JASA ARTICLE

Perceiving Southernness: Vowel categories and acoustic cues in Southernness ratings

Kaylynn M. Gunter,[a] Charlotte R. Vaughn, and Tyler S. Kendall
Department of Linguistics, University of Oregon, Eugene, Oregon 97403-1290, USA

ABSTRACT:
Southern U.S. speech has been the focus of much sociophonetic work. In terms of vowel patterns, Southern speech is often characterized by the Southern Vowel Shift (SVS, involving shifts in /e/, /ɛ/, and /aɪ/), back vowel fronting, and changes in glide dynamics. The SVS, in particular, is said to play a primary role in distinguishing the South as a unique dialect region. However, there have been few investigations of the role of various vowel quality differences in perceptions of Southern accent, particularly across the vowel space beyond /e/, /ɛ/, and /aɪ/, or that ask whether any aggregate speaker-level acoustic measures align with listeners' perceptions, despite some suggestions in the literature to this effect. The current study examines what acoustic cues contribute to non-Southern listeners' evaluations of words spoken by Southerners as sounding more or less Southern accented, looking at a range of vowels from across the vowel space. Results indicate that listeners rate the speakers' productions of /u/ and /ɔ/ as most Southern and that vowel dynamics and speaker-level measures were the acoustic factors most predictive of Southernness ratings. These results together call for further work examining vowel dynamics and a more complete set of vowel categories in perception studies of Southern speech. © *2020 Acoustical Society of America*.
https://doi.org/10.1121/10.0000550
(Received 31 March 2019; revised 20 October 2019; accepted 31 October 2019; published online 31 January 2020)
[Editor: Wendy Herd] Pages: 643–656

I. INTRODUCTION

The Southern U.S. is considered a unique dialect region by linguists and everyday listeners alike (Preston, 1993, 2015). Sociolinguists have examined the phonological system of speakers in the South, and highlighted many shibboleths of Southern speech, often focusing on vowel quality differences. Frequently, research indicates a particular vowel configuration, known as the Southern Vowel Shift (or SVS), as the predominant marker of Southernness (Fridland, 2000; Thomas, 2001; Labov *et al.*, 2006). However, less is known about what, specifically, triggers perceptions of Southernness by everyday listeners. Even studies on perceptions of Southernness that have explicitly focused on vowels (see Sec. II B) have typically investigated only a subset of vowels (e.g., /e, ɛ, aɪ/), limiting a more nuanced understanding of how other aspects of the vowel system might contribute. In addition, recent work (e.g., Kendall and Fridland, 2012) has sometimes attempted to use speaker-level summary measures (e.g., the distance between /e/-/ɛ/) to rank individual speakers' degree of overall Southernness. However, the extent to which such speaker-level measures align with listener perceptions remains relatively untested.

In this paper, we seek to better understand the role of vocalic features in the perception of Southernness, examining which vowels and which acoustic cues contribute to non-Southern listeners' evaluations of words as sounding more or less Southern. We focus on vowels because vowels are the subject of much work on regional patterns in production (e.g., Labov *et al.*, 2006; Kendall and Fridland, 2012, 2017) and because there is evidence of metalinguistic commentary on Southerners' vowels by non-linguists (Preston, 1993; Albritten, 2011).[1] We also focus on vowels because, despite much prior research on Southern vowels, very little work has actually examined a broad inventory of the vowel space in terms of what makes vowels sound "Southern." As a result, our knowledge of Southernness is based primarily on just a subset of potentially important vowel categories (e.g., /e, ɛ, aɪ/) and on a subset of potentially important acoustic cues (e.g., monophthongization for /aɪ/).

Our paper presents the results of an experiment where listeners in the Western U.S. were asked to rate single word tokens (e.g., *bad*) on their degree of Southernness (on a scale of 1–9). The stimuli were produced by Southern speakers (ranging in their degree of participation in the SVS shift) and Western speakers, who were used as a baseline. Word stimuli include vowels from a wider range of vowel categories (/i, ɪ, e, ɛ, æ, u, ʊ, ʌ, o, ɔ, ɑ, aɪ/) than those typically considered in perceptual studies of Southernness (e.g., /e, ɛ, aɪ/) in order to look at the contributions across the vowel space to listeners' perceptions of Southernness. Unlike many prior studies of perceptions of (Southern) vowels, which have used manipulated speech to investigate the influence of specific acoustic cues (e.g., Fridland *et al.*, 2004; Plichta and Preston, 2005), we presented listeners with unaltered, natural productions, in order to assess which vowel categories are perceived as most Southern, and which acoustic cues, both static and dynamic, correlate with

[a]Electronic mail: kgunter@uoregon.edu

perceptions of Southernness, without presupposing which cues might be important. Additionally, we explore the nature of speaker-level measures in listener perceptions of Southernness ratings.

To set up the present study, we first review what is known about the vowel systems of Southern speakers, providing relevant background for the acoustic features we consider (both token-level and speaker-level). Because of our interest in a comprehensive picture of what contributes to Southernness ratings, we review production studies of U.S. Southern vowels in detail. From there we turn to what is known about the perception of Southernness in the current literature, illustrating how our study adds to this body of work.

II. BACKGROUND

A. Production patterns

The Atlas of North American English (ANAE; Labov *et al.*, 2006) outlines the South as a distinct dialect region with great within-regional variability and distinct subregions. In general, though, many of the same phonological features can be seen throughout the broader region of the South (Labov *et al.*, 2006). The most well-documented feature of Southern speech is the SVS, characterized by movement in the front vowels and /aɪ/ monophthongization (Feagin, 1986; Labov, 1991; Fridland, 2000, 2003, 2012; Fridland and Kendall, 2015). The SVS is most often described as a shift in the nuclei (e.g., midpoint, steady-state, or 50% point) of the front tense and lax vowels by which the lax vowels raise and front, moving toward the periphery, and the tense vowels lower and back, resulting in acoustic overlap or reversal (Fridland, 2000; Labov *et al.*, 2006). The most widespread feature of the SVS across subregions of the South is the movement of the mid front vowels /e/ and /ɛ/, often described as one of the most characteristic features of Southern speech (Fridland, 2000; Labov *et al.*, 2006; Kendall and Fridland, 2012). In rare cases, /e/ and /ɛ/ show complete reversal in $F1$ and $F2$ space; however, they are most often realized with close spectral proximity of the nuclei rather than a complete reversal. Recent work has also found that /e/ and /ɛ/ are distinguished by glide dynamics despite proximity or overlap of the nuclei (Farrington *et al.*, 2018). Descriptions of the SVS often include the high front vowels as well as the mid vowels, but overlap/reversal of the high front vowels is rarer and limited to the Inland South (Labov *et al.*, 2006; Fridland, 2012).

Beyond static spectral measures of vowel quality in the SVS, many researchers argue that dynamic qualities are defining features of Southern speech production that largely contribute to notions of the Southern "drawl" (Feagin, 1987; Wetzell, 2000). One such feature discussed at length in the literature is the presence of glide weakening (monophthongization) of /aɪ/. Previous work indicates /aɪ/ monophthongization occurs across Southern varieties preceding voiced stops and word-finally (Feagin, 2000; Thomas, 2003; Fridland, 2003), and less often in pre-voiceless contexts, the latter being sensitive to within-region variability and social class differences (Feagin, 2000; Thomas, 2003). As such, many researchers often describe /aɪ/ as another defining characteristic of Southern speech (Wolfram *et al.*, 1998; Feagin, 2000; Thomas, 2003; Plichta and Preston, 2005; Reed, 2016) and non-linguists are aware of its status as a regional identifier (Plichta and Preston, 2005). Additionally, /aɪ/ monophthongization is traditionally described as phase one of the SVS, initiating shifts in the mid front vowels (Labov *et al.*, 2006). As a result, the SVS-shifted /e, ɛ, aɪ/ are thought to be characteristic of Southern speech and thus often explored in both production and perception studies (see Sec. II B).

In addition to /aɪ/ glide weakening, another frequently discussed aspect of glide dynamics is glide breaking and lengthening, often labeled by researchers as the Southern "drawl" and most canonically associated with /æ/ (Sledd, 1966; Feagin, 1987, 2000; Thomas, 2003; Koops, 2014). The "Southern drawl" is often impressionistically described by everyday listeners as having long, drawn-out vowels (Preston, 1993). Research examining regional differences in vowel duration has found that Southern speakers' vowels are longer than other regional varieties (Wetzell, 2000; Clopper *et al.*, 2005; Jacewicz and Salmons, 2007; Fox and Jacewicz, 2009) and these differences are most prominent in the front lax vowels of Southern speakers who show most evidence of the SVS (Clopper *et al.*, 2005; Fridland *et al.*, 2014).

Recent work on vowel dynamics has aimed to understand the role that vowel inherent spectral change (see Nearey and Assmann, 1986) has on defining vowel qualities of regional U.S. dialects (Fox and Jacewicz, 2009; Farrington *et al.*, 2018). By quantifying more precise measures of change over the course of the vowel, finer-grained phonetic distinctions arise that could potentially differ between regions. Fox and Jacewicz (2009) introduce a methodology to capture glide length (referred to as vector length; VL), internal movement over the course of the vowel (trajectory length; TL), and the rate at which the spectral information changes (rate of change, ROC), comparing these measures across regions. Their findings illustrate that the South often has more formant trajectory change over the course of front vowels than other regions. Farrington *et al.* (2018) utilize similar measures, with the addition of glide directionality (vector angle; VA), to examine the relationship between SVS participation and vowel dynamic qualities in the front vowels. In general, Farrington *et al.* find that the mid front vowels of SVS-shifted speakers have stronger dynamic distinctions than the mid front vowels of non-shifted and Western speakers.

Back vowel fronting is also a commonly described feature of Southern speech, as it is pervasive throughout the South. Back vowel fronting affects the /u/ and /o/ classes in tandem but often with /o/ being less advanced (Fridland and Bartlett, 2006). The back vowels are not frequently referenced as defining features of contemporary Southern speech in part because similar shifts can be seen across regions and are thus not thought to be uniquely Southern. Additional

https://doi.org/10.1121/10.0000550

research suggests that /u/ fronting can be further distinguished in terms of glide dynamics, with Southerners having a more monophthongal /u/ with a fronter offglide than other regions (Koops, 2010; Albritten, 2011). However, this has not yet been explored in depth.

