

Assessing merged status with Pillai scores based on dynamic formant contours

Jonathan Jibson*

Abstract. When Pillai scores are used to study vowel mergers, formants are typically sampled from the midpoint. This study compares alternative methods for calculating Pillai scores: methods that incorporate dynamic spectral information. Eighteen speakers produced 20 tokens of *Hodd* and *hawed*. Formants were sampled at 20–35–50–65–80% duration. Seven Pillai scores were calculated, each based on a different subset of those samples with temporal pooling: (i) onsets, (ii) heads, (iii) midpoints, (iv) onsets + offsets, (v) heads + tails, (vi) onsets + midpoints + offsets, and (vii) all five. Subjects also completed a vowel identification task, and the rate of identifying one low-back vowel as the other was calculated. The results of the identification task were regressed on each Pillai score separately to identify the one with the highest correlation, through model selection. Dynamic formant contours performed better than static formant values, with midpoint sampling performing worst of all. Directions are discussed for basic research on Pillai scores in phonetics.

Keywords. Pillai score; merger; vowel inherent spectral change; vowel dynamics low-back merger; sociophonetics; phonetics; individual differences

1. Introduction. The low-back merger, or *cot-caught* merger, is perhaps the best studied vowel merger in American English (Labov 1994; Labov et al. 2006). It involves the merger of the /ɑ/ and /ɔ/ phonemes. Participation in this merger helps to signal regional dialect (Majors 2005; Doernberger & Cerny 2008; Kennedy & Grama 2012; inter alia) as well as ethnic dialect (Eberhardt 2008). Southeast Wisconsin has traditionally been described as unmerged (Labov 1966), but recent work has argued for a “stable mess” status (Purnell et al. 2017: 307), showing that merged and unmerged speakers have dotted the region for over a century.

When measuring mergers in production, the Pillai score has recently been championed as a best practice (Hay et al. 2006; Nycz & Hall-Lew 2014). It is a metric of distributional overlap that can be straightforwardly calculated for multivariate data, making it ideal for F1–F2 analyses. Pillai scores are an improvement over Euclidean distance, which only compares the center of two distributions with no regard for category size (i.e. token variability). Hall-Lew (2010: 8) also explains some drawbacks to Pillai scores in sociophonetic analysis. Most relevant here is the observation that off-glide formant movement is ignored, which can be problematic if vowel dynamics are among the acoustic cues in a contrast.

This last observation touches on a recurring concern with any metric for measuring vowel merger: traditionally, vowels are sampled only once. That is, the basis of evaluating merged status is a single F1–F2 point from a vowel’s temporal midpoint or spectrotemporal nucleus. But it is well known that vowel dynamics are essential in vowel identification (Hillenbrand et al. 1995) and in dialect differentiation (Jacewicz et al. 2011; Jacewicz & Fox 2012, 2020). Indeed, vowel inherent spectral change research has consistently found that at least two samples – most often at 20% and 80% of duration – are necessary in modeling vowels (Hillenbrand 2013; Morrison 2013). Dialectology has always incorporated dynamic information with categorical and impres-

* I am grateful for helpful conversation with Joey Stanley early in this research. I would also like to thank the attendees at LSA 95 for feedback, especially Valerie Freeman, Meg Cychosz, and Jason Shaw. All errors are my own. Author: Jonathan Jibson, University of Wisconsin–Madison (jibson@wisc.edu).

sionistic off-glide diacritics (e.g. Allen 1973–1976), and a host of modern sociophonetic work quantifies vowel dynamics (Plichta & Preston 2005; Gunter et al. 2020; Renwick & Stanley 2020; inter alia). It seems important, then, to incorporate vowel dynamics into analyses involving Pillai scores.

