Contrast maintenance in California low back vowels*

Christian Brickhouse
brickhouse@stanford.edu

Rev. March 2, 2020

1 Introduction

Previous work has defined the variety of English spoken in California as being characterized in part by a merger of the low back vowels (D'Onofrio et al., 2016; Eckert, 2008; Hall-Lew, 2009; Holland, 2014; Kennedy & Grama, 2012; Labov, 1991; Labov et al., 2006; Moonwomon, 1991; Podesva et al., 2015). Younger Californians in these studies have shown LOT and THOUGHT which overlap heavily in formant space and that this overlap has been increasing over apparent time. This pattern, both the spectral overlap and the increasing spectral overlap over time, has been used as evidence that a merger is or has occurred in the LOT and THOUGHT of Californians.

Evidence of this spectral overlap comes from midpoint measurements of vowel formants. For example, D'Onofrio et al. (2016), represents vowels as a set of F1 and F2 measures taken from the vowel midpoints. These can be understood as a point in a coordinate system with the second format as the x-axis, and the first formant as the y-axis. Representing vowels as points in a two-dimensional Cartesian coordinate system is a common operationalization, and it has the advantage that these points in acoustic space strongly correlate with the

*This working paper was submitted as my second qualifying paper at Stanford. Thanks goes to my committee: Meghan Sumner, Robert Podesva, and Arto Antilla. Mistakes are my own.

position of the vowel in articulatory space. The correlation with articulatory position adds strength to the hypothesized merger as the evidence in the literature can be said to point to both acoustic and articulatory convergence in the LOT and THOUGHT vowels.

Representing vowels as coordinate points defined by their F1 and F2 midpoint values is further motivated by the way in which listeners identify vowels. In general, the primary cue to vowel identification is the first and second formant and listeners attend to this acoustic cue to determine what phoneme was intended by the speaker. As the formant values of two vowel classes approach each other, as is the case for the LOT and THOUGHT vowels in California, the ability to distinguish between the two vowels is diminished. In some cases this results in a push chain shift so that the margin of security (Labov, 1994) is maintained, while in other instances this increased confusability results in a merger. Using the primary cue to vowel class as the primary representation of the vowel provides a quantification of their margins of security.

Vowels, like all segments of speech, are time-varying signals with multiple potential axes of differentiation. Despite the advantages of two-dimensional representations, they alone only quantify a portion of the acoustic signal, and alternative cues to contrast maintenance can be overlooked. Languages encode redundant information so that listeners can reliably retrieve the intended meaning even when noise diminishes the legibility of the primary cue. Secondary cues are aspects of the signal which aid in phoneme identification but which are not the primary means by which phonemes are differentiated. For example, while post-vocalic stops in English are primarily distinguished by differences in voicing, if this primary cue is not available (whether due to noise, word-final devoicing, or experimental design) the length of the preceding vowel serves as a secondary cue to stop voicing allowing listeners to distinguish between minimal pairs. In the Inland North dialect of American English, Labov & Baranowski (2006) find that despite overlap in formant space of BET and BAT, speakers systematically use a difference of about 50ms to maintain the phonemic contrast. While vowels are generally differentiated by their location, other dimensions of the vowel

may become more reliable indicators of the phonemic contrast as the primary cue weakens diachronically.

We investigate two features, vowel dynamism and vowel length, which represent two axes along which the low-back vowels may be differentiated. Due to phonetic effects that provide redundant information, these features are likely secondary cues which aid identification when the primary cue—location—is weakened, and may be transphonologized (Hyman, 2013) by the shift of phonemic contrast from location to an extant secondary cue. To evaluate whether transphonologization has occured in the case of California's low back vowels, three acoustic analyses are presented which evaluate the overlap in their primary and secondary acoustic cues. The first replicates previous findings that the primary cue is weakening as the LOT and THOUGHT vowels approach each other in formant space over apparent time. The second evaluates the secondary cue of vowel dynamism using a novel methodology for the analysis of formant trajectories over apparent time. The final analysis evaluates the secondary cue of vowel length by analyzing change in duration over apparent time. Combined, these analyses paint a more comprehensive picture of the degree of spectral overlap in California's low back vowels.

1.1 Secondary cues and contrast maintenance

The purpose of speech is communication, and successful communication requires that a speaker reliably categorize the acoustic signal into discrete units. These discrete units—phonemes on the level of individual sounds—are identifiable by some distinctive feature in the acoustic signal.[1] In the case of vowels this primary cue is their position in articulatory space: height and backness. Speech rarely occurs in ideal conditions, and environmental noise can diminish or eliminate the ability of a listener to recover primary cues from the acoustic signals. Because of this noisy communicative channel, language users identify and encode secondary cues to contrasts in order to aid the identification of phones during discourse.

[1] *Mutatis mutandis* for sign languages and other modalities.

As a diachronic change reducing the reliability of the primary acoustic cue makes a merger more likely, further changes in the system may occur which have the effect of preventing mergers. Chain shifting, particularly push chains, is a well-known example. As phoneme A encroaches on the articulatory space of phoneme B, the articulation of phoneme B shifts in order to maintain its margin of security and in so doing may cause the movement of phoneme C by the same principles and so on and so forth until the shift is terminated, often by a merger (Martinet, 1952; Labov, 1991). However margins of security encompass the entire articulation, and encroachment along one axis of difference does not necessitate that phoneme B maintain its margin of security by moving along the same axis. Just as one airplane can avoid another by climbing, vowels approaching in formant space can avoid each other by moving apart along a third dimension.

Secondary cues are often the result of coarticulatory or other phonetic effects, and may eventually become part of the articulatory specification of the phoneme. An example of this phenomenon is vowel length before plosives. Because of articulatory constraints, vowels are phonetically lengthened before voiced stops. In English, however, vowels before voiced stops are significantly longer than would be predicted by articulatory effects alone (Beckman, 1986; Peterson & Lehiste, 1960). What was originally an articulatory effect was recognized by language learners as redundant information. This redundancy could be utilized to ensure proper transmission of the following segment's voicing, and so speakers began to attend to it as a secondary cue. This secondary cue was then learned by children as part of the phonology of the language and thus part of the articulatory specification of vowels. Phonologically these cases are interesting in that these alternations are in complementary distribution—length is conditioned by the following segment—however these differences are distinctive in that they are able to serve as the only signal to a linguistic opposition.

The development of phonological secondary cues represents the first stage in the reanalysis of a contrast marking the development of quasi-phonemes (Kiparsky, 2016). The development of nasalized vowels is one instance of this reanalysis that results in a split.

Due to anticipatory coarticulation, vowels before nasals have some degree of nasalization which serves as a secondary cue to the quality of the following stop. In some languages this secondary cue becomes phonologized as part of the specification of the vowel; there is an alophonic alternation between non-nasal and (partially) nasal vowels conditioned by nasality of the following stop. If this conditioning environment is lost—in this case, nasals are eliminated from the lexical specification—then the phonologization of this secondary cue can result in a phonemic split between nasalized and non-nasalized vowels. This is one account of the genesis of the nasal vowel system in French.

In the case of the California low back vowels, the conditioning environment is the distribution of the original LOT and THOUGHT vowel classes. The phonetic implementation of these phonemes creates coarticulatory effects based upon their location in articulatory space. These coarticulatory differences such as length or dynamism may be learned as secondary cues and phonologized as quasi-phonemes. When these vowels begin moving together in formant space the original conditioning environment for the quasi-phonemes is lost and the language is faced with two paths similar to French: split along the existing quasi-phonemic boundary or merge the primary and secondary cues. If the low back vowels in California have also merged along secondary cues, this provides evidence in favor of the merger hypothesis, whereas differentiation along an existing secondary cue would indicate transphonologization of the LOT-THOUGHT contrast.

1.2 Distinguishing mergers and non-mergers

A merger is a change in the linguistic system marked by the loss of a contrast between phonemes in a language. Two types of phenomena referred to as merger ought be distinguished: complete mergers and near mergers. Complete mergers are those where two (or more) phonemes lose all contrast between them and are indistinguishable acoustically and perceptually. Mergers of this kind are impossible to reverse as the once separate phonemes—which by definition appeared in contrastive environments—have no unique conditioning en-

vironment that would allow a regular sound change to split them into their original classes. Near mergers, on the other hand, are a phenomenon whereby speakers cannot reliably distinguish an acoustic contrast which they produce. That is, where a complete merger is indistinguishable in production and perception, near mergers are distinguished in production but not perception.

While these phenomena are easy to distinguish perceptually, when looking at production data they are harder to disentangle. Complete merger is easily distinguished because there is by definition no articulatory difference between merged vowels; if we identify any dimension along which the LOT and THOUGHT vowels differ then it cannot be a complete merger. The question becomes harder when considering near-mergers and contrast maintenance which both, by definition, have articulatory differences. Near-merger has already been defined as two or more segments which are distinct in their production but whose speakers cannot reliably tell apart. Under this broad definition it would be impossible to distinguish this from contrast maintenance without perceptual data.[2] While these phenomena make nearly identical synchronic predictions, they can be adequately distinguished if considered diachronically.