In addition to the mid and high back vowels, the low back vowels (/ɑ/ and /ɔ/) historically remain distinct in many parts of the South (Thomas, 2001; Labov *et al.*, 2006; Fridland and Kendall, 2015; Kendall and Fridland, 2017). Recent work has demonstrated, however, that speakers' distributions of these vowels are rather variable, with many Southern speakers exhibiting or approaching merger (Kendall and Fridland, 2017). For speakers with distinct spectral positions of /ɑ/ and /ɔ/, /ɔ/ is characterized by a diphthongal upglide and maintenance of rounding (Irons, 2007; Thomas, 2001; Labov *et al.*, 2006). Further, work by Majors (2005) considers the different trajectories that set apart these vowel categories for Southern speakers, noting that upgliding for /ɔ/ is greater than for /ɑ/, whereas in speakers with merged low back vowels their trajectories are identical.

The literature discussed thus far has typically focused on the vowel quality of individual vowel categories. However, some work has explored how more holistic measures of a speaker's vowel space (i.e., speaker averages over numerous tokens, such as a speaker's average /aɪ/ glide length) may be used to capture an individual's degree of Southernness more generally. Speaker-level summaries thus provide a quantitative measure of individual speakers' degree of participation in Southern norms (i.e., the SVS), rather than relying on a speaker's Southern regional affiliation, or relative degree of shift in individual tokens, as a stand-in for SVS participation.

Kendall and Fridland (2012) suggest that the degree of proximity of the mid front vowels /e/ and /ɛ/, as measured by Euclidean distance (ED), can serve as a useful diagnostic when assessing individual participation in the SVS. They argued that /e/-/ɛ/ ED correlated impressionistically with a degree of participation in SVS (such that more proximate /e/ and /ɛ/ correlated with more use of other Southern features), and also demonstrated that this ED measure patterned with their participants' perception behavior in /e/–/ɛ/ vowel classification. This measure has since been used as a metric of Southernness and its utility has been demonstrated in predicting individual behavior in Southern speech norms beyond the proximity of a speaker's /e/–/ɛ/ classes (Farrington *et al.*, 2018). However, the extent to which this ED measure corresponds with actual, everyday listeners' perceptions has remained untested. Thus, we take a primary interest in whether /e/–/ɛ/ ED correlates with listeners' perceptions of Southernness. In addition to /e/–/ɛ/ ED, we investigate several other potential speaker-level measures that may capture speakers' degree of Southernness. Namely, we examine low back vowel overlap (using the Pillai statistic for the overlap between the distributions for /ɑ/ and /ɔ/; see Kendall and Fridland, 2017), /aɪ/ glide length (see Fridland, 2003), and mean /u/ frontness. Altogether, these four speaker-level measures represent attested characteristics of Southern speech. Given the production literature on Southern speech, we expect that these four measures might correlate with a speaker's overall level of perceived Southernness, since individual Southerners may conform to these patterns to a greater or lesser extent. Thus, we test whether speakers' overall values for these four measures are predictive of how Southern their individual vowel productions are perceived to be.

As we have observed in this section, there have been a number of production studies examining features of Southern vowels across the vowel space. The comprehensive overview of both static (e.g., shifts in *F*1 and *F*2) and dynamic properties (e.g., monophthongization) of vowel quality differences suggests a number of potential cues that may be available to listeners when evaluating Southernness. Thus, in this paper we utilize both static measures (*F*1 and *F*2 at multiple time points) and vowel dynamic measures (VL, TL, VA, ROC, and vowel duration), as well as speaker-level aggregate measures, to explore the nature of these cues in perceiving Southernness across the larger vowel space.

B. Perception of Southern speech

Work on the production of Southern U.S. English has provided insight into how Southern vowels differ from other regional varieties of American English. Meanwhile, there has been less work on the perception of Southernness, and much has focused on only a subset of vowels. From the existing work on perception, however, it is clear that listeners are sensitive to a range of information when categorizing speech as Southern (Clopper and Pisoni, 2004; Fridland *et al.*, 2004; Albritten, 2011).

Most directly compatible with the questions we explore in this paper are the studies by Fridland *et al.* (2004) and Fridland (2008). Fridland *et al.* (2004) examined the role of Southern shifted front (/e/, /ɛ/, /i/, /ɪ/) and back vowels (/o/, /u/, /ʊ/) in identification of Southernness within listeners in Memphis, TN. Their two-alternative forced-choice experiment asked listeners to select the token that was produced by a Southerner. Each trial contained two resynthesized versions of the same word spoken by a Memphian, a shifted version and a non-shifted version (e.g., along both *F*1 and *F*2 dimensions). They found that listeners identified shifted mid front vowels as Southern more often than fronted back vowels. In a follow-up study, Fridland (2008) examined the same items but with listeners in the Western U.S. (Reno, NV), establishing a basis for our approach collecting perception ratings of Southernness by non-Southerners. This study found that Western listeners were more accurate overall than Memphians at identifying the Southern shifted tokens, and Western listeners also were more likely to correctly identify front vowels with shifted variants as Southern than back vowels.

While Fridland (2008) found that Westerners are not as sensitive to the shifts in the back vowels, Torbert (2004) found contrasting results for a similar study with a sample of non-Southern listeners. In this experiment, he provided both Southern and non-Southern listeners with three types

of speech stimuli: unmodified, monotonized, and low-pass filtered speech. Each of these types contained tokens of /o/ and /u/ that were either fronted or backed, and tokens of /aɪ/ that were either monophthongal or diphthongal. After each stimulus, participants were asked to rate how Southern the speaker was on a scale of 1–5. Torbert found that listeners did not rate stimuli with fronted /o/ to be more Southern than backed /o/, but rated stimuli with monophthongal /aɪ/ and fronted /u/ to be more Southern than diphthongal /aɪ/ or backed /u/.

Further examining the role of /aɪ/ in evaluations of Southernness, Plichta and Preston (2005) demonstrated that listeners from various U.S. regions are also sensitive to the degree of /aɪ/ monophthongization when categorizing tokens along a continuum from North to South. Listeners were exposed to resynthesized tokens of /aɪ/ that varied in their degree of monophthongization. The results showed that listeners were able to categorize each token along a North-middle-South continuum with the most monophthongal tokens categorized as more South-like and least monophthongal as most North-like. Plichta and Preston (2005) argue that monophthongal /aɪ/ acts as a stereotype for Southern speech to non-Southern listeners.

Some work has sought to address questions of how both static and dynamic vowel differences work in concert in the perception of Southernness. Albritten (2011) examined the role of four different features in a within-subjects matched guise task: velar fronting of *-ing* (e.g., *walkin'* for *walking*), /aɪ/ monophthongization, shifted /e/, and breaking of /æ/. In her study, Albritten exposed listeners to 16 versions of the same sentence that varied by the combination of these four variables. In a within-subjects matched-guise task, listeners from various regions across the U.S. rated each of the 16 versions according to Southernness. The results showed that listeners (across regions) rated stimuli with shifted /e/ as most Southern, followed by stimuli with /æ/ breaking. Monophthongal /aɪ/ and velar fronting were rated as least Southern in the stimuli. However, shifted /e/ paired with any other feature (e.g., shifted /e/ + velar fronting) increased Southernness ratings compared to the stimulus version with shifted /e/ alone.

Exploring the relationship between regional categorization and acoustic features, Clopper and Pisoni (2004) examined which acoustic cues people use when categorizing a speaker as Southern. Listeners from Indiana were played one of three sentences produced by speakers across different regions and were asked to bin the speakers into one of six dialect regions. Listeners were still able to categorize Southern speakers despite the absence of vowel categories typically associated with Southern speech (e.g., mid front vowels) and only one token of /aɪ/. Results showed that the most reliable cues to Southernness were the presence of a voiced fricative in the lexical item *greasy* and fronting of /u/, suggesting that listeners are able to utilize other cues than those most often associated with Southern speech when labeling a speaker as Southern, and once again suggesting /u/ fronting to be associated with Southernness. However, this study did not examine a wide range of vowels, but instead limited its analyses to features the authors had previously posited as characteristic of Southern speech.

C. Present study

Prior work thus indicates that listeners are sensitive to several acoustic cues that may trigger judgments of Southernness. However, the perception studies thus far have limited investigations to a relatively small set of highly researched variables in Southern speech production, traditionally focusing on the mid front vowels (Fridland *et al.*, 2004; Albritten, 2011) and /aɪ/ monophthongization (Plichta and Preston, 2005; Albritten, 2011). Some studies have explored other features (Clopper and Pisoni, 2004; Fridland *et al.*, 2004; Albritten, 2011), but little work has explored the extent to which the larger vowel system is involved, and no work has explicitly tested whether individual speaker-level metrics correlate with perceptions of Southernness across a range of vowels. Further, the vocalic acoustic cues examined in a lot of perception research have often been limited to single static spectral measures; dynamic measures have been relatively underexplored in the perception of Southernness (aside from /aɪ/ monophthongization), even though they are often claimed to be defining features of Southern speech in production studies (e.g., Feagin, 1987).

Finally, much of the prior work has used manipulated speech, through the synthesis or resynthesis of aspects of the stimuli (e.g., Fridland *et al.*, 2004; Fridland, 2008; Plichta and Preston, 2005). Although synthesis affords precision in isolating individual cues, it does not model the task that listeners would use when assessing the Southernness of naturally produced speech. That is, the manipulation of individual characteristics of stimuli using synthesis provides useful insight about whether certain aspects of a stimulus can be used by listeners in the perception of Southernness, but cannot determine which cues are weighted more heavily if multiple potential cues are present.

In the present study, we address these limitations in the body of prior work on the perception of Southernness, examining a larger set of vowel categories from throughout the vowel space, and using unmodified, natural tokens as stimuli. We also test the same set of acoustic cues for each vowel category. We ask the following research questions: RQ1: Which vowel categories' tokens are perceived as most Southern? That is, which vowel categories' productions by our speakers, on average, receive the highest Southernness ratings? RQ2: What acoustic cues, at both the token- and speaker-level, contribute to Southernness ratings? That is, what acoustic aspects of speakers' vowel productions correlate with listeners' ratings?

