

Preliminary Paper

Introduction

Treated as a social construct, self-identification along ethnic lines can be viewed through the lens of Social Identity Theory which argues that human categories are prototypes of integrated attributes (Hogg, 2016). In a nutshell, Social Identity Theory is defined as an individual's knowledge that they belong to a social group and they know the emotional and social significance of the group, i.e. they know what it means to "be" a member of a group. This self identification forms an individual's identity which is then mediated with the use of language particularly between two or more interlocutors (Ochs, 1993). This is all done by the interlocutors communicating by shared conventions that signify their identity and generate social meaning. One of the many forms of communicating conventions is through the production of specific sounds (Eckert & Labov, 2017) and stylistic choices in speech (Eckert & Podesva, 2011).

In the paper *Phonological Correlates of Ethnic Identity: Evidence of Divergence?* by Matthew J. Gordon, we find evidence of vowel production differences between four ethnic groups, White American, Mexican American, African American and biracial Americans. These differences are found along ethnic lines with regards to the Northern City Shift (NCS). African American and Mexican American speakers did not follow the general trend of /æ/ raising the same way that White Americans did. There were also fewer instances of "innovative" forms of production, i.e. White Americans led the change in vowel shifts while African and Mexican Americans tended to be more conservative in their speech. In Canada, there has been evidence suggesting that Chinese Canadians had a vowel pattern that shifted differently when compared to White Canadians (Hoffman & Walker, 2010), while in a separate vein, Asian Americans in San Francisco largely followed the Low Back Merger as part of general Californian trend (Hall-Lew, 2009) though perhaps not to the same degree as White Californians (Cardoso et al, 2016).

This previous research suggests to me that vowel differences will be present along ethnic lines in California. California is experiencing a shift known as the Low-Back Merger (LBM) which involves the merge of the /ɔ/ and /ɑ/ vowels (Doernberger & Cerny, 2008), the raising of /æ/, the fronting of /u/ and the lowering of /I/ and /ε/ (Sheydaei, 2024). This paper, by conducting a Linear Discriminant Analysis of "White" vowels and the "East Asian" vowels from university students in California, will provide preliminary evidence of production differences in line with Gordon (2000), Cardoso et al. (2016), and Sheydaei (2024). A Signal Detection Theory analysis will also be conducted to check the reliability of the results of the LDA.

Participants

The data for this study comes from 22 undergraduate students at the University of California, Davis. 12 self identified as East Asian (i.e. ethnically descendant of Japanese, Chinese, and Philippine peoples) and the remaining 10 self identified as White

(i.e. ethnically descendant of European peoples). They were between the ages of 18 and 20. Each participant were recruited through the UC Davis SONA system.

Data Collection

The procedure the participants followed was a straightforward one. After reading over information related to the study and signing a waiver allowing me to record them, they entered into a recording studio. After testing the sensitivity of the microphone I presented them with a Google Slides presentation containing the words I asked them to read. They were told that they would see a word on the screen and have 2 to 3 seconds to read the word before I switched to the next. They were told that not every word was a “real” word and to try their best.

Each word was typed in bright orange and presented against a white background. From my position outside of the recording booth, I clicked from slide to slide, word to word about every two seconds and made notes of words that the participant had difficulty with pronouncing. In a couple of instances, I asked participants to repeat a word after they had read through the entire list twice. For example, many participants would pronounce the word “head” in the same way the word “heed” is pronounced. Some words were unfamiliar to the participants even though they were real words. A couple were unsure about the pronunciation of the word “gin” and would produce the “g” with the voiced velar /g/ instead of the affricate /dz/. In total, each participant would repeat every word in the list twice. The list of words included HvD, BvD, BvT, pool/pull merger, pen/pin mergers, and two diphthongs the /oi/ before /l/ found in soil and the /ai/ found in sigh. The HvD words were to find neutral uses of the 10 English vowels (including /r/).