Mergers, and by extension near-mergers, arise from one of at least three diachronic patterns. In MERGER BY APPROXIMATION, two phonemes move together in articulatory space merging anywhere along the continuum between the two, including the end points (Labov, 1981; Trudgill & Foxcroft, 1978). In MERGER BY EXPANSION (Herold, 1990; Labov, 1994), the articulatory ranges of two phonemes expand until they are completely overlapping, resulting in a single phoneme which can be articulated as either original phone or intermediate articulations. For both of these cases, the diachronic pattern would be that the distance between the mean articulatory targets of two merging phonemes would decrease as they

[2]It is so broad that even perceptual data may not be sufficient, that is, one could interpret near-merger as one strategy of contrast maintenance. Since a contrast is not necessarily lost—they are still contrasted in production—it can be argued that the contrast is maintained by phones stopping just short of a complete merger. This is unsatisfying as the motivating factor in contrast maintenance is being able to reliably distinguish phonemes which the merger-in-perception aspect of near-mergers fails to accomplish.

move or expand towards each other. In MERGER BY TRANSFER (Trudgill & Foxcroft, 1978), instances of one phoneme are replaced by instances of another in a word-by-word fashion. This pattern is unique in that the phonemes do not necessarily move, rather, one phoneme is attrited until it is completely replaced. Despite this difference, so long as the investigator knows the original class of the merging words and groups tokens by that classification, the apparent pattern would be identical to the two forms of merger previously discussed. This is due to the nature of the arithmetic mean. As words from one vowel class move into another, it pulls the mean articulation of its original vowel class towards the merging vowel class. As this process continues, and as more words move into the other vowel class the mean moves closer to the other phoneme until all words are in one vowel class and the means of the two original classes are identical. Thus in all cases the mean articulation of two merging phonemes move together over time.

Contrast maintenance, on the other hand, shows a different pattern. In the simplest case, the mean articulations of two phonemes keep their distance from each other. This does not mean that they are not moving, rather that they move at the same time keeping their distance from one another. An example of this would be chain shifts where the movement of one phoneme pushes or pulls another so that their distance from each other remains relatively stable. This pattern is unhelpful in cases of apparent merger, however, as in cases like California's apparent low back merger, the distance between primary cues is decreasing. While one could look to alternative cues and show that on those dimensions the phonemes are not converging, that evidence could support either a contrast maintenance or near-merger hypothesis, and it may be the case that speakers are not attending to that cue at all.

1.3 An apparent low back merger in California

A large body of production evidence has been used to argue for an apparent merger of the LOT and THOUGHT vowels in California English. While the specifics vary, over apparent time the two vowels are showing increasing overlap in formant space . (DeCamp, 1953)

first speculated that the low back merger was occurring in California based on evidence from San Francisco speakers. This hypothesis was partially supported by Hinton et al. (1987) who found that THOUGHT moved towards LOT, but noted that this movement was "not especially vigorous." Moonwomon (1991) found that among her youngest speakers, there was almost complete overlap of LOT and THOUGHT and argued that the merger was in fact progressing rapidly. Hall-Lew (2009) finds a similar pattern, with a trend over apparent time towards increasing spectral overlap. Outside of San Francisco, the evidence is similar. D'Onofrio et al. (2016); Podesva et al. (2015) have found that in rural areas of California, the LOT and THOUGHT vowels show increasing overlap in apparent time, though with movement of LOT to THOUGHT. While these studies are consistent in finding evidence for an apparent merger in production, the evidence for a merger in perception is not nearly as robust.

Since the production evidence suggests an apparent merger, we would expect to find similarly robust results in perceptual tasks. Unfortunately, few perceptual studies have been conducted and this present study will not remedy that. However those studies which do exist do not show the same robust pattern in perception as they do in production. Of the 28 Californians interviewed between 1980 and 2000, fewer than 6 showed a merger in perception (Hinton et al., 1987; Labov et al., 2006). While this may be consistent with a merger, the low proportion of speakers with a merger in perception suggests that a merger may not have taken place. If speakers are able to reliably distinguish LOT and THOUGHT despite increasing overlap in formant space , then a merger is unlikely. An alternative explanation for this pattern of data is that the LOT and THOUGHT vowels in California are undergoing cue reweighing. Because the evidence suggests speakers can reliably distinguish between LOT and THOUGHT despite their overlap in formant space , the results of previous perceptual tasks suggest that speakers are attending to some alternative cue which is not the first or second formant.

The phonetic differences between the unmerged phonemes implicates two potential secondary cues—length and vowel dynamism—that may be recruited to maintain phonemic

contrast. The first potential cue investigated is vowel dynamism, representing the degree of movement throughout the vowel. English maintains a system of dipthongs which are identified in part by the change in formants over the course of the vowel. The second potential cue investigated is length. As discussed above this is already a salient secondary cue in other contexts in English and seems likely to be present as a secondary cue here as well. Such a process has been observed in other cases of apparent mergers such as in the Inland North dialect of English Labov & Baranowski (2006). Low vowels are longer than higher vowels due to phonetic processes presenting the opportunity for phonologization as a secondary cue in much the same way as length before stops. As LOT was initially lower than THOUGHT, it would be expected to be phonetically longer. If speakers began to attend to this length as a secondary cue to the LOT and THOUGHT contrast then as the primary height contrast was lost speakers may have begun to rely on what was previously a phonetic difference to maintain the contrast resulting in a phonological length contrast.

The results of these analyses will provide evidence for the constraints on quasi-phonemic change in a case of potential transphonologization. In cases where secondary cues are due to phonological environments, such as intervocalic voicing or pre-nasal vowel nasalization, the loss of the phonological environment is the trigger for merger or secondary split. The California case presents a situation in which the conditioning environment is not distributional but inherent to the phonemes. The LOT and THOUGHT overlap in their phonological environments, and so quasi-phonemic contrasts should arise out of the articulatory location specified by the phonology rather than the distribution of phonemes to which the vowels are adjacent. Evidence of transphonologization would thus motivate an expansion of the types of conditioning environment changes which can cause quasi-phonemic splits or mergers.

These results will further provide a rigorous foundation for perceptual work. To the degree that near-merger is a plausible explanation for the data, the results will inform hypotheses on what axes of the vowel to manipulate in a perceptual experiment. If, for example, speakers are attending to a dimension other than location to identify these contrasts, manipulating

location in an experimental setting may not give coherent results. That is, speakers with a contrast may not exhibit a contrast simply because the dimension they are attending two is remaining stable prompting uniform responses. By isolating acoustic cues to which speakers may be attending, perceptual experiments can have targeted controls and test variables providing more robust results. Specific recommendations and directions for future research are provided in the discussion of the following three analyses.

2 Methods

Three analyses are presented to test the hypothesis that the LOT and THOUGHT vowels are merging in production over apparent time in California English. The first is a replication of previous studies which found increasing overlap in formant space . The second and third experiments use production data to test the hypothesis that listeners are attending to (and thus speakers are producing) acoustic cues other than F1-F2 in order to distinguish the LOT and THOUGHT vowels; each tests a particular cue to which listeners may attend. The second analysis evaluates whether LOT and THOUGHT show similar formant dynamics—whether the changes in articulatory posture during production are similar for the two vowels. The third analysis evaluates whether LOT and THOUGHT show similar durations—whether the articulatory posture is held for similar lengths of time for the two vowels. All three analyses use wordlist data collected at the end of a sociolinguistic interview in which the tokens were not adjacent.

2.1 Data collection

Data were collected between 2012 and 2018 from five field sites as part of the Voices of California project. The field sites (and year of collection) used in these analyses were Bakersfield (2012), Sacramento (2014), Salinas (2016), Humboldt Bay (2017), and Redlands (2018). Participants were over the age of 18 at the time of the interview and were excluded

	Site	*Total*	White	Hispanic	Multiracial	A/B/I
F1-F2 Mean	Bakersfield	*66*	37	15	2	12
	Sacramento	*115*	74	7	10	24
	Salinas	*46*	5	35	4	2
	Humboldt	*84*	64	4	7	9
	Redlands	*75*	50	12	7	6
	Total	***386***	**230**	**73**	**30**	**53**
Dynamics	Bakersfield	*98*	57	21	3	17
	Sacramento	*131*	87	10	9	25
	Salinas	*42*	6	31	3	2
	Humboldt	*95*	73	4	9	9
	Redlands	*70*	46	11	7	6
	Total	***436***	**269**	**77**	**31**	**59**
Duration	Bakersfield	*107*	61	23	3	20
	Sacramento	*136*	89	10	11	26
	Salinas	*54*	6	42	4	2
	Humboldt	*95*	73	4	9	9
	Redlands	*79*	51	13	9	6
	Total	***471***	**280**	**92**	**36**	**63**

Table 1: Sample by racial grouping, analysis, and field site. Cells in bold are totals by column; those in italics are totals by row. Racial groups that did not have at least one speaker per cell were combined so that the model was not rank deficient. This resulted in the combination of the Asian, Black, and Native American (Indigenous) racial groups into the A/B/I group.

if they lived outside the field site for more than 8 years.[3] They participated in an hour long sociolinguistic interview for which they were not paid, and were asked to read a wordlist at the end of the interview.

The wordlist contained the same items for all fieldsites, however the ordering of the words changed across field sites. Participants from Bakersfield read a wordlist whose order was randomized and unique to them. Later field sites[4] had two wordlists differing in the word order, and participants were randomly assigned one of those two orders to read. For all wordlists there is a single token of the LOT and THOUGHT vowels in the lexical items *cot* and *caught* which were not adjacent in any wordlist ordering.

Acoustic analysis was carried out automatically by script[5] with some manual measurements of tokens where the scripts failed. The wordlists were force aligned using the Penn Force Aligner (Rosenfelder et al., 2014). Stressed vowels were extracted and measurements of F1 and F2 were taken at 10 equidistant points throughout the vowel using PraatSauce (Kirby, 2019). For normalization purposes, the multiple measures for each formant were averaged to yield one formant value per formant per token. These tokens were normalized using the Nearey normalization method as implemented in the vowels package (Kendall & Thomas, 2018) in R (R Core Team, 2018). Analyses were conducted using the base R functions and for mixed effects regressions the lme4 package (Bates et al., 2019) was used with p-values calculated by the lmerTest package (Kuznetsova et al., 2019).

2.2 Spectral overlap in F1 and F2 space

This analysis serves to replicate previous studies that have found the LOT and THOUGHT vowels converging in formant space over apparent time (D'Onofrio et al., 2016; Hall-Lew, 2009; Podesva et al., 2015). To evaluate the hypothesis that these vowels are moving together

[3]Specifically, they were excluded if they lived outside the field site for more than 2 years before 18 or more than 6 years after 18.