In this study, non-Southern listeners were asked to rate individual words on a Southernness rating scale (from 1 to 9) in an experimental task. As introduced above, we use naturalistic speech tokens to allow for greater exploration of potential acoustic cues, including dynamic and speaker-level measures. While using natural tokens in some sense

https://doi.org/10.1121/10.0000550

restricts the conclusions we can draw, it has the advantage of not presupposing the acoustic cues that matter to listeners. If our study follows prior work, we expect that /e, ɛ, aɪ/ will be ranked as most Southern among listeners (RQ1), and that vowel dynamics and /e/-/ɛ/ ED will meaningfully correlate with Southernness ratings (RQ2).

III. METHODS

A. Participants

Participants for the perception study were undergraduate students recruited through the Human Subjects Pool at the University of Oregon and were given partial course credit for their participation. A total of 84 individuals participated in the study. Participants who did not complete all portions of the study ($N = 17$, most failing to complete the demographic survey following the experiment) or reported having a native language other than English ($N = 6$) were removed, leaving a total of 61 participants. Out of the remaining participants, 39 identified as female, 21 identified as male, and one identified as non-binary. The majority of participants identified as European American ($N = 40$). The remaining number of participants identified as African American ($N = 4$), Asian American ($N = 6$), Hispanic/Latino ($N = 6$), Native American ($N = 1$), Pacific Islander ($N = 1$), and unspecified "Other" ($N = 3$). The mean age of participants was 19, with ages ranging from 18 to 22 years of age. All participants were raised in the U.S. with most participants raised in the West or Northwest region ($N = 53$). The participants ranged in their exposure to Southern U.S. English, but most reported having little to none ($N = 38$) or some exposure ($N = 21$), with the remaining participants reporting having a lot of exposure ($N = 2$).

B. Stimuli

The stimuli were selected from a database of recorded European American speakers collected for a larger project examining speech across regions of the U.S. (Kendall and Fridland, 2012, 2017). All speakers were either recorded using a TASCAM digital recorder with a Shure WH30XLR head-mounted microphone or Marantz digital recorder and a Shure SM93 lavalier microphone in a quiet setting (see Kendall and Fridland, 2017, for full methodology). From the larger database, we selected 12 Southern speakers balanced by gender from Memphis, TN. Speakers were selected based on the data reported in Fridland and Kendall (2012) categorizing individuals as shifted or non-shifted in terms of participation in the SVS. We selected six speakers who were characterized as shifted Southerners and six speakers who were categorized as non-shifted Southerners. Figure 1 presents mean vowel plots for two speakers (created from tokens examined in the current paper; see Sec. III D below) as illustrative of a speaker who was categorized as shifted (Isaac815), and a speaker who was categorized as non-shifted (Brittany371).[2] We also selected five Western speakers (three male, two female), from Reno, NV, from Kendall and Fridland (2017). As our primary interest in the present paper is the perception of Southernness, we presented Westerners to listeners to act as a baseline group to help anchor Southernness ratings, but do not include them in most analyses.

Stimuli were monosyllabic words[3] selected from recorded word list elicitations, with phonological environments consisting of preceding and following voiced or voiceless stops or fricatives (e.g., *deaf, date, soap*). The same words were selected for each speaker.[4] Across words, the stimuli included 12 vowel categories to encompass most

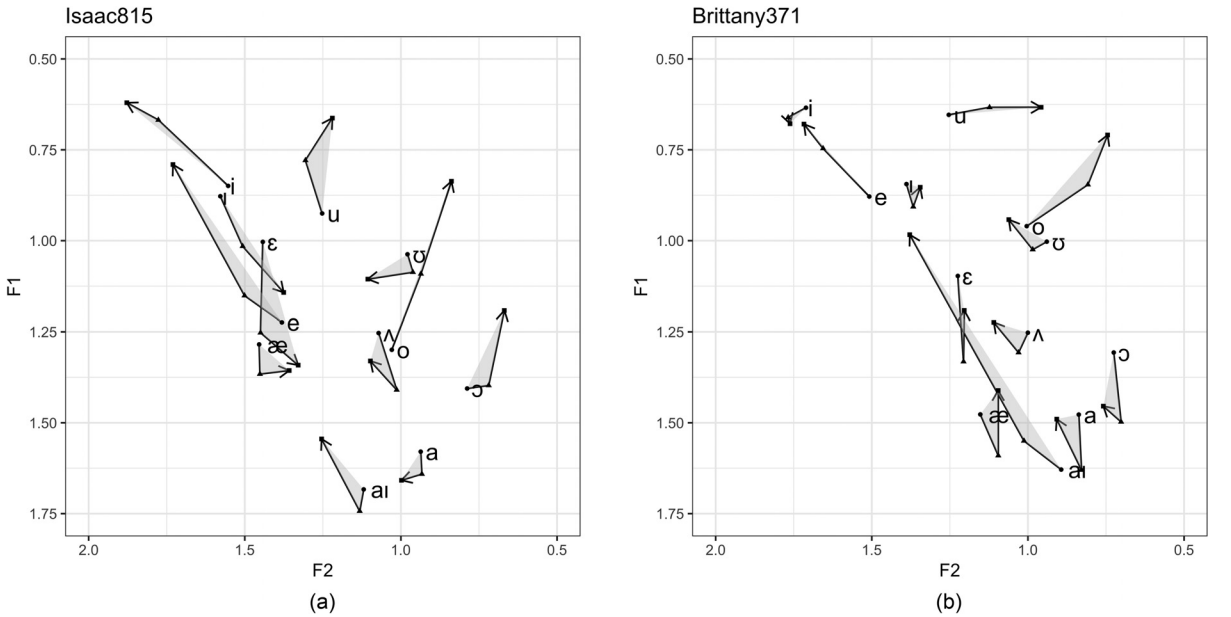

FIG. 1. Mean vowel plots for two Southern speakers representing a shifted speaker (Isaac815, on left) and a non-shifted speaker (Brittany371, on right). Vowel data presented in normalized units, arrows represent glide direction and offglide position. (See the supplemental material for vowel plots for all speakers.[2])

of the speakers' vowel spaces: /i, ɪ, e, ɛ, æ, ɑ, ɔ, ʌ, o, ʊ, u, aɪ/. Additionally, the words used ranged in their grammatical categories (e.g., function words, nouns, etc.), but were all produced in a word list elicitation and thus were fully stressed productions. We selected four words per vowel category with the exception of /ʊ/, for which there were only three available words per speaker (4 word tokens × 11 vowel categories × 17 speakers = 748 + 51 tokens of /ʊ/ = 799 tokens; see the Appendix for a full list of stimuli). All stimuli were amplitude normalized at 70 db sound pressure level (SPL) using a custom Praat script and began with 50 ms of silence.

C. Procedure

The task was conducted remotely (outside of the lab setting) using a custom online interface and took approximately 20 min to complete. Participants were requested to wear headphones. The instructions asked participants to respond to the question "How Southern does this speaker sound? 1 being not Southern at all to 9 being very Southern" for each word, with a numerical scale from 1 to 9 running from left to right. Prior to the task, participants received the following instructions to encourage them to use the entire scale: "Some speakers may not sound Southern at all; please feel free to use the left end of the scale to indicate that. The opposite is also true; some speakers may sound very Southern, and you may use the right end of the scale to indicate this."

No single participant heard all 12 vowel categories. Instead, each participant was randomly assigned to one of three conditions, each consisting of six vowel categories. In all conditions, the three vowel categories strongly implicated in the SVS, /e, ɛ, aɪ/ (as described in Sec. II A), were included because of their importance in prior work on the perception of Southernness (Plichta and Preston, 2005; Fridland, 2008; Albritten, 2011). The remaining three vowels in each condition varied (Table I). These conditions were not meant to represent predictions about the combination of vowel categories but rather served to minimize the length of the task. Each participant heard all 17 speakers and all word tokens (three or four depending on the vowel category) from each speaker for the six vowel categories in their respective conditions. The stimulus order was randomized by participant. A demographic survey was presented to participants following the experiment.

D. Acoustic measures

In order to extract acoustic measures of stimuli, all word tokens were transcribed and then force aligned using the Montreal Forced Aligner (McAuliffe *et al.*, 2017) and all vowel boundaries were subsequently checked by the authors and adjusted if necessary. Vowel measurements were made in Praat (Boersma and Weenink, 2017) using a Praat script from Farrington *et al.* (2018). Each token was visually inspected and formant-tracking estimation settings were adjusted as necessary prior to measuring formants. The subsequent output was inspected for accuracy by the authors and remeasured if necessary. Following Farrington *et al.* (2018), formant values were normalized using the Watt and Fabricius modified method (Fabricius *et al.*, 2009).

TABLE I. Experimental conditions.

	Participants heard:	
Condition	SVS vowels	Other vowels
1	/e/ /ɛ/ /aɪ/	/i/ /ɪ/ /ʌ/
2	/e/ /ɛ/ /aɪ/	/æ/ /ɑ/ /ɔ/
3	/e/ /ɛ/ /aɪ/	/u/ /ʊ/ /o/

1. Token-level vowel measures

Static spectral measurements of *F*1 and *F*2 were recorded at 20, 50, and 80% points in the vowel duration, corresponding to vowel onsets, nuclei, and glides, respectively. Further analysis examining the role of dynamics and directionality followed Farrington *et al.* (2018) and included VL, TL, VA, and spectral ROC. Previous work examining vowel dynamicity (e.g., Fox and Jacewicz, 2009; Farrington *et al.*, 2018) has primarily focused on front vowels. However, in order to examine all 12 vowel categories, we applied these same acoustic measures to every token included in our stimuli to ask whether static or dynamic measures of each vowel token are relevant for listeners' Southernness ratings of those tokens. In addition, we also measured vowel duration, normalized on a per-speaker basis by dividing each vowel token's duration by the speaker's mean duration of all stimulus vowels, following Fridland *et al.* (2014). Altogether, we considered 11 total measures for each individual vowel token: three measurement points for each of *F*1 and *F*2, four dynamic measures, and vowel duration.

