

[22] D'Onofrio, A. 2015. Persona-based information shapes linguistic perception: Valley Girls and California vowels. *J. Socioling.* 19, 241–256.

[23] Hinton, L. et al. 1987. It's not just the Valley Girls: A study of California English. *Proc. 13th Ann. Mtg. Berkeley Ling. Soc.* 117–128.

[24] Holland, C. L. 2014. Shifting or shifted? The state of California vowels. PhD dissertation, University of California, Davis.

[25] Pratt, T., D'Onofrio, A. 2017. Jaw setting and the California Vowel Shift in parodic performance. *Lang. in Soc.* 46, 283–312.

[26] Podesva, R. J. 2011. The California Vowel Shift and gay identity. *American Speech* 86, 32–51.

[27] Jurafsky, D., Martin, J. H. 2022. *Speech and language processing*, draft of 3rd ed.

[28] Boersma, P., Weenink, D. 2022. Praat. Version 6.2.23

[29] Dickson, G., Durantin, G. 2019. Variation in the reflexive in Australian Kriol. *Asia-Pacific Lang. Var.* 5, 171–207.

[30] Kursa, M. B., Rudnicki, W. R. 2010. Feature selection with the Boruta package. *J. Stat. Soft.* 36, 1–13.

[31] R Core Team 2022. R: A language and environment for statistical computing. Version 4.2.1

[32] Eckert, P., Labov, W. 2017. Phonetics, phonology and social meaning. *J. Socioling.* 21, 467–496.

[33] Cheng, A. 2016. A Survey of English vowel spaces of Asian American Californians. *UC Berkeley PhonLab Annual Report* 12, 348–384.

[34] Gardner, M. H., Roeder, R. V. 2022. Phonological mergers have systemic phonetic consequences: PALM, trees, and the Low Back Merger Shift. *Lang. Var. and Change* 1–24.

[35] Gumperz, J. J. 1982. *Discourse strategies*. Cambridge U. Press.

[36] Kiesling, S. F. 2009. Style as stance. In: Jaffe, A. (ed), *Stance: Sociolinguistic perspectives*. Oxford U. Press, 171–194.

[37] Eckert, P. 2008. Variation and the indexical field. *J. Socioling.* 12, 453–476.

[38] Guy, G. R. 2013. The cognitive coherence of sociolects: How do speakers handle multiple sociolinguistic variables? *J. Prag.* 52, 63–71.

[39] Montgomery, C., Moore, E. 2018. Evaluating S(c)illy voices: The effects of salience, stereotypes, and co-present language variables on real-time reactions to regional speech. *Language* 94, 629–661.

[40] Hay, J., Drager, K. 2010. Stuffed toys and speech perception. *Linguistics* 48, 865–892.

[1] We thank Nicole Holliday, Carina Ahrens, and Anne-Marie Mölders, as well as two anonymous reviewers, for valuable feedback on an earlier version of this paper. This research was funded in part by the UC Davis Department of Linguistics.