[4]Sacramento, Salinas, Humboldt, and Redlands

[5]See the GitHub repository at `https://github.com/chrisbrickhouse/california-vowels` for the scripts

in formant space , their euclidean distance in normalized formant space for each speaker is calculated and modeled by a linear regression. If previous findings are reflected in this data, we should observe a decrease in euclidean distance between the LOT and THOUGHT vowels as age increases.

2.3 Vowel dynamics

While vowels have particular, and in some cases identical, articulatory targets, it is not sufficient to show that two vowels have the same goal. The previous analysis of spectral overlap in formant space evaluates whether the point in space (articulatory or acoustic) that vowels attain is different, however even if their goals are identical the way they achieve it may not be. Because the tongue has multiple muscles and degrees of freedom, it is possible that there is a one-to-many mapping between targets and the movements required to reach those targets. These movements are captured not by the vowel's target, but by the transitions to that target. By analyzing the change in articulatory posture over time, we can evaluate whether the entire articulatory specifications of the vowels are converging, not just their targets.

As the articulatory posture can be inferred from the acoustics, changes in articulatory posture can be infered from change in the acoustics. Thus the changes in F1 and F2 position during articulation—their trajectories—represents the postural changes in which we are interested, in the same way as static F1-F2 measurements represent articulatory targets. As mentioned previously, the LOT and THOUGHT vowels were measured at 10 equidistant points throughout the vowel providing a 10-point formant track. This presents an analytical problem: how do we operationalize similarity in trajectory.

To exemplify these issues, consider the 10-point ordered lists A and B:

(1) a. $A = \{2, 4, 6, 8, 10, 12, 14, 16, 18, 20\}$

b. $B = \{102, 104, 106, 108, 110, 112, 114, 116, 118, 120\}$

In one sense, these trajectories are similar, they both show a constant increase of 2 units and thus represent parallel lines. In another sense they are different, as B is shifted upwards by 100 units. We can represent A and B as functions:

(2) a. $A(t) = 2t$

b. $B(t) = 2t + 100$

Because the previous F1-F2 analysis already evaluated the targets themselves, we are interested in how those targets are attained. So while the targets of A and B differ by 100 units, the way they attain their targets—their trajectories—is similar.

Just as evaluating the similarity of A and B was made easier by converting them to the functions in (2), we can evaluate the similarity of LOT and THOUGHT tokens by representing the 10 point formant measurements as functions as well. Each formant of a given token has 10 data points which can be represented as an ordered list like in (1). While those lists showed a linear relation, formant trajectories are not usually linear, and so should not be modeled by a linear function.

There are thus three operationalizations required for the analysis: a way to model formant trajectories, a way to evaluate the similarity of those models, and—because we are interested in diachronic change—a way to evaluate changes in their similarity over time. We will consider these in turn.

Modeling formant trajectories The ideal family of functions to model the trajectories is one which is similar in shape to the pattern we expect given the known phonetic context. If the function naturally takes a similar shape to the data, then the parameters of that function are more meaningful in that they are less reflective of the known broad pattern (phonetic context) and more reflective of the subpatterns of interest (individual and social factors). We know that for the two tokens in our data set—*cot* and *caught*—the tongue will go from a velar closure, to a low vowel, to an alveolar closure; the tongue moves down then up while moving forward. This will be reflected in the formants: the first formant will increase then

decrease; the second formant will generally increase with some variation in the middle due to the vowel target. We use the cosine function as our basis function given that its full period matches the expected pattern of the first formant, and its half period matches the expected pattern of the second formant.

The formants are modeled using a generalized additive model (GAM). A GAM represents a series of data points as the sum of various basis functions:

$$Y(x) = \epsilon + \beta_1 f_1(x) + \beta_2 f_2(x) + ... + \beta_u f_u(x) \tag{1}$$

The basis functions for our GAM are cosine waves. For a model of N data points with u terms, the function for the kth term is:

$$f_k(x) = cos\left[\frac{\pi k}{N}\left(x + \frac{1}{2}\right)\right] \tag{2}$$

The coefficients of the model, β_k, are fit to the data set, $X(n)$, and are determined using the discrete cosine transform:

$$\beta_k = \sum_{n=0}^{N-1} X(n) \cdot cos\left[\frac{\pi k}{N}\left(n + \frac{1}{2}\right)\right] \tag{3}$$

And so the entire trajectory is modeled by the function:

$$Y(x) = \frac{1}{2}\beta_0 + \sum_{k=1}^{u-1} \beta_k \cdot f_k(x) \tag{4}$$

Like all GAMs, the accuracy of Y(x) is dependent on the value of u—how many terms are used in the model. If $u = N$ then this model reproduces the data to which it was fit (perfectly if multiplied by the correct scaling factor). This is undesirable as there is noise in our data that we do not want to model, and so we must determine the optimal value of u so that our model does not over- or under-fit the data.

We determine u by finding the value at which we see diminishing returns using the elbow method. As with any model, we can quantify the error using the sum of squared differences. As the value of u increases, the sum of squared differences will decrease, but it will not do so at a constant rate. There is a point where the rate of change begins to diminish, and so the benefit of an additional term is outweighed by the risk of overfitting. If we were to plot the error of a model against the number of terms used in that model, we would get graphs like those in the top panel of figures 1 and 2. Graphs of this sort have an "elbow" which is the point at which there are diminishing returns. Looking at the top panels in figures 1 and 2, it is rather difficult to determine the elbow point impressionistically.

To guide the selection of u we use a quantitative measure to find the elbow point of these graphs. If we model the sum of squared errors as function of the number of terms, $\epsilon(u)$, then there is a function, $\epsilon'(u)$, which returns how much error has been removed; in formulaic terms, $\epsilon'(u) \approx \epsilon(u-1) - \epsilon(u)$. For values of u towards 0 and N, the value of $\epsilon'(u)$ will not change much, but at the elbow point it will quickly go from consistently big changes to consistently small changes. Thus the elbow point is the value of u such that $\epsilon'(u)$ has the greatest rate of change. We were able to quantify the rate of change of $\epsilon(u)$ by taking it's derivative, yielding $\epsilon'(u)$. Since we are interested in the rate of change of $\epsilon'(u)$, we take the second derivative of $\epsilon(u)$, yielding $\epsilon''(u)$. Because we want to know at what point $\epsilon''(u)$ is greatest, we find the value of u that returns the greatest absolute value of $\epsilon''(u)$. However the actual data we have are described by discrete functions, not differentiable functions as we had been tacitly assuming. We will therefore need to estimate what the second derivative would be if it were differentiable, which we can determine using the central difference equation in (5). These estimated values are presented in the bottom panels of figures 1 and 2 and the maximum absolute value for was determined by visual inspection. From these graphs, u for F1 was decided to be 4, and u for F2 was decided to be 3. As can be seen from figure 1, the actual absolute second derivative was at $u = 2$ however visual inspection of the fit showed an unacceptable fit, and so the second highest absolute value, $k = 4$, was used.

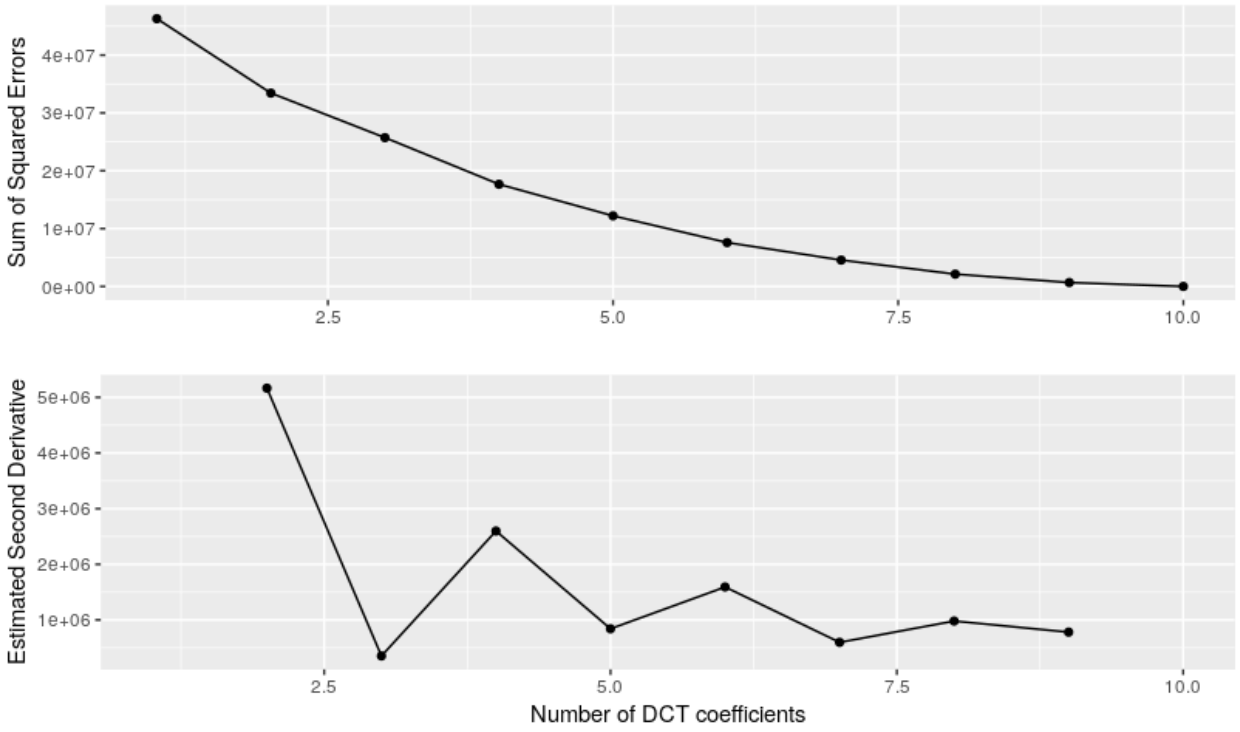

Figure 1: The sum of squared errors for Discrete Cosine Transform (DCT) models of F1 trajectories plotted against the number of coefficients used in that model (top) along with the estimated curvature of the function at that point (bottom).

$$\epsilon''(u) \approx \epsilon(u-1) + \epsilon(u+1) - 2\epsilon(u) \tag{5}$$

Evaluating similarity of trajectories The second desideratum of our analysis of vowel trajectories is a way to evaluate how similar two vowels are. Knowing that we are ultimately interested in how their similarity changes over time, this evaluation metric should lend itself well to temporal analysis. The euclidean distance measurement provides a quantitative measure of similarity that has already proven itself useful for diachronic studies of vowels.