After the participants had completed their time in the recording booth, they were asked to answer a short questionnaire to gather specific information on their demographics. These included: what is your gender identity, what is your ethnic background, what area of California did you grow up in? These questions in an open-ended format because I wanted them to think about how they identified and I felt that yes or no questions were inadequate. I recognize my role in influencing the answers however as they could only answer the questions I asked them narrowing the ways in which they could identify. I asked these questions after they completed their time in the recording booth because I wanted them to read the provided stimuli as naturally as possible.

I recorded the participants using Praat and sampled at 44100 hertz. Each participant recording was done in a single wav file unless I had to re record an individual’s production. In that case, I created a separate wav file that contained only the repeated word or words. I then used the cut and paste feature in Praat to paste the repeated word into the original wav file. After the recording had finished, I took the wav file recordings and hand annotated them to align the spectrographs with the written words. Following this, I used DARLA (Dartmouth Linguistic Automation) to properly align the segments of the words to their respective slices in the spectrographs.

Once I was satisfied with the alignments, it was time to turn to vowel extraction and analysis. Using the FastTrack plugin for Praat (Barreda, 2021) I extracted the vowels from each individual wav file using the aligned text grids and kept them stored as wav files for

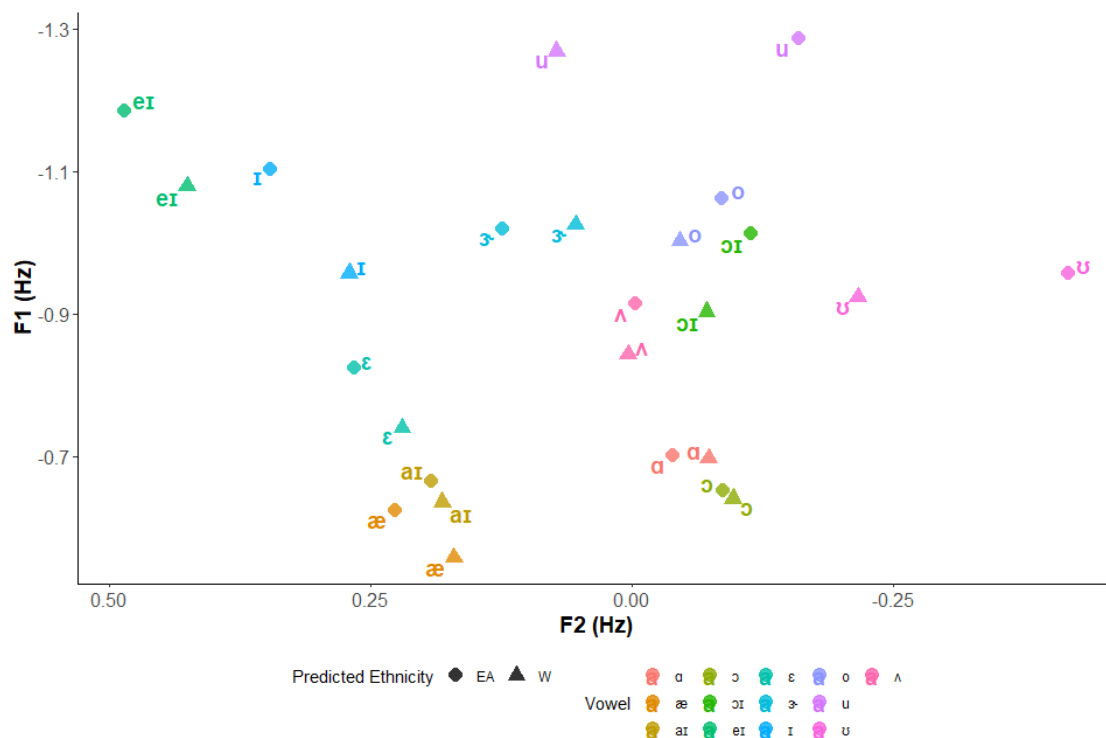
each production of a vowel and diphthong. Once completed, I used the TrackFolder command under the FastTrack plugin in Praat to run the LPC analysis to find the frequencies of the first three formants. I then normalized everyone's formants following log-mean normalization (Barreda et Nearey, 2018).

