THE UNIVERSITY of EDINBURGH

Edinburgh Research Explorer

Improved representation of variance in measures of vowel merger

OPEN ACCESS

Improved representation of variance in measures of vowel merger

Lauren Hall-Lew

University of Oxford, Phonetics Laboratory, 41 Wellington Square, Oxford OX1 2JF, United Kingdom

Abstract. The difference between two vowel clusters, such as the distance between two phonemes involved in merger, has most commonly been measured using techniques such as taking Euclidean distances (e.g., between averages or between minimal pairs). Relying on averaged data or minimal pair data only captures some of the variability between two given vowel clusters; reliance on averages obscures the amount of variability within a given vowel class, while calculating distance between minimal pairs relies on few tokens per speaker, which may be difficult to obtain from naturalistic data. Hay, Warren, and Drager (2006) introduced an alternative approach to calculating the distance between two vowel clusters that accounts for the variability between those clusters, taking token-specific formant values, rather than averages, as input. The measure is the *Pillai-Bartlett* statistic (see Baayen 2008:158), an output of a Multivariate Analysis of Variance (MANOVA), which represents the proportion of one variance that can be predicted by another variance. A higher Pillai value indicates a lower degree of overlap between two vowel clusters in F1/F2 space. Since the value is derived from a MANOVA, Pillais can account for known internal factors influencing the production of merger, such as phonological environment, thereby reducing the need to obtain minimal pair lists. This paper argues for using Pillais as measures of merger by comparing results from low back merger in California English (Hall-Lew 2009) with the analysis of front vowel merger in Hay, Warren, and Drager (2006), and further suggests that the Pillai statistic is a useful measure for measuring any vowel change-in-progress, such as the fronting of the mid- and high back vowels.

Keywords: vowels, formant space, methodology, sound change
PACS: 43.70.Jt, 43.72.Ar

1. INTRODUCTION

Common measures of vowel merger, such as the Euclidean distance in F1/F2 space between averages or between minimal pairs, have only been able to capture some of the variability between two given vowel clusters. Reliance on averages obscures the amount of variability within a given vowel class, so that if a speaker produces one vowel class highly consistently but another vowel class highly variably, that difference between classes is totally lost. What's more, if the analyst averages across phonological environments, any effects of phonological environment on the position of those vowels will be obscured. The issue of phonological environment has been important for all studies documenting vowel mergers in progress, and so one methodological solution has been to separate out vowel production according to minimal pairs, calculating the distance between pairs, so that each speaker is represented by a list of values corresponding to separate phonological environments. Carving up the data in this way, however, not only leaves the analyst with multiple numerical values per speaker, but also reduces the statistical power of subsequent tests for effects of social variables (for example), because of the decreased number of tokens per environment per speaker. The problem of representing cluster variance also remains unaccounted for.

Hay et al. [1] introduced an alternative approach that accounts for the variability between two vowel clusters, as well phonological environment, by taking token formant values as input, rather than averages. The measure is the *Pillai-Bartlett statistic*, an output of a Multivariate Analysis of Variance (MANOVA), which represents the proportion of one variance that can be predicted by another variance, given any known conditioning. Hay, *et al.*, refer to the statistic as the 'Pillai score,' a term which I will adopt here. In short, the higher the Pillai scoreâ the lower the degree of overlap between two vowel clusters in F1-F2 space. In this paper I will go into detail about the implementation of the Pillai score as a measure of vowel merger, arguing that it is not only a more accurate representation of vowel cluster difference, but that it is also easy to use and straight-forwardly interpretable. I will further argue that it is a reliable measure for any difference between any two vowel clusters, therefore making it ideal for calculating other changes in progress, such as the fronting of the nuclei of the mid- and high back vowels that is taking place across many varieties of U.S. and U.K. English. The paper concludes with implications for interpreting the Pillai score in the

context of different kinds of merger-in-progress, as well as the general limitations of the Pillai score method.