While we have defined a model of formants, we have not determined a way to represent vowels. In typical euclidean distance analyses of vowels, a vowel is represented as a two dimensional point in formant space with the coordinates determined by point measurements of a vowel's first and second formants. However, our model of formants returns a set of values

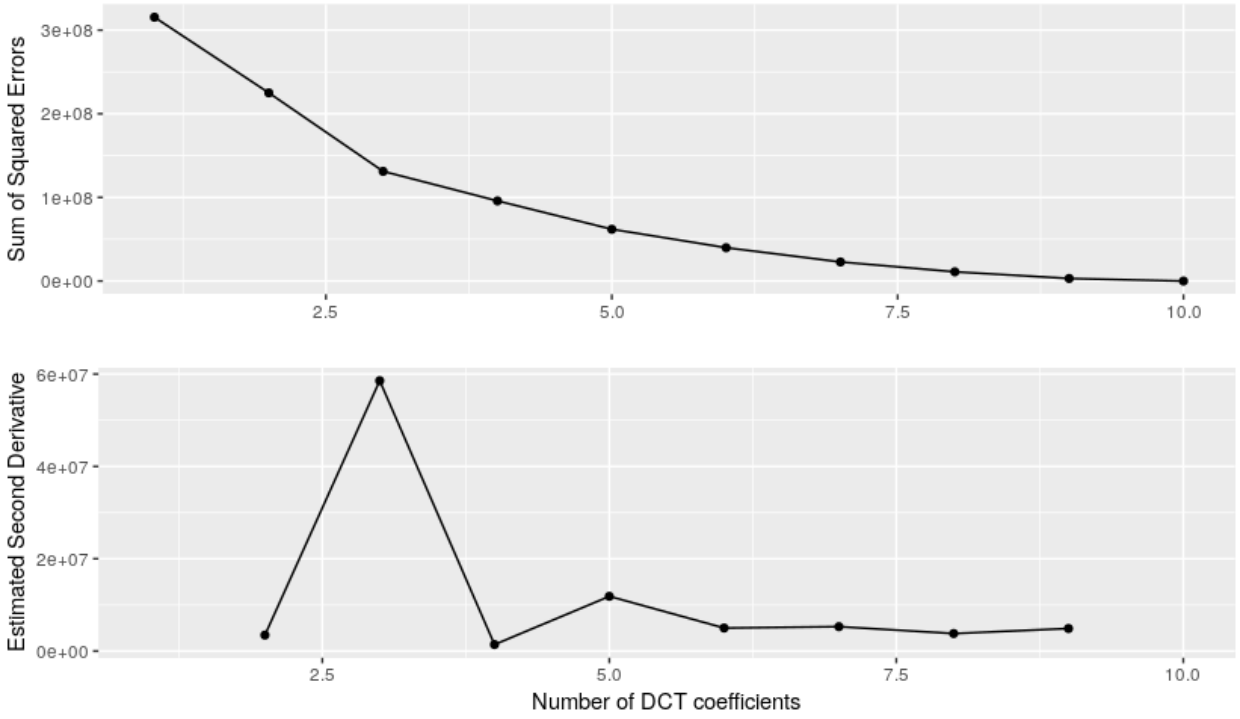

Figure 2: The sum of squared errors for DCT models of F2 trajectories plotted against the number of coefficients used in that model (top) along with the estimated curvature of the function at that point (bottom).

rather than a single value. So while we will still represent vowels as points in formant space , our space will have more dimensions than the familiar two dimensional representation using F1-F2.

A data set with too many dimensions can cause the data to become sparse and create problems for the analysis. We assume that if two vowels differ along any dimension, then they are different. As the number of dimensions increases, there are more ways for vowels to be different, not all of which are equally important or meaningful. Because the euclidean distance measurement takes all dimensions into account equally, difference along any dimension, even a relatively unimportant one, will be captured by the distance. As the number of dimensions increases, the chance that two similar vowels have a dimension which spuriously differentiates them increases, and so the chance of an unreliable euclidean distance measurement also increases. To avoid this, we want a representation with as few dimensions as possible.

The number of dimensions used in this analysis is based on the value of u found in the previous section. While we have a model of our formants, the output of that model is 10 autocorrelated formant values. A vowel representation using model outputs would thus be a 20-dimensional representation of autocorrelated values which is far from ideal. Instead, because of the way we constructed our models, the model coefficients are an equivalent representation of the formant. For every model, the functions and terms used in the function are identical; the only difference between models is their coefficients, and since there are fewer coefficients than outputs, the dimensionality of the data can be reduced.

If all coefficients were used, then based on our determined u values, the representation should be 7-dimensional, however we can further reduce the dimensionality of the data. Recall example (2). The two functions are similar; they differ only in their displacement along the y-axis. Now recall the model described in equation (4). The first coefficient, β_0, is used only to account for displacement along the y-axis. In fact, β_0 is proportional to the

mean formant measurement.[6] This is already investigated explicitly in the first analysis, and so using the first model coefficient would simply add noise to the model which we are already analyzing separately. Thus, just like in (2), the constant proportional to the first coefficient of each formant model can be disregarded from our representation to improve our ability to evaluate similarity leaving us with a 5-dimensional vowel representation.

The similarity between two vowels can then be determined by the 5-dimensional euclidean distance between their two representations. To be explicit, for a vowel, V, it's representation is the 5-dimensional vector:

$$V = \langle {}^{F1}\beta_1,\ {}^{F1}\beta_2,\ {}^{F1}\beta_3,\ {}^{F2}\beta_1,\ {}^{F2}\beta_2 \rangle \tag{6}$$

And the similarity, S, between the vowels v and w, is given by:

$$S_{v,w} = \sqrt{\sum_{n=0}^{4} (v_n - w_n)^2} \tag{7}$$

where v_n and w_n are the nth item in the vectors v and w respectively. If $S_{v,w}$ is 0, then the vowels have identical formant trajectories;[7] the value of $S_{v,w}$ increases, the similarity of v and w decreases.

Change over time With a method for evaluating the similarity of two vowels, the final desideratum is a method of evaluating changes in this similarity over time. For each speaker there are two tokens, one *cot* and one *caught*. We model the first and second formants as described above for both tokens and create vowel representations using the each token's formant model coefficients. We can calculate the euclidean distance between these representations as described above providing us with a single similarity measurement for each speaker. This

[6] Recall equation (3). β_0 reduces to $\sum X(n) \cdot cos(0)$ Since $cos(0) = 1$, β_0 is equal to the sum of all 10 formant measurements for that vowel formant. Thus $\frac{\beta_0}{2} \propto \frac{\beta_0}{N} = \mu$.

[7] It is important to note that identical formant trajectories do not entail identical vowels. v and w could have different average formant values, but have the trajectory of the formants identical.

is analyzed using the familiar method of linear regression, predicting the euclidean distance from the speaker's age.

2.4 Duration

The length of an articulation can serve to distinguish between two phonemes as can be seen in languages like Finnish and Arabic. Labov & Baranowski (2006) argue that some Inland North speakers had acquired a length contrast that separated the BET and BOT vowels. We will evaluate whether a similar length contrast between LOT and THOUGHT has been acquired by speakers of California English.

As previously discussed, when two phonemes approach each other in acoustic space, acoustic features which had carried redundant information may be recruited as a means of contrast preservation; this same pattern applies equally well to length. There is an inverse correlation between a vowel's height and its duration as it takes longer to achieve a low vowel target than a high vowel target. These phonetic effects can be phonologized as a secondary cue and thus the length differences can remain despite the vowel's articulatory position. When the phonemes move together, their lengths will remain different and the contrast is maintained by promoting the length contrast.

If the length difference is phonologized as a secondary cue, the difference should be emphasized and thus longer than would be expected by phonetic effects alone. While the durational differences are not compared to expected phonetic effects, the hypothesis that the differences are due to phonetic effects is falsified by a pattern of divergence. Because the phonetic effects of height on length should be relatively constant, the length difference between LOT and THOUGHT should be relatively constant if they are not moving. We know that they are in fact moving, but they are moving together; if the effects were purely phonetic we would expect the durations to converge over time. If the difference is relatively constant, the outcome is equivocal between near merger and contrast maintenance as previously discussed. However if the durational differences are becoming more pronounced while the two

vowels move together, then this is the exact opposite pattern of what would be expected from phonetic effects and is evidence in favor of a newly phonologized length contrast.

To test these hypothesis, a mixed effects linear regression was conducted on the vowel durations to evaluate the change in length over apparent time. Because raw duration data are strongly, positively skewed, the vowel durations were log transformed. The duration data was collected as part of the initial PraatSauce data extraction, however in some cases measurement errors[8] were corrected by hand.

