Low-Back Merger in California English: Working Paper

Christian Brickhouse

March 27, 2019

Caveat

This draft represents a work in progress and therefore **you should not cite this work for any reason**. The analysis is still ongoing and so this document will change substantially and without notice. Further work may demonstrate significant errors in present or past texts, and so readers should be aware that this analysis is subject to change as more work is done. This document is not intended as a scholarly publication but as living documentation of the ongoing analysis.

Methods

Data were collected as part of the Voices of California project which conducts sociolinguistic interviews with life-long California residents. At the end of the interview, participants are asked to read a wordlist to ensure that vowels of itnerest are captured. This study analyzes that word list data from 570 California English speakers. The selection criteria excluded any person who lived outside of the fieldsite of interest for more than 3(?) years between the ages of 8(?) and 18, or for more than 6(?) years after 18. The field sites, with year of fieldwork, were Merced (2010), Redding (2011), Bakersfield (2012), Sacramento (2014), Salinas (2016), Humboldt Bay (2017), and Redlands (2018).

The wordlists were force aligned using the Penn Forced Aligner (citation), extracted by automated script (see extract_vowels_cj.praat), and analyzed using PraatSauce (citation). Each vowel was measured at 10 equidistant points providing change in values over time. Because each measurement represents one tenth of the vowel, it is time normalized and represents a position in the vowel rather than an absolute time into the vowel. The multiple measures for each vowel were collapsed for normalization as well as for analyses that do not look at vowel dynamics. This was done by taking the mean of all 10 measurements from the duration of the vowel, yielding a mean F1 and mean F2 within the vowel. These took the place of what would otherwise have been F1 and F2 measurements from the midpoint of the vowel. These means were used as the inputs to the Nearey normalization (citation) method implemented in the `library(vowels)` package.

Results

The sample used in this analysis comprises 570 speakers from 7 field sites.

```
## [1] "Number of participants per fieldsite:"
```

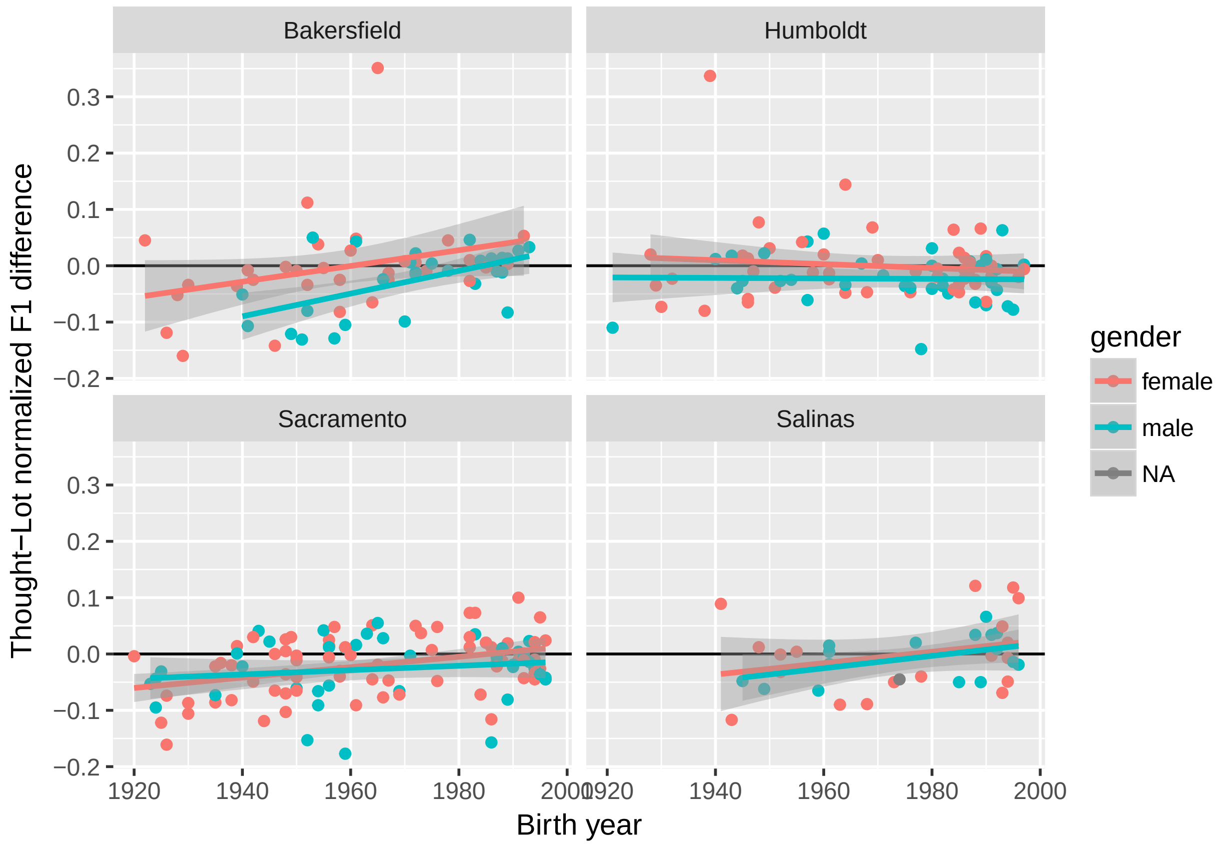
```
##
## BAK HUM MER RDL RED SAC SAL
## 111  96  59  81  97 139  80
```

Due to measurement errors, incomplete responses, or functional limitations each individual analysis may use a subset of this data.

