The new normal: Multi-modal distributions signifying loci of vocalic stylization

Janneke Van Hofwegen

Qualifying Paper

June 16, 2014
Abstract
This paper proposes a new model for how stylization can be explored quantitatively. Specifically, the analysis tests three hypotheses: 1) that intra-speaker vowel distributions along the F1 and F2 axes are not necessarily normal; in fact, they can be multi-modal; 2) that distributions on the outside edges of multi-modal distributions are loci for stylization; and 3) that vowel classes likely to vary multi-modally are also those carrying sociolinguistic salience. This is demonstrated through an exhaustive study of the vocalic repertoire of a young female speaker (age 19), a lifelong resident of Redding, an inland-non-urban city in Northern California, and an advanced participant in the California Vowel Shift (CVS). For this speaker, all vowel tokens for front lax vowels BIT, BET, and BAT, and back vowels TOO/BOOT and TOE/BOAT—monophthongal vowel classes implicated in the CVS—were extracted from forced-aligned sociolinguistic interviews, and measurements were taken at F1 and F2 midpoints (~ 3000 tokens). Goodness-of-fit tests for distributions of vowel tokens along the F1 and F2 axes revealed them to be far from normal in any statistical sense. In fact, for these CVS-shifted vowels, the distributions were bi-modal or tri-modal, indicating smaller distributions of tokens on the extreme edges of the “normal” distribution, also in the primary acoustic dimension and direction of each vowel class’s movement in the CVS. I argue these multi-modal distributions depict clusters of stylized tokens, not only because they represent extreme F1/F2 values relative to a speaker’s norm, but also because they correlate with other factors shown to be perceptually/stylistically salient, namely prosodic prominence (Mo 2008; Cole et al. 2010) and creak (Podesva 2007; Mendoza-Denton 2011). In short, this speaker uses extreme distributions of CVS-shifted vowels to stylize, which suggests a link between stylistic performance and sound change. This supports findings by others (e.g., Podesva 2011a; Drager, Eckert, and Moon 2008) showing that features already in flux via sociolinguistic and structural factors are also prime resources for stylization.
1 Introduction

Style as an integral component of language variation is a major tenet of third-wave variation theory. In her review of the three major waves of variation theory, Eckert (2012) explains, “variation does not simply reflect, but also constructs, social meaning” (p. 87). The contrast between a variation that constructs rather than reflects social meaning is an important one. If the first two waves considered speakers to be guileless participants in linguistic processes and social structures largely outside of themselves and their control, the third wave emphasizes the processes and structures inside the purview of the individual. To summarize the third-wave enterprise further, Eckert says “The emphasis on stylistic practice in the third wave places speakers not as passive and stable carriers of dialect, but as stylistic agents, tailoring linguistic styles in ongoing and lifelong projects of self-construction and differentiation (2012, p. 97-98).

In this vein, third-wave studies have shown how individuals construct styles and identities via linguistic variation, but it has largely been up to the discretion of the researcher to ascertain how this is done analytically. If first- and second-wave variationists have started with structures (i.e., specific linguistic and social factors) and leave style largely out of the picture, third-wave variationists start with styles (Eckert 2012) and then seek to identify the linguistic structures that correspond meaningfully to them. And, of course, what those meanings are is another integral part of the analysis. Both approaches place onus on the analyst to identify the relevant structures or styles under study in the first place, but—largely because the third-wave enterprise is so oriented to individual stylistic patterns (and thus, not on broad demographic categories)—the third-wave analytical approach has been understandably difficult to operationalize, and hence to replicate in broad strokes. ‘How-to’ textbooks for conventional (i.e. first- and second-wave) sociolinguistic analyses are not difficult to come by, but such relatively straightforward formulae for stylistic analyses do not as-yet exist. This may be because of the relative novelty of the third-wave approach, but it is also likely due to its idiosyncratic nature as well.

At the same time, it cannot be denied that speakers are not only idiosyncratic but are also part of communities that participate in larger linguistic systems. It is true that speakers are in many ways “passive and stable carriers of dialect” (Eckert 2012, p. 97) regardless of the individual styles they embody. Thus, in order to understand a speaker’s stylistic potential, we also need to know the larger linguistic system she engages within as a basis for comparison.

Keeping in mind both the larger linguistic system and the idiosyncrasies of personal style, this paper proposes a new model for how stylization—specifically vocalic stylization—can be explored quantitatively. In doing so, it provides a toolkit for analysts who want to assess a speaker’s idiosyncratic stylistic repertoire in a quantitatively principled, systematic way, which can be automated and broadly
applied to multiple speakers as well. In order to do this, the paper tests three main hypotheses:

1) Intra-speaker vowel distributions along the F1 and F2 axes are not necessarily normal; in fact, they can be multi-modal.

2) Distributions on the outside edges of multi-modal distributions are loci for stylization.

3) Vowel classes most likely to vary multi-modally are also those carrying sociolinguistic salience. Specifically, vowel classes implicated in sound changes in progress are more likely than others to be resources for extreme vocalic stylization.

Through an in-depth case study of the vowel system of one speaker, this analysis confirms all three hypotheses in turn, suggesting that multi-modal distributions can be a quantitative means for identifying stylized tokens. While the particular dimensions utilized here are the acoustic axes F1 and F2 for vocalic data, theoretically the multi-modal picture is one that can represent any quantitative linguistic feature used for stylistic purposes, including consonantal variation, voice quality, and rhythm to name a few.

1.1 Social meaning in vowel variation in the acoustic space
Like many linguistic features, vowel tokens can be carriers of social meaning. It is not news to sociolinguists that differences in the placement of vowels in the acoustic space—in terms of the acoustic dimensions F1 (relative vowel height) and F2 (relative vowel frontness)—can be important social signifiers on a community-wide level. Indeed, Labov (1963) established this in one of the first sociolinguistic studies ever conducted. In his classic work on the speech of Martha’s Vineyard residents, Labov found that the relative height, or F1, of the nucleus of the /ay/ (BITE) and /aw/ (BOUT) vowels reliably indexed a speaker’s orientation to the local culture and industries of Martha’s Vineyard.

Similar sets of findings have been replicated in studies of smaller speech communities and still smaller communities of practice. Eckert’s (2000) study of the “jocks” and “burnouts” of “Belten High” found significant differences in the advancement of the Northern Cities Vowel Shift between the two social groups. The burnouts were more advanced in the most innovative components of the shift: exhibiting farther backing (lower F2) of BET and BUT and raising (lower F1) of the nucleus of BITE. Eckert calls these “urban variables” because they link the burnout identity to the urban context with which the burnouts associate themselves.

Getting smaller in scope, Mallinson and Childs (2007) find significant differences in the relative frontness (F2) of the back vowels BOAT and BOOT for two groups of African Americans living in a small Appalachian town. The “Porch Sitters” and the
“Church Ladies” are two communities of practice, each comprised of four women. Mallinson and Childs describe in great detail the differing social and semiotic practices of these two groups—women who actually share kinship ties across groups—and show how their linguistic features reflect these differences. In terms of their vowels, the Church Ladies produce significantly fronter (higher F2) BOAT and BOOT vowels than do the Porch Sitters. The authors attribute these patterns to different affiliations with regional (Appalachian English) versus ethnic (African American English) identities. The Church Ladies, with their fronted Appalachian-English-style BOAT and BOOT vowels, align themselves with a model of a nice southern white lady, while the Porch Sitters align themselves with supra-regional African American language and identity. These alignments are evident not only in linguistic patterns, but also in terms of personal style (dress, hair) and manner.

Finally, there are studies that look at stylistic vowel variation in the speech of individuals (i.e., intra-speaker variation). Schilling-Estes (1998) describes the performative register of Rex O’Neal, an iconic resident of Ocracoke Island, North Carolina. In his performative register, Rex O’Neal significantly raises (lower F1) and backs (lower F2) the nucleus of BIDE and BYE tokens (so they sound like BOID and BOY, respectively), revealing a strong association for this vowel with local island identity and Rex’s agency in portraying that identity.