3 Results

3.1 Spectral overlap in F1 and F2 space

To replicate previous findings of increasing spectral overlap in F1 and F2 space, the euclidean distance between the two vowels was evaluated using a mixed effects model. If the vowels are becoming closer over time, the euclidean distance should asymptotically approach zero, and so the distances were log transformed so that they could be evaluated using a linear model. The model predicted the log transformed euclidean distance from the fixed effects of gender, age, and their interaction along with by field site random intercepts.

There were a total of 386 speakers from 5 field sites included in this analysis; the number by field site can be seen in table 1. Speakers were excluded if they were more than two standard deviations from the mean.

The model corroborates previous findings that the vowels are increasing in their degree of spectral overlap in formant space over apparent time. For multiracial speakers, there is a marginal effect of age (β=0.022, SE=0.012, t(357)=1.85, p<0.066) such that the LOT and THOUGHT vowels are moving apart over apparent time. Given that this analysis included only 30 multiracial speakers, it is unclear how reliable this result is. For Asian, Black, or Native American speakers, there appears to be no effect of age, however the model shows

[8] which were excluded from previous analyses

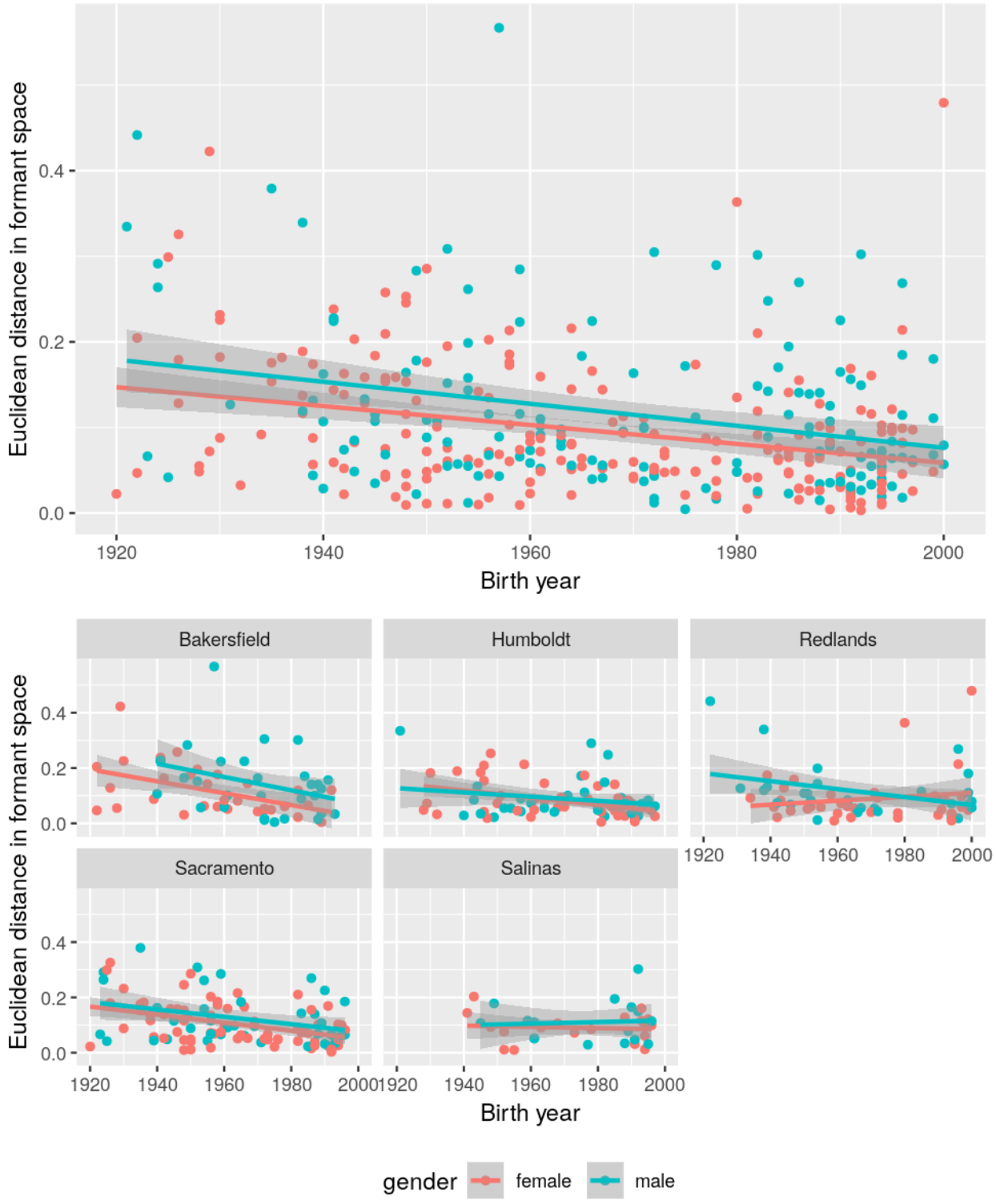

Figure 3: LOT and THOUGHT are becoming more similar in formant space over apparent time, though this pattern varies across site.

Factor	Estimate	Std.Er.	Df	p
White speakers				
Intercept	−2.62	0.006	619	< 0.0001
Birth year	−0.014	0.003	345	< 0.0001
Asian, Black, or Native American speakers				
Intercept	0.307	0.136	353	0.025
Multiracial speakers				
Birth year	0.022	0.012	357	0.066

Table 2: Significant and marginal results of the F1-F2 Euclidean distance model by factor and analysis group. The intercept represents the log distance between the LOT and THOUGHT vowels for a speaker with mean age, gender, etc. Negative parameter estimate values (other than the intercept) represent convergence.

that these speakers generally have LOT and THOUGHT vowels which are further apart than for white speakers (β=0.307, SE=0.136, t(353)=2.26, p<0.025). For white speakers there is a significant main effect of age (β=-0.014, SE=0.003, t(345)=-5.35, p<0.0001) such that LOT and THOUGHT vowels are becoming closer over apparent time. Though the plot in figure 3 suggests a main effect of gender, the model reveals no such effect. Rather, it seems that a participant's race explains most of the variation that appears to be due to gender.[9] While the model reveals LOT and THOUGHT convergence over apparent time, this pattern is apparent primarily among white speakers.

3.2 Vowel dynamics

To evaluate whether LOT and THOUGHT are becoming more or less similar in their formant trajectories over apparent time, the euclidean distance between the vowels in DCT space was evaluated using a generalized linear model. As the first DCT coefficient is proportional to the mean formant frequency, it was not included in this calculation given that the convergence of mean formant values was explicitly tested in section 3.1. The DCT coefficients of both formants were considered dimensions in the distance measurement allowing for distance

[9] A model run without race as a factor finds a significant main effect of gender (β=0.229, SE=0.088, t(357)=2.6, p<0.010) which is not present for any racial group when race is added as a factor in the model suggesting that the gender factor was picking up on variation better explained by race.

Factor	Estimate	Std.Er.	Z-score	p
White speakers				
Intercept	6.53	0.038	170	< 0.0001
Birth year	-0.003	0.002	-1.70	0.090
Asian, Black, or Native American speakers				
Birth year	-0.006	0.004	-1.67	0.093

Table 3: Significant and marginal results of the vowel trajectory difference model by factor and analysis group. The intercept represents the log distance between the vowel trajectories of LOT and THOUGHT for a speaker with mean age, gender, etc. Negative parameter estimate values (other than the intercept) represent convergence.

of the whole vowel trajectory, not just a single formant, could be analyzed. As such the euclidean distance between LOT and THOUGHT represents the distance between the vowels as represented in a 5-dimensional space: the second, third, and fourth DCT coefficients of F1 and the second and third DCT coefficients of F2. A linear model predicting the Euclidean distance was constructed to investigate whether the vowels are converging in apparent time.[10] The model considered the fixed effects of birth year and gender as well as their interaction.

A total of 438 speakers were included in this analysis. The number of speakers per field site can be seen in table 1. Because the combination of 5 measurements could compound the effects of measurement errors, euclidean distance measured as more then two standard deviations from the mean were removed.

The model reveals marginal effects of age such that LOT and THOUGHT appear to be converging over time. For Asian, Black, or Native American (β=-0.006, SE=0.004, z=-1.67, p<0.093) and White speakers (β=-0.003, SE=0.002, z=-1.70, p<0.090) the model reals marginal main effects of age such that the LOT and THOUGHT vowels are converging over apparent time. Given that the magnitude and direction of these effects are similar, the effect seems reliable.[11] There are no significant effects of gender.

As can be seen from the bottom panel of figure 4, patterns vary by field site.[12] Humboldt

[10] A generalized mixed effect linear regression was attempted, but even the most minimal random effect structures resulted in a boundary fit, and so no random effects were included.