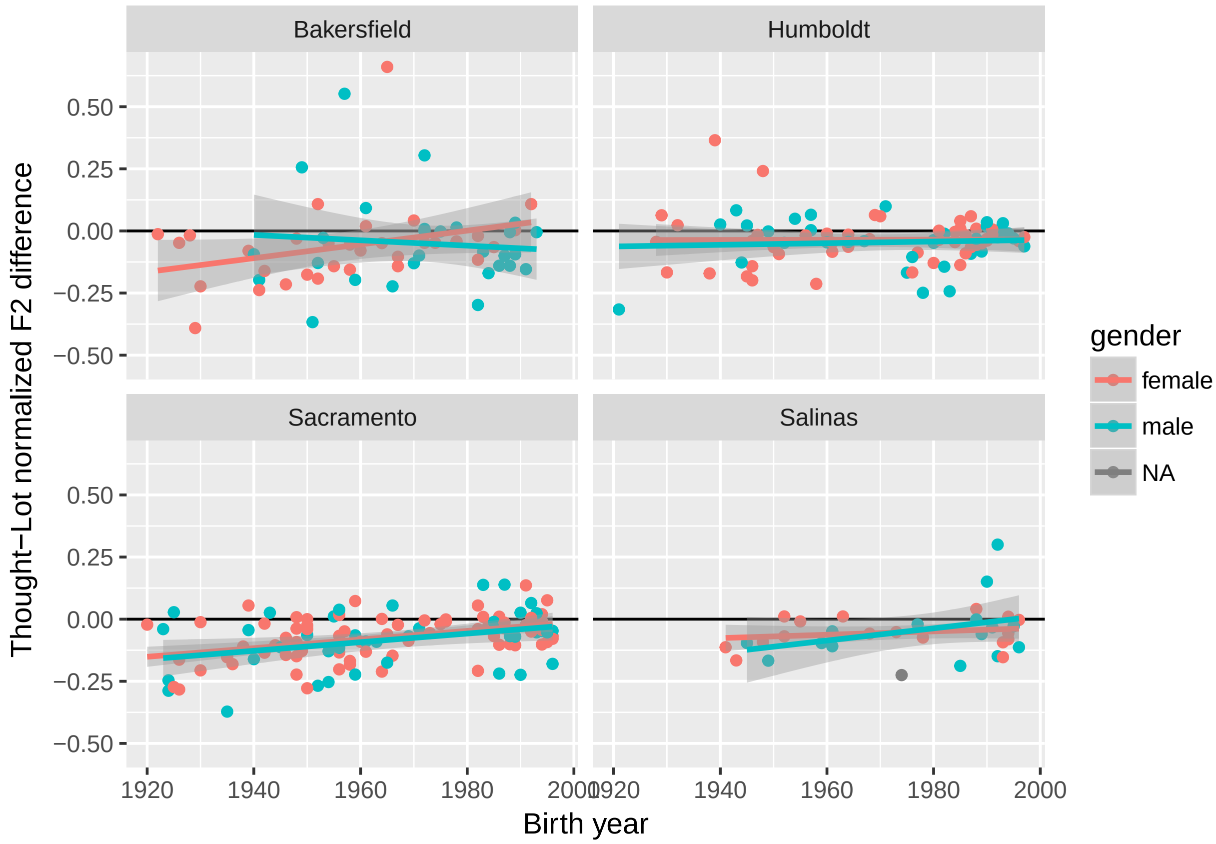
F1 and F2 space

Measurements of the F1 and F2 space have previously been used to argue for an apparent merger and so the data here should show similar patterns of decreasing distance between LOT and THOUGHT vowels in F1-F2 space. For each participant their normalized formant values for LOT were subtracted from the normalized formant values for THOUGHT so that the degree of overlap can be easily quanitified. If the value is 0 then they overlap each other perfectly.

```
## Warning: Removed 33 rows containing non-finite values (stat_smooth).
## Warning: Removed 33 rows containing missing values (geom_point).
```




```
## Warning: Removed 33 rows containing non-finite values (stat_smooth).
## Warning: Removed 33 rows containing missing values (geom_point).
```



In line with previous work we observe substantial similarity between LOT and THOUGHT vowels. The distance between F1 is NA in normalized formant space. The distance between F2 is NA. There is reason to consider that this is not indicative of a complete merger (i.e., not a near-merger) as both these means are significantly different from 0, however the difference may be caused by some areas which lack the merger and some that do.

```
model.f1 = lm(F1diff ~ site*gender*birthyear,data=data.f1f2)
summary(model.f1)$coefficients
```

```
##                                  Estimate   Std. Error    t value
## (Intercept)                 -2.7411801171 1.0537178754 -2.6014365
## siteHUM                      3.4258129973 1.3162822034  2.6026433
## siteSAC                      0.9080507145 1.2097392435  0.7506169
## siteSAL                      0.7414332884 1.6773761690  0.4420197
## gendermale                  -1.2651091078 1.7170860246 -0.7367768
## birthyear                    0.0013983487 0.0005382899  2.5977612
## siteHUM:gendermale           0.6462471733 2.1328237138  0.3030007
## siteSAC:gendermale           2.3182749554 1.9688150007  1.1774976
## siteSAL:gendermale           1.0824998972 2.7169489580  0.3984248
## siteHUM:birthyear           -0.0017461672 0.0006710378 -2.6021890
## siteSAC:birthyear           -0.0004750222 0.0006174665 -0.7693085
## siteSAL:birthyear           -0.0003863738 0.0008524689 -0.4532409
## gendermale:birthyear         0.0006204578 0.0008733123  0.7104650
## siteHUM:gendermale:birthyear -0.0003176577 0.0010837979 -0.2930968
## siteSAC:gendermale:birthyear -0.0011604467 0.0010013798 -1.1588478
## siteSAL:gendermale:birthyear -0.0005319080 0.0013777249 -0.3860771
##                                Pr(>|t|)
```

```
## (Intercept)                   0.009780623
## siteHUM                       0.009747099
## siteSAC                       0.453518101
## siteSAL                       0.658818618
## gendermale                    0.461879416
## birthyear                     0.009883359
## siteHUM:gendermale            0.762115904
## siteSAC:gendermale            0.240004027
## siteSAL:gendermale            0.690622892
## siteHUM:birthyear             0.009759707
## siteSAC:birthyear             0.442363322
## siteSAL:birthyear             0.650728835
## gendermale:birthyear          0.478011429
## siteHUM:gendermale:birthyear 0.769666825
## siteSAC:gendermale:birthyear 0.247512712
## siteSAL:gendermale:birthyear 0.699734885
```

In order to separate out the effects of demographic and social factors on degree of overlap, a linear model was constructed to predict the normalized F1 difference from field site, gender, and birthyear as well as their interactions. The results of the model show that younger Californians produce LOT and THOUGHT vowels closer together in the F1 dimension. There is a significant main effect of birth year such that younger Californians produce closer first formants. There is also a main effect such that speakers from Humboldt tend to produce LOT vowels higher than THOUGHT vowels. There is also a significant interaction such that the effect of birth year in Humboldt is in the direction of greater overlap over time (it is significant because Humboldt has a very positive difference, meaning their LOT vowel was higher than THOUGHT and so the effect of birth year, which would usually increase scores to bring them closer to zero needs to decrease the scores in Humboldt).

```
model.f2 = lm(F2diff ~ site*gender*birthyear,data=data.f1f2)
summary(model.f2)$coefficients[,1:3]
```

```
##                                   Estimate  Std. Error    t value
## (Intercept)                   -5.482851391 2.086394448 -2.6279074
## siteHUM                        5.392502427 2.606280054  2.0690418
## siteSAC                        1.996504341 2.395321651  0.8335016
## siteSAL                        4.092588626 3.321257433  1.2322407
## gendermale                     7.560992071 3.399884193  2.2238969
## birthyear                      0.002769712 0.001065831  2.5986414
## siteHUM:gendermale            -8.169730535 4.223057859 -1.9345533
## siteSAC:gendermale            -7.568897548 3.898315463 -1.9415816
## siteSAL:gendermale           -11.087751100 5.379644167 -2.0610566
## siteHUM:birthyear             -0.002741816 0.001328676 -2.0635700
## siteSAC:birthyear             -0.001032727 0.001222603 -0.8446952
## siteSAL:birthyear             -0.002092338 0.001687915 -1.2395990
## gendermale:birthyear          -0.003849187 0.001729186 -2.2260114
## siteHUM:gendermale:birthyear   0.004152834 0.002145954  1.9351924
## siteSAC:gendermale:birthyear   0.003847764 0.001982763  1.9406067
## siteSAL:gendermale:birthyear   0.005636558 0.002727938  2.0662335
```

Another linear model was constructed to predict the normalized F2 difference from field site, gender, and birth year. There is a main effect of site such that speakers in Humboldt have overlapping second formants and Redding speakers produce their LOT vowel much lower than speakers at other field sites (though this result should be interpreted with caution due to the very low number of speakers in that sample). There is a main effect of gender such that men seem to have more similar second formants than do women. There is also a small effect of birth year such that younger speakers have closer second formants. The effect of birth year however doesn't hold in Humboldt where similarity of second formants seems to be unaffected by age.