Studies vary in terms of how the extent of merger-in-progress, for a given speaker, is measured. In terms of the low back vowel merger, which is perhaps the most well-studied merger in U.S. English, previous methods of calculating vowel distance include the following: non-acoustic, auditory coding (as in early Atlas data; see DeCamp [2]); qualitative, visual descriptions of F1-F2 plots of cluster overlap (*e.g.*, Moonwomon [3]); quantitative, acoustic measures of token cluster means, based on F1 and F2 separately or Euclidean/Cartesian distances (*e.g.*, Irons [4], Baranowski [5]); and quantitative, acoustic measures of minimal pairs, again based on separate formant values or Euclidean/Cartesian distances (Di Paolo [6], Majors [7], Johnson [8], Dinkin [9]). Some studies explicitly include multiple measures (see Labov et al. [10]). Each of these approaches, even in combination, has some weaknesses that the Pillai score method improves upon.

Basing any measure on averages eliminates potentially important information about the distribution of the token cluster. If the production of one vowel class is much more variable than the other, then this is descriptively central to the account of that process of merger. Relying on minimal pairs similarly diminishes this information, although presumably one could take the respective distances between each of the minimal pairs to approximate a representation of the variance; this is not ideal, however, because the result is a list of distances, rather than a single measure of variance. Furthermore, while some minimal pairs occur naturally in conversation or in interviews, others do not, which them means that the analyst often bases measures of merger on data elicited in minimal pair wordlist style. It has long been shown that word list style differs markedly from other speaking styles (Labov [11]), and for sound changes in progress, we would like a measure of vowel distance than can be implemented for any dataset, regardless of the context in which the vowel data were obtained.

The use of minimal pairs in measure of merger stems from an attempt to account for the well-known effects of phonological conditioning (e.g., Labov et al. [12]) on sound change. Statistically, studies of merger deal with the factor of phonological environment either in the calculation of the means (as in the vowel plotting program *Plotnik*; see Dinkin [9]), which is a good method if one is going to use means, or else are entered into the statistical model at the same stage as social factors. This latter technique means that token effects are combined in the same model as speaker effects, when the two are presumably independent – *i.e.*, the retraction effect of a following liquid on the low and back vowels is an effect of coarticulation that operates similarly across all speakers in a given sample. An ideal way to contend with phonological effects is to model them at an earlier stage, such as is done in *Plotnik*, but to then represent vowel distance according to the entire variance of production, rather than averages. The Pillai score does this.

2. ABOUT THE PILLAI SCORE

The Pillai score as a method of merger was introduced by Hay et al. [1] in their analysis of the merger of NEAR and SQUARE (see Wells [13] for vowel class keywords) in New Zealand English. The Pillai score maintains information about the vowel token cluster distribution, accounts for phonological environment, and allows for the use of unbalanced (interview/conversation) data. This section describes the function of the method in detail and includes a step-by-step introduction on the implementation of the method for any dataset using the *R* statistical analysis environment (see Baayen [14]).

The *Pillai-Bartlett statistic*, here called the Pillai score, is one of the four common MANOVA tests. Multivariate analysis of variance (MANOVA) is a type of analysis of variance (ANOVA) that is used when there are more than two dependent variables. In the case of vowel clusters, both the first and second formants are dependent variables. The Pillai statistic was the one chosen by Hay, *et al.*, due to the comparative analysis in Olsen [15]. It is also the default statistic used in the *R* statistics analysis environment for MANOVA (Baayen [14], p158). When the MANOVA considers two vowel classes, then the "higher Pillai scores indicate greater distance (in F1 and/or F2) between the two vowels ... As a summary of the degree to which two distributions are kept distinct, this is superior to taking Euclidean distances between means, because it takes account of the degree of overlap of the entire distribution" (Hay et al. [1], p467). In other words, when the two variables under question are the two vowel classes specifically involved in a merger-in-progress, then "the lower the Pillai score, the more advanced the merger" (Hay et al. [1], p467).

By using the Pillai score as a measure of merger, phonological environment can be included in the MANOVA. Each speaker's Pillai score can be then used as the dependent variable in a model testing for the social correlates of the degree of merger, with the understanding that phonological environment as a known conditioning variable has already been accounted for. Hay et al. [1] based their calculations were based on minimal pair vowel midpoint data obtained in experimental settings. By doing this, they were able to obtain a single Pillai value for each of their speakers, as well as calculating individual Pillai scores for each minimal pair, confirming that each pair was distinct from the next across all speakers, and showing how some environments (*really/rarely*) were more merged than others (*beer/bare*).