Previous researchers have undertaken the same task for other analyses, working to incorporate vowel dynamics into traditionally static methods. Most relevant for this paper would be Jacewicz & Fox (2017: 448). The goal was to better delimit the vowel space (in formant space). When examining the corner vowels, formants were sampled not just at the midpoint but at five times throughout duration (20–35–50–65–80%), and all five samples were pooled without regard to temporal order. This effectively quintupled the number of points on the plot over which the vowel space would be drawn. This kind of vowel space reflects a vowel’s entire spectral territory, not just a snapshot at one sample. That same kind of rethinking might benefit Pillai scores.

Here the central questions of this paper can be asked. First, can formant dynamics enrich Pillai score analyses? Second, can we validate the answer to the first question by using individual differences in perception – relating perception to production? It will be shown that dynamic Pillai scores correlate better with perception behavior than do the traditional static Pillai scores, though no model achieved particularly high correlation.

2. Methods. Students at the University of Wisconsin–Madison were recruited for this study. In-person announcements were made in introductory linguistics courses and communication sciences & disorders courses, and emails were sent to various language departments for distribution to declared majors. The requirements were (i) to have normal hearing and (ii) to have lived in Wisconsin from age 4 to age 12. The second criterion was used to control for dialect without listing complex inclusion criteria in recruitment materials. Subjects were compensated with \$10 for their efforts (and some received extra credit in one of their courses, as determined solely by the instructors of the various classes).

The number of subjects for this experiment was 17. Their ages ranged from 18 to 24, with a mean of 20.2. Two had completed a bachelor’s degree, and the other 15 had completed some college. Two reported their ethnicity as *Native American + Caucasian* and *Asian + Caucasian*, respectively, while the other 15 reported *Caucasian*. All reported their gender as male.

Participation involved two portions – reading followed by listening. The typical subject took just under 20 minutes for the reading portion and just under 30 minutes for the listening portion. Subjects completed their assigned experiment in a sound-attenuated booth in the phonetics lab of the University of Wisconsin–Madison. All reading tasks and listening tasks were done using Sennheiser PC8 USB headphones. The recording of speech was done using Audacity, with a bit rate of 24 bit and a sampling rate of 44.1 kHz. The listening portion was administered using Praat. For all tasks, subjects operated a 2017 MacBook Air with a 2.2 GHz Intel Core i7 processor and 8 GB of 1600 MHz DDR3 RAM.

2.1. PRODUCTION TASK. The experiment first involved a production task. This consisted of a word-in-phrase reading list. The target words were the 11 non-rhotic monophthongs in the context /hVd/ – *heed*, *hid*, *hey’d*, etc. Those target words were embedded in five different carrier phrases: *Please say ___ for me*; *We asked if ___ was the word*; *You thought that ___ sounded right*; *I’m reading ___ on the paper*; *They said that ___ was good*. Each word-in-phrase combination was repeated 6 times, for a total of 30 tokens of each target word. The first and last sentences on each page of the reading consisted of the target word *herd* in the various sentence frames. This was done because pilot subjects were prone to coughing, whispering, and creaking on the first and last sentences of a page, and some subjects would turn the page before finishing

the final sentence (despite instruction not to do so). The target words in those sentences were not analyzed.

Before the reading task began, subjects were shown the target words, and the experimenter used each one in a sentence and defined it if it was unknown to the subject. Subjects read seven practice sentences as well in this training phase. Feedback was given on speaking rate in cases of particularly careful or hasty speech, though the controlled conditions for all subjects meant that variation in speaking rate would be low (cf. Quené 2008).

Not all tokens that were recorded were analyzed. Some words were misread, and others proved difficult for Praat to formant-track. Thus, only 20 tokens for each vowel were analyzed. If a subject produced more than 20 viable tokens of a vowel, tokens with outlier durations were discarded. This was chosen as the criterion because the formant trajectories of a given token depend in part on duration, with shorter tokens showing undershoot (Lindblom 1963; Flemming 2004; Gendrot & Adda-Decker 2005).