There is an interaction such that younger Redding speakers seem to have closer second formants but the same caveat about sample size applies there.

Gender was an important predictor in F2 distance as it interacted with a number of other predictors, so in order to better interpret those interactions a simple effects analysis was conducted.

```
model.f2.simple = lm(F2diff ~ site*gender*birthyear-site,data=data.f1f2)
summary(model.f2.simple)$coefficients[,1:3]
```

```
##                                  Estimate  Std. Error    t value
## (Intercept)                  -5.482851391 2.086394448 -2.6279074
## gendermale                    7.560992071 3.399884193  2.2238969
## birthyear                     0.002769712 0.001065831  2.5986414
## siteHUM:genderfemale          5.392502427 2.606280054  2.0690418
## siteSAC:genderfemale          1.996504341 2.395321651  0.8335016
## siteSAL:genderfemale          4.092588626 3.321257433  1.2322407
## siteHUM:gendermale           -2.777228108 3.322878566 -0.8357898
## siteSAC:gendermale           -5.572393206 3.075597118 -1.8118086
## siteSAL:gendermale           -6.995162474 4.231999577 -1.6529214
## siteHUM:birthyear            -0.002741816 0.001328676 -2.0635700
## siteSAC:birthyear            -0.001032727 0.001222603 -0.8446952
## siteSAL:birthyear            -0.002092338 0.001687915 -1.2395990
## gendermale:birthyear         -0.003849187 0.001729186 -2.2260114
## siteHUM:gendermale:birthyear  0.004152834 0.002145954  1.9351924
## siteSAC:gendermale:birthyear  0.003847764 0.001982763  1.9406067
## siteSAL:gendermale:birthyear  0.005636558 0.002727938  2.0662335
```

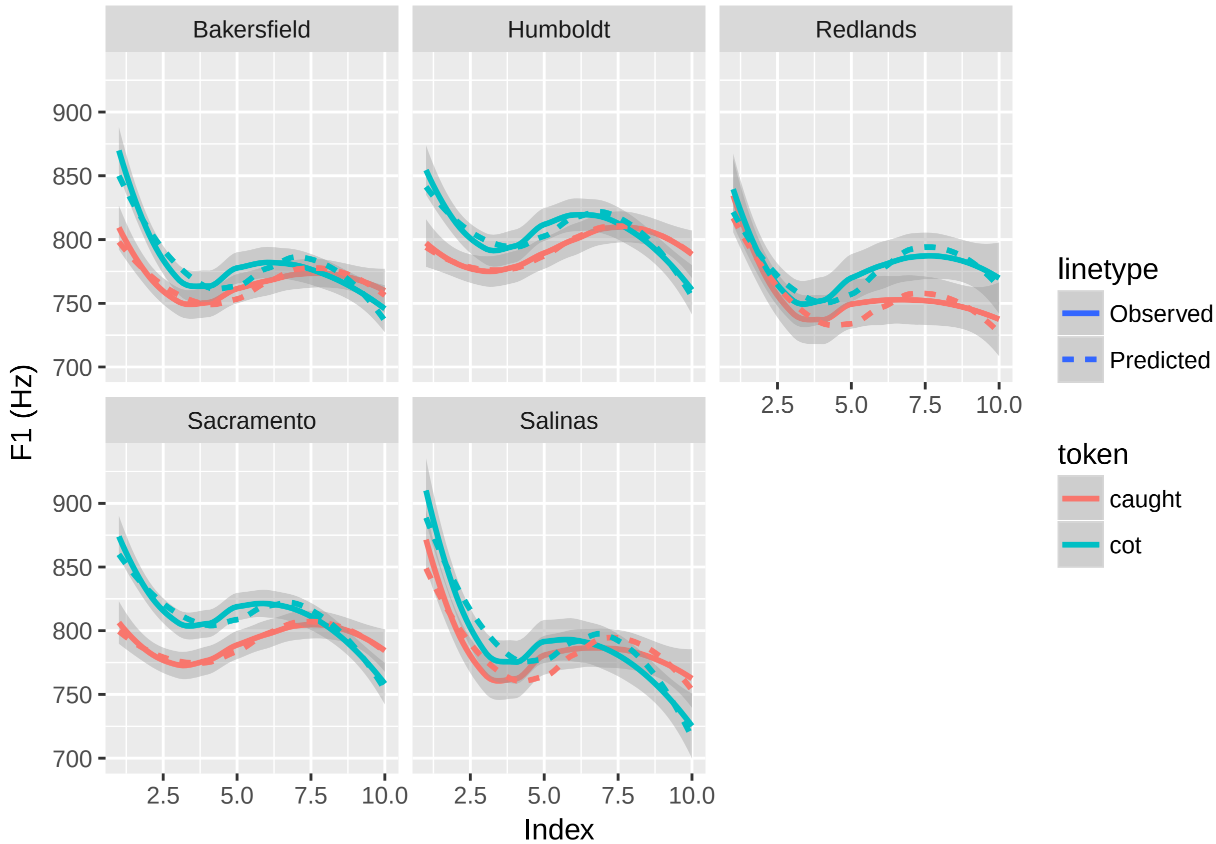
The marginal interaction between Humboldt and gender seems to be driven by women who have a great deal of overlap in their second formants. The interaction between Salinas and gender on the other hand seems to be driven by men who produce more distinct second formants.

Vowel Dynamics

Vowels are rarely static over time, and the pattern of movement during articulation can serve as a cue by which speakers distinguish vowel phonemes. Previous work has found that the F1 and F2 values of LOT and THOUGHT vowel midpoints are converging, and the previous analysis replicated this finding using averages across the vowels. Despite this apparent overlap it may be possible that the two vowel classes are still separate and distinguished by their dynamics. To test this hypothesis the differences in F1 and F2 paths between LOT and THOUGHT vowels were compared.