However, another advantage of the Pillai score is that MANOVAs can be calculated over spontaneously produced, naturalistic data, which is unbalanced for precise phonological environment, given the assumption that the nature of this unbalanced distribution will be roughly similar across all the speakers in the sample. In other words, assuming that one speaker is not more likely than another speaker to produce pre-liquid tokens of the vowel in question, then minimal pairs are not necessary – what is necessary is a large enough sample of speech such that a reasonably comprehensive sample of phonological environments will naturally arise. This attention to eliciting large quantities of speech from individuals has always been a goal of sociolinguistic fieldwork and presents nothing particularly new to the data collection process; rather, the Pillai score method is already well-suited to accepted methodologies.

Besides Hay et al. [1], the only other known uses of the Pillai as a measure of merger are Kennedy [16] and Hall-Lew [17], the latter which is discussed further on in this paper. Kennedy [16], uses the Pillai score to examine the merger of the THOUGHT and FOOT vowel classes (see Wells [13]) before /l/ in New Zealand English. In her analysis she points out one of the weaknesses of the Pillai measure: "A limitation of the Pillai score is that it does not provide a meaningful measure of significance alongside the distance measurement" (Kennedy [16], p63). In other words, the range of Pillai values across a speaker sample is continuous and provides a means of representing the relative extent of merger between any two speakers – but while the MANOVA will identify those speakers with a clear vocalic distinction, providing a *p* value corresponding to the difference between the vowel clusters for a given speaker, it will not provide statistical discrimination between those with near-merger and those with complete merger. However, this drawback is not unlike any other measure of merger currently used, and is therefore arguably still a preferable choice over these other methods.

2.1. Calculating the Pillai score

◇	A	B	C	D	E	F
1	spkr	class	word	phon	F1N	F2N
2	Aaron	LOT	a_lot	t	1.811	0.11
3	Emily	LOT	adopted	p	1.255	-0.404
4	Aaron	THOUGHT	along	ng	0.925	-0.655
5	Amy	THOUGHT	August	g	1.699	-0.341
6	Sal	THOUGHT	August	g	0.882	-0.679
7	Sal	THOUGHT	authors	th	1.38	-0.465
8	Maya	THOUGHT	awful	f	1.738	-0.347
9	Maya	THOUGHT	awful	f	1.358	-0.403
10	Molly	THOUGHT	awful	f	2.094	-0.841
11	Aaron	LOT	block	k	1.572	-0.375
12	Amy	LOT	block	k	1.686	-0.295
13	April	LOT	block	k	1.707	-0.184
14	Bruce	LOT	block	k	1.404	-0.281
15	Bruce	LOT	block	k	1.512	-0.148
16	Bruce	LOT	block	k	1.447	-0.469
17	Cheri	LOT	block	k	1.177	-0.333
18	Cheri	LOT	block	k	1.749	-0.528
19	Enid	LOT	block	k	1.791	-0.174

FIGURE 1. Example input of normalized vowel midpoint data testing for the extent of merger between LOT and THOUGHT (Wells [13]), according to phonological environment, for each speaker.

Given familiarity with the *R* statistical analysis environment, calculating the Pillai score for any dataset is quite straightforward. As an example, assume that the input data exists in a .csv file with a separate vowel token on each row and the following information in each column: speaker, vowel class, word, phonological environment, F1, and

F2, where each row of the file represents a particular vowel token. An example of this, measuring the merger of the LOT and THOUGHT vowel classes (also known as the low back vowel merger), is given in Figure 1.

In this example, taken from Hall-Lew [17], phonological environment is defined maximally conservatively, as the phoneme immediately following the relevant vowel (coded here as: [p, b, t, d, g, k, f, v, s, z, m, n, 'ng', 'sh', 'ch', 'th', 'dh', 'dz'] and #, or word final). The formant measurements represent Lobanov-normalized Bark values. But this is an artifact of the decisions in this particular study; phonological environment can be defined in whatever way is preferable to the researcher, and the normalization of the vowel formant data occurs at a prior stage, orthogonal to the analysis of merger.