With the 20 tokens per vowel identified, acoustic values were gathered. F1 and F2 were sampled 5 times throughout duration: at 20–35–50–65–80% of duration, with the sampling window being 6% of token duration (i.e. the width of the window varied from token to token). Hertz values were converted to the Bark scale. A median F1–F2 point was calculated for each of the 5 samples, producing a course-grained formant contour for each speaker. This process was done for both of the low-back vowels.

From the 5 samples came several subsets or combinations. These subsets correspond to different temporal sampling methods used in different studies, as follows. With static sampling, the conventional point is at 50%. The approach taken in Holt (2011) is 35%, which was meant to more closely approximate the spectral nucleus of the vowel. Using the 20% is conventionally accepted in vowel inherent spectral change research as necessary (Morrison 2013). As for dynamic models, the dual-target hypothesis holds that 20–80% samples are the optimal model of vowels (Hillenbrand 2013). A similar approach is taken by Assmann & Katz (2000: 1858, 1865–1866) and by Purnell (2007: 16), where sampling is done one-third and two-thirds through vowel duration. Farrington et al. (2018: 195) used 20–50–80% in order to minimally capture potential triphthongal movement. Finally, using all five points is done in several large studies (Jacewicz et al. 2006, et seq; Renwick & Olsen 2017; Renwick & Stanley 2020). Thus, the seven subsets for this study are as follows:

- Onsets alone (20%)
- Heads alone (35%)
- Midpoints alone (50%)
- Onsets and offsets together (20–80%)
- Heads and tails together (35–65%)
- Onsets, midpoints, and offsets together (20–50–80%)
- All 5 samples together (20–35–50–65–80%)

The goal was to implement the temporal pooling method of Jacewicz & Fox (2017) but with varying levels of fidelity or detail. Figure 1 below shows a comparison of two of these subsets: both panels report on the same tokens, all of which are raising and fronting. A Pillai score between the low-back vowels was calculated for each of the seven subsets separately. As described below, the seven resulting Pillai scores served as the different indices in model selection.

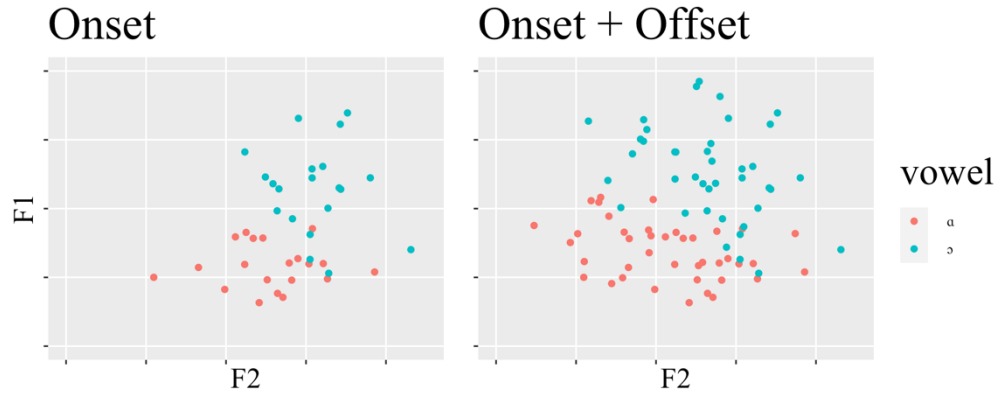


Figure 1. Two subsets of the same 20 /a/ and /ɔ/ tokens, with different subsetting.

2.2. PERCEPTION TASK. The stimuli in this study were generated using the Berkeley Phonetics Machine (Sprouse & Johnson 2016), an implementation of the Klatt synthesizer (Klatt 1980). The values for F1 and F2 were taken from Jacewicz et al. (2011: 80–86). That study reports F1 and F2 values at 20–35–50–65–80% of vowel duration for male and female speakers of three generations in three dialect areas, with words being spoken in the context /hVd/. For the present study, the male values for the Parent generation in Wisconsin ($n = 14$) served as the source.