Data Analysis

This short paper will review the Linear Discriminant Analysis (LDA) of the “White” vowels and the “East Asian” vowels. Following that, a Signal Detection Theory (SDT) Analysis is conducted for each vowel. This steps are done for three reasons. First, if a model is able to correctly discriminate between the ethnicities using vowel formant data, its suggests that there exist differences in production. Second, If the sensitivity of the model (d') is 1.5 or higher for each vowel, the model does a pretty good job of distinguishing between the two ethnicities. Third, if the c value is at or near zero for each vowel, then any confusion found by the model is likely due to phonetic overlap. The lda model (lda_model) is $\text{lda_model} \leftarrow \text{lda}(\text{vowel_ethnicity} \sim \text{f11-f15} + \text{f21-f25} + \text{f31-f35})$.

In plain English, I expect the model to predict the relative location of each vowel for each ethnicity within the vowel space. I will make these predictions given the start and end of the first, second, and third formants of each vowel.

LDA Predictions



LDA Prediction of Vowels by Ethnicity

The plot above shows the location of each vowel for the two ethnicities based on what the model predicted. Before I report the results of the SDT analysis I will highlight a three generalizations that I note from the plot. First, between White speakers and East Asian speakers, the front vowels (front of \mathfrak{z}) are predicted to be different from each other mainly in terms of height. The most dramatic difference appears to be /eI/ and /I/ while /ε/ is less dramatic but still noticeable.

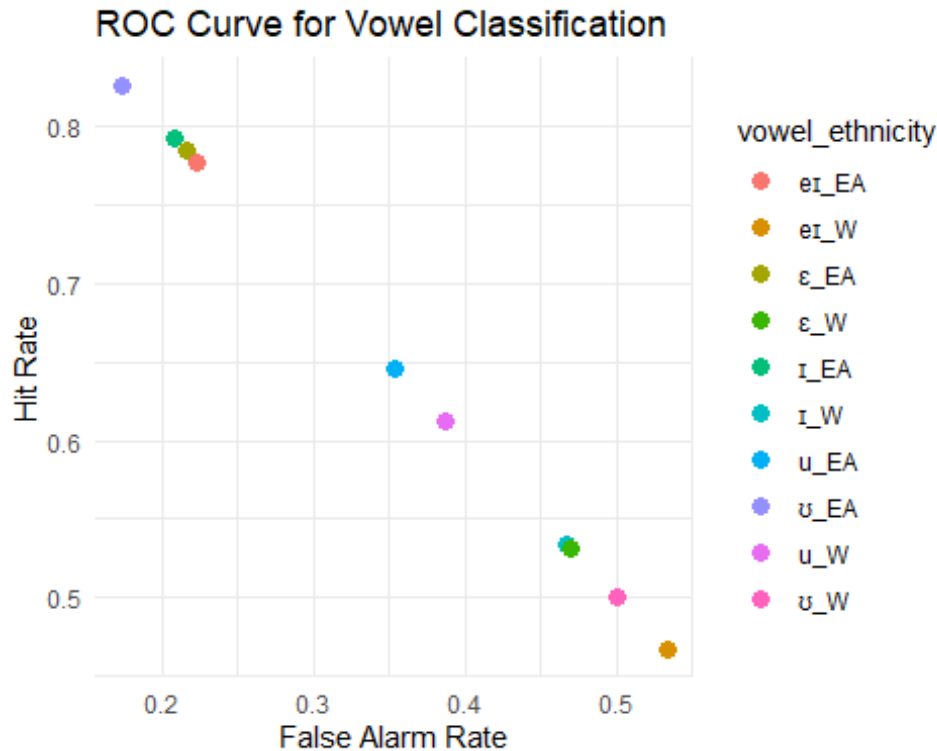
Second, the central vowels (between \mathfrak{z} and \mathfrak{u}) are not predicted to be different from each other. The model probably finds these vowels to be confusable. While it was generally clear that the East Asian speakers raised the front vowels more than the White speakers, I cannot find a generalizable pattern for the middle vowels.