2. Speaker-level measures

To test the claim that a speaker's ED between their /e/ and /ɛ/ classes serves as a proxy for their degree of Southernness more generally (Kendall and Fridland, 2012), we calculated /e/–/ɛ/ ED measures for each speaker, and included three other speaker-level measures to compare against /e/–/ɛ/ ED: mean /aɪ/ glide length, low back vowel Pillai, and mean /u/ *F*2. Table II gives summary information

TABLE II. Summary statistics for speaker-level measures for stimuli.

Speaker-level	Mean (μ)	Std. dev. (σ)	Min	Max
		South		
/e/–/ɛ/ ED	0.49	0.24	0.08	0.79
/aɪ/ glide	0.67	0.28	0.25	1.07
/u/ mean F2	1.25	0.07	1.12	1.38
/ɑ/–/ɔ/ Pillai	0.79	0.20	0.30	0.99
		West		
/e/–/ɛ/ ED	0.74	0.14	0.58	0.9
/aɪ/ glide	0.86	0.13	0.74	1.07
/u/ mean F2	1.04	0.21	0.85	1.43
/ɑ/–/ɔ/ Pillai	0.32	0.25	0.01	0.77

https://doi.org/10.1121/10.0000550

of the speaker-level measures for Southern and Western speakers.

The /e/–/ɛ/ ED measure captures the distance between the mean of the normalized first and second formants for the midpoints (50%) of /e/ and /ɛ/ (Kendall and Fridland, 2012). Normalizing the formant values converts raw $F1$ and $F2$ to equivalent scales, making them appropriate for use in ED (which is sensitive to differences in scale between dimensions; see Fridland *et al.*, 2014, p. 343). An individual /e/–/ɛ/ ED score closer to 0 indicates more proximal /e/–/ɛ/ classes, with larger values indicating more distant classes. Since ED is unsigned, speakers with a full reversal of /e/ and /ɛ/ (where /ɛ/ is tenser than /e/) will still have positive ED values. However, the speakers in our data for whom /e/ and /ɛ/ are completely reversed have quite proximal /e/ and /ɛ/ classes and their EDs are near 0. Speaker-level /aɪ/ monophthongization was calculated by averaging the VL of /aɪ/ across all four tokens used in the stimuli. The shorter the /aɪ/ glide length, the more monophthongal productions a speaker has on average. Low back vowel merger was calculated using Pillai scores, which capture the degree of overlap between vowel categories (see Nycz and Hall-Lew, 2013; Fridland *et al.*, 2014). The Pillai scores were calculated using the midpoint of the vowel, using the MANOVA statistical test in R. A score closer to 1 indicates speakers have more distinct /ɑ/–/ɔ/ classes, and a score closer to 0 indicates more overlapping classes. Mean /u/ $F2$ is simply the mean of each speaker's /u/ $F2$ values, with larger values indicating a more fronted average /u/ position for the speaker.

IV. RESULTS

Data included 24 411 Southernness ratings, coming from 61 participants across the three conditions (condition 1 = 406 items × 20 participants, C2 = 407 × 19, C3 = 389 × 22). Southernness ratings were z-score normalized on a by-participant basis (making the grand mean across participants 0 and standard deviation 1). Positive scores for a participant then represent higher Southernness ratings for that participant than the average, and negative scores represent lower Southernness ratings. Southernness ratings of stimuli for Western speakers were on average below zero ($\mu = -0.35$, $\sigma = 0.82$), and for Southern speakers were on average above zero ($\mu = 0.15$, $\sigma = 1.03$). A two-sample t-test indicates there was a significant difference in Southernness ratings between Southern and Western speakers ($t = 45.25$, $p < 0.01$), confirming that Western speakers indeed were perceived as sounding less Southern than Southerners. Again, since our research questions center around what listeners perceive as Southern, and our inclusion of Westerners was meant to serve as an anchor for listeners, the remainder of our analyses focuses on Southern speakers only, with Westerners removed.

We start by examining the distribution of Southernness ratings across stimuli by vowel categories for Southern speakers. Figure 2 shows the average Southernness rating across tokens of each vowel category, sorted by rating. The

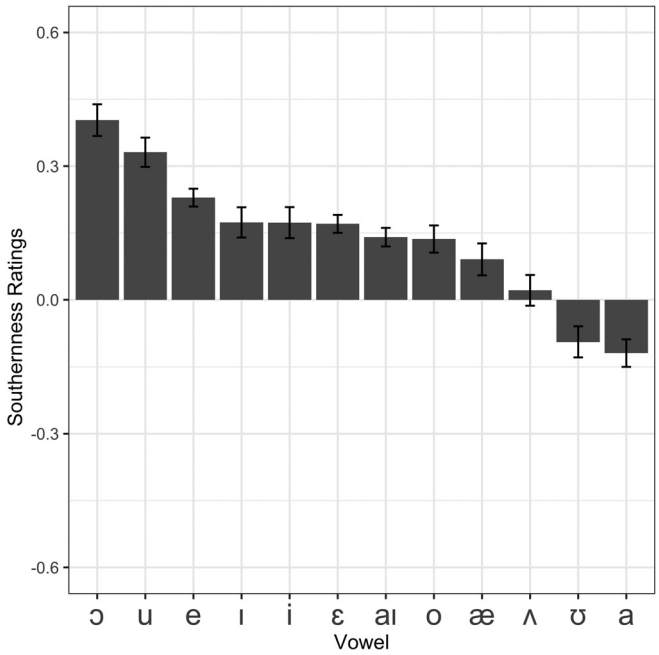

FIG. 2. Raw Southernness ratings of tokens produced by Southern speakers, averaged by vowel category, rank ordered from highest to lowest.

figure depicts a general trend for Southerners' tokens to be rated as sounding Southern across the vowel space, with the stimuli from most of the vowel categories showing positive values, indicating higher ratings. The Southernness ratings for words containing /e/, /ɛ/, and /aɪ/ are not at the top of the ranking but dispersed throughout the rank order. Instead, we see that productions of the back vowels /ɔ/ and /u/ are rated as most Southern sounding. In addition, tokens from many vowel categories are given roughly equivalent Southernness ratings, as noted by overlapping error bars. However, productions of three vowel categories, /ʌ/, /ʊ/, and /ɑ/, appear to be generally perceived as less Southern. Turning to how individual speakers pattern within the data, Fig. 3 depicts by-item Southernness ratings averaged for individual Southern speakers. The figure illustrates that, unsurprisingly, some speakers are rated across the vowel space as more Southern (e.g., Dustin312, Isaac815) and others less Southern (e.g., AngelaTN201, Abbey503). These ratings generally align with the labels Fridland and Kendall (2012) provided for these speakers as "shifted" and "non-shifted," although there is a range of Southernness ratings within the shifted speakers as shown in Fig. 3.

A. Research question 1: Looking across a wide range of vowels, what vowel categories' tokens are perceived as most Southern?

We entered this experiment with a hypothesis that the productions of a subset of vowels, namely, /e/, /ɛ/, and /aɪ/, are most likely to be rated as highly Southern sounding due to their associations with the SVS and attention in prior work on the South. As such, our design presented stimuli from all three of these vowel categories to all participants along with stimuli from three other vowel categories *not*

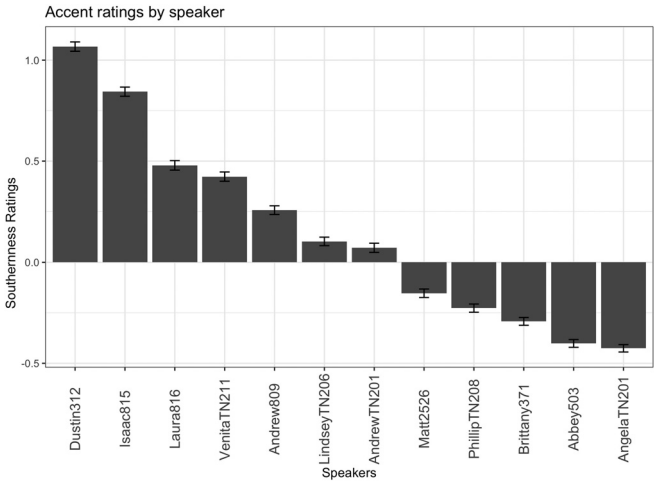

FIG. 3. By item Southernness ratings of tokens averaged by speaker (for all Southern speakers), rank ordered from highest to lowest.

implicated in the SVS, which were varied by condition. Although Fig. 2 suggests that productions of these three SVS-implicated vowels were in fact *not* perceived as most Southern in our data, we tested our *a priori* hypothesis using a linear mixed effect regression with /e/, /ɛ/, and /aɪ/ coded as one level, and all other vowels coded as another level in the predictor. We used effects coding, making the SVS vowel values of −0.5 and non-SVS vowel values of 0.5. This method facilitates ease of interpretation with coefficients being interpreted around the conditions mean. The model tested the main effects of SVS status, experimental condition, and the interaction of SVS status and condition. In addition to these fixed effects, the model included random intercepts for speaker, participant, vowel category (all 12 levels), and word. The model results, shown as type II Wald Tests in Table III, do not identify significant differences for SVS status (SVS vowels or not), for condition, or for the interaction. In other words, the findings confirm what is suggested in Fig. 2, that tokens with /e/, /ɛ/, and /aɪ/ as a group are not perceived differently than the other group of tokens with non-SVS-implicated vowel categories by condition.

While the model does not identify significance for the tested fixed effects, the random effect structure provides further insight into the data. Likelihood ratio tests using the ranova function from the lmerTest library in R (Kuznetsova *et al.*, 2017) confirmed the significance of all random effects (all at the level of $p < 0.001$). The random effects indicate that the most variability is across speakers ($\sigma = 0.53$),

TABLE III. χ^2 values from Type II Wald Tests for the linear model of Southernness ratings by SVS vowels, condition, and interaction of SVS vowel by condition.