In a series of studies using ethnographic data collected in two elementary schools in Northern California, very different from each other in terms of the ethnic and socioeconomic make up of their student populations, Eckert (1996; 2010; 2011) finds several examples of preadolescent girls using vowel variation for affective and stylistic effect. For example, in the ethnically diverse, relatively low-income school, she observes girls to be backing (not raising) BAN (Eckert 1996). For white speakers of California English, this vowel raises/tenses in this context. Eckert argues that these girls are backing instead of raising this feature as a stylistic and ethnic marker in discourse. In the whiter, wealthier school, Eckert finds similar stylistic innovation in the vowel space. Specifically, she notes how girls often push tokens of BOT and BITE back (low F2) and up (low F1) in the vowel space for negative affect (Eckert 2010). Additionally, she observes that “trendy” speakers, those keenly asserting themselves in the heterosexual marketplace (Eckert 2011), produce very fronted (higher F2) BOAT tokens in particularized lexical and discourse contexts.

In another study, Podesva (2011a) analyzes the speech of one speaker, Regan, to show that his vowel space (particularly for vowels implicated in the California Vowel Shift) changes significantly as he changes contexts. For example, when Regan adopts a ‘partier’ persona, his BOAT and BOOT vowels are fronter (F2), his BAN vowels are higher (F1) and his BAT tokens are backer (F2) than they are for other contextually bound personae. This, Podesva argues, is evidence for how regional identity (exemplified in Regan’s California vowels) interacts crucially with social identities through the media of different stylistic performances.
And these are just a few representative studies that show the link between vowel variation and social and stylistic meaning, all of them basing their findings on changes in the F1 and F2 acoustic dimensions. These studies substantiate this paper’s assumption that variation along the F1 and F2 axes can be exploited for stylistic purposes on both a community-wide and intra-speaker level.

1.2 Stylization as extreme-taking

After reviewing various studies that analyze stylistic variation, one unifying fact among them is that stylization is special. That is, it often extends beyond a speaker’s “normal” style for a particular context. Coupland (2007) identifies *stylization* as the phenomenon of “…projecting personas, identities and genres other than those…current in the speech event…” (p. 154), which suggests that when a speaker is stylizing, she is pushing the limits of what is the typical for her at a particular moment. Linguistically speaking, these limits may be the limits of her vowel space (specific to her dialect), pitch envelope, rhythmic patterns, voice quality, morphosyntactic features, etc., but in order to linguistically embody the various personas, identities, and genres that Coupland refers to, she must create something either totally different from, or at least sufficiently different from, what her range may be when she is not stylizing.

One assumption being made here is that speakers do indeed have a stylistic norm particular to them in a given context, one from which they can deviate for extreme stylistic purposes. This norm is itself a style – but it is the medium through which a speaker presents her most consistent self to whatever degree it exists in a particular interactional context. A speaker’s various contextually-specific stylistic norms and their respective extreme stylizations are both grounded in a speaker’s native grammar, which is a product of how a speaker is situated in the broader language systems and social fabric of her world.

This stylistic norm should not be confused with the *vernacular* in the Labovian sense. Labov’s (1972) *vernacular principle* holds that “the style which is more regular in its structure and in its relation to the evolution of the language is the vernacular, in which the minimum attention is paid to speech” (p. 112). As Eckert (2003), Coupland (2003), and Bucholtz (2003) assert, the variationist’s quest for the vernacular, assumed to be the locus of *authentic or real* language, is laden with problematic ideologies and misses a huge picture of real and potential variation, as well as a whole lot of social meaning as a consequence. To get the full picture, they argue, analysts must consider the extra stuff as well. Stylistic practice is a huge part of this missing picture, one that should be embraced, not ignored, by language scientists.

In concordance with these views, I likewise attest that a stylistic norm has little to do with whether a speaker is paying attention to speech. Rather, it is simply a characterization of the overarching style a speaker is utilizing for a given interaction or part of that interaction. And, I argue that it is this stylistic norm which is the jumping-off point for extreme stylistic variation. As Bucholtz (2003) says, “it is only
when we are surprised out of our assumptions that we can truly appreciate the creative and innovative sociocultural work that social actors regularly accomplish with language” (p. 407). Stylization is one way in which these linguistic surprises are manifest.

1.3 Stylization via outlying tokens
But, how are they manifest in real linguistic data? Following this notion that stylized tokens are moments of linguistic surprise, one branch of thought theorizes that stylistic stance-taking occurs in the outermost extremes of a normal repertoire. In linguistic terms, this can happen in the production of outlying tokens. What a particular outlier is relative to a speaker’s norm has been variably defined for different types of linguistic features. For Kiesling (2012), utilizing a discourse analytic approach, outliers are defined as metapragmatic, performative tokens of interaction. They are outliers in the typical discourse frame because they are meta, in that they are specifically produced for the speaker to take an explicit stylistic stance, and for the listener to notice. Because these tokens are metapragmatically acknowledged in the discourse, they are sufficiently different from the (non-meta) discourse norm as to warrant the name outlier, though as they are not quantitative they cannot adhere to any statistical parameter. Regardless, these tokens, Kiesling argues, do special stylistic work.

In Podesva (2011b), we see the outlier analytical tool applied to a quantitative feature of intonation, namely phrase-final intonational contours in declarative contexts. While differentiating between those contours that rose, fell, or remained level (in terms of f0), Podesva identifies outliers as those phrase-final tokens whose f0 values lay outside two standard deviations from the mean (a standard statistical delineation for outliers, also called the ‘2α’ method, cf. Leys, Ley, Klein, Bernard and Licata 2013). As he says, “If an axis of phonetic variation indexes a particular social meaning, then outliers on that axis can be understood as the strongest indicators of meaning” (p. 254). Podesva argues that his speaker’s acoustically extreme falling contours (manifest as f0 outliers) are especially perceptually salient and hence serve as particularly stylistic tokens.

The outlier analytical tool for stylistic analysis has also been applied to vowel data. Van Hofwegen (2013) identified outlying vowel tokens along the F1 and F2 acoustic axes for a prototypical jock and a burnout from Eckert’s (1989; 2000) ethnography of 1980’s-era “Belten High” in the Detroit suburbs. After extracting all vowel tokens from the entire sociolinguistic interviews, Van Hofwegen (like Podesva 2011b) identified outlying tokens to be those found beyond two standard deviations from the mean in either F1 or F2 dimensions. Overall, she found the burnout to produce more outlying tokens than the jock, largely occurring in words iconic for the burnout identity: words like cigarette, pot (marijuana), john (bathroom), and jock (in a sarcastic reference to the other social group in school). The higher proportion of outlying tokens for the burnout is consistent with a picture of her utilizing more linguistic and stylistic extremes generally (Eckert 2000). Moreover, the burnout’s...
outlying vowel tokens were most often found for vowel classes implicated in the prevailing Northern Cities Vowel Shift (NCS). That is, vowels currently undergoing change in the broader (regional) linguistic system were also those used by the burnout speaker for stylistic variation.

Table 1 provides a breakdown for outlying tokens versus non-outlying tokens in the burnout’s speech from Van Hofwegen (2013). Notable are two main findings. First, data for most of the vowel classes indicate a percentage of outlying tokens that is larger than the expected for a normal distribution. In a normal distribution, tokens beyond two standard deviations from the mean should comprise roughly five percent of the distribution. But, for many of the burnout’s vowel classes, we see outlying proportions that are above five percent. This is suggestive of a pattern whereby the speaker’s vowel distributions are not entirely normal. Second, the vowel classes with the highest percentage of outlying tokens are those classes implicated in the Northern Cities Vowel Shift—the exception being BOAT. This finding is suggestive of a link between stylistic variation and linguistic change.