Third, the back vowels (especially /u/) are predicted to be dramatically different. The most striking prediction is the difference in the fronting of /u/. /u/ fronting, a relatively common feature in North American English (Havenhill, 2004)(Kataoka, 2010), appears to be much more pronounced in white speakers than in East Asian speakers. The fronting of /u/ in the White population is very surprising. In contrast to the front vowels, the back vowels are predicted to shift more in the White population than the East Asian population.

With specific regards to the LBM, the fronting of /u/ fits the predicted pattern while the lowering of /I/ and /ε/ don't. I suspect that the White population is leading the shift more so than the East Asian population.

SDT Analysis

The SDT Analysis is to test the reliability of the model. I specifically want to see the hit rate and miss rate for the five vowels mentioned above. Those vowels are, /eI/, /I/, and /ε/ in the front and /u/ and /ʊ/ in the back. I will also report the sensitivity (d') and the response bias in an appendix.



As can be seen in the plot above, most of the vowels had a hit rate higher than that of the false alarm rate. The /u/ of the White population was split evenly while the /eɪ/ in that population had more false alarms than hits. Both vowels had a d' and response bias at or just slightly below zero. This suggests that there exists confusion in the categories and that the model does not have any systemic bias.

In contrast the /eɪ/ and /u/ in the East Asian population had a very high Hit Rate and a moderately reasonable d' values (1.5 and 1.8 respectively). This could mean that the two populations have true differences in production though some overlap still exists. It could also mean that the White Population had more differences within itself which made the prediction more difficult.

The /ε/ vowel for both groups had high hit rates with d' values of 1.57 for the East population and 1.52 for the White population. The response bias of the East Asian population is -2 suggesting that the model is more liberal in its predictions. However, given the false alarm rate is low, I think this suggests that while overlap of the categories still exists, the model's bias doesn't strongly sway the prediction. The /u/ of the two groups had a relatively high hit rate, around .6. However, the d' values (.75 for the East Asian population and .57 for the White population) suggest that there may be considerable overlap between the two populations. The response biases were also negative (-5.5 and -8 for the East Asian and White respectively) making the model very liberal in its predictions.

Finally, the /ɪ/ hit rate values for both groups were high (.79 for the East Asian population and .53 for the White population) but the d' for the White population is only .16 compared

to the East Asian population of 1.6. This could mean that there is more internal variation in the White population than in the East Asian population.

Overall, the model has done a decently good job at discriminating and predicting vowels along the ethnic lines. However, it has to be noted that the model has been 100% accurate in its predictions and some of these predictions are very liberal. In fact, I would wager that the White population (based on the d' values) have a higher degree of internal variation. This variation could be the reason for the varying hit rates across the five vowels focused on.

Discussion

Based on the predictions made by the model, it appears that the White population and the East Asian population had different vowel spaces. These spaces aren't different in terms of shape, that is impossible, but in terms of scale. The White population appears to front their back vowels much more than the East Asian population while the latter produces higher front vowels. Importantly, these differences are phonetic in nature and not phonemic, i.e. the /u/ of one population will be identified as an /u/ even though the formant values of it will be different.

In terms of the LBM, it appears that the White population follows that general trend. Their /ε/ and /ɪ/ were lower than that of the East Asian population suggesting that they are lowering those vowels with respect to the other group. Their /u/ was also predicted to be much more fronted, again following the trend of /u/ fronting.

This fits the previous work done showing that ethnic minorities produce their vowels in phonetically different ways. In the Northern cities, the African American and Mexican American populations followed the general patterns of the Northern City Shift while White Americans began to differ from it (Gordon, 2000). In Toronto, Chinese Canadians went against the features of the Canadian Vowel Shift (Hoffman & Walker, 2010). From my perspective, without having a non-California with which to compare, I cannot make a claim as to whether or not (or to what degree) the East Asian population follows the LBM trend. I can make a claim they appear to follow it less so than the White population.