Factor	χ^2	*p-value*
SVS Vowels	0.24	0.62
Condition	3.18	0.20
SVS Vowels * Condition	0.14	0.93

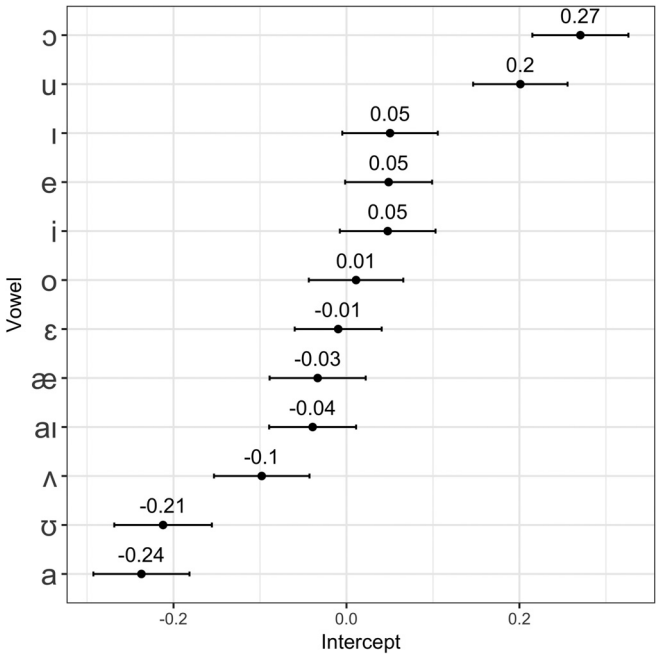

FIG. 4. Random intercepts with the standard error of vowel category from the linear mixed effects model examining the role of SVS vowels as a centered predictor on z-scored Southernness ratings.

followed by vowels ($\sigma = 0.17$), and the least variability is by participants ($\sigma = 0.05$).

Considering the individual random effect estimates for each vowel category highlights why SVS status and condition are not significant predictors of Southernness ratings and aligns with the raw Southernness ratings presented in Fig. 2. Figure 4 depicts the random effect estimates of all vowel categories. From this figure, it is clear that /e/, /ɛ/, and /aɪ/ productions have values close to the mean of Southernness ratings. Further, many of the productions of other vowel categories have near equivalent values to these SVS-implicated vowels, in line with the raw ratings in Fig. 2. The stimuli that show the highest Southernness ratings contain the vowels /ɔ/ ($u = 0.27$) and /u/ ($u = 0.20$). In addition, three vowel categories have negative estimates, /ʌ/ ($u = -0.1$), /ʊ/ ($u = -0.21$), and /ɑ/ ($u = -0.24$), confirming that listeners generally rate productions of stimuli containing these vowel categories as less Southern sounding.

B. Research question 2a: What token-level acoustic cues are predictive of listeners' Southernness ratings?

We now investigate which acoustic cues are predictive of listeners' Southernness ratings across vowel categories. We address RQ2 in two parts, first examining token-level measures (2a) and then turning to speaker-level measures (2b).

As described in Sec. III D 1, we consider the following token-level acoustic predictors: *F*1 and *F*2 at 20, 50, and 80% of the vowels' durations, vowel duration, and a series of dynamic measures including VL, TL, ROC, and VA. To examine the factors in a comparable way across all 12 vowel

https://doi.org/10.1121/10.0000550

categories, we apply the same simple modeling strategy to the data for each vowel category: we build a single model containing all of the acoustic predictors as main effects with a maximal random effect structure. Before proceeding, we wish to note that a full analysis of how all these acoustic measures affect the Southernness ratings for stimuli across 12 vowel categories is a complex undertaking and doing so in detail goes beyond the scope of this paper. Acoustic predictors are expected to be correlated with one another (e.g., $F1$ and $F2$ are known to be correlated in different parts of the vowel space, measurements at different time points within a vowel are not independent of one another, and dynamic measures like ROC are a function of other measures, namely, duration and TL) and this is a problem for most statistical modeling methods. More generally, providing full, detailed analyses for each of the 12 vowel categories would take more space than available in the present paper. Further, using the same baseline model for each vowel category has advantages for the consideration of the addition of speaker-level measures, which we turn to in Sec. IV C.

Thus, to begin examining which acoustic cues are predictive of Southernness, we limit our focus to assess how well the token-level acoustic predictors account for the variance in Southernness ratings on a per vowel category basis. Due to the large number of collinear predictors and the fact that we are interested in which features are relevant for each of the 12 vowel categories separately, we focus on two aspects of the models: which predictors are significant as fixed effects and the conditional R^2 (amount of variance explained; see Nakagawa and Schielzeth, 2013) for each model. At this stage, we do not attend to directions of effects or magnitudes of coefficients, or test for interactions between predictors. Assessing the effects more closely is left for future work, although after presenting the overall results across the entire dataset, we include closer analyses for two vowel categories, both to shed further light on our analyses and to chart what a more complete analysis might look like for each vowel category.

Models were built using the lmerTest package for R, with z-scored Southernness ratings as the dependent variable, and included random intercepts for the participant, speaker, and word. All fixed effects were centered and scaled and included as by-participant random slopes. That the token-level predictors were all included as fixed effects and as by-participant random slopes is important; much of each predictor's contribution to each model thus arises in the random slope terms, which motivates our broader interest in overall variance explained by each model (conditional R^2) rather than simply examining the importance of fixed effects.

Table IV displays χ^2 values based on type II Wald tests from Analysis of Deviance tables as well as the conditional R^2 values. R^2 values were calculated by testing the predicted values from each vowel category model against the z-scored Southernness ratings observed in the data, following Baayen (2008) and Nakagawa and Schielzeth (2013). Vowel categories range in the amount of variance accounted for by the models, from best predicted /æ/ ($R^2 = 0.51$) to worst predicted /ɪ/ ($R^2 = 0.28$). Many of the fixed effect terms arise as significant across vowel categories. However, for some vowels, like /e/, very few fixed effects emerge as significant although the model still accounts for a large portion of the variance (R^2 for /e/ $= 0.46$). It is possible that too much of the variance is being parceled out over the random effect structure, resulting in none of the fixed effects emerging as significant predictors for /e/. Further, it is possible that collinearity among the fixed effects makes it difficult for the model to determine which of the predictors is more important since much of the variance is shared across predictors in addition to the random effects. Overall, taking into account both fixed and random effects, the models suggest that token-level predictors account for a large portion of the variance in Southernness

TABLE IV. χ^2 values based on Type II Wald tests from Analysis of Deviance tables and conditional R^2 values for token-level predictors in mixed-effect linear models, for each vowel category. Vowel categories ordered by conditional R^2 (highest to lowest), and predictors ordered by number of vowels with effect. *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, . $p < 0.1$, — $p \geq 0.1$.

Vowel	Spec ROC	Traj Len	$F2$ 50%	Vec Len	$F2$ 20%	$F2$ 80%	Vec Ang	$F1$ 80%	F1 20%	Dur	$F1$ 50%	Model R^2
/æ/	8.55 **	7.65 **	15.47 ***	2.87 .	11.86 ***	—	—	3.33 .	—	—	—	0.51
/ɔ/	—	3.33 .	6.60 *	9.84 **	—	—	—	—	7.62 **	—	—	0.48
/aɪ/	12.03 ***	6.79 **	—	7.89 **	—	6.60 *	3.41 .	4.11 *	10.28 **	—	—	0.47
/ɛ/	—		5.80 *	—	—	—	—	—	—	—	6.29 *	0.46
/e/	—	3.51 .	—	—	—	—	—	—	—	—	—	0.46
/i/	11.03 ***	5.29 *	3.99 *	—	—	—	—	4.37 *	-	6.02 *	—	0.42
/u/	4.81 *	—	—	—	—	9.11 **	—	—	—	—	—	0.37
/ʌ/	—	—	—	—	3.99 *	—	3.58 .	—	—	—	—	0.32
/o/	5.30 *	—	13.36 ***	3.53 .	—	—	—	—	—	—	2.91 .	0.31
/ɑ/	6.98 **	9.43 **	—	—	4.74 *	9.10 **	—	—	—	—	—	0.30
/ʊ/	6.12 *	—	—	—	—	—	—	—	—	13.84 ***	—	0.29
/ɪ/	9.78 **	11.43 ***	5.09 *	—	3.26 .	—	4.41 *	—	—	—	—	0.28
No. of vwls w/ effect	8	7	6	4	4	3	3	3	2	2	2	
$\sum\chi^2$	64.60	47.43	50.31	24.13	23.85	24.81	11.40	11.81	17.90	19.86	9.20	

ratings across vowel categories, but still leave a lot of unexplained variance (a point we explore further below).

Having fit each of these models, we can now directly consider RQ2a, asking which acoustic cues predict Southernness ratings across vowel categories. We first focus on the overall trends in the fixed effects of the models presented in Table IV. Following this broad treatment, we examine two vowel categories more closely to refine our insights, but, as mentioned above, do not examine each vowel category in depth for sake of space.

Overall, the results reported in Table IV show that spectral ROC and TL, two factors capturing dynamic properties, are the most common significant predictors, obtaining significant χ^2 values for eight and seven of the 12 vowel categories, respectively. We also note an overall pattern that $F2$ across time points tends to be more important than $F1$ in listeners' Southernness judgments, with significant χ^2 values for 10 and 6 of the 12 vowel categories, respectively. While these factors appear most often as significant in the models, we also must remember that many of the acoustic measures are collinear (e.g., $F1$ at 25% and $F1$ at 50%, TL and ROC). Thus, we do not claim that non-significant items are unimportant to evaluations of Southernness, but rather are simply *less* important, and likely work in concert with other cues. For example, and surprising given the findings of prior production-based research (Clopper *et al.*, 2005; Fridland *et al.*, 2014), duration does not come out as significant in most models. This may be because duration is necessarily tied to other dynamic measures (e.g., spectral ROC) so the models are accounting for the role of duration through those other measures. Further, this may, in fact, be evidence that spectral ROC is more important than duration alone as a cue to Southern vowel variants, as the findings in Farrington *et al.* (2018) suggest. Spectral ROC may arise as important because it is a function of both duration and trajectory length, capturing overall more dynamic components than the single predictors of either duration or glide information alone.