The outlier analytical framework appears to have real utility for assessing stylization for a variety of linguistic features. However, as Ley et al. (2013) attest, the two standard deviation method for identifying outliers may at times be problematic, primarily because the baseline upon which an outlier is determined (the mean) has a value that is largely biased by the presence of outliers. They advise a different method for outlier detection—one that is based on absolute deviation and from the median, not the mean. However, this is just one of many statistical methods used for gauging outliers, including Chauvenet’s criterion, Grubb’s test for outliers (used by Labov, Baranowski, and Dinkin 2010), and Pierce’s criterion. A theory of stylization based on the identification of statistical outliers may find any of these approaches useful.
However, it’s important to remember the theoretical assumptions which underlie the outlier method, as well as to acknowledge that the particularities of the mathematical formulas used or the cut-off thresholds for what counts as “extreme” enough are still up to the discretion of the analyst. Hence, regardless of what outlier identification method an analyst chooses, there still remains the possibility that the outlier-as-stylization model (by the statistical definition) is not adequate for characterizing the true breadth of a speaker’s tokens of stylization. Depending on the goals of a study, this may or may not be a problem.

In fact, one anecdotal observation arising from the Van Hofwegen (2013) study was that there seemed to be more perceptually stylistic vowel tokens in the data than just the strictly-calculated outlying ones. In their study on the same jock/burnout speech data, Eckert, Lee, and Venkatesh (2013) decided the $2\alpha$ outlier threshold in Van Hofwegen (2013) actually underrepresented the number of stylistic tokens in the data. So, they choose a $1.5\alpha$ threshold and called the $1.5-2\alpha$ additions “extreme” vowel tokens and the $2\alpha+$ tokens were “outliers.” Regardless of whether we deem some tokens “extreme” and some tokens “outliers,” and whether we choose a $1.5\alpha$ or $2\alpha$ threshold, the main take-away from this example is that it can be difficult to find the most appropriate principled, quantitative approach for identifying stylization.

This discussion is not in any way intended to discredit the methods of others. Instead, it offers an alternative: if there is good reason to hypothesize that a particular speaker exploits certain features for stylistic effect, then there need not be predetermined statistical thresholds for what would determine these stylized tokens. Instead, all the available data should be considered, to get a sense for when and how stylistic extremes emerge, resulting in an inductive, not deductive, approach.

1.4 Vocalic extremes and sound change

As mentioned above, there has been some work suggesting that stylization and sound change may go hand in hand (Eckert 2000; Podesva 2011a; Van Hofwegen 2013; Eckert et al. 2013). If we assume an extreme-taking basis for stylization, is there additional evidence outside of third-wave studies to suggest that extreme-taking is linked to sound change?

Indeed there is. Labov et al. (2010) discuss the “problem” of outlying vowel tokens in intra-speaker vowel data, and examine the extent to which these tokens are perceptually problematic (i.e., unparsable or ambiguous) for listeners in a speech community undergoing sound change. In a perception experiment whereby listeners were asked to rate various productions of the word ‘bad’ by a Philadelphia-area female speaker—productions that were synthesized to vary along a diagonal axis from fronted/raised to backed/lowered—they found that listeners were largely not affected by statistical outliers (extremes) along these dimensions and were able to parse and rate the tokens consistently. In fact, they found that a small sub-sample of participants actually interpreted outlying tokens to be perceptual targets for the
BAT vowel. They speculate that it may be these very individuals, those who readily assign extreme tokens as targets for exemplars in their phonemic perceptual systems, who are the driving forces for sound change in speech communities.

In sum, there is much to suggest that vowels which are in flux for structural reasons also provide utility for stylistic purposes. The Labov et al. (2010) data suggest that some listeners are especially prone to attending to outlying tokens implicated in sound change. My proposal is in accordance with theirs, but with an added caveat: extreme tokens are not only instrumental in sound change, but it is because they are implicated in sound change that they carry stylistic potential.

2 The Study

When we think about examining intra-speaker vowel variation, what becomes immediately clear is that we cannot employ traditional dialectological methods. There have been no shortage of vowel studies conducted over the last decades (e.g., Labov 1991; Thomas 2001, among many, many others) aimed not only at describing major dialect features, but also assessing various sound changes through the lenses of apparent and (more rarely) real time. These studies typically employ a fairly controlled sampling method for vowel tokens, and are also often limited to a predetermined number of tokens per vowel class, an approach recommended by experienced sociophonetic practitioners (cf. Di Paolo, Yaeger-Dror, and Wassink 2010; Baranowski 2013).

However, as intentioned, the result of this approach is a very circumscribed picture of a speaker's variation in the acoustic space. In fact, when there is extreme variation evidenced in controlled studies, it is often viewed as problematic (Labov et al. 2010). With a limited number of tokens, it is impossible to tell whether a particular outlying token is an error/anomaly or whether it is itself indicative of systematic variation within the individual. Figure 1 shows the vowel space of AR, a 19-year-old white female speaker from Redding, California, based on a highly controlled approach consistent with that used by dialectologists. The plot shows mean Hertz values for all vowel classes, determined from midpoint formant measurements from twenty-five vowel tokens for each vowel class. Control parameters were duration (i.e., no tokens with duration less than 75 milliseconds), preceding environment (i.e., no tokens following vowels, glides or /r/), following environment (i.e., no tokens preceding vowels, glides, liquids, except /l/ where indicate), content words (i.e., no function words), lemma (i.e., no more than two tokens of the same lemma, unless in cases of paucity), and observer effect (i.e., majority of tokens extracted after fifteen minutes of the sociolinguistic interview).

Looking at this picture from a dialectologist’s perspective, we can make some observations. Primarily, we can note that this speaker participates in the California Vowel Shift (CVS), which is not surprising given general findings for speakers from Redding (Podesva, Calder, Chen, D’Onofrio, Flores-Bayer, Kim and Van Hofwegen
The CVS, schematized in Figure 2, is characterized by a merger between BOT and BOUGHT (Labov, Ash & Boberg 2006), with back vowel (BOOT/BOAT/BOOK) fronting (Hinton, Bremner, Corcoran, Learner, Luthin, Moonwomon, and van Clay 1987, Hall-Lew 2009) and front lax vowel (BIT/BET/BAT) lowering and centralizing (Hagiwara 1997, Eckert 2008, Kennedy and Grama 2012). As the vowel plot indicates, AR’s COT and CAUGHT are merged, her BOOT and BOAT are quite fronted relative to their pre-/l/ counterparts, and her front lax vowels are quite low in the space, with BAT playing the role of the lowest anchor vowel.

![Figure 1](image1.png)

**Figure 1.** Vowel means plot for AR, female from Redding, CA, age 19. Controlled vowel sample.

![Figure 2](image2.png)

**Figure 2.** The California vowel shift

Presenting vowel spaces in terms of vowel means is the norm for dialectologists (cf. Thomas 2001), but as Figure 3 shows, a quick glimpse into the distributional
patterns of even a controlled sample shows us that there is much more to the variation than meets the eye. Moreover, the extent of the variation in the space can lead one to wonder how representative these mean values are in the first place.

![Mean vowel formant values non-normalized](image)

**Figure 3. Vowel means+2α plot for AR. Controlled vowel sample.**

Figure 3 shows the same vowel plot as above, this time with bars indicating the distribution two standard deviations (2α) from the mean. Even from this controlled sample, we can see that this speaker exhibits wide distributional patterns, primarily in the F1 dimension for front and low vowels, and in the F2 dimension for high and back vowels. With such broad distributions for such tightly controlled data, it is difficult to determine what their breadth is indicative of a speaker’s idiolectal patterns or just a particularly spread sample of tokens. After all, a sample of 25 tokens out of what can be as many as 1000 tokens/vowel class in a single sociolinguistic interview is a small and potentially non-representative sample.