[11] A model run without race as a predictor finds a significant main effect of age (β=-0.003, SE=0.001, z=-2.3, p<0.02)

[12] Due to limited sample size for some racial groups in each field site, race was not included as a predictor

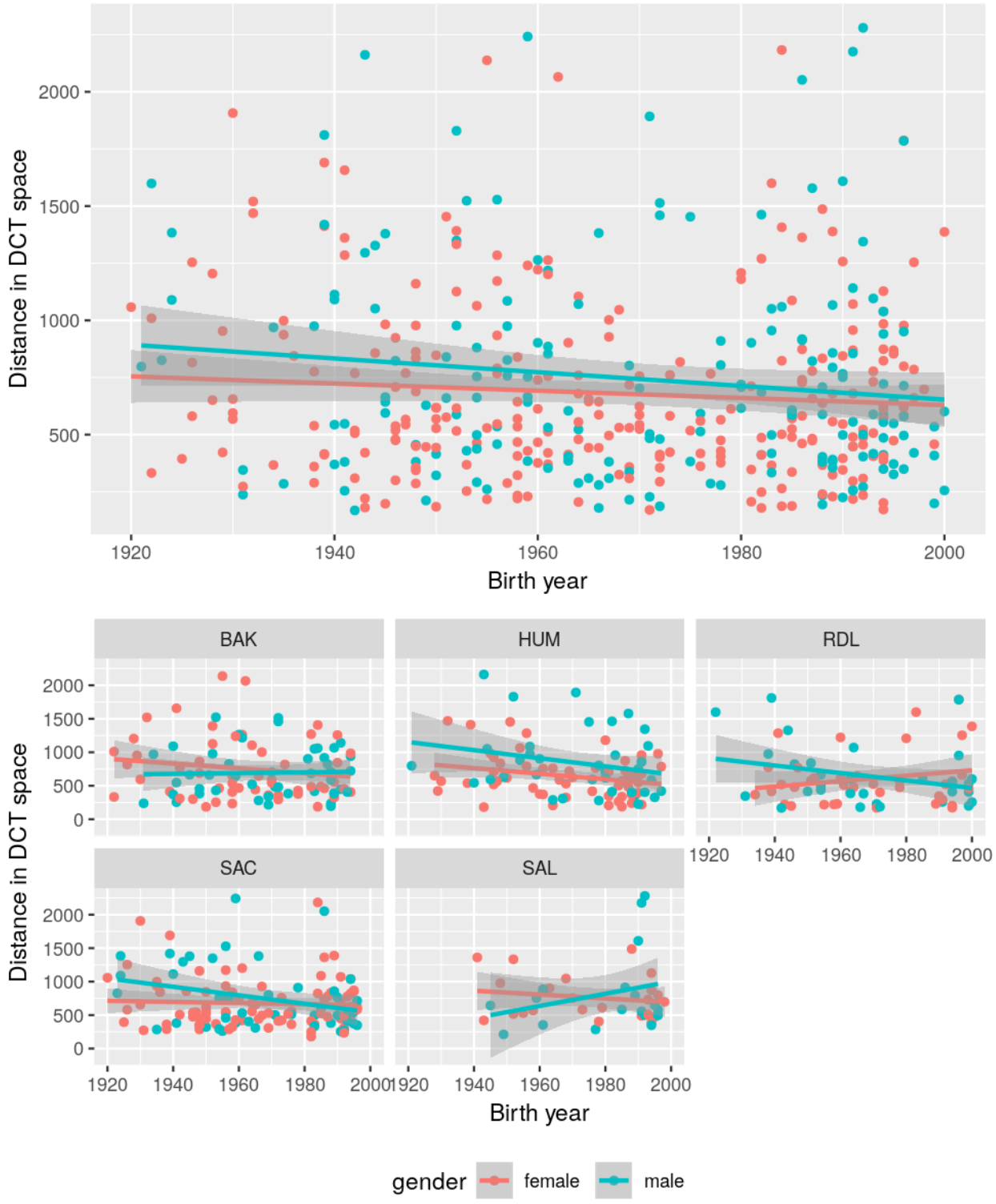

Figure 4: Euclidean distance of LOT and THOUGHT formant trajectories by age. The top panel shows the pattern across the data, and the bottom panel shows the data broken out by field site.

Factor	Estimate	Std.Er.	df	p
White speakers				
Intercept	−1.49	0.029	4.56	< 0.0001
Token	0.116	0.016	419	< 0.0001
Gender	0.046	0.023	425	0.035
Token×Gender	−0.073	0.032	418	0.024
Token×Birth year	0.002	0.0007	419	0.010
Asian, Black, or Native American speakers				
Birth year	−0.003	0.001	432	0.005
Token×Birth year	−0.003	0.002	425	0.060

Table 4: Estimates for significant and marginal effects from a mixed effects regression model of vowel duration by factor and analysis group. The intercept represents the log of the difference of the LOT and THOUGHT vowel durations for a speaker with mean age, gender, etc. Negative parameter estimate values (other than the intercept) represent convergence.

is the only field site which shows a significant main effect of age (β=-0.006, SE=0.003, z=-2.4, p<0.02) though Sacramento shows a marginal effect in the same direction (β=-0.004, SE=0.002, z=-1.8, p<0.07) with younger speakers having slightly closer formant trajectories. Humboldt also shows a significant main effect of gender (β=0.29, SE=0.11, z=2.6, p<0.01) such that women have closer formant trajectories than men. Redlands shows a different gender pattern. While no main effect of gender is observed in Redlands there is a significant gender and age interaction (β=-0.02, SE=0.008, z=-2.4, p<0.02) with women diverging over apparent time while men converge. Bakersfield and Salinas show no significant effects.

3.3 Duration

To evaluate whether BOT and BOUGHT vowels differ in length, a linear mixed effects model predicting the log duration of the segments was fit to the data. The model predicted the log duration from the fixed effects of vowel, age, and gender as well as their interaction. The maximal random effects structure was included but terms were removed in order for the model to converge on a non-singular fit resulting in the following random effect structure: random by site and by participant intercepts.

in field site specific models.

In total, 473 speakers were included in the analysis. Of these 473, 28 were represented by only one token while the remaining 445 had a LOT and a THOUGHT token. The specifics of the sample for each site is listed in table 1. Participants were excluded from this analysis if their log duration was more than two standard deviations away from the mean *and* they were excluded in one of the previous analyses.

The model reveals significant effects for Asian, Black, Native American, and White speakers, though the nature of these effects differs between racial groupings. For Asian, Black, and Native American speakers there is a main effect of age (β=-0.003, SE=0.001, t(432)=-2.85, p<0.005) such that both LOT and THOUGHT vowels are becoming shorter over time. A marginal interaction between token and age (β=-0.003, SE=0.002, t(425)=-1.89, p<0.060) suggests that these vowels are also converging in duration over apparent time. The model is unable to discern a difference in the average length of LOT and THOUGHT vowels as shown by the lack of a main effect of token for speakers in this racial grouping (β=-0.048, SE=0.037, t(430)=-1.30, p<0.195). The pattern for White speakers, however, seems to be moving in a different direction.

Among White speakers, the model shows that the LOT and THOUGHT vowels are different and growing more different over time. A main effect of token (β=0.116, SE=0.016, t(419)=7.40, p<0.0001) shows that THOUGHT is on average 25ms shorter than LOT. The interaction between token and age (β=0.002, SE=0.0007, t(419)=2.57, p<0.010) shows that this difference is increasing over apparent time.

The model reveals that there are also gender effects among White speakers. There is a main effect of gender (β=0.046, SE=0.023, t(425)=2.11, p<0.035) such that the LOT and THOUGHT vowels of White men are on average shorter than for White women. The interaction between token and gender (β=-0.073, SE=0.032, t(418)=-2.26, p<0.024) reveals that the length difference is not the same for LOT and THOUGHT vowels. While the difference between men and women's THOUGHT durations is about 17ms, the difference between men and women's LOT vowel is about 2ms. As can be seen in the top panel of figure 5 younger

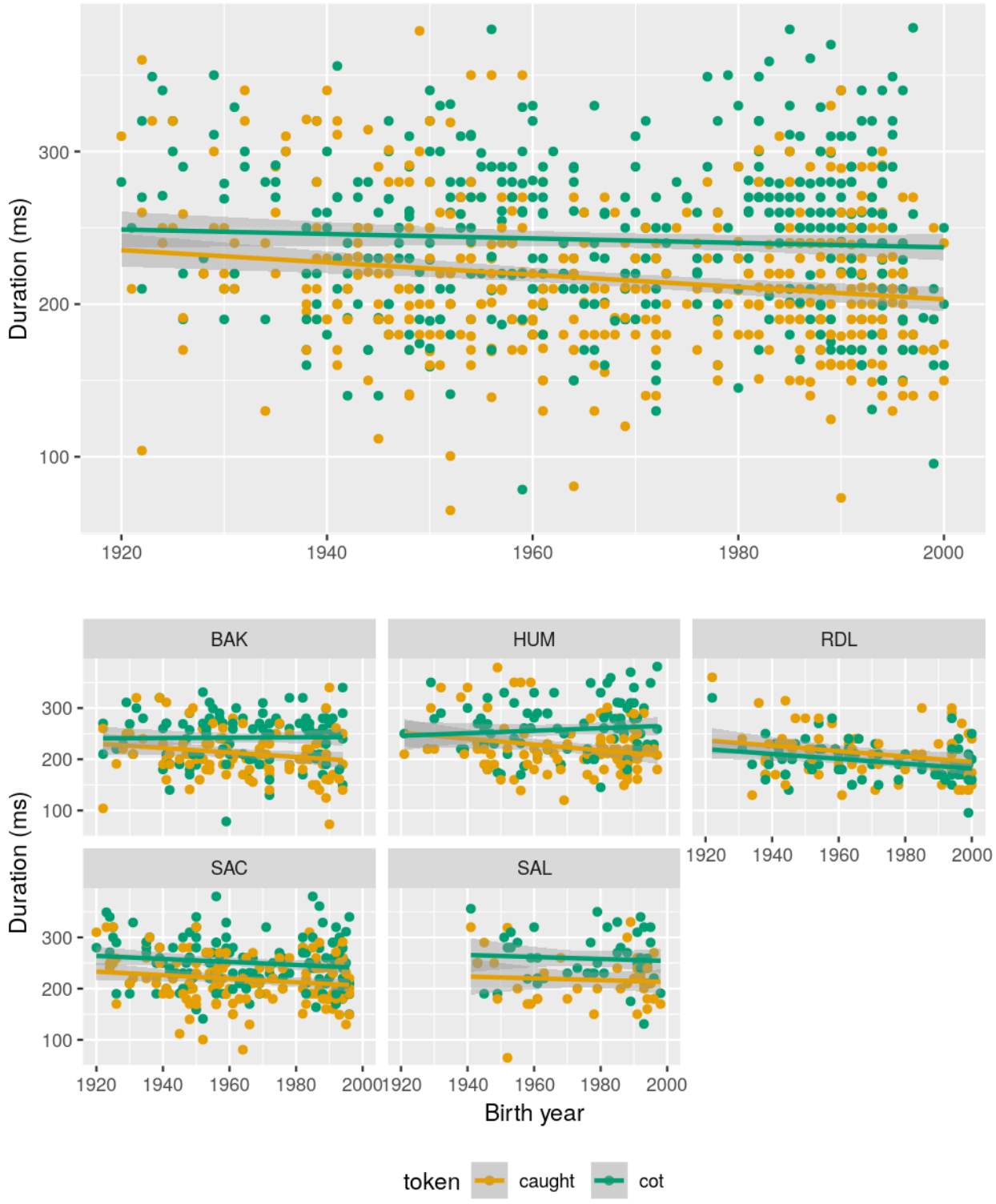

Figure 5: The duration, in milliseconds, of the LOT and THOUGHT vowels for speakers born in a given year. The bottom panel shows the same data broken out by field site.