We now turn to a closer examination of two of the vowel categories to provide further insights into how the measured acoustic cues influenced listeners' perceptions of Southernness and to map out what a deeper statistical analysis of these data might involve than is possible here for all 12 vowel categories. We focus on /ɔ/, the vowel perceived to be most Southern in our results, and /aɪ/, a vowel category for which we have strong *a priori* hypotheses from the literature. Here, we transition from examining the χ^2 and R^2 values in Table IV to examining individual model coefficients in order to examine the directionality of effects. While the models above are based on the expectation of collinearity and do not attempt to combat collinearity, for these closer considerations, we focus on reducing collinearity and finding the best model in order to examine only the most robust effects for each vowel category.

To develop best models for the Southernness ratings for these two vowel categories, we began by assessing the collinearity using the variance inflation factor (VIF) and a correlation matrix for each model presented above. We then conducted model comparisons by removing single predictors with the highest VIFs (VIF > 10; see O'Brien, 2007, for discussion) and high correlation coefficients and comparing them against the previous model (e.g., those presented in Table IV). Following this process, we chose the best fit model determined by model comparisons via likelihood ratio tests (following Baayen, 2008) with a significance of $p < 0.05$ from the reduction of collinear predictors. The final model from this process now acted as our baseline model from which we increased model complexity to include all theoretically motivated interactions. We then reduced the models again and tested model comparisons using likelihood ratio tests, and arrived at the final best fit models. We also tested for random effects significance using the rand function from the lmerTest package in R (Kuznetsova *et al.*, 2017) and removed any random effects that were not significant ($p > 0.05$).

In the case of /ɔ/, the best model accounted for 67% of the variance ($R^2 = 0.67$, notably larger than the 0.48 achieved in the main model above) with a number of fixed effects arising as significant: VL ($t = 2.28$, $p < 0.05$), $F1$ at the onset (20% point; $t = -2.87$, $p < 0.01$), $F2$ at onset (20% point; $t = 2.92$, $p < 0.01$), the interaction of $F1$ onset and $F2$ onset ($t = -4.23$, $p < 0.001$), $F1$ at midpoint (50% point; $t = -2.36$, $p < 0.05$), and $F2$ at offset (80% point; $t = -2.13$, $p < 0.05$). Table V displays the model coefficients for /ɔ/. The significant predictors largely align with the significant predictors in the main model (Table IV) even after the removal of collinear terms, and the results in these two models are not substantively different. In terms of the directions of effects, the model here (Table V) shows a clear pattern that lower and backer /ɔ/ tokens with longer glides received higher Southernness ratings. The two stimulus speakers who were displayed earlier, in Fig. 1, usefully illustrate these patterns. Figure 1 shows Isaac815 has strong upgliding diphthongal productions (long VL) compared to Brittany371 who has a shorter VL. As evident in Fig. 3, these two speakers ranged in the overall Southernness ratings they received. In line with these speakers' overall patterns of ratings, Isaac815's /ɔ/ tokens were given high

TABLE V. Linear mixed effects model estimates for /ɔ/. ** $p < 0.01$, *$p < 0.05$, .$p < 0.1$.

Factor	Estimate	Std. Error	t-value
Intercept	0.14	0.57	0.25
$F1$ (20%) **	−1.33	0.46	−2.87
$F2$ (20%) **	0.80	0.27	2.92
$F1$ (50%) *	−0.49	0.21	−2.36
$F2$ (50%)	−0.57	0.44	−1.30
$F1$ (80%)	0.06	0.11	−0.57
$F2$ (80%) *	−0.47	0.22	−2.13
Duration	−0.08	0.08	−0.96
VL *	0.31	0.14	2.28
TL	−0.24	0.15	−1.59
$F1$ (20%) × $F2$ (20%) ***	−1.44	0.34	−4.23

https://doi.org/10.1121/10.0000550

TABLE VI. Model coefficients for /aɪ/. *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, . $p < 0.1$.

Factor	Estimate	Std. Error	t-value
Intercept	−0.42	0.30	−1.42
$F1$ (20%) ***	0.65	0.18	3.66
$F2$ (20%) *	−0.72	0.31	−2.33
$F1$ (50%)	−0.02	0.08	−0.24
$F2$ (50%)	0.14	0.12	1.13
$F2$ (80%) ***	−0.60	0.12	−4.98
Duration	0.05	0.07	0.72
TL	0.11	0.09	1.16
Spectral ROC ***	−0.27	0.08	−3.32
VA .	0.06	0.04	1.70
$F1$ (20%) × $F2$ (20%) ***	0.70	0.21	3.40
$F2$ (80%) × TL **	0.11	0.04	2.58

ratings ($\mu = 1.20$, $\sigma = 1.00$), and Brittany371's /ɔ/ received low ratings ($\mu = -0.16$, $\sigma = 0.77$).

For /aɪ/, the best model accounts for 68% of the variance ($R^2 = 0.68$, again notably higher than the 0.47 in the main model above), with a number of significant predictors: $F1$ at the onset (20% point; $t = 3.66$, $p < 0.001$) and $F2$ at the onset (20% point; $t = -2.33$, $p < 0.05$), $F2$ at the offset (80% point; $t = -4.96$, $p < 0.001$), Spectral ROC ($t = -3.31$, $p < 0.001$), the interaction of $F1$ and $F2$ at the onset ($t = 3.40$, $p < 0.001$) and the interaction of $F2$ offset and TL ($t = 2.58$, $p < 0.01$). Table VI displays the model coefficients. Once again, the model here (Table VI) aligns with the model above in most ways; however, VL drops out of the model with the addition of interactions and removal of collinear terms. Additionally, the model here identifies that the role of TL is conditioned by $F2$ at offset. Overall, the results from this model indicate that glide offsets closer to the nucleus position of /aɪ/ and vowels with slower ROC received higher Southernness ratings. These results are in line with what we would expect from more monophthongal productions of /aɪ/ in Southern speech. As with the model for /ɔ/, these results are congruent with the view provided by the vowel plots of Fig. 1, where Isaac815, whose /aɪ/ class has high Southernness ratings ($\mu = 1.05$, $\sigma = 0.90$) shows clear monophthongal productions with fronter and lower onset position, while, in comparison, Brittany371, who has low ratings for /aɪ/ ($\mu = 0.07$, $\sigma = 0.90$), maintains a more diphthongal /aɪ/ with backer and higher onset.

In closing, a primary outcome of the closer analyses of /ɔ/ and /aɪ/ is that while we see improved model fits and more nuance to the results (e.g., interactions), the best models largely reinforce the patterns that emerge from the coarser analysis across all categories (Table IV). The conditional R^2 values reported in Table IV for models across all vowel categories can be taken as reasonable floors, which can be improved upon through closer, more vowel-specific analyses.

C. Research question 2b: What speaker-level acoustic cues are predictive of listeners' Southernness ratings?

Token-level acoustic measures captured some of the variances in the Southernness ratings, and now we ask whether the addition of speaker-level measures improves these models. As described in Sec. III D 2, we take particular interest in /e/–/ɛ/ ED, but also test /aɪ/ glide length (mean VL), /u/ mean $F2$ position, and /ɑ/–/ɔ/ Pillai. As with the token-level predictors discussed above, all factors were scaled and centered and each factor was added to the models simultaneously as both a fixed effect and a by-participant random slope.

Table VII expands on the results presented above in Table IV by adding each of the four speaker-level measures individually to the models reported in Table IV. Here again, the table presents the χ^2 values for statistically significant results from type II Wald tests and the overall variance explained by the model, displayed both as conditional R^2 values and in terms of percentage change from the token-level models in Table IV.

TABLE VII. χ^2 values (and percent change over token-level model) for speaker-level predictors added to mixed-effect linear models from Table IV, for each vowel category. Vowel categories ordered by overall model fit (highest to lowest) from Table IV. *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, . $p < 0.1$, — $p \geq 0.1$.

Vowel	Tbl IV Model R^2	/e/–/ɛ/ ED χ^2	/e/–/ɛ/ ED R^2	/aɪ/ mean VL χ^2	/aɪ/ mean VL R^2	/ɑ/–/ɔ/ Pillai χ^2	/ɑ/–/ɔ/ Pillai R^2	/u/ mean $F2$ χ^2	/u/ mean $F2$ R^2
/æ/	0.511	30.06 ***	0.531 (3.8%)	4.22 *	0.533 (4.3%)	—	0.53 (3.6%)	—	0.521 (1.9%)
/ɔ/	0.482	13.23 ***	0.503 (4.3%)	5.44 *	0.486 (0.9%)	—	0.48 (0%)	6.43 *	0.488 (1.2%)
/aɪ/	0.467	22.33 ***	0.498 (6.8%)	12.15 ***	0.481 (3.0%)	—	0.47 (1.6%)	—	0.467 (0%)
/ɛ/	0.462	50.12 ***	0.475 (2.8%)	15.90 ***	0.478 (3.4%)	3.65 .	0.46 (0%)	—	0.472 (2.1%)
/e/	0.462	28.89 ***	0.483 (4.6%)	6.88 **	0.472 (2.2%)	—	0.46 (0.3%)	—	0.475 (2.9%)
/i/	0.422	6.71 **	0.431 (2.2%)	—	0.428 (1.4%)	—	0.43 (1.5%)	—	0.423 (0.2%)
/u/	0.368	—	0.390 (6.2%)	—	0.381 (3.7%)	—	0.39 (6.7%)	—	0.376 (2.3%)
/ʌ/	0.321	8.24 **	0.342 (6.4%)	3.63 .	0.346 (7.8%)	—	0.33 (1.8%)	—	0.328 (2.2%)
/o/	0.309	3.29 .	0.328 (6.1%)	—	0.322 (4.2%)	—	0.31 (0.6%)	—	0.320 (3.6%)
/ɑ/	0.300	3.64 .	0.322 (7.3%)	2.76 .	0.316 (5.3%)	—	0.30 (-0.2%)	—	0.307 (2.4%)
/ʊ/	0.290	5.64 *	0.350 (20.8%)	13.94 ***	0.318 (9.5%)	—	0.29 (0.2%)	—	0.299 (3.0%)
/ɪ/	0.279	5.99 *	0.307 (10.1%)	5.79 *	0.297 (6.7%)	4.77 *	0.29 (5.4%)	—	0.280 (0.4%)
No. of vowels with effect		11		9		2		1	
Ave. R^2 percent improvement		6.78%		4.37%		1.79%		1.85%	

These results demonstrate that /e/–/ɛ/ ED is the most influential of the four speaker-level measures tested. It is significantly predictive of Southernness ratings for nine of the 12 vowel categories (all but /ɑ/, /o/, and /u/) and as marginally predictive for an additional two (/ɑ/ and /o/). Only the model for /u/ does not yield a χ^2 value for /e/–/ɛ/ distance with $p < 0.1$, but even here the model including /e/–/ɛ/ ED obtains an improved model fit over the model without it ($R^2 = 0.39$ as opposed to 0.37). Across vowel categories, the model fits improved an average of 6.77% over the token-level only models with the addition of /e/–/ɛ/ ED.