In order to understand the patterning of intra-speaker variation, we need an approach that considers all the tokens, not only in the interests of being as representative as possible, but also because when we have sufficient token counts, we can use various goodness-of-fit tests to determine significant statistical characteristics of the distribution set as well.

### 2.1 The sampling method and measurements

As mentioned above, this approach gathers *every single vowel token* available per vowel class for a given speaker. While just five years ago this task would have been arduously time-consuming and impractical, automated tools now make this not only feasible but also relatively painless. For this particular analysis, one sociolinguistic interview approximately 75 minutes long sufficed. But first, it had to be orthographically transcribed and time-aligned. Then, word and phone segment boundaries for all segments in the interview were determined using the Penn forced
alignment software package FAVE (Rosenfelder, Fruehwald, Evanini, & Yuan, 2011). Finally, a script developed for Praat software (Boersma and Weenink 2012) was used to extract midpoint measurements for all sonorant segments in the interview, including midpoint F1 (Hz), midpoint F2 (Hz), midpoint F3 (Hz), overall intensity (dB), f0 (Hz), H1-H2 (dB). Additionally each segment’s duration (milliseconds), stress (primary, secondary, unstressed) previous and following segment, and word (in which the segment was uttered) were also noted.

Because the second component of the analysis (described below in 2.4) involved a band pass filtered measure of intensity called spectral emphasis (Cole, Mo, Hasegawa, Johnson 2010; Heldner 2003; Tamborini 2005), the recording was also band pass filtered at three different bandwidths: 0-500 Hz, 501-2000 Hz, and 2001-4000 Hz. Intensity measurements for each of these bandwidths at the midpoints of all sonorant segments were also gathered and added to the data set.

Of particular focus for this analysis were the vowels BIT, BET, BAT, and the back vowels TOO/BOOT and TOE/BOAT. These vowel classes were selected not only because they were vowel classes implicated in the CVS, but also because they were monophthongal in nature. Given their relatively minimal dynamism, the automated nature of segment boundary identification, and the reality of alignment error (Yuan 2009), monophthongs ensure more accurate automated measurements. For example, if, due to alignment error, a measurement is not taken exactly at a segment’s midpoint, the monophthongal nature of vowel ensures that most of the vowel duration is in a portion of steady state. Altogether, roughly 12,000 vowels were extracted and measured, with about 3000 (N = 2,922) considered for this in-depth analysis of distributions.

2.2 The speaker

The nature of this analysis is that it presents a model for gauging intra-speaker variation and potential stylization. Thus, in theory, naturalistic speech data of sufficient length from any speaker would suffice. There need be no pre-defined guidelines for selecting a particular speaker; any will do. However, in order to make sense of intra-speaker variation, some knowledge of the speaker and of her speech community is necessary.

This paper explores all the tokens in the vowel envelope for AR, a young (age 19), white, female speaker from Redding, California, whose (albeit controlled) vowel space is shown in Figure 1 and Figure 3. The primary data on which this analysis was conducted was a sociolinguistic interview from the Voices of California corpus, conducted by a male fieldworker in 2011. The Voices of California project is a large-scale dialectology project out of Stanford University, the goal of which is to document and explore the range of variation found across the state, which—save for its large, liberal, coastal metropolises—has been under-studied from a dialectology perspective.
Redding is a mid-sized town (population ~90,000 as of the 2010 US Census) located in Shasta County, on the northernmost edge of California’s Central Valley. For many historical, cultural, and geographical reasons, Redding inhabitants view themselves as quite distinct from, and in fact often oppositional to, the big cities of Sacramento, San Francisco and Los Angeles to the south. However, within Redding, there is a salient distinction between speakers who orient themselves towards rural/country livelihoods and those who orient toward the town proper. Previous work on the community (Podesva et al. 2013; Podesva and Van Hofwegen 2013) has found these country versus town orientations to significantly pattern the linguistic variation. Specifically, town-oriented individuals are more advanced in some components of the CVS, namely COT/CAUGHT lowering and backing/lowering of BAT (Podesva et al. 2013).

As a young town-oriented female, AR has moderately-shifted CVS vowels. In fact, as a member of the larger sample used in Podesva et al.’s (2013) study, AR proved to be quite shifted with respect to her front lax vowels, much more so than her country-oriented counterparts. Her TOO/BOOT vowels are also shifted substantially forward, while her TOE/BOAT classes are also fronted (though not as much as TOO/BOOT) with respect to BOWL. Additionally, given her role as a young female, we might expect AR to not only be advanced in terms of changes from below (i.e., the CVS), but also more innovative generally in her speech (Labov 1990). In sum, there is reason to believe AR to have some interesting patterns in her speech, so she makes as fine a subject as any for a case study of this kind.

2.3 The analysis

Figure 4 shows AR’s vowel space with means for all stressed tokens (i.e., primary and secondary stress) for the vowels BIT, BET, BAT, TOE/BOAT, and TOO/BOOT. Dashed lines indicate the extent of the distributions 2α away from the mean. These tokens are not controlled, which means their distribution clouds include both content and function words, and all preceding/following environments (save for where we’ve isolated post-coronal and pre-/l/ tokens as indicated). Unstressed tokens were left out of the analysis, as the CMU pronunciation dictionary quite irregularly assigns vowel classes to unstressed segments. Moreover, some vowels like BIT have a disproportionately high number of unstressed tokens, while other vowels like BAT have none. Unstressed tokens were, however, included in determining the measure for spectral emphasis, which is discussed below in Section 2.4. Not under analysis, but to provide reference for comparison for the high back vowels are BOWL and POOL. Anchor vowels also shown are BEET (front) and HULL and BULL (back).

Keeping in mind the first hypothesis presented in this paper (see below), the analysis proceeded by studying the distributions for F1 and F2 for each of the vowels under study.

**Hypothesis 1**: Intra-speaker vowel distributions along the F1 and F2 axes are not necessarily normal (in fact, they can be multi-modal).
As discussed in Section 1.1, there is ample precedent for considering vowel variation in terms of unidimensional axes. As we can see from this plot, for some of the vowels the bulk of the variation is clearly unidimensional (e.g., F1 for BAT), while for others the variation seems to proceed in both directions (e.g., F2 and F1 for BIT). A more in-depth examination of the distributions will uncover the patterning of tokens making up these spreads.

![Mean vowel formant values non-normalized](image)

**Figure 4. Vowel means+2α plot for AR. All stressed vowels under analysis.**

Distributional analyses were conducted of midpoint measurements for F1 and F2 for all tokens of the vowels under study with the help of JMP statistical software (JMP 2012). JMP’s *Distribution Platform* enables the user to plot a data set against an axis and determine the best distribution from a variety of options (normal, bi-modal, tri-modal, lognormal, Johnson, Poisson, Gamma Poisson, etc.). The best distribution is determined primarily via the Akaike Information Criterion (AICc) value (Akaike 1974), which is a measure of the relative goodness-of-fit of a particular model for a given set of data, determined via the model’s log-likelihood and the number of parameters inherent in the model. The AICc additionally incorporates a given sample’s size. The particular parameters relevant for a normal distribution are μ (mean) and σ (standard deviation). The *Continuous Fit* application in JMP’s *Distribution Platform* fits several types of distributions to the parameters particular to them and compares AICc values for each distribution type. The distribution type with the lowest AICc value is the model with the best fit of the data.

If, as was often the case, an AICc value indicated the best fit for a distribution was non-normal (specifically, bi- or tri-modal), then JMP’s *Distribution Platform* can
helpfully provide the resulting μ (mean) and σ (standard deviation) values as well as the curve for each of the modes of the best-fit distribution.