These formant values were used as the baseline for creating the vowel stimuli. The 5-point trajectory was re-created by linear interpolation from one sample to the next. Formant values at 5–95% duration were estimated by extrapolation, since the source data omitted nearly half the vowel’s duration. Specifically, the contour from 65–80% duration was continued but with half the length, so that direction was identical but spectral rate of change was halved. These stimuli were used in two separate vowel identification tasks which were completed by the same subjects. In one task there were two repetitions of each, and in the other there were three repetitions of each. Thus, this study reports on the identification rates for five tokens of both vowels.

In the vowel identification task, the visual display consisted of 10 options, namely the 10 /hVd/ words tested in this study. It was decided that *hud* would not be included because no /ʌ/ stimuli were actually created, even though speakers may have heard *hud* on a given trial. The options were arranged alphabetically (by column) in two columns and five rows. Subjects advanced to the next trial immediately upon selecting a vowel category. The instructions were to choose which vowel was heard. Subjects were familiar with the words because of the reading task they completed beforehand. Trials were fully randomized for each speaker.

Subjects were first presented with a 20-item training session (Jacewicz & Fox 2012: 1415; Oder et al. 2013: 28). This training session operated precisely as in the test session, but the stimuli consisted of vowels using female phonetics from the same source. This familiarized subjects with the nature of the Klatt stimuli without revealing any stimuli used in the test session.

3. Results. Pillai scores based on the seven subsets are given in Figure 2. The distributions are similar because they come from the same underlying data combined in different ways. Two noteworthy differences emerge, though. First, the three static models skew higher than the four dynamic models – all quartiles are higher, but the difference across models is largest in the higher quartiles. Second, the distributions are bimodal, with a primary peak around 0.75 and a secondary peak around 0.20. Bimodality indicates that these speakers were either unmerged or merged rather than somewhere in between.

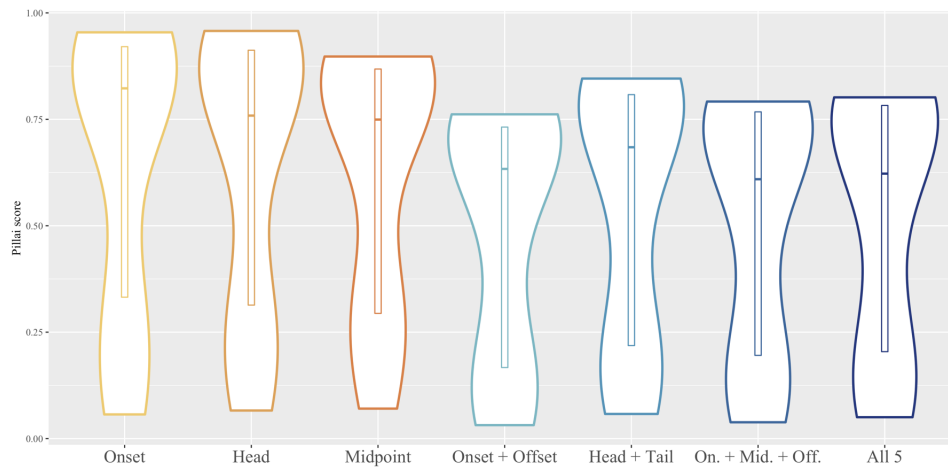


Figure 2. Pillai scores for the seven subsets.

The basic method used to address this paper’s central questions is model selection. In this approach, the same dependent variable is regressed on different independent variables, and the variable with the highest correlation is chosen as the optimal model. Here, the dependent variable is the average rate of categorizing one low-back vowel as the other (the mutual confusion rate). This rate reflects merged status better than the correct categorization rate because it isolates the vowels relevant to the low-back merger. If the phonemes are merged for a speaker, it could be that both vowels are categorized at chance levels for each other, or it could be that both vowels are identified as one or the other. This distinction should not matter for the present study – different patterns would suggest different mechanisms of merger (i.e. approximation versus expansion; see Trudgill & Foxcroft 1978; Herold 1990) but not different merged statuses. This mutual confusion data was the same in each model. The independent variables used in the different models are the seven subset-based Pillai scores. Again, these scores varied by a small margin since they were different subsets of the same underlying data.