So, in this example, the Pillai score will calculated for each speaker, given in column A, based on the distributions in the two vowel classes, given in column B, with respect to phonological environment, in column D, for the dependent variables F1, in column E, and F2, in column F. Column letters are rendered as numbers in R. The code entered into R in order to run the MANOVA is:

```
data$subj = as.factor(data$spkr)
subs=levels(as.factor(data$spkr))
for(sub in subs) {
tmp<-subset(data,spkr==sub)
deps=cbind(tmp[,5],tmp[,6])
type<-factor(tmp[,2])
labls<-factor(tmp[,3])
phon<-factor(tmp[,4])
fit<-manova(deps ~ type + phon)
print(sub)
print(summary(fit, test="Pillai"))
```

This then produces an output similar to Figure 2:

```
[1] "Aaron"
          Df Pillai approx F num Df den Df   Pr(>F)
type       1 0.4794    9.2091      2     20 0.001462 **
phon       7 0.5854    1.2414     14     42 0.283565
Residuals 21
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
[1] "Abby"
          Df  Pillai approx F num Df den Df Pr(>F)
type       1 0.29063  2.04856      2     10 0.1796
phon       4 0.63144  1.26881      8     22 0.3088
Residuals 11
```

FIGURE 2.   Example output of the MANOVA, showing Pillai scores for the first two speakers in the sample

Figure 2 shows the results of the MANOVA for the first two of the 29 speakers in the analysis by Hall-Lew [17]. 'Aaron' has a Pillai of 0.479 and 'Abby' as a Pillai of 0.291, telling us that Aaron has a greater LOT/THOUGHT distinction than Abby does. Furthermore, Aaron's type difference reaches significance, where Abby's does not. This tells us that Aaron is producing a LOT/THOUGHT distinction, however, it does not tell us that Abby is necessarily producing a merger – she could be near-merged. The interpretation of the Pillai scores for each speaker become more meaningful when seen in the context of a greater speaker sample, as shown in the following section.

3. METHOD APPLICATION: VARIATION IN SAN FRANCISCO ENGLISH

3.1. Low Back Vowel Merger

Hall-Lew [17] used the Pillai score as a measure of the low back or LOT/THOUGHT vowel merger in San Francisco English. In the Western United States there is a change-in-progress whereby the vowels in these classes are merging to a single vowel class. According to the Atlas of North American English (Labov et al. [10]), San Francisco is behind in apparent time, in that fewer speakers produce a completed merger relative to other urban locations across the Western states. Hall-Lew [17] analyzed a socially stratified sample of 29 speakers, ages 16–76. Unlike Hay et al. [1], the data for the analysis came from loosely structured sociolinguistic interviews, so although phonological environments were similar across speakers, an exact balance of equal numbers and types of tokens (with respect to following consonant) was not obtained. Format data was obtained from both the vowel midpoint, defined as the highest point of F1 and/or the approximate midpoint of the steady state of the vowel, and the off-glide, taken at approximately 20 milliseconds inward from the end of phonation. However, Pillai scores were calculated based only on the midpoint, and not the offglide, values (though separate calculations could easily have been run on the offglide values). All midpoint values, along with values for the vowels in FLEECE, GOOSE, and GOAT were converted to Barks and normalized using the speaker-intrinsic Lobanov algorithm implemented by the NORM vowel normalization suite (Adank et al. [18], Thomas and Kendall [19]).

The MANOVA was run in the *R* statistical environment with respect to phonological environment as defined in the most conservative way possible, at the level of segment (as opposed to place and/or manner of articulation), as shown in Figure 1. The results for this set of 29 speakers ranged from Pillai scores of 0.71 (the most distinct speaker) to 0.01 (the most merged merged). This contrasts with the wider range found for the NEAR/SQUARE merger by Hay et al. [1], which ranged from 0.97 to 0.00. This difference may reflect the closeness of LOT and THOUGHT in speakers vowel spaces generally (given articulatory constraints) and/or a more advanced overall progression towards merger.