/aɪ/ mean VL also arises as significantly predictive of Southernness ratings for seven of the vowel categories and as marginally predictive for another two. Its influence is thus less important than that of /e/–/ɛ/ ED, but it still improves models an average of 4.37% over token-level only models. The other two speaker-level factors are much less influential and usefully demonstrate that it is not simply the inclusion of *any* regionally variable speaker-level information that helps to improve the statistical models.

Our initial interest in speaker-level measures was motivated by prior work using the speaker-level measure /e/–/ɛ/ ED as a way to rank order speakers according to their participation in the SVS, as a "proxy" for Southernness (Kendall and Fridland, 2012; Farrington *et al.*, 2018). Given our listener rating results, we can now test how well /e/–/ɛ/ ED aligns with Southernness as rated by listeners. That is, we test the extent to which the rank ordering of speakers according to their average perceived Southernness (i.e., the ordering plotted in Fig. 3) correlates with the rank orderings of speakers according to each of our four speaker-level measures, using Spearman's Rho rank order correlations. The rank ordering of average speaker Southernness ratings and /e/–/ɛ/ ED are highly correlated ($\rho = -0.94$, $p < 0.001$). The other speaker-level measures correlate less well, with /aɪ/ VL ($\rho = -0.63$, $p < 0.05$) followed by /u/ $F2$ ($\rho = 0.60$, $p < 0.05$) also obtaining significant correlations, and /ɑ/–/ɔ/ Pillai ($\rho = 0.43$, $p = 0.17$) not. The very high correlation for /e/–/ɛ/ ED provides some confirmation that the measure may indeed be a reasonable proxy for a speaker's perceived Southernness.

In sum, speaker-level measures, and particularly /e/–/ɛ/ ED, do capture something more general about a speaker's vowel space that listeners use in making Southernness ratings, and this extends beyond the vowel categories they directly measure. However, these results do not explain the reason *why* /e/–/ɛ/ ED correlates with Southernness ratings across the vowel space. This question awaits future experimentation, though we offer some speculation in the discussion.

V. DISCUSSION

Our analyses have sought to explore which vowel categories' phonetic variants, as produced by the speakers in our experiment, were perceived as most Southern by our (Western) listeners, and which acoustic cues give rise to this perception. Based on prior literature about the production and perception of Southern vowels (Thomas, 2001; Plichta and Preston, 2005; Labov *et al.*, 2006; Albritten, 2011; Fridland, 2012), we hypothesized that the productions of vowels most discussed and associated with the SVS in prior research (/e/, /ɛ/, and /aɪ/) would be rated as most Southern. However, in our data, words containing these vowels were not rated as most Southern, with that distinction going instead to words containing /u/ and /ɔ/. We believe this is important to highlight, as it suggests that (non-Southern) listeners are attending to more of the vowel space than is typically given attention in perception studies. We are not the first to observe this, as Clopper and Pisoni (2004) illustrated that in the absence of SVS-implicated vowels, participants used other available cues in the stimuli to classify speakers as Southern. Taken together, our findings call for more studies to consider other aspects of Southern speech when examining what cues give rise to judgments of Southernness.

Our findings that productions of back vowels were rated as sounding on average the most Southern to our non-Southern listeners aligns with some prior findings and diverges from others. Our results are congruent with those of Torbert (2004) and Clopper and Pisoni (2004), using naturalistic stimuli, who found that /u/ fronting was used by listeners in the identification of Southern speech. On the other hand, Fridland (2008), using resynthesized speech varying only in $F1$ and $F2$, found that Western listeners were not as sensitive as Southern listeners to shifts in back vowels when selecting Southern tokens, in particular /u/. Taking our results together with prior results, then, it appears that listeners use multiple cues, beyond just the $F1/F2$ dimensions, to identify Southern variants of back vowels (e.g., as further discussed below, our results and Koops, 2010 suggest that back vowel dynamics may be important cues).

Further, we do not suggest that productions of /u/ and /ɔ/ would *always* be the vowels rated as most Southern for all speakers and by all listeners. In fact, our finding that /e/, /ɛ/, and /aɪ/ were not rated as among the most Southern vowels when averaging across all of our Southern speakers is likely related to the degree to which our Southern speakers actually participated in the SVS. Speakers, of course, were variable for these features across their productions of the stimuli, and our selection of speakers intentionally included Southerners who had been identified in prior work (Kendall and Fridland, 2012) as not fully engaged in the SVS, so those acoustic cues were not as available to listeners for those speakers. Further, /u/ may have been rated as more Southern because more of our speakers in the stimuli may participate in /u/ fronting more uniformly, providing stronger acoustic cues available to listeners. However, we also point to our results that suggest that Southern /u/ is distinguished from other regional /u/ fronting by glide dynamics, in line with Koops (2010), and it may be those cues that listeners use to rate /u/ as more Southern in our data.

Relatedly, the speakers rated as most Southern in our study (e.g., Isaac815 in Fig. 1) do receive higher Southernness ratings for /e/, /ɛ/, and /aɪ/ compared to the larger set of speakers. In addition, these most Southern rated speakers indeed

https://doi.org/10.1121/10.0000550

show more monophthongal /aɪ/ and more proximal mid front vowels acoustically, in line with a traditional SVS system. However, even for these speakers, whose /e/, /ɛ/, and /aɪ/ are SVS-like in production and rated highly accordingly, participants still rated /ɔ/ tokens as the most Southern. Thus, we emphasize that the explanation for unexpected vowels like /ɔ/ having high Southern ratings cannot *only* be that our speakers do not happen to participate strongly in expected Southern shifted patterns (e.g., less monophthongal /aɪ/ than might be expected). Rather, we suggest that other vowels beyond /e/, /ɛ/, and /aɪ/ are indeed also important for non-Southern listeners in the evaluation of Southernness. These findings point towards future work to explore the broader vowel space more closely in the perceptions of Southern speech.

Our results also highlight the importance of a wider range of acoustic cues in Southernness ratings than simply spectral position in terms of $F1$ and $F2$. In particular, vowel dynamics (specifically spectral rate of change and trajectory length) play a large role in accounting for Southernness ratings across vowel categories at the token level. Folk linguistic perceptions often comment on Southern vowels being long and "drawn-out" (Preston, 1993), which some researchers attribute to vowel dynamic properties of lengthening and glide movement (e.g., Feagin, 1987). Our findings reinforce the notion that listeners utilize dynamic cues when assessing Southernness, and this extends to other vowel categories beyond those typically associated with dynamic changes (like diphthongs and /æ/). The results of our study are in line with work suggesting the importance of vowel dynamics in Southern speech (e.g., Fox and Jacewicz, 2009; Farrington *et al.*, 2018) and highlight the need for more work in production and perception on vowel dynamics. Surprisingly, as we note above, duration was not a significant predictor, contrary to expectations. We hypothesize that this may be the result of ROC being a better metric capturing overall duration and glide dynamics in one measure, as discussed above. In addition, we also note one other possible explanation is that duration may be less of a strong perceptual cue for participants listening to individual read words, randomized across talkers, than it would be in other more naturalistic settings.

Beyond token-level factors, we found that speaker-level measures, and /e/–/ɛ/ ED in particular, are predictive of Southernness ratings across much of the vowel space, perceptually confirming suggestions by Kendall and Fridland (2012). That /e/–/ɛ/ ED did predict ratings of other vowel categories suggests that it is capturing more than just the distance between the means of a speaker's /e/ and /ɛ/ class, and may represent something more holistic about a speaker. This result may be due to /e/–/ɛ/ ED being correlated with many of the other token-level acoustic cues (e.g., vowel dynamics) and thus it just happens to capture many relevant cues in one measure. Alternatively, it could also be that /e/–/ɛ/ ED correlates with token-level cues that listeners attend to beyond those we measured in this study. A third possibility is that listeners attend to speakers holistically as they make judgments of Southernness across individual tokens, accumulating information about the individual speakers over the course of the experiment, and calling on their experiences of a speaker's other vowels more holistically when rating tokens. Although we attempted to mitigate this possibility by using multiple speakers and randomizing stimuli, it is unclear whether this completely removed listeners' ability to anchor to individual speakers in some way. The extent to which listeners bring in prior experience when rating individual tokens is a question for future work.

Finally, just as we expect our results to differ depending upon the specific speakers chosen, we also expect differences across listener populations. For example, Southern listeners in Fridland (2008) were more sensitive to SVS-like shifts in the front vowels of Southerners than were Western listeners. We welcome future research into how listener dialect and experience may affect which vowels are rated highly, and which acoustic cues are attended to in evaluations of Southernness.

The present study adds to a larger body of work exploring how non-Southerners evaluate Southern speakers (e.g., Clopper and Pisoni, 2004; Fridland, 2008). It expands the scope of vowels investigated beyond those assessed in prior work and examines which of a wide range of acoustic features are predictive of how everyday (i.e., non-linguist) listeners rate Southernness. Our results indicate that listeners are sensitive to a variety of vocalic features in making Southernness judgments, including productions of various vowel categories across the vowel space and their dynamic cues.

ACKNOWLEDGMENTS

The stimuli for this project came from work funded by National Science Foundation Grant Nos. BCS-0518264, BCS-1123460, and BCS-1122950. We thank Sam Elliott and Charlie Farrington at the University of Oregon for support with various aspects of this research. We also thank Lisa Redford, Melissa Baese-Berk, and several anonymous reviewers for helpful comments on the paper. Finally, we thank Valerie Fridland for her continual support and helpful discussions.