speakers have a difference of about 40ms between the LOT and THOUGHT vowels.[13]

The model revealed a significant effect of segment (β=0.115, SE=0.012, t(445)=9.6, p<0.0001) such that THOUGHT is on average 25ms shorter than LOT. A significant interaction between token and birth year (β=0.001, SE=0.0006, t(449)=2.1, p<0.035) shows that this pattern is strengthening over apparent time; There is a main effect of birth year on vowel length (β=-0.001, SE=0.0004, t(450)=-3.3, p<0.001) as younger speakers tend to pronounce both vowels shorter, but THOUGHT seems to be shortening at a faster rate given the interaction. Men have a smaller difference in duration compared to women as shown by the significant interaction between token and gender (β=-0.065, SE=0.024, t(445)=-2.7, p<0.008).

In order to evaluate the ways these trends vary across field sites, the data was subsetted based upon site and a mixed effects linear model similar to the one above was fit to the subsetted data. The model for each field site predicted log duration of the vowel from the fixed effects of vowel, gender, and age with random intercepts by participant.[14] As can be seen in the bottom panel of figure 5, not every community seems to follow the general pattern identified above and this is corroborated by the models run for each site.

Only Bakersfield and Humboldt show a significant interaction between age and token. The interaction effects in Bakersfield (β=0.003, SE=0.001, t(101)=2.3, p<0.025) and in Humboldt (β=0.003, SE=0.001, t(91)=2.6, p<0.012) are both about twice as strong as the effect identified in the general model above. These two field sites differ in the influence of gender on language patterns. Bakersfield shows no main effect of gender and only a marginal interaction between token and gender (β=-0.102, SE=0.052, t(101)=-1.96, p<0.053). Humboldt however

[13] With an intercept of -1.5 for centered predictors and a main effect of token of 0.116, the log duration of LOT is $-1.5 + \frac{0.116}{2} = -1.44$ and $-1.5 - \frac{0.116}{2} = -1.56$ for THOUGHT. The distance from the intercept to the youngest speaker age is 32 years meaning a $0.001 \times 32 = 0.039$ increase in the effect of token for the youngest speaker. Thus the log duration of LOT for the youngest speaker is expected to be $-1.5 + \frac{0.116+0.039}{2} = -1.42$ and $-1.5 - \frac{0.116+0.039}{2} = -1.58$ for THOUGHT. As these are log durations, the exponentiation of them times 1000 gives us milliseconds, so the duration of LOT is 241ms and THOUGHT is 207ms giving a difference of 34ms which includes rounding errors. These calculations are corroborated graphically in figure 5 which shows a gap between the LOT and THOUGHT regression lines for the youngest speakers of between 30 and 50ms.

[14] Because of the small sample for some racial groups in each field site, race was not included as a predictor in individual field site models

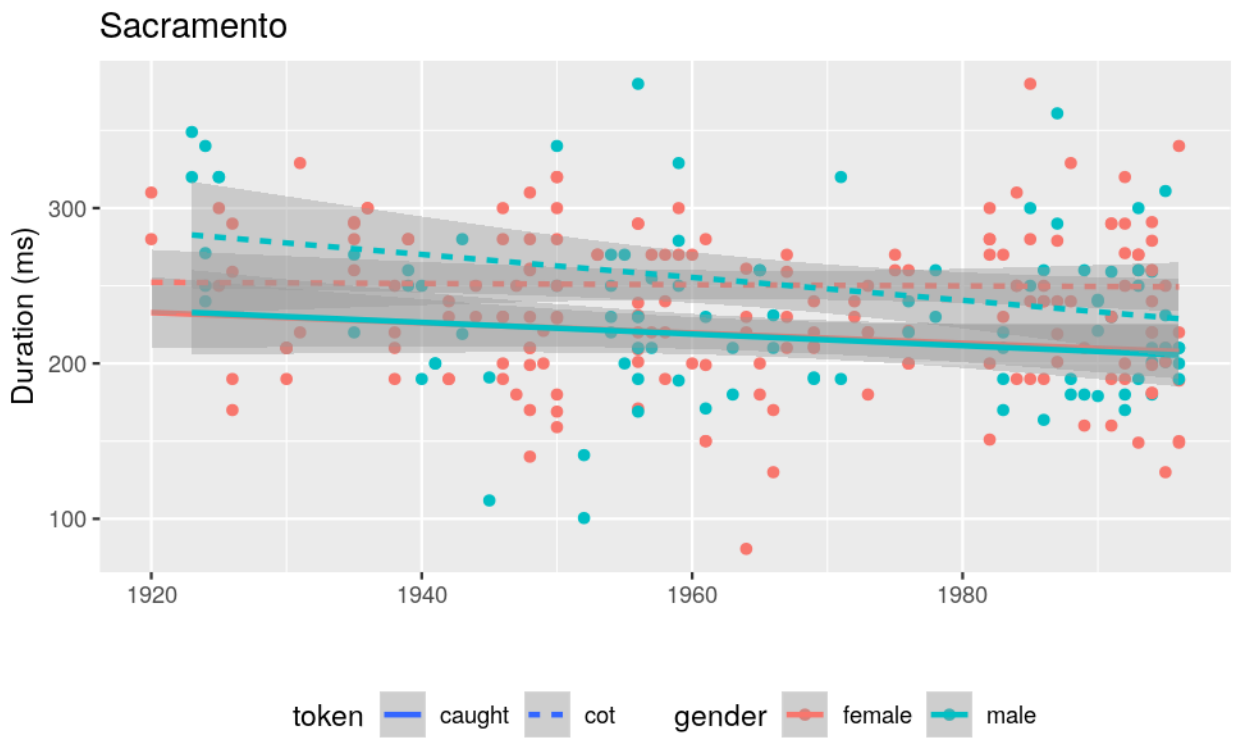

Figure 6: A graph of the marginal three way interaction between age, gender, and token in Sacramento.

shows a significant main effect of gender (β=0.07, SE=0.03, t(91)=2.0, p<0.045) such that women have shorter vowel durations in general, and a significant interaction between gender and token (β=-0.13, SE=0.05, t(91)=2.5, p<0.01) such that women have shorter THOUGHT durations than men.

Both Redlands and Sacramento show a main effect of age with the LOT and THOUGHT vowels in both sites getting shorter over time, but neither shows an interaction. Redlands shows a main effect of age (β=-0.002, SE=0.0009, t(75)=-2.5, p<0.01) about twice as strong as the effect across California and the main effect of age in Sacramento. For Sacramento the main effect of age (β=-0.001, SE=0.0007, t(133)=-2.0, p<0.047) is comparable to the main effect observed in the general model above. There is a marginal three way interaction between token, gender, and age (β=-0.003, SE=0.002, t(132)=-1.7, p<0.093) shown in figure 6. The marginal interaction suggests that the divergence pattern seen in Bakersfield and Humboldt seems to be occuring among Sacramento women, while men maintain or even possibly close

distance between the length of the vowels.

All the field sites show a significant effect of token though the nature of the effect differs across them. Humboldt was the only field site with an estimated main effect of token (β=0.14, SE=0.03, t(91)=5.5, p<0.0001) within two standard errors of the general effect. The estimated effects for Bakersfield (β=0.162, SE=0.027, t(101)=6.1, p<0.0001), Sacramento (β=0.15, SE=0.020, t(132)=7.5, p<0.0001), and Salinas (β=0.195, SE=0.052, t(30)=3.7, p<0.0008) were all more than two standard errors away from the general effect estimate, though Salinas was the only site whose estimate was above three standard errors.[15] Redlands was *below* two standard errors and had a significant main effect of token in the opposite direction of the other field sites (β=-0.059, SE=0.025, t(73)=-2.3, p<0.02) which shows that the LOT vowel in Redlands seems to be shorter than THOUGHT despite the opposite pattern occurring in the other four field sites.

4 Discussion

This study investigates the apparent merger in production of the LOT and THOUGHT vowels in California English and shows that the data are inconsistent with a definition of merger. By analyzing additional acoustic dimensions such as duration and formant dynamics, the data presented paint a more complicated picture of spectral overlap than previously thought. While there is ample evidence that the low back vowels in California are converging in formant space, this study presents evidence of divergence over apparent time in the duration of these vowels which is not predicted by a merger hypothesis. To account for this data, we propose that this pattern is better understood as a case of transphonologization.

If two vowels are merging, then they approach each other in acoustic space over time. Previous studies have supported the hypothesis by providing evidence that the LOT and THOUGHT vowels are approaching each other in formant space . Hall-Lew (2009) used Pillai

[15]Salinas was the smallest sample size by a rather large margin and this may explain the result. There were no other significant effects for Salinas.

scores, a measure of distributional overlap, to show that the acoustic range of San Franciscan LOT and THOUGHT vowels are increasing in their overlap. D'Onofrio et al. (2016) used the euclidean distance between formant point measurements to show that these vowels are approaching each other in formant space as well. The present study replicates these findings with a third methodology, euclidean distance between the vowels' formant means across their duration. Numerous other studies have found results similar to those previously discussed, and so there is strong evidence that the two vowels are converging in formant space across time.