If a distribution was determined to be multi-modal, then the analysis proceeded to a secondary phase. In this phase, the μ and σ values for each mode were used to isolate tokens particular to that mode for post-hoc descriptive and logistic regression analysis. That is, if a distribution was determined to be multi-modal, then tokens falling one standard deviation above and below the mean for each mode were extracted into a separate data set in which these mode-specific tokens were categorized as high/mid/low (in the F1 dimension) or front/center/back (in the F2 dimension). Logistic regression was then conducted on these categorical data to see which factors significantly predicted the tokens’ respective placements in the various modes. Predictor variables incorporated into the logistic regression included: spectral emphasis, duration, f0, H1-H2. These predictors were selected as such because they are theorized here to be associated with stylization (see below).

2.4 Orthogonal parameters for assessing stylized modes
For a multi-modal distribution, in order to say that one or more of the modes is a locus for stylization, we need more tools at our disposal. Here is where the second hypothesis for the analysis comes in:

**Hypothesis 2:** Distributions on the outside edges of multi-modal distributions are loci for stylization.

The distribution analysis alone can tell us that a given distribution is not normal and that the modes are sufficiently different from each other (i.e., little or no overlap). While there is some reason to assert that extra modes on the extreme edges of a distribution are clusters of stylized tokens—simply taking the stylization-as-extreme point of view—extreme modes may still be artifacts of inherent linguistic factors not related to stylization. Thus, I argue that additional measures are needed to test whether these extreme modes contain clusters of stylized tokens.

The orthogonal parameters used for determining relative stylization are prosodic prominence and creak. Prosodic prominence is a measure used by Mo (2008) and Cole et al. (2010) which contributes to perceived prominence for vowels at the phrase level. Different segmental characteristics associated with prosodic prominence have been tested in several languages, but Tamborini (2005) finds that duration, spectral emphasis, and to a much lesser extent, f0 contribute to prosodic prominence in English. Spectral emphasis is a measure of band-pass filtered intensity (dB). While relative overall intensity has long been associated with phrasally-prominent syllables across many languages (Heldner 2003), intensity measures for certain mid-level bandwidths have been found to more significantly differentiate phrasally-prominent syllables than others.

These bandwidths can be speaker specific. So, in order to determine which bandwidth predicted the most phrasally-prominent syllables for AR, her recording
was band-pass filtered according to the following bandwidths: 0-500 Hz, 501-2000 Hz, and 2001-4000 Hz (following Mo 2008), leading to intensity measurements for each of these bandwidths 0-500dB, 501-2000dB, and 2001-4000dB, alongside overall intensity which had already been measured from the non-bandpass-filtered original recording.

At this point, several regression analyses were conducted, one for each of these intensity measures, and also incorporating the other factors shown to contribute to prosodic prominence (duration and f0), to see which collection of factors most significantly predicted stress for all of AR’s vowels. Consistent with Tamborini (2005), duration quite significantly predicted AR’s stress, as did also f0 (though marginally). The intensity measure that most significantly predicted stress was 0-500dB (henceforth to be called spectral emphasis). In fact, the overall best model for predicting stress for all vowels was one that incorporated all three of these factors: duration, spectral emphasis, and f0. Under the assumption that tokens utilized for stylization will also be more perceptually prosodically prominent, these factors were used as predictors in the logistic regression described above.

But, there are other ways to be perceptually prominent than simply duration, intensity, and f0. Another feature that has been shown to have stylistic salience in American English is creak (cf. Podesva 2007; Mendoza-Denton 2011). A quick listen to AR’s speech is enough to convince that she is a prolific user of this feature, and likely for stylistic effect. Yet, given its very acoustic nature (relatively low intensity, relatively low f0), creak is likely not to come out as prosodically prominent given the measure outlined above. So, creak was considered separately alongside prosodic prominence, both as orthogonal predictors of stylization. For the purposes of this analysis, creak was operationalized as H1-H2 (Thomas 2011), where creaky voice has low (negative) values.

3 Results
A look into the distributions of the vowels under analysis reveals a complex series of patterns, but with one unifying characteristic: the distributions of vowel tokens along the F1 (for front lax vowels) and F2 (for high back vowels) axes are far from normal in any statistical sense. Goodness-of-fit tests confirmed bi-modal or tri-modal distributions for components of most of the CVS-implicated vowel classes, indicating smaller distributions of tokens outside of the “normal” distribution, even after phonetic and word frequency factors were considered. Secondary regression analyses on stylistic predictors of these various distributions revealed complex interplays of stylization and extreme acoustic variation, which will be discussed in greater detail below.

3.1 BAT
As we can see from Figure 4, most of the breadth of variation for BAT can be found in the F1 dimension, so it seems plausible for the BAT analysis to focus on this dimension. However, to justify this decision, both F1 and F2 dimensions were
plotted and subjected to distribution analysis. The F1 and F2 distributions for all 619 BAT tokens are shown in Figure 5 and Figure 6, respectively. In these and all following figures of distributions, yellow lines indicate what would be a normal fit line. In some cases, the normal distribution is the best fit, but if additional modes better fit the data, they are also indicated with fit lines: green lines for bi-modal distributions and blue for tri-modal distributions.

Thus, in Figure 5, we see that the best fit for the distribution is a tri-modal pattern (blue line) with modal means of 659 Hz (σ = 85), 870 Hz (σ = 78), and 1066 Hz (σ = 51), respectively. In Figure 6, we see the best fit for the data is a bi-modal distribution (green line), with modal means of 1665 Hz (σ = 110) and 1961 Hz (σ = 116) respectively. Because the modal means plus/minus one standard deviation extend to distinct clusters of tokens (i.e., no overlap), it was possible to extract discrete clusters of tokens centering around the respective means of the various modes, which were subsequently labeled 'high,' ‘mid,’ or ‘low’ (F1) and ‘front’ or ‘back’ (F2) according to their corresponding place in the vowel space. The extraction of tokens from these discrete clusters enabled a closer descriptive look at what tokens make up these distributions as well as the logistic regression discussed in Section 2.3.

Considering Hypothesis 1, it is encouraging to see that neither distribution turns out to fit a normal curve. But, a closer look at the F2 spread reveals that the vast majority of tokens in the front cluster are function words such as at, that, has, have, etc. Even though these vowels are annotated as stressed by the aligner due to their
phrase position, it turns out they are still comparatively reduced in terms of duration and intensity. On the theoretical premise that phonetically reduced tokens are not likely to be stylized (because they cannot be prosodically prominent), it seems clear that tokens in the front distribution are likely not candidates used by AR for stylization. But, if we take the front reduced distribution out of the equation, we are left with a normal distribution centering around a mean that makes sense given AR's idiolect. Hence, it seems clear that the bi-modal character of the F2 distribution for BAT can be attributed to structural phonetic factors, not stylization.

The tri-modal spread for BAT’s F1, though, tells a different story. What kinds of tokens make up these three humps? Table 2 shows how the extracted F1 distributions break down in terms of function/content words, duration, and spectral emphasis. We can quickly see that the high distribution is different from the other two. Out of the 132 tokens in the high distribution, for example, 114 of them are non-content words. Duration-wise, the mean for tokens in this distribution (72 milliseconds) is 60% of the means for the other distributions. In terms of spectral emphasis, the high/front distribution has a mean of 51 dB, roughly 3-5 dB lower than the means for the other distributions.

Table 2. Descriptive characteristics for extracted tokens from the three F1 distributions

<table>
<thead>
<tr>
<th>F1 Distribution</th>
<th>N</th>
<th>% function words</th>
<th>Mean Spectral emphasis (dB)</th>
<th>Mean Duration (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>high</td>
<td>133</td>
<td>85.7</td>
<td>52.3</td>
<td>72.48</td>
</tr>
<tr>
<td>mid</td>
<td>286</td>
<td>57.3</td>
<td>56.0</td>
<td>123.63</td>
</tr>
<tr>
<td>low</td>
<td>53</td>
<td>24.5</td>
<td>58.2</td>
<td>194.26</td>
</tr>
</tbody>
</table>

As we can see from Table 2, the majority of the high distribution is comprised of reduced function tokens, which seem to largely correspond to the front F2
distribution discussed above. However, one important observation is that not all function words are found in the top distribution, suggesting that not all function words are phonetically reduced and are hence potential candidates for stylization.