APPENDIX: STIMULI

/i/	deep, deed, seat, bead
/ɪ/	bit, did, kid, tick
/e/	date, take, gate, bait
/ɛ/	bet, deaf, bed, debt
/æ/	dad, sad, half, pad
/ʌ/	dutch, tuck, does, but
/u/	dude, boot, soup, booed
/o/	soap, boat, poke, doze
/ʊ/	took, foot, hood
/aɪ/	tide, type, tie, dive
/ɑ/	cop, dock, pod, pot
/ɔ/	bought, dog, hawk, paw

[1]We recognize there are many acoustic properties of speech that may contribute to the perception of Southernness, such as prosodic features (Clopper and Smiljanic, 2011; Reed, 2016) and consonantal features (Clopper and Pisoni, 2004; Jacewicz *et al.*, 2010) but we do not focus on them here.

[2]See supplementary material at https://doi.org/10.1121/10.0000550 for vowel plots for each speaker used in the stimuli creation and for the aggregate region and group (shifted/non-shifted) vowel plots.

[3]Prior to selecting stimuli, all words present in the word list elicitations were rated for how much they were associated with the Southern U.S. Undergraduate students ($N = 17$) in an introductory linguistics course at the University of Oregon were presented with the orthographic words and were asked to rate each word on a scale of 1–6 for how much it made them think of the Southern U.S. (1 = not at all, 6 = very strongly). The words selected for use as stimuli were not rated as highly Southern, with the lowest score a 1.2 and the highest score a 3.8 ($M = 2.3$, $SD = 0.69$).

[4]There were a few speakers who did not produce every word we included, thus the total number of tokens across conditions was unequal. Overall, we were missing one token of *booed*, *tied*, and *bead* across three speakers. Additionally, the original word lists did not include /ɔɪ/ and /aʊ/ tokens, and thus those vowels could not be included in this study. Future work could benefit from exploring these diphthongs as an addition to understanding percepts of Southernness across the vowel space.

Albritten, R. (**2011**). "Sounding Southern: Phonetic features and dialect perceptions," Ph.D. thesis, Georgetown University, Washington, D.C.

Baayen, H. (**2008**). *Analyzing Linguistic Data: A Practical Introduction to Statistics using R*, 1st ed. (Cambridge University Press, New York).

Boersma, P., and Weenink, D. (**2017**). "Praat: Doing phonetics by computer, version 6.0.28 [computer program]," http://www.praat.org/ (Last viewed 1/10/2020).

Clopper, C. G., and Pisoni, D. B. (**2004**). "Some acoustic cues for the perceptual categorization of American English regional dialects," J. Phon. **32**, 111–140.

Clopper, C. G., Pisoni, D. B., and de Jong, K. J. (**2005**). "Acoustic characteristics of the vowel systems of six regional varieties of American English," J. Acoust. Soc. Am. **118**, 1661–1676.

Clopper, C. G., and Smiljanic, R. (**2011**). "Effects of gender and regional dialect on prosodic patterns in American English," J. Phon. **39**, 237–245.

Fabricius, A. H., Watt, D., and Johnson, D. E. (**2009**). "A comparison of three speaker-intrinsic vowel formant frequency normalization algorithms for sociophonetics," Lang. Var. Change **21**, 413–435.

Farrington, C., Kendall, T., and Fridland, V. (**2018**). "Vowel dynamics in the Southern vowel shift," Am. Speech **93**, 186–222.

Feagin, C. (**1986**). "More evidence for major vowel change in the South," in *Diversity and Diachrony*, edited by D. Sankoff (John Benjamins, Amsterdam), pp. 83–95.

Feagin, C. (**1987**). "A closer look at the Southern drawl: Variation taken to extremes," in *Proceedings of the Fifteenth Annual Conference on New Ways of Analyzing Variation*, October 17–19, Stanford, CA, pp. 137–150.

Feagin, C. (**2000**). "Sound change in the South," Am. Speech **75**, 342–344.

Fox, R. A., and Jacewicz, E. (**2009**). "Cross-dialectal variation in formant dynamics of American English vowels," J. Acoust. Soc. Am. **126**, 2603–2618.

Fridland, V. (**2000**). "The Southern Shift in Memphis, Tennessee," Lang. Var. Change **11**, 267–285.

Fridland, V. (**2003**). " 'Tie, tied and tight': The expansion of /aɪ/ monophthongization in African-American and European-American speech in Memphis, Tennessee," J. Socioling. **7**, 279–298.

Fridland, V. (**2008**). "Regional differences in perceiving vowel tokens on Southerness, education, and pleasantness ratings," Lang. Var. Change **20**, 67–83.

Fridland, V. (**2012**). "Rebel vowels: How vowel shift patterns are reshaping speech in the modern south," Lang. Ling. Compass **6**, 183–192.

Fridland, V., and Bartlett, K. (**2006**). "The social and linguistic conditioning of back vowel fronting across ethnic groups in Memphis, Tennessee," Eng. Lang. Linguist. **10**, 1–22.

Fridland, V., Bartlett, K., and Kreuz, R. (**2004**). "Do you hear what I hear? Experimental measurement of the perceptual salience of acoustically manipulated vowel variants by Southern speakers in Memphis, TN," Lang. Var. Change **16**, 1–16.

Fridland, V., and Kendall, T. (**2012**). "Exploring the relationship between production and perception in the mid front vowels of US English," Lingua **122**, 779–793.

Fridland, V., and Kendall, T. (**2015**). "Within-region diversity in the Southern vowel shift: Production and perception," in *Proceedings of the 18th International Congress on Phonetic Science*, August 10–14, Glasgow, UK.

Fridland, V., Kendall, T., and Farrington, C. (**2014**). "Durational and spectral differences in American English vowels: Dialect variation within and across regions," J. Acoust. Soc. Am. **136**, 341–349.

Irons, T. L. (**2007**). "On the Southern shift in Appalachian English," Univ. Pennsylvania Work. Pap. Linguist. **13**, 121–134.

Jacewicz, E., Fox, R. A., and Lyle, S. (**2010**). "Variation in stop consonant voicing in two regional varieties of American English," J. Int. Phon. Ass. **39**, 313–334.

Jacewicz, E., Fox, R. A., and Salmons, J. (**2007**). "Vowel duration in three American English dialects," Am. Speech **82**, 367–385.

Kendall, T., and Fridland, V. (**2012**). "Variation in perception and production of mid front vowels in the US Southern vowel shift," J. Phon. **40**, 289–306.

Kendall, T., and Fridland, V. (**2017**). "Regional relationships among the low vowels of U.S. English: Evidence from production and perception," Lang. Var. Change **29**, 245–271.

Koops, C. (**2010**). "/u/-Fronting is not monolithic: Two types of fronted /u/ in Houston Anglos," Univ. Pennsylvania Work. Pap. Linguist. **16**, 111–122.

Koops, C. (**2014**). "Iconization and the timing of Southern vowels: A case study of / æ / iconization and the timing of Southern vowels: A case study of / æ /," Univ. Pennsylvania Work. Pap. Linguist. **20**, 81–90.

Kuznetsova, A., Brockhoff, P. B., and Christensen, R. H. B. (**2017**). "lmerTest package: Tests in linear mixed effects models," J. Stat. Softw. **82**(13), 1–26.

Labov, W. (**1991**). "The three dialects of English," in *New Ways of Analyzing Variation in English*, edited by P. Eckert (Academic Press, San Diego, CA), pp. 1–44.

Labov, W., Ash, S., and Boberg, C. (**2006**). *The Atlas of North American English: Phonetics, Phonology and Sound Change* (De Gruyter, Berlin).

Majors, T. (**2005**). "Low back vowel merger in Missouri speech: Acoustic description and explanation," Am. Speech **80**, 165–179.

McAuliffe, M., Socolof, M., Mihuc, S., Wagner, M., and Sonderegger, M. (**2017**). "Montreal Forced Aligner: Trainable text-speech alignment using Kaldi," in *Proceedings of Interspeech*, August 20–24, Stockholm, Sweden.

Nakagawa, S., and Schielzeth, H. (**2013**). "A general and simple method for obtaining R^2 from generalized linear mixed-effects models," Methods Ecol. Evol. **4**(2), 133–142.

Nearey, T. M., and Assmann, P. F. (**1986**). "Modeling the role of inherent spectral change in vowel identification," J. Acoust. Soc. Am. **80**, 1297–1308.

Nycz, J., and Hall-Lew, L. (**2013**). "Best practices in measuring vowel merger," Proc. Mtgs. Acoust. **20**, 060008.

O'Brien, R. M. (**2007**). "A caution regarding rules of thumb for variance inflation factors," Qual. Quant. **41**(5), 673–690.

Plichta, B., and Preston, D. (**2005**). "The /ay/s have it: The perception of /ay/ as a North–South stereotype in United States English," Acta Linguist. Hafniensia **37**, 107–130.

Preston, D. R. (**1993**). "Folk dialectology," in *American Dialect Research: Celebrating the 100th Anniversary of the American Dialect Society, 1889–1989* (John Benjamins Publishing, Philadelphia), pp. 333–378.

Preston, D. R. (**2015**). "The South: Still different," in *New Perspectives on Language Variety in the South: Historical and Contemporary Approaches* (University of Alabama Press, Tuscaloosa, AL), pp. 311–326.

Reed, P. E. (**2016**). "Sounding Appalachian: /ai/ monophthongization, rising pitch accents, and rootedness," Ph.D. thesis, University of South Carolina, Columbia, South Carolina.

Sledd, J. H. (**1966**). "Breaking, umlaut, and the Southern drawl," Language (Baltimore) **42**, 18–41.

Thomas, E. R. (**2001**). *An Acoustic Analysis of Vowel Variation in New World English* (Duke University Press Books, Durham, NC), pp. 15–58.

Thomas, E. R. (**2003**). "Secrets revealed by Southern vowel shifting," Am. Speech **78**, 150–170.

Torbert, B. (**2004**). "Southern vowels and the construction of social salience," Ph.D. thesis, Duke University, Durham, NC.

Wetzell, W. B. (**2000**). "Rhythm, dialects and the Southern drawl," Master's thesis, North Carolina State University, Raleigh, North Carolina.

Wolfram, W., Hazen, K., and Schilling-Estes, N. (**1999**). "The shifting status of outer banks /aɪ/," Publ. Am. Dialect Soc. **81**, 104–119.