The overall finding was that this sample of speakers exhibited a change in apparent time towards merger, as shown by a significant correlation between the age of the speaker and that speaker's Pilliai score. This correlation is shown in Figure 3, and is significant at $p < 0.01$. The figure shows that the older speakers have higher Pillai scores, indicating greater difference between their productions of the LOT vowel class and the THOUGHT vowel class. Younger speakers are more likely to have Pillai scores closer to zero. A few of the speakers have *adjusted Pillai scores*, represented by negative values. The negative polarity of these scores was not obtained from the MANOVA, but was adjusted post-hoc, a slight downside to the use of the Pillai score method which is discussed further in section 4.1.
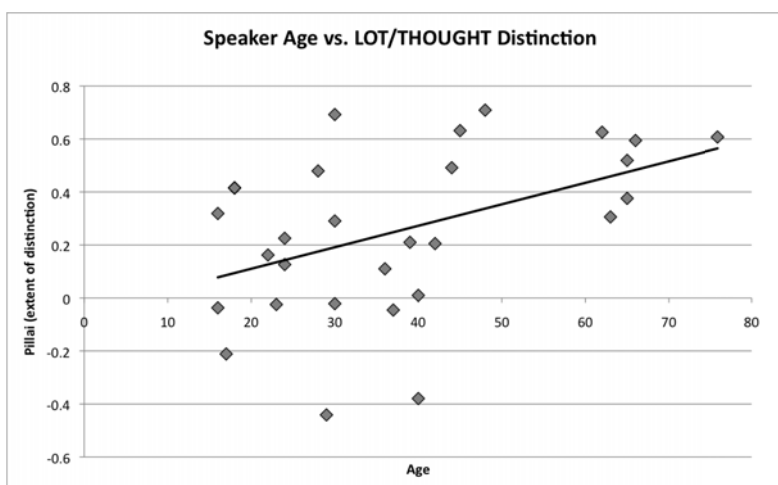

FIGURE 3.   Speaker age *vs.* Pillai score, showing change-in-progress towards merger (Hall-Lew [17])

3.2. Using Pillai to quantify non-merger changes in progress

As a representation of the difference between any two clusters, the Pillai score can logically represent any measure of variable distance in vocalic space, not just mergers. Another example taken from Hall-Lew [17] is the fronting of the nuclei of the mid- and high back vowel classes, also known as the GOAT and GOOSE classes (Wells [13]). Back vowel fronting is a well known change-in-progress across much of the United States, and a robust feature of Western U.S. English varieties (Hinton et al. [20], Luthin [21], Conner [22], Hagiwara [23], Ward [24], Hall-Lew [25]). Previous findings that the environment of a following /l/ inhibits the fronting of back vowel nuclei (Luthin [21], Di Paolo and Faber [26], Thomas [27]) provide a means for calculating the extent of nuclei fronting in environments other than preceding /l/. An index of back vowel fronting can be calculated for each speaker based on the normalized F1/F2 distance between their GOAT of GOOSE class token distribution and the corresponding distribution of tokens in pre-/l/ environments (here called 'COAL' or COOL, respectively). The Pillai score thus represents the difference between the distribution of productions of those vowels without final /l/ and the distribution of productions of tokens with final /l/. Higher Pillai scores indicate greater fronting, and values close to zero indicate no difference between the vowel distribution and its distribution before /l/, in other words, no fronting.

For this analysis, the same speakers, Bark conversion, and normalization were used as in the analysis of low back vowel merger. The Range of Pillai scores for GOAT/COAL ran from 0.86 (fronted) to 0.11, while the range of Pillai scores for GOOSE/COOL was larger, from 0.92 (fronted) to 0.10. This slight difference in range was not surprising, since GOOSE vowels are known to front further than GOAT, in general. In comparison again to the merger analysis by Hay et al. [1], the Pillai scores for GOAT and GOOSE indicate that San Francisco English back vowels are still slightly less distinct from their pre-/l/ allophones, even for the most fronted speakers, than the NEAR and SQUARE vowel classes are for the most distinct New Zealand English speaker. Among those San Francisco English speakers with a strong LOT/THOUGHT distinction, the difference in F1/F2 space with respect to cluster variance is still not as great as the differences found for back vowel fronting, a result that is unsurprising given articulatory constraints.

There are two advantages to using the Pillai trace rather than more traditional methods that calculate fronting based on F2 distance measures between averaged data points (*e.g.* Hall-Lew [25]). First, using Pillai scores derived from MANOVAs results in measures of fronting that account for the amount of spread within each allophonic class, an important feature of vocalic variability that cannot be accounted for by using averages. Second, the Pillai score measure also allows the analyst to easily move beyond a definition of fronting as movement in F2 space, and accounts for any F1 change as well.