Figure 3 shows the data that was fitted with linear models. Each panel plots the same mutual confusion rate against one of the Pillai scores from Figure 2. Simple linear regression was performed with two fixed factors: mutual confusion rate and Pillai score. Table 1 lists the coefficient of determination for each of the seven models. It is clear that the dynamic model Onset + Offset has the largest value, whereas the Midpoint model has the lowest value. In fact, with only one exception, all dynamic models outperformed all static models. By broad convention, all these correlation values are usually considered moderate or medium in psychology research (Cohen 1992: 157) and small in social science research (Ferguson 2009: 533), though these are only rules of thumb.

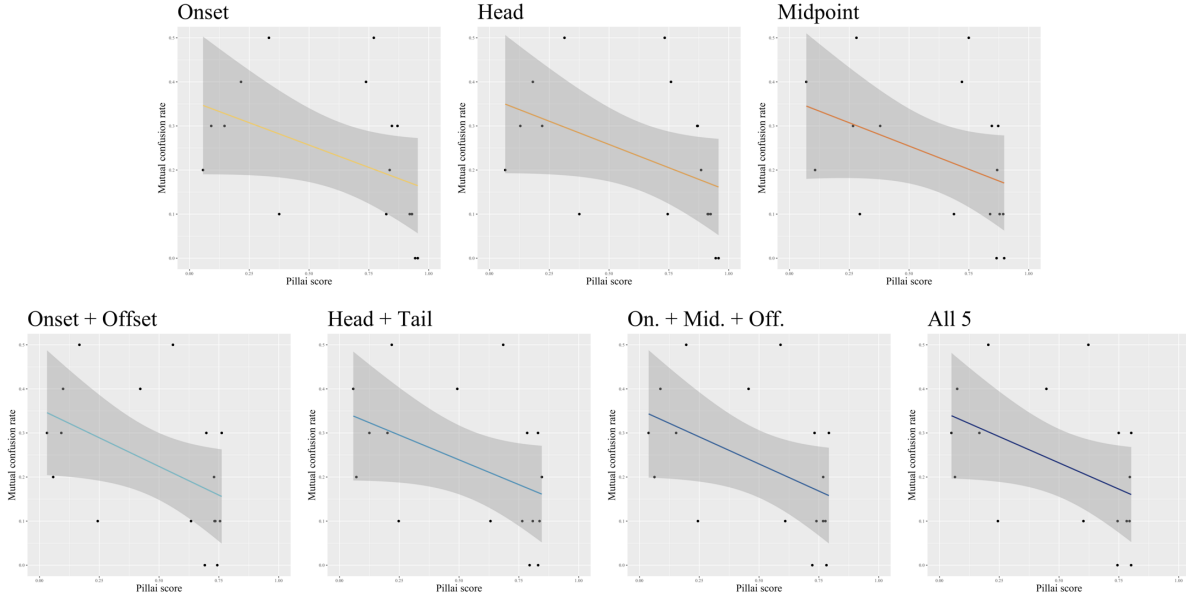


Figure 3. Regression models for the seven subset-based Pillai scores.

Table 1. Coefficients of determination (adjusted R^2) for the models in Figure 3.

Onset	Head	Midpoint	Onset + Offset	Head + Tail	On. + Mid. + Off.	All 5
0.130	0.136	0.106	0.172	0.135	0.154	0.149

4. Discussion. Using dynamic spectral information in calculating Pillai scores improved their correlation with perception behavior. Pillai scores based on onset and offset samples performed best, following the dual-target hypothesis of vowel inherent spectral change. The model with the lowest correlation was in fact the midpoint model. Since dynamic models almost always led to higher correlation with perception, it is recommended that researchers calculate Pillai scores based on dynamic sampling rather than static sampling.