After reviewing the SDT analysis, it is fairly common that the five vowels mentioned are produced differently between the two groups even though some overlap exists. Both populations are the same age group and both live in a small college town in the Sacramento Valley. If there were no overlap between the two, the differences would instead be phonemic and not phonetic. Ultimately, in order to be more confident in the results, I will need more participants to gather more data.

Next Steps and Conclusion

The next steps in this project are to 1: gather more speakers to train a more confident model and 2: take the predictions of that model and present them to listeners. If listeners are able to confidently discriminate between the two groups given a produced vowel, I can be fairly confident that these productions can be markers of ethnic identity.

While I did not discuss the meaning of ethnicity in this paper, I cannot end it without noting the its fluid nature. Ethnicity is not a concrete object, it is a construction and as such no two people will define it the same. In fact, no two people who identify as the same ethnicity will perform that ethnicity in the exact same way. This paper proposes that despite this fluidity, individuals will perform their ethnicity in similar ways.

References

- Austen, M. (2020). Production and perception of the Pin-Pen merger. *Journal of Linguistic Geography*, 8(2), 115–126. doi:10.1017/jlg.2020.9
- Barreda, S. Nearey, T. (2018). A regression approach to vowel normalization for missing and unbalanced data. *Journal of the Acoustical Society of America*, 144, 500-520.
- Barreda, Santiago. "Fast Track: fast (nearly) automatic formant-tracking using Praat" *Linguistics Vanguard*, vol. 7, no. 1, 2021, pp. 20200051.
- Cardoso, A., Hall-Lew, L., Kementchedjhieva, Y. & Purse, R. 2016. 'Between California and the Pacific Northwest: The front lax vowels in San Francisco English.' *American Speech*, 101(1), 33–54.
- Eckert, P., & Podesva, R. J. (2011). Sociophonetics and sexuality: Toward a symbiosis of sociolinguistics and laboratory phonology. *American Speech*, 86(1), 6-13.
- Eckert, P. and Labov, W. (2017), Phonetics, phonology and social meaning. *J Sociolinguistics*, 21: 467-496. <https://doi.org/10.1111/josl.12244>.
- Foulkes, P. Docherty, G. The social life of phonetics and phonology, *Journal of Phonetics*, Volume 34, Issue 4, 2006, Pages 409-438, ISSN 0095-4470, <https://doi.org/10.1016/j.wocn.2005.08.002>.
- GeeksforGeeks. (n.d.). Linear Discriminant Analysis in R Programming. GeeksforGeeks. <https://www.geeksforgeeks.org/linear-discriminant-analysis-in-r-programming/>
- Gordon, Matthew. Phonological Correlates of Ethnic Identity: Evidence of Divergence?. *American Speech*, Volume 75, Number2, Summer 2000, pp.115-136.
- Hall-Lew, L. (2009). Ethnicity and phonetic variation in a San Francisco neighborhood. Stanford University.
- Havenhill, J. (2024). Articulatory and acoustic dynamics of fronted back vowels in American English. *The Journal of the Acoustical Society of America*, 155(4), 2285–2301. <https://doi.org/10.1121/10.0025461>
- Hoffman, M. F. & Walker, J. A. 2010. 'Ethnolects and the city: Ethnic orientation and linguistic variation in Toronto English.' *Language Variation and Change*, 22(1), 37–67. doi:10.1017/S0954394509990238
- Hogg, M. A. (2016). Social identity theory (pp. 3-17). Springer International Publishing.