Results from Hall-Lew [17] are shown in Figure 4. In contrast to the analysis of merger, here a greater Pillai score indicates a greater degree of fronting. Overall, the figure shows that San Francisco English is undergoing a change-in-progress toward the fronting of back vowel nuclei, where younger speakers have more fronted productions than older speakers. One insight to emerge from this study was that this change in apparent time was significant for the GOAT class ($p < 0.002$) but not the GOOSE class ($p = 0.057$).[1]

Using the Pillai score to calculate back vowel fronting relies on the assumption that the mid- and high back vowels before /l/ are not undergoing sound change, *i.e.*, that they are not involved in back vowel fronting, which was grounded on the supporting claims made for San Francisco vowels by Luthin [21] and corroborated by other dialect studies in the West (*e.g.*, Hagiwara [23], Hall-Lew [25]). In dialect areas where back vowel fronting occurs in pre-/l/ environments, degree of fronting could be calculated based on reference to a (relatively) more stable vowel. This solution could have been taken for the present analysis as well; Pillai scores could have been calculated based on the FACE vowel class, for GOAT, and the FLEECE class, for GOOSE, where *lower* Pillai scores would then indicate *greater* fronting towards those front vowels. Ultimately, the use of the Pillai score as a measure of vocalic sound change-in-progress is promising, but only under the assumption that there is a stable point in the speaker's vowel space to which productions of the changing vowel can be compared. For some language varieties undergoing complete chain shifts, identifying such a stable vowel may not be possible.

---

[1] When the GOOSE class is separated according to phonological environment, which was not done *a priori* for the calculation of Pillai scores, fronting of those vowels following an anterior coronal correlates significantly with speaker age, showing change in apparent time, while correlation with GOOSE in other environments is not significant.
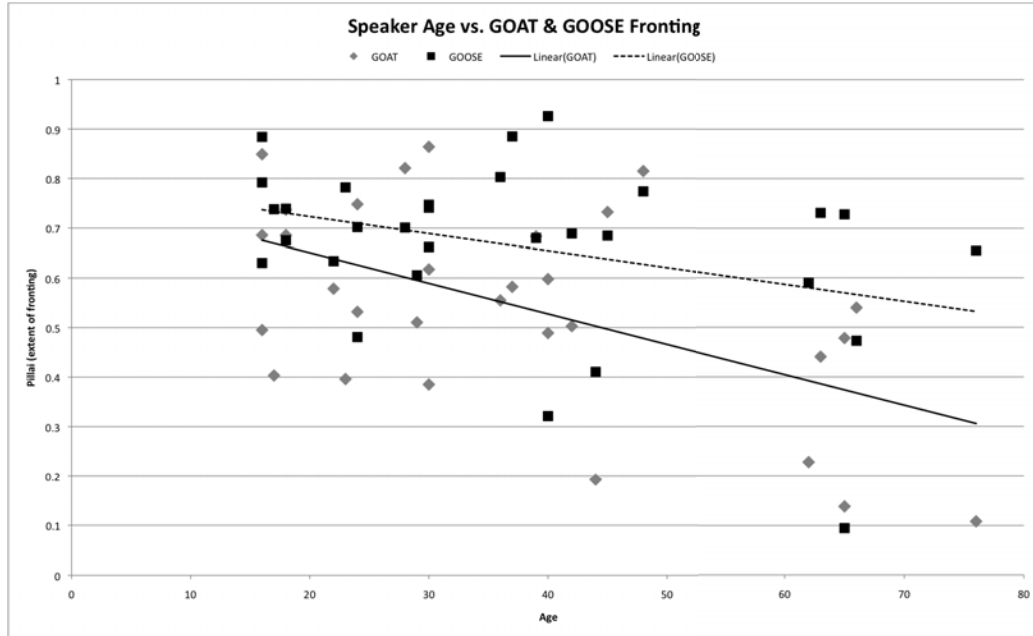
FIGURE 4. Speaker age *vs.* Pillai score, showing change-in-progress towards fronting (Hall-Lew [17])

4. DISCUSSION

4.1. Drawbacks

The Pillai score measure shares the same weakness of most other measures of merger in not representing directionality. For example, a speaker with a high Pillai score may appear to have distinct vowel classes when, in fact, the speaker is producing THOUGHT *lower* and/or *fronter* than LOT, displaying a flip-flop distribution (Labov [28], p143). In the data from San Francisco English, this problem was compensated for in the same way that someone relying on Euclidean distances would compensate for it, that is, by assigning any speaker with a flip-flopped pattern an adjusted score that converted their raw score to negative values. Determination of 'flip-flop' was based on a visual inspection of the distributions shown in vowel plots of normalized averages and normalized token distributions: a speaker was determined to be flip-flopped if their average THOUGHT production was fronter than their LOT distribution. Note that the extent of difference between THOUGHT and LOT averages is orthogonal to classification of a speaker as flip-flopped. In other words, the labeling of a speaker as flip-flopped (the decision about whether to retain a positive or convert to a negative Pillai value) is not a claim about extent of merger or distinction (the size of Pillai value). Although this modification is less than ideal, it would remain a necessarily modification for all other measures of merger currently in use.