It should be noted immediately, though, that there is one case in which temporal pooling would not improve the analysis. It is when direction is contrastive, as noted by Hall-Lew (2010: 8). This is demonstrated well with the front vowels in the Southern Vowel Shift (see Thomas 2020 and references therein). In particular, /e/ and /ɪ/ almost have equal but time-reversed formant values (Jacewicz et al. 2011: 58). Since temporal pooling does not preserve temporal structure, an onset + offset method would end up with two equivalent regions for these two vowels. Still, in the traditional method of sampling from the midpoint, there is simply one equivalent formant region. Thus, temporal pooling is no worse than the traditional method, just no better.

A follow-up analysis could attempt to preserve temporal structure. The Pillai score is an output of a MANOVA model, and the convention has been to use two factors, namely F1 and F2. But the model can handle a third variable, such as for time point. In principle, the Pillai score of such a model should represent three-dimensional overlap of a series of formant space plots being

stacked on top of each other. This was not done in the present analysis out of an abundance of caution, due to a frank lack of familiarity with such statistical modeling.¹

The analysis of this study involves portions of two separate vowel identification tasks, and the stimuli were not naturalistic. This was done strictly for convenience, gleaning insight from another study with different aims. If the stimuli had consisted of naturalistically recorded tokens, perception behavior might have been different, and it is unclear what impact this would have on the correlation of the various models above. It would be going too far, though, to ignore the results of this study entirely. Though no listener would be fooled into thinking that the stimuli are from humans, they are still perceived as speech (Jibson 2019, 2020). Future work will include as stimuli the naturalistic productions that were recorded from the speakers in the present study.

The relationship between production and perception was modeled linearly here, but it is not clear that this is appropriate for such data. Most sociophonetic work has instead supported sigmoidal trends (see D'Arcy 2015: 589–591 and references therein). In any event, basic descriptive statistics for individual differences in this correlation are lacking. Replicating the current study with a larger number of subjects would help reveal the typical shape of the correlation between production and perception. It might be the case, for instance, that the bimodality seen here is representative of most speakers. But preliminary analysis of other speakers (not reported on here) suggests that Pillai scores are more continuously distributed across speakers. Such future work should also explicitly ask subjects to report their awareness of the merger (Di Paolo 1992; Labov et al. 2006) in order to further enrich our understanding of merger in progress.

To go further on this point, more research is needed to establish typical Pillai scores for canonical or unambiguous situations. What, for instance, is the lowest Pillai score we should expect for a fully merged speaker – approximately zero? And what is a normal score for the difference between /i/ and /e/ in the Midlands or Inland North dialects, which are in stable contrast? What is a normal score for that difference in Spanish dialects? Much work has been done on dispersedness (see Hall 2011 and references therein), but it rests almost exclusively on Euclidean distance of centroids, which is inferior to measures of overlap. Until those baseline descriptive statistics are known, the interpretation of any given analysis using these tools will be ad hoc.

The model selection paradigm in this paper can test a variety of regressors to answer different questions. For instance, mutual confusion can be regressed on Euclidean distance and Bhattacharyya's affinity as well as Pillai score in order to identify the best correlate with perception behavior. Such empirical evidence, if found, would serve as strong validation for the more theoretical basis for using Pillai scores described above. And the paradigm would work with the results of other perceptual tasks, as well, such as sensitivity indices from discrimination tasks.

The conclusion at present is that merger in perception, as measured by mutual confusion, correlates to a small degree with merger in production, as measured by Pillai scores. Furthermore, there is evidence that Pillai scores are more veridical and behaviorally valid when based on dynamic rather than static formant values. It is recommended that studies using Pillai scores to assess merger incorporate dynamic formant values.

References

Allen, Harold B. 1973–1976. *The linguistic atlas of the Upper Midwest*. 3 vols. Minneapolis: University of Minnesota Press.