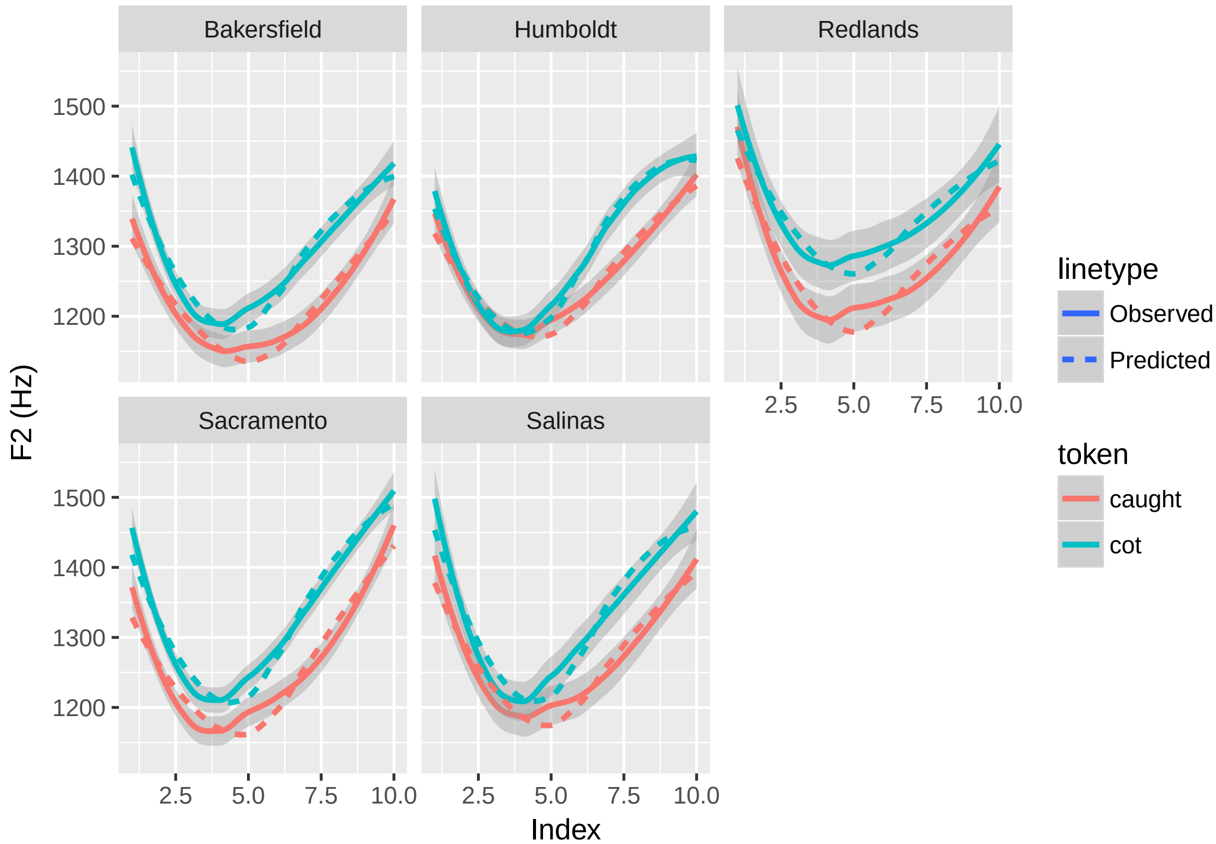
```
## Warning: Removed 119 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 230 rows containing non-finite values (stat_smooth).
```




```
## Warning: Removed 119 rows containing non-finite values (stat_smooth).

## Warning: Removed 230 rows containing non-finite values (stat_smooth).
```



In order to analyze the differences in vowel dynamics the trajectories of each speaker's vowels were modelled using a discrete cosine transform (DCT). The trajectory over time can be conceptualized as a wave, and this wave can be described by the superposition of multiple simple waves. These trajectories then, instead of being represented by a series of measurements, can be represented by the amplitudes of each of those simple waves. The DCT applies this process using cosine waves and provides a number of coefficients which represent the amplitudes of these waves. By using the DCT the representaionts of vowel trajectories can be reduced from 10 point measurements to a few coefficients.