Johnson, K. (2006). Resonance in an exemplar-based lexicon: The emergence of social identity and phonology. *Journal of Phonetics*, 34(4), 485–499.
<https://doi.org/10.1016/j.wocn.2005.08.004>

Linear discriminant analysis. *Nat Rev Methods Primers* 4, 71 (2024).
<https://doi.org/10.1038/s43586-024-00357-9>

Ochs, E. (1993). Constructing Social Identity: A Language Socialization Perspective. *Research on Language and Social Interaction*, 26(3), 287–306.
https://doi.org/10.1207/s15327973rlsi2603_3

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.
doi:10.1017/S095439451500006X

Podesva, Robert J; D’Onofrio, Annette; Van Hofwegen, Janneke; Kim, Seung Kyung. *Language Variation and Change*; Cambridge Vol. ` 27, Iss. 2, (Jul 2015): 157-186.
DOI:10.1017/S095439451500006X

Thomas, E.R. (2013). Sociophonetics. In *The Handbook of Language Variation and Change* (eds J.K. Chambers and N. Schilling). <https://doi.org/10.1002/9781118335598.ch5>

Sheydaei, I. (2024). The Low-Back-Merger Shift: Evidence from MENA Americans in the Upper Midwest and southern California: MENA Americans and the Low-Back-Merger Shift. *English Today*, 40(1), 22–31. doi:10.1017/S0266078423000299

Sumner, M., Kim, S. K., King, E., & McGowan, K. B. (2014). The socially weighted encoding of spoken words: A dual-route approach to speech perception. *Frontiers in Psychology*, 4, 1015.

Villarreal, D. (2018). The Construction of Social Meaning: A Matched-Guise Investigation of the California Vowel Shift. *Journal of English Linguistics*, 46(1), 52-78.
<https://doi.org/10.1177/0075424217753520>

Appendix

Signal Detection Theory Data

	vowel_ethnicity	hit_rate	false_alarm_rate	d_prime	response_bias
1	aI_EA	0.8326996	0.1673004	1.9297766	0.000000e+00
2	aI_W	0.5909091	0.4090909	0.4597682	-5.551115e-17
3	eI_EA	0.7777778	0.2222222	1.5294193	0.000000e+00
4	eI_W	0.4666667	0.5333333	-0.1673035	0.000000e+00
5	o_EA	0.8000000	0.2000000	1.6832425	-1.110223e-16
6	o_W	0.5762712	0.4237288	0.3847266	0.000000e+00
7	u_EA	0.6458333	0.3541667	0.7481908	-5.551115e-17
8	u_W	0.6125000	0.3875000	0.5716817	-8.326673e-17
9	æ_EA	0.8750000	0.1250000	2.3006988	0.000000e+00
10	æ_W	0.5666667	0.4333333	0.3357880	0.000000e+00
11	ɑ_EA	0.8800000	0.1200000	2.3499736	0.000000e+00
12	ɑ_W	0.6500000	0.3500000	0.7706409	0.000000e+00

13	\mathfrak{C}_{EA}	0.8125000	0.1875000	1.7742931	0.000000e+00
14	\mathfrak{C}_W	0.5609756	0.4390244	0.3068864	-6.938894e-17
15	\mathfrak{CI}_{EA}	0.8402778	0.1597222	1.9912004	0.000000e+00
16	\mathfrak{CI}_W	0.6115702	0.3884298	0.5668280	8.326673e-17
17	ε_{EA}	0.7844991	0.2155009	1.5749567	-2.220446e-16
18	ε_W	0.5304740	0.4695260	0.1529231	0.000000e+00
19	\mathfrak{Z}_{EA}	0.6619718	0.3380282	0.8357012	0.000000e+00
20	\mathfrak{Z}_W	0.6000000	0.4000000	0.5066942	0.000000e+00
21	\mathfrak{I}_{EA}	0.7924188	0.2075812	1.6296851	-5.551115e-17
22	\mathfrak{I}_W	0.5334773	0.4665227	0.1680279	-4.857226e-17
23	\mathfrak{U}_{EA}	0.8260870	0.1739130	1.8776286	0.000000e+00
24	\mathfrak{U}_W	0.5000000	0.5000000	0.0000000	0.000000e+00
25	\wedge_{EA}	0.8264463	0.1735537	1.8804294	0.000000e+00
26	\wedge_W	0.6200000	0.3800000	0.6109616	0.000000e+00