In order to assess the extent to which F1 distributions were predicted by factors of prosodic prominence, logistic regression was conducted. Predictor variables were spectral emphasis, f0, logDuration. The best-fit model for the data included an interaction term spectral emphasis*logDuration, the precise relationship of which is presented in graphic form in Figure 7. It should also be noted here that H1-H2 (creak) was included in earlier regression runs but never emerged as a significant predictor for BAT F1 distribution. Table 3 summarizes the logistic regression results.

Table 3. Logistic regression results for prosodic prominence factors predicting BAT F1 distribution

<table>
<thead>
<tr>
<th>Term</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>Chi Square</th>
<th>Prob &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept [high]</td>
<td>-0.705398</td>
<td>1.2312568</td>
<td>0.33</td>
<td>0.5667</td>
</tr>
<tr>
<td>Intercept [low]</td>
<td>-0.101679</td>
<td>1.2310335</td>
<td>0.01</td>
<td>0.9342</td>
</tr>
<tr>
<td>Spectral emphasis</td>
<td>-0.06516</td>
<td>0.0199805</td>
<td>10.64</td>
<td>0.0011**</td>
</tr>
<tr>
<td>f0</td>
<td>0.00440255</td>
<td>0.0019137</td>
<td>5.29</td>
<td>0.0214*</td>
</tr>
<tr>
<td>logDuration</td>
<td>-0.987261</td>
<td>0.1855498</td>
<td>28.31</td>
<td>&lt; 0.0001***</td>
</tr>
<tr>
<td>Spectral emphasis * logDuration</td>
<td>0.11107708</td>
<td>0.0284555</td>
<td>15.24</td>
<td>&lt; 0.0001***</td>
</tr>
</tbody>
</table>

As Table 3 displays, all measures of prosodic prominence emerged as significant predictors of F1 distribution location, the most dramatic being spectral emphasis and logDuration. As it turns out, the furthest down tokens in the vowel space (the low F1 distribution) have significantly higher spectral emphasis, f0, and logDuration, than do the other distributions, meaning the lowest distribution is the most prosodically prominent.

And, this pattern is intensified further when we look more closely at the low distribution, for it is here that the relationship between spectral emphasis and logDuration is most intense. See Figure 7 for a graphic representation of this interaction effect. In this figure we see that the linear relationship between spectral emphasis and log duration is most dramatic in the lowest distribution, suggesting a much more dramatic effect of prosodic prominence for the tokens in this distribution. This, I argue, makes this lower distribution a locus of stylization for AR.
3.2 BET

Plots for the F1 and F2 distributions for the 446 tokens of stressed BET are presented in Figure 8 and Figure 9. Again, as with BAT, the distributions are non-normal. The F1 distribution is bi-modal, with means of 747 (σ = 91) and 798 (σ = 19), respectively. The F2 distribution is also bi-modal, with means of 1291 (σ =
138) and 1845 (σ = 156), respectively. In contrast to the distribution spreads for the BAT vowel, the distributions for BET overlap quite a bit. Thus, it was impossible to extract mean plus/minus one standard deviation clusters of tokens for each mode and conduct logistic regression. However, a descriptive look at the tokens which center around the means reveals some notable patterns.

The backed, smaller distribution for F2 are comprised entirely of pre-/l/ tokens. The vast majority of these tokens are words like help, well, twelve, etc. In fact, when we remove the pre-/l/ tokens from the picture, we are left once again with an F2 distribution that meets the statistical definition of normalcy. Thus, as with BAT, we do not see multi-modal distributions on the F2 dimension outside of predictable phonetic patterning.

When we look to BET’s F1 distribution, though (as with BAT), the pattern is still suggestive of stylistic patterning. Because discrete clusters from the modes were impossible to extract for logistic regression, it cannot be said definitively that a particular distribution is best predicted by creak or prosodically prominent factors. However, a general trend could be noted via mixed effects linear regression modeling applied to the entire BET sample. Here, the outcome variable considered was BET F1, while predictor variables were spectral emphasis, logDuration, f0 and H1-H2. Regression results are reported in Table 4.
Table 4. Linear mixed effects regression results for prosodic prominence factors predicting BET F1

<table>
<thead>
<tr>
<th>Term</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>Denominator DF</th>
<th>t Ratio</th>
<th>Prob &gt; [t]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>779.85259</td>
<td>46.95437</td>
<td>413.7</td>
<td>16.61</td>
<td>&lt; 0.0001***</td>
</tr>
<tr>
<td>Spectral emphasis</td>
<td>2.357347</td>
<td>0.685911</td>
<td>393.7</td>
<td>3.44</td>
<td>0.0007***</td>
</tr>
<tr>
<td>f0</td>
<td>0.067362</td>
<td>0.057283</td>
<td>403.5</td>
<td>7.49</td>
<td>0.2403</td>
</tr>
<tr>
<td>logDuration</td>
<td>60.204408</td>
<td>8.038251</td>
<td>432.8</td>
<td>1.18</td>
<td>&lt; 0.0001***</td>
</tr>
</tbody>
</table>

As with BAT, H1-H2 (creak) did not turn up as a significant predictor. Additionally, f0 was likewise not significant, although it strengthened the model’s AICc to leave it in the regression. The significant predictors in the model were spectral emphasis and logDuration. As with BAT, the simple pattern was that as F1 increased (i.e., got lower in the vowel space), both logDuration and spectral emphasis also increased. This suggests that lower BET tokens are more prosodically prominent and hence greater candidates for stylization.

3.3 BIT

The patterning for AR’s 759 BIT tokens shares some characteristics with BAT and some with BET. Like BAT, BIT’s F1 distribution is tri-modal in character, with means at 526 (σ = 43), 616 (σ = 38), and 702 (σ = 58), respectively. Additionally, like BAT, the distribution spreads are discrete enough (in terms of one standard deviation) to be extracted into separate modes and subjected to logistic regression. Like BET, the BIT F2 distribution is bi-modal, with a farther back mode containing most of the pre-/l/ tokens. Without the pre-/l/ tokens, the F2 distribution is normal. Figure 10 and Figure 11 depict the distributions along their respective axes.
The tri-modal F1 distribution for BIT lends itself nicely to logistic regression, results for which are presented in Table 5. Here, we see that both the topmost and bottommost distributions are significantly predicted by stylistic factors, but different ones in turn. BIT is the first variable for which creak is a significant predictor. Lower H1-H2 values (i.e., more creak) significantly predict the highest F1 distribution.

<table>
<thead>
<tr>
<th>Term</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>Chi Square</th>
<th>Prob &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept [high]</td>
<td>3.93188819</td>
<td>0.8738241</td>
<td>20.25</td>
<td>&lt; 0.0001***</td>
</tr>
<tr>
<td>Intercept [low]</td>
<td>4.68365416</td>
<td>0.8791591</td>
<td>28.38</td>
<td>&lt; 0.0001***</td>
</tr>
<tr>
<td>Spectral emphasis</td>
<td>-0.0839525</td>
<td>0.0151097</td>
<td>30.87</td>
<td>&lt; 0.0001***</td>
</tr>
<tr>
<td>H1-H2</td>
<td>-0.0312082</td>
<td>0.0091156</td>
<td>11.72</td>
<td>0.0006***</td>
</tr>
<tr>
<td>f0</td>
<td>0.0021474</td>
<td>0.0013755</td>
<td>2.44</td>
<td>0.1185</td>
</tr>
<tr>
<td>logDuration</td>
<td>0.23552492</td>
<td>0.127198</td>
<td>3.43</td>
<td>0.0641</td>
</tr>
</tbody>
</table>
However, spectral emphasis significantly predicts the lowest BIT distribution (see Figure 13). Duration and f0 play no significant role in determining any distributions.