Our second analysis further clarifies the nature of this convergence and provides evidence that the entire articulations of the LOT and THOUGHT vowels are converging over apparent time. This convergent pattern is not particularly strong or complete. For the youngest speakers, the formant trajectories of LOT and THOUGHT are still far apart. The remaining distance between the two vowels may close, or it may be the result a phenomenon not analyzed.[16] The gap between the formant trajectories of LOT and THOUGHT for the youngest speakers is not fatal to a merger hypothesis as it can be adequately explained based on existing data. Our definition of merger does not require complete overlap of the acoustic signal, rather, a near-merger can occur when two vowels are sufficiently close that their perceptual distinction is lost. Given the results in the formant space analysis, it is possible that there is enough of a merger in production that they are perceptually merged, even if the onset and offset of the vowels differ. While the second analysis complicates the apparent merger hypothesis, the results of the first two analyses can still be explained by a merger hypothesis.

Our third analysis shows that these vowels are diverging over apparent time, contrary to a merger hypothesis. Simultaneously with the formant space convergence, the duration difference between LOT and THOUGHT vowels has been increasing. This pattern suggests that the contrast between LOT and THOUGHT in formant space has shifted to a contrast

[16] We suspect it is the result of phonetic rounding, which differs between the two vowels and would result in different formant dynamics. Further investigation is needed, however, so we refrain from speculating further.

in length similar to Labov & Baranowski (2006). As LOT and THOUGHT began as vowels distinguished, in part, by height, the phonetic effects of the articulation resulted in a pattern whereby LOT was slightly longer than THOUGHT. Speaker-listeners recognized that this small difference in length provided redundant information to the vowel phoneme, and eventually phonologized it as a secondary cue producing a quasi-phonemic contrast (Kiparsky, 2016). Over time, the pattern first identified by DeCamp (1953) began to diminish the F1-F2 differences between the LOT and THOUGHT vowels in California removing the conditioning environment for the quasi-phonemes. As these vowels continued approaching each other in formant space , the speaker-listeners transfered the contrast to a more reliable dimension to maintain the distinction despite the loss of the formant space contrast in LOT and THOUGHT .

This account is consistent with previous empirical findings, despite coming to an alternative conclusion. We hypothesize that the length cue originated from phonetic length differences, and the direction of the increasing length distinction is in line with the predictions of that hypothesis. We would expect the shorter vowel to be the higher vowel, and the lower vowel to be the longer vowel in line with the expected phonetic effects. This is indeed the case. As would be expected in a situation of phonologization, the phonetic length difference has been enhanced with the shorter vowel—*caught*—becoming even shorter over apparent time. We suspect that this may represent a case of transphonologization (Hyman, 2013) whereby the formant contrast has been replaced by a length contrast, but variation evident from the data suggest a more complicated picture when race, gender, and location are taken into account.

Future work should test whether our hypothesized length distinction is perceptually salient. If a study were to show that speakers cannot reliably distinguish between LOT and THOUGHT vowels regardless of vowel duration, that would contradict these claims. Such a result would be highly unexpected. Firstly, the most recent perceptual work (Labov et al., 2006) tended to *not* find evidence of the low back merger in California English. The

results presented here suggest further divergence in the two decades since those experiments. Similarly the duration difference identified here is long enough to serve as a distinctive cue. Labov & Baranowski (2006) showed that for some Inland North speakers, what appeared to be a merger was in fact distinguishable to native speakers. Their distinctions were made based on duration differences of 50ms, which is comparable to the differences seen in our youngest speakers. Given previous perception results, the continued divergence since then, and evidence that differences this small can be distinctive, the length difference here is likely to be noticeable.

This work can be further improved upon through analyses of vowels not from wordlists and including more tokens per vowel class. The two main limitations of the work here are that the LOT and THOUGHT vowel classes were each represented by a single token extracted from a wordlist. Because of this restriction in our data set, effects such as phonological context were able to be abstracted away from. For example, the analysis of vowel dynamics presented here relies upon all tokens being in the same phonetic context. The use of wordlist data also aids in the rapid collection of data which has allowed the large sample size used in this analysis. While this has provided us with a number of advantages, it also raises questions about the use of this distinction in colloquial speech and the regularity of change across the lexicon.

5 Conclusion

This paper complicates previous claims of a merger in production of the California low back vowels by presenting evidence of contrast maintenance through development of a length contrast. We argue that the pattern of data is better explained as a transphonologization of a quasi-phonemic length contrast (Hyman, 2013; Kiparsky, 2016) rather than a merger. Previous studies which find convergence in formant space for the LOT and THOUGHT vowels are replicated, but simultaneous to this convergence we observe that the duration of these

vowels are diverging. This pattern of divergence suggests a new length-based contrast as a replacement for the formant-based contrast.

Our data suggest that further investigations into length contrasts in California will be fruitful. Further work should confirm these findings using additional tokens and elicitation tasks. Our findings show considerable variation among Californians, and data from less formal tasks may reveal additional sociolinguistic patterning. Because of the sustained interest of native speaker researchers in California English, there exists a wealth of data to explore these questions and further develop our understanding of language change.

References

Bates, D., Maechler, M., Bolker, B., Walker, S., et al. (2019). Linear Mixed-Effects Models using 'Eigen' and S4 [Computer software manual]. Comprehensive R Archive Network. Retrieved from `https://cran.r-project.org/web/packages/lme4/lme4.pdf`

Beckman, M. E. (1986). *Stress and non-stress accent.* Walter de Gruyter.

DeCamp, D. (1953). *The pronunciation of English in San Francisco.* University of California, Berkeley.

D'Onofrio, A., Eckert, P., Podesva, R. J., Pratt, T., & Van Hofwegen, J. (2016). The low vowels in California's central valley. *The Publication of the American Dialect Society*, *101*(1), 11–32.

Eckert, P. (2008). Where do ethnolects stop? *International Journal of Bilingualism*, *12*(1-2), 25–42.

Hall-Lew, L. (2009). *Ethnicity and phonetic variation in a San Francisco neighborhood* (Unpublished doctoral dissertation). Stanford University.

Herold, R. (1990). *Mechanisms of merger: The implementation and distribution of the low back merger in eastern pennsylvania* (Unpublished doctoral dissertation). University of Pennsylvania.

Hinton, L., Moonwomon, B., Bremner, S., Luthin, H., Van Clay, M., Lerner, J., & Corcoran, H. (1987). It's not just the Valley Girls: A study of California English. In *Annual Meeting of the Berkeley Linguistics Society* (Vol. 13, pp. 117–128). doi: doi:10.3765/bls.v13i0.1811

Holland, C. (2014). *Shifting or Shifted? The state of California vowels* (Unpublished doctoral dissertation). University of California, Davis.

Hyman, L. (2013). Enlarging the scope of phonologization. In A. C. L. Yu (Ed.), *Origins of sound change: Approaches to phonologization* (pp. 3–28). Oxford University Press Oxford, UK.

Kendall, T., & Thomas, E. (2018). Vowel Manipulation, Normalization, and Plotting [Computer software manual]. Comprehensive R Archive Network. Retrieved from `https://cran.r-project.org/web/packages/vowels/vowels.pdf`

Kennedy, R., & Grama, J. (2012). Chain shifting and centralization in California vowels: An acoustic analysis. *American Speech*, *87*(1), 39–56.

Kiparsky, P. (2016). Labov, sound change, and phonological theory. *Journal of Sociolinguistics*, *20*(4), 464–488.

Kirby, J. (2019). PraatSauce. Retrieved from `https://github.com/kirbyj/praatsauce/tree/aa9c2a4f29c6a29e18f782ae78a695525b54047b`

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2019). Tests in Linear Mixed Effects Models [Computer software manual]. Comprehensive R Archive Network. Retrieved from `https://cran.r-project.org/web/packages/lmerTest/lmerTest.pdf`

Labov, W. (1981). Resolving the neogrammarian controversy. *Language*, 267–308.

Labov, W. (1991). The three dialects of English. In P. Eckert (Ed.), *New Ways of Analyzing Sound Change* (pp. 1 – 44). New York: Academic Press.

Labov, W. (1994). *Principles of linguistic change: Volume 1: Internal factors. vol. 20 of.*

Labov, W., Ash, S., & Boberg, C. (2006). *The atlas of North American English: Phonetics, phonology and sound change.* Walter de Gruyter.

Labov, W., & Baranowski, M. (2006). 50 msec. *Language variation and change*, *18*(3), 223–240.

Martinet, A. (1952). Function, structure, and sound change. *Word*, *8*(1), 1–32.

Moonwomon, B. (1991). *Sound change in San Francisco English* (Unpublished doctoral dissertation). University of California, Berkeley.

Peterson, G. E., & Lehiste, I. (1960). Duration of syllable nuclei in English. *The Journal of the Acoustical Society of America*, *32*(6), 693–703.

Podesva, R. J., D'onofrio, A., Van Hofwegen, J., & Kim, S. K. (2015). Country ideology and the California vowel shift. *Language Variation and Change*, *27*(2), 157–186.

R Core Team. (2018). R: A Language and Environment for Statistical Computing [Computer software manual]. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from `https://www.R-project.org/`

Rosenfelder, I., Fruehwald, J., Evanini, K., Seyfarth, S., Gorman, K., Prichard, H., & Yuan, J. (2014). *FAVE (Forced Alignment and Vowel Extraction) Program Suite v1.2.2.* doi: doi:10.5281/zenodo.22281

Trudgill, P., & Foxcroft, T. (1978). On the sociolinguistics of vocalic mergers: Transfer and approximation in east anglia. *Sociolinguistic Patterns in British English*, 69–79.