¹ My thanks to Joey Stanley and Valerie Freeman for independently raising this possibility, which will be addressed in future work.

- Assmann, Peter F. & William F. Katz. 2000. Time-varying spectral change in the vowels of children and adults. *Journal of the Acoustical Society of America* 108. 1856–1866.
- Cohen, Jacob. 1992. A power primer. *Psychological Bulletin* 112. 155–159.
- D’Arcy, Alexandra. 2015. Variation, transmission, incrementation. In Patrick Honeybone & Joseph Salmons (eds.), *The Oxford handbook of historical phonology*, pp. 583–602. Oxford: Oxford University Press.
- Di Paolo, Marianna. 1992. Hypercorrection in response to the apparent merger of (a) and (ɔ) in Utah English. *Language & Communication* 12(3). 267–292.
- Doernberger, Jeremy & Jacob Cerny. 2008. The low-back merger in Miami. *University of Pennsylvania Working Papers in Linguistics* 14(2). 11–15.
- Eberhardt, Maeve. 2008. The low-back merger in the Steel City: African American English in Pittsburgh. *American Speech* 83(3). 284–311.
- Farrington, Charlie, Tyler Kendall & Valerie Fridland. 2018. Vowel dynamics in the Southern Vowel Shift. *American Speech* 93. 186–222.
- Ferguson, Christopher J. 2009. An effect size primer: A guide for clinicians and researchers. *Professional Psychology: Research and Practice* 40(5). 532–538.
- Flemming, Edward. 2004. Contrast and perceptual distinctiveness. In Hayes et al. (eds.), *Phonetically based phonology*, 233–276. Cambridge: Cambridge University Press.
- Gendrot, Cédric & Martine Adda-Decker. 2005. Impact of duration on F1/F2 formant values of oral vowels: An automatic analysis of large broadcast news corpora in French and German. *INTERSPEECH* 2005. 2453–2456.
- Gunter, Kaylynn M., Charlotte R. Vaughn & Tyler S. Kendall. 2020. Perceiving Southernness: Vowel categories and acoustic cues in Southernness ratings. *Journal of the Acoustical Society of America* 147. 643–656.
- Hay, Jennifer, Paul Warren & Katie Drager. 2006. Factors influencing speech perception in the context of a merger-in-progress. *Journal of Phonetics* 34. 458–484.
- Hall, Daniel Currie. 2011. Phonological contrast and its phonetic enhancement: Dispersedness-without dispersion. *Phonology* 28.1–54.
- Hall-Lew, Lauren. 2010. Improved representation of variance in measures of vowel merger. *Proceedings of Meetings on Acoustics* 9. 060002.
- Herold, Ruth. 1990. Mechanisms of merger: The implementation and distribution of the low back merger in Eastern Pennsylvania. Doctoral dissertation, University of Pennsylvania.
- Hillenbrand, James M., Laura A. Getty, Michael J. Clark & Kimberlee Wheeler. 1995. Acoustic characteristics of American English vowels. *Journal of the Acoustical Society of America* 97. 3099–3111.
- Hillenbrand, James M. 2013. Static and dynamic approaches to vowel perception. In Geoffrey Steward Morrison & Peter F. Assmann (eds.), *Vowel inherent spectral change*, 9–30. Heidelberg: Springer.
- Holt, Yolanda. 2011. A cross generational dialect study in Western North Carolina. Doctoral dissertation, The Ohio State University.
- Jacewicz, Ewa, Robert Allen Fox & Joseph Salmons. 2006. Prosodic prominence effects on vowels in chain shifts. *Language Variation and Change* 18. 285–316.
- Jacewicz, Ewa, Robert Allen Fox & Joseph Salmons. 2011. Cross-generational vowel change in American English. *Language Variation and Change* 23. 45–86.