To test the hypothesis that vowel dynamics are used to distinguish between LOT and THOUGHT vowels, a logistic regression was fitted to the data predicting the token from the first 4 DCT coefficients. As the model predicts whether the token is LOT or THOUGHT, a significant predictor is able to reliably distinguish between the vowels. The predictors of the model were the first four DCT coefficients as well as their interactions with gender, birthyear, and field site. While the model allowed a four way interaction between the three social variables and a given DCT coefficient, it did not consider the main effects of or interactions between the social variables in predicting the token.

```
model.dct.F1 = glm(token~(DCT1+DCT2+DCT3+DCT4)*(birthyear*gender*site)-(birthyear*gender*site),
                   data=F1_dct_coeffs,
                   family="binomial")
summary(model.dct.F1)$coefficients[,1:3]
```

```
##                                       Estimate Std. Error      z value
## (Intercept)                        0.008606313  0.0967287  0.088973725
## DCT1                              -0.267575711  0.2336319 -1.145287534
## DCT2                               0.531257973  0.2632743  2.017887750
## DCT3                              -0.397664102  0.2739829 -1.451419596
## DCT4                               0.394384800  0.2926605  1.347584675
```

```
## DCT1:birthyear                   -0.241364625  0.2819621 -0.856017957
## DCT1:gendermale                   0.879853268  0.3783805  2.325313167
## DCT1:siteHUM                      0.513802419  0.3273682  1.569494043
## DCT1:siteRDL                      0.244478101  0.2966062  0.824251462
## DCT1:siteSAC                      0.549631924  0.3065872  1.792742450
## DCT1:siteSAL                     -0.312905121  0.5000159 -0.625790331
## DCT2:birthyear                    0.509290327  0.3370561  1.510995693
## DCT2:gendermale                   0.236699181  0.4362783  0.542541755
## DCT2:siteHUM                      1.319996526  0.5209970  2.533597218
## DCT2:siteRDL                     -0.918428502  0.3966343 -2.315555014
## DCT2:siteSAC                      0.628481210  0.3881455  1.619189697
## DCT2:siteSAL                      0.379523217  0.7254874  0.523128605
## DCT3:birthyear                   -0.424797304  0.2631327 -1.614384387
## DCT3:gendermale                   0.021852705  0.4564764  0.047872590
## DCT3:siteHUM                     -0.417430238  0.4018422 -1.038791312
## DCT3:siteRDL                      0.138292557  0.4045521  0.341841167
## DCT3:siteSAC                     -0.140155890  0.3416502 -0.410232094
## DCT3:siteSAL                      0.037119757  0.5155023  0.072006963
## DCT4:birthyear                   -0.173403426  0.2612962 -0.663627866
## DCT4:gendermale                   0.089453628  0.4851408  0.184386958
## DCT4:siteHUM                     -0.008476944  0.4007342 -0.021153531
## DCT4:siteRDL                     -0.155566943  0.4247459 -0.366258813
## DCT4:siteSAC                     -0.079697116  0.3828812 -0.208151042
## DCT4:siteSAL                     -0.426285937  0.5455588 -0.781374918
## DCT1:birthyear:gendermale        -0.016804941  0.4101412 -0.040973546
## DCT1:birthyear:siteHUM            0.474936700  0.4027057  1.179364379
## DCT1:birthyear:siteRDL            0.316773229  0.3312416  0.956320909
## DCT1:birthyear:siteSAC           -0.087263955  0.3577329 -0.243936031
## DCT1:birthyear:siteSAL            0.045839311  0.5486812  0.083544520
## DCT1:gendermale:siteHUM          -0.968152038  0.5901664 -1.640472868
## DCT1:gendermale:siteRDL          -0.242758241  0.5336310 -0.454917822
## DCT1:gendermale:siteSAC          -0.998945390  0.4840401 -2.063765612
## DCT1:gendermale:siteSAL           0.053343991  0.7957572  0.067035511
## DCT2:birthyear:gendermale        -0.166479963  0.5055513 -0.329303787
## DCT2:birthyear:siteHUM           -0.200750044  0.5805230 -0.345808924
## DCT2:birthyear:siteRDL           -0.394724508  0.4442545 -0.888509957
## DCT2:birthyear:siteSAC           -0.420312783  0.4378878 -0.959864152
## DCT2:birthyear:siteSAL           -0.268845012  0.8511305 -0.315868160
## DCT2:gendermale:siteHUM          -0.773782045  0.9175307 -0.843330946
## DCT2:gendermale:siteRDL          -0.355402495  0.6889495 -0.515861463
## DCT2:gendermale:siteSAC          -0.390414523  0.6391204 -0.610862267
## DCT2:gendermale:siteSAL          -0.967817607  1.1910544 -0.812572103
## DCT3:birthyear:gendermale         0.409202963  0.5407754  0.756696733
## DCT3:birthyear:siteHUM            0.671228414  0.4168547  1.610221703
## DCT3:birthyear:siteRDL            0.376983732  0.3982598  0.946577438
## DCT3:birthyear:siteSAC            0.587246749  0.3199840  1.835237740
## DCT3:birthyear:siteSAL            0.595034940  0.5285078  1.125877284
## DCT3:gendermale:siteHUM          -0.315019259  0.7607934 -0.414066737
## DCT3:gendermale:siteRDL           0.420422305  0.7272366  0.578109354
## DCT3:gendermale:siteSAC          -0.018601733  0.5830448 -0.031904464
## DCT3:gendermale:siteSAL           0.253484445  0.9024042  0.280899013
## DCT4:birthyear:gendermale        -0.378036463  0.5571573 -0.678509379
## DCT4:birthyear:siteHUM           -0.068683092  0.4070346 -0.168740166
## DCT4:birthyear:siteRDL            0.158925080  0.4062326  0.391216912
```

```
## DCT4:birthyear:siteSAC            0.112137567  0.3491658  0.321158504
## DCT4:birthyear:siteSAL            0.310883881  0.5736330  0.541956076
## DCT4:gendermale:siteHUM           0.114024514  0.6651667  0.171422469
## DCT4:gendermale:siteRDL          -0.246877944  0.7795542 -0.316691205
## DCT4:gendermale:siteSAC          -0.050452731  0.6584002 -0.076629270
## DCT4:gendermale:siteSAL           0.135321798  0.9953048  0.135960155
## DCT1:birthyear:gendermale:siteHUM  0.282263714  0.6609540  0.427055030
## DCT1:birthyear:gendermale:siteRDL  0.254063606  0.4982901  0.509870868
## DCT1:birthyear:gendermale:siteSAC  0.567223937  0.5225471  1.085498122
## DCT1:birthyear:gendermale:siteSAL -0.428360903  0.8487845 -0.504675717
## DCT2:birthyear:gendermale:siteHUM -0.449666791  0.9641260 -0.466398357
## DCT2:birthyear:gendermale:siteRDL  0.418863041  0.6918701  0.605407060
## DCT2:birthyear:gendermale:siteSAC -0.185379589  0.6566517 -0.282310368
## DCT2:birthyear:gendermale:siteSAL  1.049516015  1.3085090  0.802070167
## DCT3:birthyear:gendermale:siteHUM  0.673683856  0.8654750  0.778397845
## DCT3:birthyear:gendermale:siteRDL -0.444487020  0.7418635 -0.599149368
## DCT3:birthyear:gendermale:siteSAC  0.006144352  0.6506291  0.009443709
## DCT3:birthyear:gendermale:siteSAL -1.208664434  0.9778606 -1.236029389
## DCT4:birthyear:gendermale:siteHUM -0.197204007  0.7633665 -0.258334642
## DCT4:birthyear:gendermale:siteRDL  0.469526380  0.7753376  0.605576719
## DCT4:birthyear:gendermale:siteSAC  0.063728338  0.6916185  0.092143766
## DCT4:birthyear:gendermale:siteSAL -0.360416162  1.1294515 -0.319107250
```