Thus, it appears as though AR may be stylizing differently along the F1 dimension. With significantly lower BIT tokens, she is likely to add more intensity (spectral emphasis); with significantly higher BIT tokens, she is likely to creak.
3.4 TOE/BOAT

As mentioned above, the post-coronal TOE tokens were analyzed separately from BOAT because previous work has shown this class to precede their non-coronal counterparts in fronting-related sound changes. When the 213 BOAT tokens were subjected to distribution analysis, it was found that their distribution was statistically normal for both the F1 and F2 dimensions.

For the 454 TOE tokens, on the other hand, only the F1 axis resulted in a normal distribution. The F2 dimension was tri-modal, with modal means at 1403 (σ = 172), 1687 (σ = 38), and 1920 (σ = 109), respectively. As with the front lax vowels, this finding is consistent with the hypothesis that multi-modal distributions occur in the acoustic dimension undergoing sound change. The F2 distribution for TOE is provided in Figure 14.

![Figure 14. F2 distribution (Hz) of all TOE tokens (axis flipped)](image)

While the modal means plus/minus one standard deviation allowed for no overlap between modes (and thus successful extraction of mode-relevant tokens), the relative paucity of tokens in some of the distributions (e.g., only 21 tokens in the front distribution and only 40 tokens in the center distribution) make the logistic regression findings merely suggestive at best. Accordingly, instead of logistic regression predicting specific distributions, these data were incorporated into linear mixed effects modeling (as was done above for BET), with the same stylistic predictors tested with the other vowels in this analysis.

Here, we see two of the prosodic prominence factors emerging as significant in the model, while creak was not a significant predictor at all. As Table 6 shows, as F2 goes up (i.e., tokens are fronted in the vowel space), spectral emphasis goes up
significantly as well. LogDuration, another significant factor, had a negative relationship with F2. Tokens further back in the vowel space were longer. Though suggestive, these findings resonate a bit with those for BIT F1. That is, we see AR utilizing different prosodically prominent factors at different poles of her vowel distribution. Specifically, for her most extreme front tokens, she adds more intensity (spectral emphasis), while for her most extreme back tokens, she adds duration.

Table 6. Linear mixed effects regression results for prosodic prominence factors predicting TOE F2

| Term               | Estimate | Std. Error | Denominator DF | t Ratio | Prob > |t| |
|--------------------|----------|------------|----------------|---------|---------| | |
| Intercept          | 834.23804| 81.08363   | 439.1          | 10.29   | < 0.001*** |
| Spectral emphasis  | 4.1894924| 1.412125   | 435.5          | 2.97    | 0.0032**  |
| f0                 | -0.181547| 0.148937   | 436.8          | -1.22   | 0.2235   |
| logDuration        | -180.7979| 10.38615   | 434            | -17.41  | < 0.001*** |

3.5 TOO/BOOT

After the flurry of non-normal distributions along the F1 dimension for the front lax vowels, and along the F2 dimension for the TOE class, comparatively, the TOO/BOOT vowel classes are quite normal. In fact, along both dimensions for both TOO and BOOT vowel classes, the distributions are statistically normal, confirmed with Shapiro-Wilk W tests. Thus, no clear patterns can be seen as to whether AR is stylizing in a particular way with these vowel classes.

4 Discussion

Through the in-depth exploration of the distributional patterns of one speaker’s entire repertoire for key CVS-implicated vowels (inasmuch as her repertoire is represented via one sociolinguistic interview), this analysis has conclusive findings for all three of the hypotheses offered in the Introduction. While these findings may be considered suggestive at this point due to the case-study format of this paper, they carry powerful implications for both the studies of stylization and of linguistic change.

First, distributional analyses of the vowels BIT, BET, BAT, and TOE revealed that an assumption of normality is not appropriate for the data. This confirms Hypothesis one, that intra-speaker vowel distributions along the F1 and F2 axes cannot necessarily be assumed to be normal, and in fact are often multi-modal. In some cases, that non-normal patterning can be attributed to normal phonetic factors, as it was for BAT’s highest mode and for BET’s and BIT’s back-est (pre-/l/) modes. Other times, however, there is no apparent structural reason for these additional extreme clusters of tokens. This, I argue, leaves their patterning open to alternative (i.e.,
stylistic) explanations.

In particular, the F1 dimension for BAT, BET and BIT in this study showed patterning consistent with a view of stylization as being an extreme manifestation of a speaker’s norm. For example, all of the non-normal distributions, whether bi-modal or tri-modal showed extra modes on the extreme edges of the overall space for that vowel. And, this stylization is composed of clusters of systematically varying tokens, which crucially number more than just the few outliers posited in previous work (Van Hofwegen 2013). Moreover, logistic regression analysis where possible (F1 for BAT and BIT) and otherwise linear mixed effect analysis (F1 for BET; F2 for TOE) revealed that the most extreme edges of the distribution correlated strongly with factors of either prosodic prominence or creak. Assuming that stylistic tokens carry more prosodic prominence than do their “normal” counterparts, and following Podesva (2007) and others in assigning stylized significance to the use of creak, we can conclude from this that these extra modes are pushed to the extremes in their F1/F2 values because they are being used by the speaker for stylistic purposes. This confirms Hypothesis two, that distributions on the outside edges of multi-modal distributions are loci for extreme stylization.

One phonetic factor not considered in this paper is phrase position. One question that could be raised is whether relative prosodic prominence and its corollary creaky voice, which turn out to be significant for certain extreme distributions of tokens in this study, are merely artifacts of phrase position, not evidence for stylization. While this contention is plausible, and can be explored in further work, I believe there is reason to assume that phrase position and stylization are themselves correlated. Specifically, if we know from English phonology that certain positions in the intonational phrase are likely to correlate with longer duration, higher intensity, higher f0, and/or more creak, it is likely that those same positions will also be places utilized by speakers for stylization.

In their examination of prosodic prominence, Cole et al. (2010) explored word frequency and information structure as predictors for acoustic measures of prosodic prominence. They found that high frequency words and words containing old information are likely to be less prosodically prominent. This, I would argue, does not contradict the notion of stylization presented in this paper. If stylized tokens are “moments of surprise” (Bucholtz 2003), then there is no reason to preclude high frequency words from the stylistic equation. In fact, as this study demonstrated with BAT, it is true that the high-frequency function words in the highest distribution were significantly less prosodically prominent than the predominantly content words in the other distributions. However, not all function words were phonetically reduced. There were many instances of the word ‘that,’ for example, in the interview with an extremely lowered F1, long duration, and high spectral emphasis.

The same goes for new information. There is reason to believe that, given its very novelty, new information would be more likely to be stylized than old. And, because new information is likely to come at the end of phrases, in the same position in
which we are more likely to see creak, it does not seem implausible that we would see a relationship between these two factors for speakers who use creak for stylistic effect.

In terms of the third hypothesis tested in this paper, the analysis confirms that the multi-modal patterns were exhibited largely on the very axes in which linguistic change is on-going. Hypothesis three posits that vowel classes most likely to vary multi-modally are those that carry sociolinguistic salience. As myriad studies have shown, vowels implicated in sound changes carry sociolinguistic salience, either above or below the level of consciousness, which is evidenced in systematically patterned social variation. Thus, it is no surprise that stylistic variation (another factor carrying sociolinguistic salience) would coincide with sound change for vowels that are in flux.

This study looked at vowels that are clearly in flux in California, and for this speaker. In Redding, the town in which AR has spend the entirety of her life, we see differing extents of California shifting for different speakers, based primarily on sex, age, and country or town orientation (Podesva et al 2013). As a young town-oriented female, we expect AR to be in the process of retracting/lowering her front lax vowels to a more significant degree than would other members of her community. Thus, it is notable that AR’s most dramatically multi-modal patterns also occur in the front lax vowels, and along the F1 dimension. That is, for these vowels we see broad multi-modal distributions with extreme extra distributions on the edges that are significant for either prosodic prominence or creak. BAT and BIT showed the most extreme patterning in this respect, with BET exhibiting more suggestive findings.