- Jacewicz, Ewa & Robert Allen Fox. 2012. The effects of cross-generational and cross-dialectal variation on vowel identification and classification. *Journal of the Acoustical Society of America* 131. 1413–1433.
- Jacewicz, Ewa & Robert A. Fox. 2017. Dialect perception by older children. *INTERSPEECH*, pp. 354–358.
- Jacewicz, Ewa & Robert Fox. 2020. Perception of local and non-local vowels by adults and children in the South. *Journal of the Acoustical Society of America* 147. 627–642.
- Jibson, Jonathan. 2019. An AX experiment exploring perceptual fidelity in vowels. Poster presented at the 24th Mid-continental Phonetics and Phonology conference, 5 October, Milwaukee, Wisconsin.
- Jibson, Jonathan. 2020. Vowel identification and goodness based on level of formant detail. *Journal of the Acoustical Society of America* 148. 2505.
- Kennedy, Robert & James Grama. 2012. Chain shifting and centralization in California vowels: An acoustic analysis. *American Speech* 87(1). 39–56.
- Klatt, Dennis. 1980. Software for a cascade/parallel formant synthesizer. *Journal of the Acoustical Society of America* 67. 971–995.
- Labov, William. 1966. *The social stratification of English in New York City*. Washington, DC: Center for Applied Linguistics.
- Labov, William. 1994. *Principles of linguistic change: Vol. 1, internal factors*. Oxford: Blackwell.
- Labov, William, Sharon Ash & Charles Boberg. 2006. *The atlas of North American English: Phonetics, phonology, and sound change*. Berlin: Mouton de Gruyter.
- Lindblom, Björn E. 1963. Spectrographic study of vowel reduction. *Journal of the Acoustical Society of America* 35, 1773–1781.
- Majors, Tivoli. 2005. LOW back vowel merger in Missouri speech: Acoustic description and explanation. *American Speech* 80(2). 165–179.
- Morrison, Geoffrey Stewart. 2013. Theories of vowel inherent spectral change. In Geoffrey Stewart Morrison & Peter F. Assmann (eds.), *Vowel inherent spectral change*, 31–48. Heidelberg: Springer.
- Nycz, Jennifer & Lauren Hall-Lew. 2013. Best practices in measuring vowel merger. *Proceedings of Meetings on Acoustics* 20(060008).
- Oder, Austin L., Cynthia G. Clopper & Sarah Hargus Ferguson. 2013. Effects of dialect on vowel acoustics and intelligibility. *Journal of the International Phonetic Association* 43(1). 23–35.
- Plichta, Bartłomiej & Dennis R. Preston. 2005. The /ay/s have it: The perception of /ay/ as a North-South stereotype in United States English. *Acta Linguistica Hafniensia* 37(1). 107–130.
- Purnell, Thomas. 2007. Formant velocities: Additional evidence of dynamic formant trajectories. Unpublished manuscript, Madison, WI.
- Purnell, Thomas, Eric Raimy & Joseph Salmons. 2017. *Upper Midwestern English*. In Raymond Hickey (ed.), *Listening to the past: Audio records of accents of English*, pp. 298–324. Cambridge: Cambridge University Press.
- Renwick, Margaret E.L. & Joseph A. Stanley. 2020. Modeling dynamic trajectories of front vowels in the American South. *Journal of the Acoustical Society of America* 147. 579–595.
- Renwick, Margaret E.L. & Rachel M. Olsen. 2017. Analyzing dialect variation in historical speech corpora. *Journal of the Acoustical Society of America* 142. 406–421.

- Sprouse, Ronald L. & Keith Johnson. 2016. The Berkeley Phonetics Machine. *UC Berkeley Phonetics and Phonology Lab Annual Report* 12.
- Thomas, Erik R. 2020. Sociophonetic trends in studies of Southern US English. *Journal of the Acoustical Society of America* 147. 529–540.
- Trudgill, Peter & Tina Foxcroft. 1978. On the sociolinguistics of vocalic mergers: Transfer and approximation in East Anglia. In Peter Trudgill (ed.), *Sociolinguistic patterns in British English*. Baltimore: University Park Press. 69–79.