The Pillai also misses several factors that may contribute to vocalic distinction: the length and/or direction of the off-glide, and the roundedness, phonation, or duration differences between vowels. Furthermore, speakers with the merger may differ with respect to the location of merged vowel, which is another feature independent of the measure of merger. The MANOVA assumes that the shape of a distribution is the same between clusters. However, for changes in progress, this is usually not the case. This may also be a problem of greater or lesser importance depending on the variable being studied. Given greater articulatory space in the front of the mouth, the distributions of NEAR and SQUARE may be relatively more similar in shape than the distributions of LOT and THOUGHT, and both pairs are more similar to one another in shape than are the distributions of GOAT versus COAL.

For the use of Pillai scores as a measure of back vowel fronting, these calculations unfortunately did not indicate the extent of unrounding, which often accompanies fronting. However, a measure of roundedness, such as F3, could be included in the model as an additional independent factor. As with merger, the position of the off-glide also does not enter into the model. For San Francisco English, this is not a huge problem, since the vast majority of off-glides appear to be more rounded than their nuclei and produced further back than their nuclei.

4.2. Types of merger

One question that remains is if the Pillai score work equally as well a measure of merger across different kinds of merger. A merger process may be classified according to (at least) one of three possible pathways to merger: merger-by-approximation, merger-by-transfer (Trudgill and Foxcroft [29]), and merger-by-expansion (Herold [30]). In merger-by-approximation, the productions of the two vowel classes become increasingly produced (across a population of speakers) at increasingly similar points in the vowel space, so that the both vowel classes 'move' towards one another, eventually forming a single class. In merger-by-transfer, the lexical members of one vowel class become gradually 'transferred' to the other vowel class, so that movement towards merger consists of movement through the lexicon, rather than movement through vowel space. Merger-by-expansion is somewhat similar to merger-by-approximation, except that the extremes of the original points of production remain, so that rather than two vowel clusters moving towards one another, the two vowel clusters increase their range of token variability to the point where they overlap one another in vowel space. Is the Pillai score an equally good measure across these three types of merger?

This is ultimately an empirical question. However, based on knowing the assumptions of the MANOVA, it seems that the Pillai score is better representation of mergers-by-approximation and -expansion than for mergers-by-transfer. The reason is that MANOVA assumes that the two distributions being compared have same shape. This is not a problem for either approximation or expansion because, in theory, the shapes of the vowel token clusters are relatively similar through the course of change. However, again in theory, merger-by-transfer involves one vowel class remaining stable, while the other class undergoes gradual lexical diffusion. So the shape of the stable vowel will regular, while the shape of the diffusing vowel will be irregular. In mergers-by-transfer, the Pillai value may overestimate the difference between the two vowels. This bias would exist equally for all speakers a particular sample, with the exception of those speakers who are completely distinct and those who are completely merged. But among those speakers involved in the change-in-progress, the relative extent of merger from speaker to speaker would be just as comparable as for other types of vowel merger.

5. CONCLUSIONS

The goal of this paper is simply to encourage and extend the application of a new measure of vocalic difference that is more elegant and statistically powerful than previous measures. In contrast to measures between vowel class averages or between minimal pair word lists, the Pillai score is a single numerical value of the difference between two vowel token clusters that can account for phonological environment over an unbalanced dataset. The measure is not only useful for calculating vowel mergers-in-progress, but can be used to represent any vocalic sound change, given at least one stable point in the speaker's vowel system.