What may account for the different patterns found for BAT, BET, and BIT in turn? There has been some question in the (albeit limited) literature on the CVS about which process is initiating the shift. Some argue that the shift is a pull-chain stimulated by the COT/CAUGHT merger (Labov, Ash and Boberg 2006). Other work has suggested that it is a push-chain stimulated by the lowering of BIT (Kennedy and Grama 2012). But the question has thus far not been satisfactorily answered. But, a look at the entire range of variation for these front vowels, as was done here for AR, can shed light on the answer to this question.

If we associate extreme variation with shifting vowels (Labov et al. 2010), then not only can we attribute these extremes to stylization, but also to the likelihood that this vowel is moving down in the vowel space—in fact leading the way for the higher lax vowels to follow. If we think about California BAT shifting as a lowering process (not a backing process as it has been described in previous literature), then there is theoretically no limit to how low it can go, save for the anatomical limit of the mouth’s opening. BET and BIT, in contrast, are limited in their lowering due to the proximity of vowels below them. Thus, while AR’s BAT has the luxury of creating stylized clusters at extremely low points in the space, BET cannot do this as dramatically. Especially given its squished position between BAT and BIT, it is then not surprising that BET (though still multi-modal) would exhibit overlapping
distributions. BIT, in the upper periphery of the space, has more room to maneuver both up and down, which is likely why the three F1 distributions for BIT were more discrete.

At the very least, however, the fact that all of these front vowels show multi-modal distributions and the fact that we know AR to be an advanced speaker in her community in terms of the shifting of these vowels, leads us to the conclusion that it is their very shifting nature that leads the distributions to pattern in this way, providing resources for AR to stylize extreme tokens along the distribution.

What about the high back vowels TOE/BOAT and TOO/BOOT? In contrast to the front lax vowels, for the high back vowels, we only see multi-modal distributions with one vowel class, TOE, along the F2 dimension. It is notable that the multi-modal effect is evident on the F2 dimension, the direction of the vowel’s movement in the CVS. However, if we think about AR’s relative progress with respect to shifting for these high back vowels, it is perhaps no surprise that she shows comparatively little in terms of multi-modal distributions and corresponding stylization for these vowels. Podesva et al (2013) found that, generally speaking, all younger Redding speakers had fronted TOO/BOOT, while country-oriented people had significantly fronter TOE/BOAT vowels than the town speakers. In particular, TOE leads the way in the fronting, with BOAT lagging behind. This is consistent with the findings of others (Kennedy and Grama 2012; Hall-Lew 2009) that TOO/BOOT leads the way in California, with TOE/BOAT following. The fact that AR’s TOE tokens exhibit multi-modal patterning suggests that it is precisely this vowel class in her high back vowel space that is right now undergoing shift. We might conclude from this that her TOO/BOOT have neared completion, and the process for TOE/BOAT has now begun for her, with TOE leading the way. This is certainly consistent with general findings for the Redding community.

5 Conclusion
In conclusion, this paper illuminates two methodological findings and one big theoretical proposal. The first methodological finding is that the analyst cannot assume linguistic features (in this case, vowels) to pattern normally across a distribution. The fact that the data in this study were robustly non-normal in many cases is an important finding. Analysts need to gather enough data points so as to carry out meaningful distribution analysis. Many statistical tests carry with them an assumption of normality. Thus, a sociolinguist carrying out quantitative analysis should be informed as to the specific patterning of the distributions in her data.

The second methodological finding presented here is that prosodic prominence can be a measure for assessing stylization. This is likely from a theoretical standpoint, notwithstanding the fact that this measure has been shown to be perceptually salient for listeners. While the social meanings of tokens in the extreme distributions found in this study have not yet been explored, there is work (e.g.,
Podesva 2007; Eckert et al. 2013) suggesting that clustering of salient features corresponds to greater degrees of stylization. Thus, this paper provided evidence for distributions that not only were acoustically extreme in terms of the F1/F2 dimensions, but were also either prosodically prominent or creaky.

The main theoretical proposal presented here is that stylization and linguistic change are inextricably linked. AR's vowel space presents us with ample evidence for this. The vowels currently undergoing shifts in her system: BAT, BET, BIT, and TOE are also the ones exhibiting multi-modal distributions. These “extra” distributions on the extreme edges of her overall variation for these vowels are loci for stylization because they correlate significantly with factors of prosodic prominence and creak.

If there is indeed a link between multi-modal distributions, stylization, and vowel shifting, there is an important implication for future analysis. If these findings play out more broadly, then analysts can assume an inextricable relationship between stylization and linguistic change. Specifically, if we know there to be features currently undergoing linguistic change (as in this case, sound change) on a community-wide level, then those features are ripe and more likely to be used for use for stylistic purposes for the individual speaker. This is not an entirely new idea (see discussion in Section 1.1), as others have also found that features undergoing change are likely to also be those used for stylization.

But, what this study also suggests is the flip-side of this finding. Namely, that if we see multi-modal (i.e., extreme stylization) patterning for certain vowel classes in a speaker’s repertoire, then we can predict/posit sound changes in progress for those classes, which are likely to also be occurring on a community-wide level. For example, the fact that AR’s TOO/BOOT classes do not exhibit multi-modal patterning, but her TOE (not BOAT) class does, provides suggestive evidence that it is the TOE class that is currently undergoing change in her community. In AR’s case, we know this to be true because we have already conducted a community-wide study on Redding incorporating all of these vowels, and because we know some of the background/context of the CVS in other prominent California communities.

Most work to date has taken an inter-speaker approach to studying community-wide linguistic change across apparent and real time. This paper proposes a model whereby intra-speaker variation, specifically in terms of features used for stylistic purposes, can shed light on features likely to be undergoing change on a community-wide level. More analysis of intra-speaker variation and its relationship to inter-speaker variation within a speech community would be helpful to substantiate the extent of this relationship.

In all, this study proposes a link between stylization and linguistic change that can be observed in operation from either the inter-speaker sound change or intra-speaker stylization side of an analysis. However, it still cannot as yet answer the chicken/egg question: what comes first, the stylization or the linguistic change?
is an important question, one which I hope to investigate further in future work, But, at the moment all I propose here is an inextricable link that goes both ways, not a solution to the actuation problem.
References


Leys, Christophe, Christophe Ley, Oliver Klein, Phillippe Bernard, and Laurent Licata. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology* 49: 764-766.


Endnotes

i  *Style* as a sociolinguistic construct has gone through a few iterations as variation theory has progressed through waves one, two, and three. First-wave theorists viewed style largely as a matter of paying varying attention to speech, due to formal or informal contextual pressures (cf. Labov 1966; 1972). Third-wave variation theory, by contrast, attributes much more speaker agency as well as a more diverse set of potential contexts, features, and outcomes to stylistic variation.

ii This paper follows Yaeger-Dror and Thomas (2010) in using a B_T vowel class frame, occasional exceptions noted for phonemic/phonetic relevance.

iii The fronting effect of a preceding coronal place of articulation on its following vowel has been well-established in phonology (cf. Flemming 2003) and for California speech (cf. Hall-Lew 2009). Accordingly, the post-coronal tokens from the BOAT and BOOT vowel classes (i.e., TOE and TOO) are considered separately from the non-post-coronal tokens here.

iv In order for these distribution plots to more closely resemble the vowel spaces they represent, the relevant axes in the figures are flipped. That is, tokens with higher F1 values (which correspond to lower points in the vowel space) are represented in these figures on the bottom of the distributions; likewise, tokens with higher F2 values (fronter in the vowel space) are illustrated at the left edge of the distributions.

v As raw duration values at the segmental level are significantly right-skewed, the convention is to log-transform them to help centralize the data (Rosen 2005).