There is no general distinction between LOT and THOUGHT in height movement, however there are a number of significant interactions. To compensate for multiple comparisons, the alpha level for these analyses has been Bonferroni corrected to a 0.025 level and so a number of trends below the 0.05 level are not considered significant here. The interaction between gender and the first DCT coefficient is significant to the 0.025 level and shows that men have a lower average first formant frequency. This effect is easily explained by the use of non-normalized vowel formants in this analysis due to men, in general, having lower formant frequencies. Additionally, there are significant interactions between the second DCT coefficient and field site. To probe these interactions further, the data from Humboldt, Redlands, and Bakersfield were analyzed separately. It appears that the interactions are driven by no difference in vowel dynamics for Redlands speakers, but differences for speakers in Bakersfield and Humboldt.

```
model.dct.F2 = glm(token~(DCT1+DCT2+DCT3+DCT4)*(birthyear*gender*site)-(birthyear*gender*site),
                   data=F2_dct_coeffs,
                   family="binomial")
summary(model.dct.F2)$coefficients[,1:3]
```

```
##                                     Estimate Std. Error     z value
## (Intercept)                       0.17153175 0.09716485  1.76536836
## DCT1                              0.31517286 0.27764359  1.13517067
## DCT2                             -0.22600650 0.37072000 -0.60964204
## DCT3                              0.60988394 0.34143520  1.78623625
## DCT4                              0.98844660 0.33673702  2.93536663
## DCT1:birthyear                   -0.33922700 0.32894811 -1.03124775
## DCT1:gendermale                   0.81387461 0.42780461  1.90244471
## DCT1:siteHUM                     -0.16507680 0.42568031 -0.38779524
## DCT1:siteRDL                     -0.27438790 0.32828342 -0.83582624
## DCT1:siteSAC                      0.26907804 0.39080462  0.68852316
## DCT1:siteSAL                     -0.40456652 0.61189812 -0.66116647
## DCT2:birthyear                   -0.24907471 0.39118592 -0.63671695
## DCT2:gendermale                   0.44680520 0.51842271  0.86185499
## DCT2:siteHUM                     -0.78251404 0.52554968 -1.48894399
## DCT2:siteRDL                      0.26973876 0.48546210  0.55563300
```

```
## DCT2:siteSAC                      -0.71385506 0.51806535 -1.37792473
## DCT2:siteSAL                      -0.80495361 0.80038643 -1.00570622
## DCT3:birthyear                     0.36674635 0.33712528  1.08786369
## DCT3:gendermale                   -1.86741265 0.55221942 -3.38164972
## DCT3:siteHUM                      -0.06359950 0.47778703 -0.13311266
## DCT3:siteRDL                      -0.80036174 0.54919235 -1.45734320
## DCT3:siteSAC                      -0.99943769 0.42753173 -2.33769244
## DCT3:siteSAL                      -1.04487632 0.72508436 -1.44104104
## DCT4:birthyear                     0.17327153 0.34976957  0.49538766
## DCT4:gendermale                   -0.08439344 0.47724399 -0.17683501
## DCT4:siteHUM                       0.99946252 0.56142649  1.78021974
## DCT4:siteRDL                      -1.14220575 0.58503996 -1.95235509
## DCT4:siteSAC                       0.54866004 0.48804790  1.12419301
## DCT4:siteSAL                       0.05601056 0.76940786  0.07279697
## DCT1:birthyear:gendermale          0.47511569 0.49749488  0.95501624
## DCT1:birthyear:siteHUM             0.32868561 0.46265973  0.71042623
## DCT1:birthyear:siteRDL             0.45840635 0.37255496  1.23043953
## DCT1:birthyear:siteSAC            -0.40249694 0.41759036 -0.96385593
## DCT1:birthyear:siteSAL            -0.06351714 0.63371264 -0.10023020
## DCT1:gendermale:siteHUM           -0.32018914 0.62116272 -0.51546740
## DCT1:gendermale:siteRDL           -0.23556242 0.53256359 -0.44231792
## DCT1:gendermale:siteSAC           -0.71040423 0.58568456 -1.21294682
## DCT1:gendermale:siteSAL            0.05172560 1.00599451  0.05141737
## DCT2:birthyear:gendermale          0.35898087 0.60141527  0.59689350
## DCT2:birthyear:siteHUM             0.11488298 0.53214315  0.21588735
## DCT2:birthyear:siteRDL            -0.03186269 0.48296606 -0.06597294
## DCT2:birthyear:siteSAC            -0.10778907 0.52570254 -0.20503815
## DCT2:birthyear:siteSAL            -0.86613657 0.89554051 -0.96716626
## DCT2:gendermale:siteHUM           -0.21211221 0.80300604 -0.26414771
## DCT2:gendermale:siteRDL           -0.76637282 0.70818356 -1.08216691
## DCT2:gendermale:siteSAC            0.02021739 0.72630042  0.02783613
## DCT2:gendermale:siteSAL            1.41763154 1.92006354  0.73832532
## DCT3:birthyear:gendermale         -0.21951984 0.63586944 -0.34522785
## DCT3:birthyear:siteHUM            -0.19499266 0.48441056 -0.40253593
## DCT3:birthyear:siteRDL            -0.18405546 0.56927932 -0.32331309
## DCT3:birthyear:siteSAC            -0.13791026 0.42125414 -0.32738019
## DCT3:birthyear:siteSAL             0.60863296 0.82470159  0.73800387
## DCT3:gendermale:siteHUM            0.90672133 0.73778091  1.22898454
## DCT3:gendermale:siteRDL            1.66759686 0.78177814  2.13308198
## DCT3:gendermale:siteSAC            2.35872524 0.65766109  3.58653609
## DCT3:gendermale:siteSAL            1.84686953 1.92078811  0.96151653
## DCT4:birthyear:gendermale         -0.37209074 0.58061925 -0.64085155
## DCT4:birthyear:siteHUM             0.46990659 0.55305533  0.84965566
## DCT4:birthyear:siteRDL            -0.28135310 0.58746077 -0.47893088
## DCT4:birthyear:siteSAC             0.27426175 0.47649524  0.57558129
## DCT4:birthyear:siteSAL             0.33871015 0.79657095  0.42521027
## DCT4:gendermale:siteHUM           -1.47660319 0.71897823 -2.05375229
## DCT4:gendermale:siteRDL            0.65213935 0.77571954  0.84068960
## DCT4:gendermale:siteSAC           -0.99849802 0.66381745 -1.50417561
## DCT4:gendermale:siteSAL           -2.01706521 1.62873086 -1.23842758
## DCT1:birthyear:gendermale:siteHUM -0.30135064 0.69553883 -0.43326214
## DCT1:birthyear:gendermale:siteRDL -0.39210138 0.56848824 -0.68972645
## DCT1:birthyear:gendermale:siteSAC -0.07276182 0.63018513 -0.11546102
## DCT1:birthyear:gendermale:siteSAL -0.95624364 1.02969198 -0.92866960
```

```
## DCT2:birthyear:gendermale:siteHUM  0.87783635 0.91532465  0.95904371
## DCT2:birthyear:gendermale:siteRDL -0.04459653 0.72074732 -0.06187541
## DCT2:birthyear:gendermale:siteSAC -0.33211798 0.79793567 -0.41622150
## DCT2:birthyear:gendermale:siteSAL -0.61587494 1.84872451 -0.33313505
## DCT3:birthyear:gendermale:siteHUM -0.21078828 0.83786254 -0.25157860
## DCT3:birthyear:gendermale:siteRDL -0.00517511 0.84272133 -0.00614095
## DCT3:birthyear:gendermale:siteSAC -0.38944273 0.72473128 -0.53736156
## DCT3:birthyear:gendermale:siteSAL -0.27565737 1.81269874 -0.15207015
## DCT4:birthyear:gendermale:siteHUM -0.23181062 0.77693359 -0.29836607
## DCT4:birthyear:gendermale:siteRDL  0.46018694 0.85034547  0.54117645
## DCT4:birthyear:gendermale:siteSAC  0.77326447 0.75097098  1.02968622
## DCT4:birthyear:gendermale:siteSAL  1.59016538 1.64201606  0.96842255
```

The distinction between cot and caught however does seem to be generally maintained in the F2 dimension by a main effect of the fourth DCT coefficient. This coefficient represents deviations from the average formant value at high frequency wavelengths and especially small deviations towards the ends of the trajectory. This can be seen in the plot above. The F2 trajectories, across field sites, seem to diverge about half way through the vowel. There also appears to be an effect of gender on the trajectories of LOT and THOUGHT vowels as demonstrated by the interaction between gender and the third DCT coefficient. This interaction appears to be driven by men having greater movement in their trajectories for THOUGHT tokens compared to women. This effect seems to not be true of Sacramento however, as the three way interaction between the third DCT coefficient, gender, and field site counteracts the effect of the coefficient and gender itneraction.