ACKNOWLEDGMENTS

REFERENCES

1. J. Hay, P. Warren, and K. Drager, *Journal of Phonetics* **34**, 458–484 (2006).
2. D. DeCamp, *The Pronunciation of English in San Francisco*, Ph.D. thesis, University of California, Berkeley (1953).
3. B. Moonwomon, *Sound Change in San Francisco English*, Ph.D. thesis, University of California, Berkeley, Berkeley, CA (1992).
4. T. L. Irons, *Language Variation and Change* **19**, 137–180 (2007).
5. M. Baranowski, *Phonological Variation and Change in the dialect of Charleston, South Carolina*, vol. 92 of *Publication of the American Dialect Society*, Duke University Press, 2007.
6. M. Di Paolo, *Language and Communication* **12**, 267–292 (1992).
7. T. Majors, *American Speech* **80**, 165–179 (2005).
8. D. E. Johnson, *Stability and Change along a Dialect Boundary: The low vowels of Southeastern New England*, Ph.D. thesis, University of Pennsylvania, Philadelphia, PA (2007).
9. A. Dinkin, *Dialect Boundaries and Phonological Change in Upstate New York*, Ph.D. thesis, University of Pennsylvania (2009).
10. W. Labov, S. Ash, and C. Boberg, *The Atlas of North American English: Phonetics, phonology, and sound change*, Mouton de Gruyter, New York/Berlin, 2006.
11. W. Labov, *Language in Society* **1**, 97–120 (1972).
12. W. Labov, M. Yaeger, and R. Steiner, *A Quantitative Study of Sound Change in Progress: Report on National Science Foundation Contract GS-3287*, U.S. Regional Survey, Philadelphia, PA, 1972.
13. J. C. Wells, *Accents of English*, Cambridge University Press, Cambridge, UK, 1982.
14. R. H. Baayen, *Analyzing Linguistic Data: A Practical Introduction to Statistics using R*, Cambridge University Press, 2008.
15. C. L. Olsen, *Psychological Bulletin* **83**, 579–586 (1976).
16. M. Kennedy, *Variation in the Pronunciation of English by New Zealand school children*, Master's thesis, Victoria University of Wellington (2006).
17. L. Hall-Lew, *Ethnicity and Phonetic Variation in a San Francisco Neighborhood*, Ph.D. thesis, Stanford University (2009).
18. P. Adank, R. Smits, and R. van Hout, *Journal of the Acoustical Society of America* **116**, 3099–3107 (2004).
19. E. R. Thomas, and T. Kendall, Norm: The vowel normalization and plotting suite: An online tool for sociophonetic vowel normalization, http://ncslaap.lib.ncsu.edu/tools/norm/ (2007).
20. L. Hinton, S. Bremner, H. Corcoran, J. Learner, H. Luthin, B. Moonwomon, and M. van Clay, *Proceedings of the Annual Meeting of the Berkeley Linguistics Society* **13**, 117–127 (1987).
21. H. W. Luthin, "The Story of California /ow/: The Coming-of-Age of English in California," in *Variation in Language NWAV-XV at Stanford: Proceedings of the Fifteenth Annual Conference on New Ways of Analyzing Variation*, edited by K. M. Denning, S. Inkelas, F. C. McNair-Knox, and J. R. Rickford, Stanford, CA, 1987, pp. 312–324.
22. R. C. Conner, *The Fronting of /uw/ and /ow/ in Native Phoenicians' Speech*, Master's thesis, Arizona State University (1997).
23. R. Hagiwara, *Journal of the Acoustical Society of America* **102**, 655–658 (1997).
24. M. Ward, *Portland Dialect Study: The fronting of /ow, U, uw/ in Portland, Oregon*, Master's thesis, Portland State University (2003), URL `http://www.pds.pdx.edu/index.html`.
25. L. Hall-Lew, *Penn Working Papers in Linguistics* **10.2**, 105–116 (2005).
26. M. Di Paolo, and A. Faber, *Language Variation and Change* **2**, 155–204 (1990).
27. E. R. Thomas, *An Acoustic Analysis of Vowel Variation in New World English*, A Publication by The American Dialect Society 85, Duke University Press, 2001.
28. W. Labov, *Principles of Linguistic Change: Internal Factors*, vol. 20 of *Language and Society*, Blackwell Publishing, Ltd., Oxford, UK, 1994.
29. P. Trudgill, and T. Foxcroft, "On the sociolinguistics of vocalic mergers: Transfer and approximation in East Anglia," in *Sociolinguistic patterns in British English*, edited by P. Trudgill, University Park Press, Baltimore, 1978, pp. 69–79.
30. R. Herold, *Language Variation and Change* **9**, 165–189 (1997).