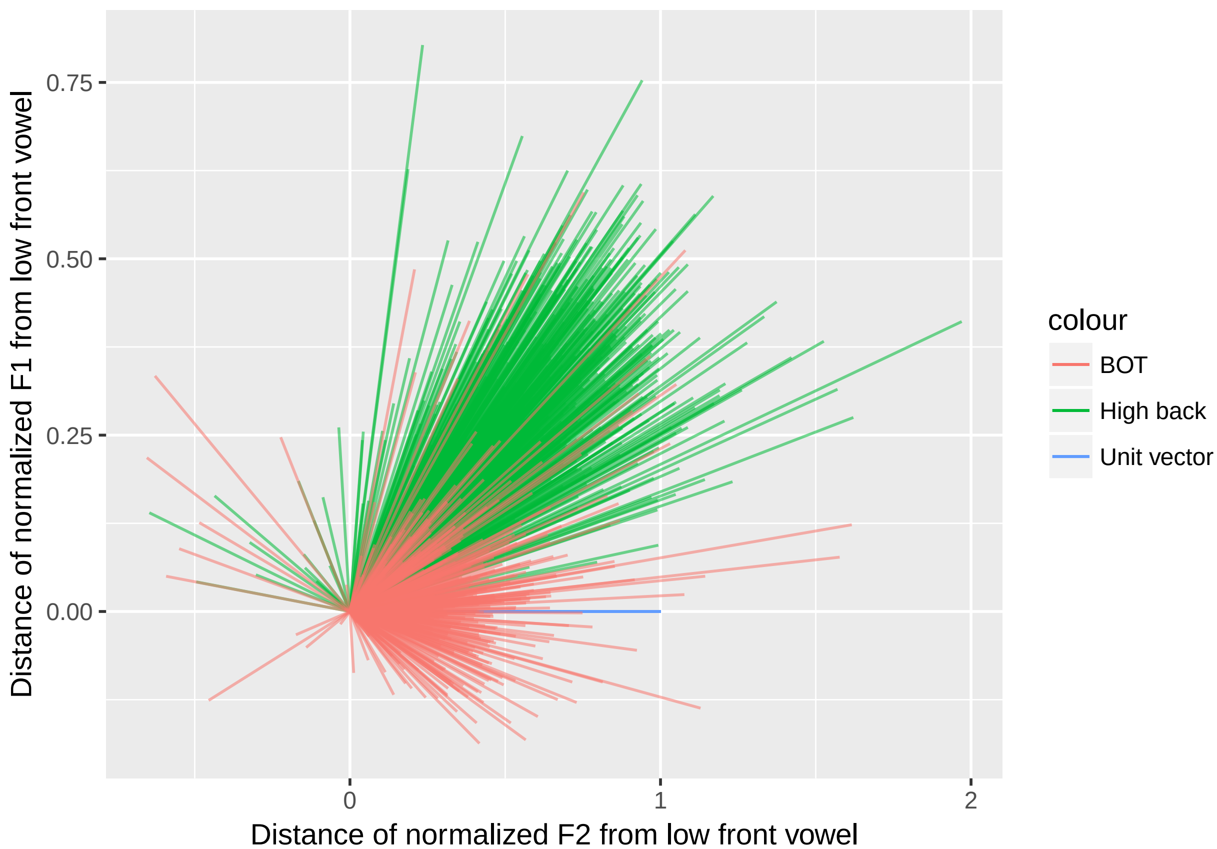
Given this analysis two possibilities present themselves: the LOT-THOUGHT merger is in fact a near merger or a phonological distinction is maintained by differences in vowel trajectories. The differences between these two possibilities can be resolved through a study of Californians' abilities to perceptually distinguish between these tokens when systematically manipulated by trajectory. If participants in this study are able to accurately distinguish between LOT and THOUGHT tokens based upon vowel trajectory then this would be evidence against the heretofore hypothesized merger.

This also represents an advantage of using the DCT for analyzing vowel trajectories as synthetic vowels can be created based upon the distributions observed. As the DCT has a known inverse function, the parameters found here can be used to create vowels with formant patterns at points of interest in the distribution. For example tokens in the tail of the distribution, which empirically are rare, can be created in great number to increase statistical power. Similarly, in areas where the distributions overlap, vowels at intervals not attested in the data can be synthesized to investigate in a more controlled fashion the point at which vowel identification begins to shift (should these be perceived as different vowels).

Vowel Space Shape

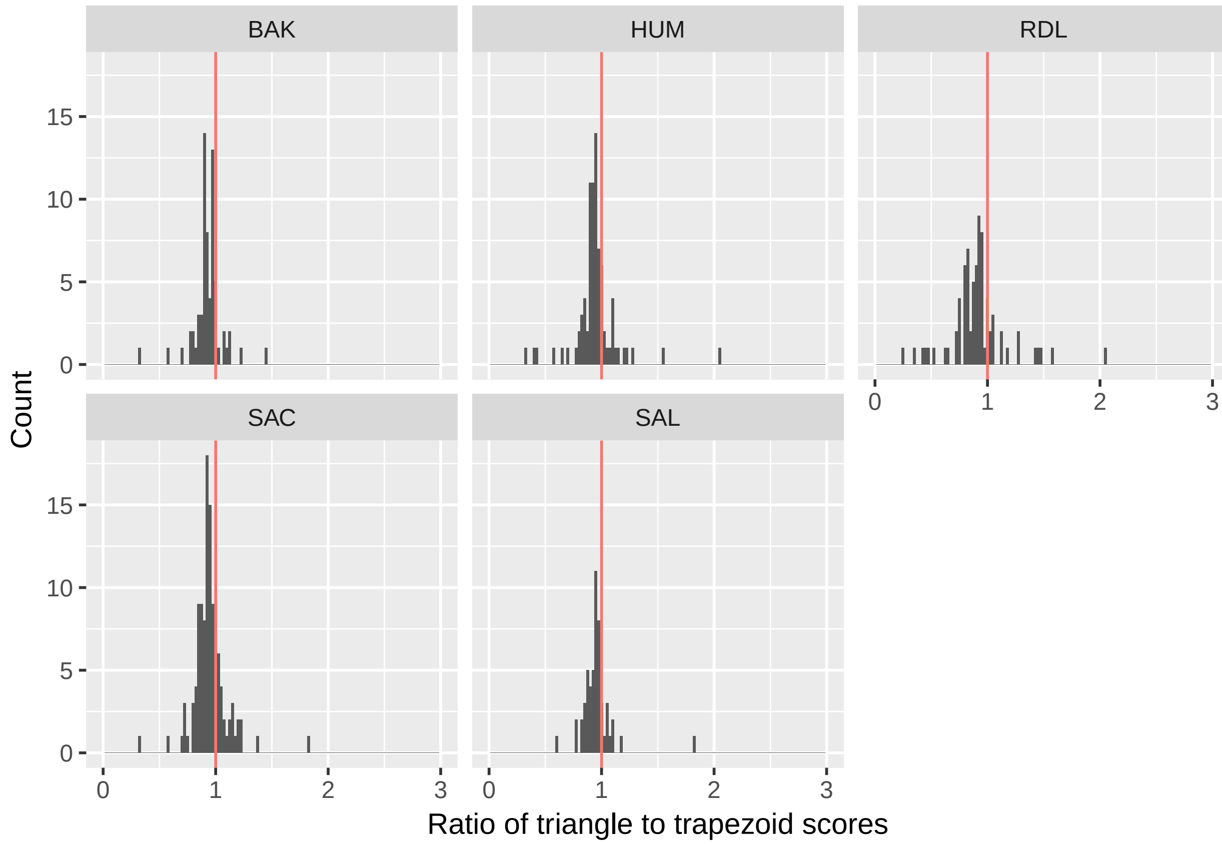
To investigate whether the phonetic movement of the LOT and THOUGHT vowels is causing a shift in the shape of the vowel space, the vowels were treated as vectors and their angles measured. If the vowel space were triangular then the LOT-THOUGHT vowels should lie on the line between the lowest vowel and the highest backest vowel. For each speaker an algorithm identified both the lowest vowel that was not LOT or THOUGHT and the highest backest vowel. The high back vowel was defined as a vowel in the vowel classes of BOAT or POOL that had the lowest Euclidean distance from the origin in Hz space (i.e., closest to the origin).

The high back vowel and the LOT vowel were then converted to vectors from the low front vowel in order to test the hypothesis that LOT lies along the line between the low vowel and high back vowel. If this hypothesis were true then the angle between the two vectors should be zero. The alternative hypothesis—that the vowel space is trapezoidal—would predict that the LOT vector would have an angle of zero with the horizontal (F2) unit vector. As the plot below shows, the LOT vector appears to have minimal overlap with the high back vector, suggesting that there is not a consensus for triangular vowel spaces.

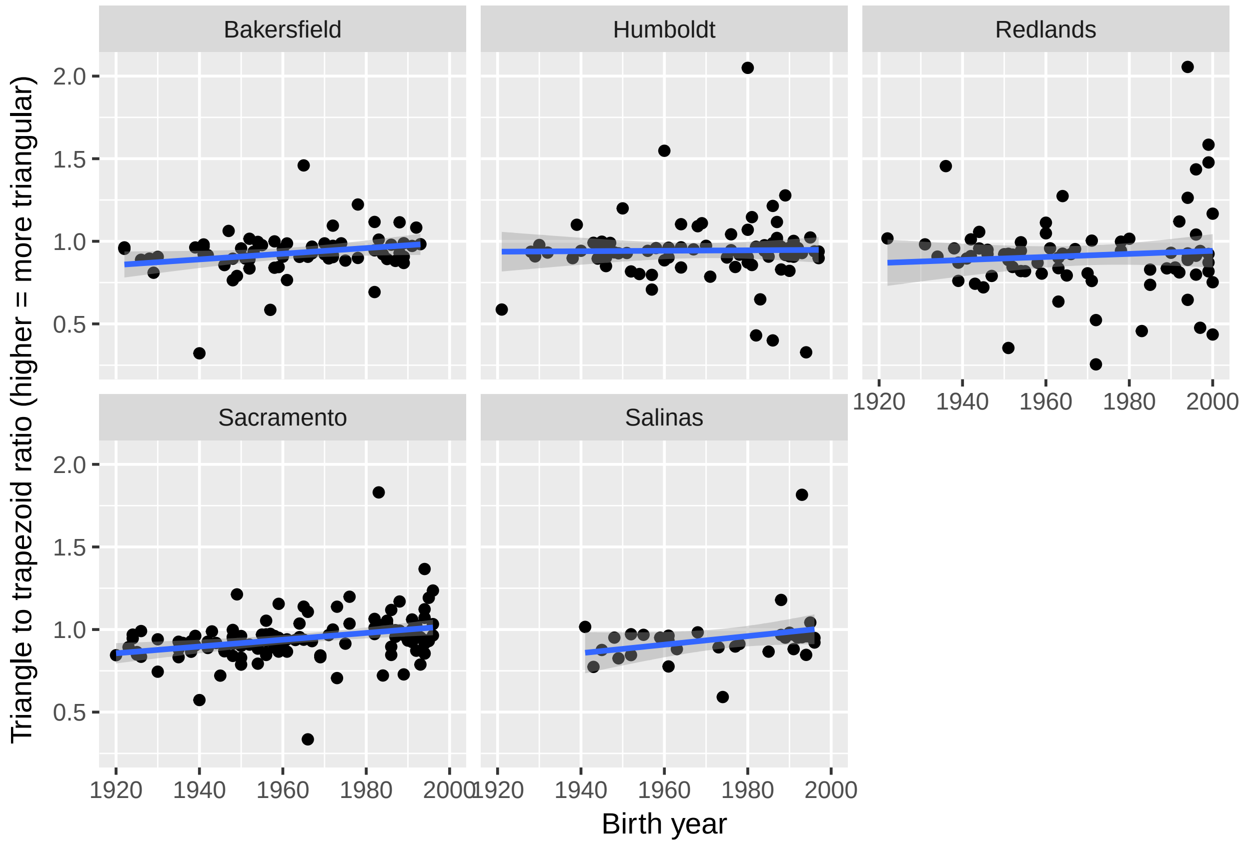


To more precisely test this hypothesis, a metric for how triangular or trapezoidal a speaker's vowel space is was computed. The triangularity of a vowel space was defined as the angle between the LOT vector and the high back vector. The trapezoidality of a vowel space was defined as the angle between the horizontal unit vector and the LOT vector. The ratio of triangularity to trapezoidality allows for the comparison of which state, triangle or trapezoid, best describes a speaker's vowel space shape. If this ratio has a value of 1 then the LOT vowel lies perfectly between triangular and trapezoidal states. If the value is greater than 1 then the vowel space is more triangular than it is trapezoidal supporting the hypothesis that the LOT-THOUGHT movement is causing a more triangular vowel space. However if the ratio is less than 1 then the vowel space is more trapezoidal supporting the alternative hypothesis that the LOT-THOUGHT movement is not causing a change in the vowel space shape. The plot below corroborates the intuitions from the previous figure as for each field site (Merced and Redding were not included due to small sample size) the bulk of speakers have a more trapezoidal than triangular space, though a notable few have more triangular vowel space.

```
## Warning: Removed 21 rows containing non-finite values (stat_bin).
```



However because the California Vowel Shift seems to be a change in progress, this analysis should consider how the vowel space has been changing over time. The figure below shows the triangle to trapezoid ratio for a participant by their birth year, and there appears to be a trend in apparent time whereby younger participants have a more triangular vowel space. For this analysis the data was further subsetted to remove those whose birthyear was unknown and any participant with negative ratio values as they were few and (from the first figure) likely to be measurement errors. The pattern seems strongest in Bakersfield and Sacramento. Redlands looks to have an increasing trend however younger speakers there seem to have a wider envelope of variation which may indicate an interesting social patterning of the low back vowels.



Observations

Given the pattern of data in both analyses it seems that over time and across California the LOT and THOUGHT vowels are moving closer together and perhaps upwards or inwards.

Humboldt also seems to be a very interesting location as it appears the merger in production took place before the other field sites.