# Perception of local and non-local vowels by adults and children in the South

Ewa Jacewicz, and Robert Allen Fox

Citation: The Journal of the Acoustical Society of America **147**, 627 (2020); doi: 10.1121/10.0000542
View online: https://doi.org/10.1121/10.0000542
View Table of Contents: https://asa.scitation.org/toc/jas/147/1
Published by the Acoustical Society of America

JASA ARTICLE

# Perception of local and non-local vowels by adults and children in the South

Ewa Jacewicz[a] and Robert Allen Fox
*Department of Speech and Hearing Science, The Ohio State University, 1070 Carmack Road, Columbus, Ohio 43210, USA*

**ABSTRACT:**
This study assessed the ability of Southern listeners to accommodate extensive talker variability in identifying vowels in their local Appalachian community in the context of sound change. Building on prior work, the current experiment targeted a subset of spectrally overlapping vowels in local and two non-local varieties to establish whether adult and child listeners will demonstrate the local dialect advantage. Listeners responded to isolated target words, which minimized the interaction of multiple linguistic and dialect-specific features. For most vowel categories, the local dialect advantage was not demonstrated. However, adult listeners showed sensitivity to generational changes, indicating their familiarity with the local norms. A differential response pattern in children suggests that children perceived the vowels through the lens of their own experience with vowel production, representing a sound change in the community. Compared with the adults, children also relied more on stress cues, with increased confusions when the vowels were unstressed. The study provides evidence that identification accuracy is dependent upon the robustness of cues in individual vowel categories—whether local or non-local—and suggests that the bottom-up processes underlying phonetic vowel categorization in isolated monosyllables can interact with the top-down processing of dialect- and talker-specific information. © *2020 Acoustical Society of America*.
https://doi.org/10.1121/10.0000542
(Received 23 April 2019; revised 23 July 2019; accepted 5 September 2019; published online 31 January 2020)
[Editor: Wendy Herd] Pages: 627–642

## I. INTRODUCTION

Phonetic research on regional variation in the Southern United States has predominantly focused on vowel production across sub-regional varieties of the Southern American English (SAE). Acoustic studies have provided great detail about the variation and sound change within the Southern vowel system, including sub-regional manifestations of the Southern Vowel Shift (SVS; e.g., Dodsworth and Kohn, 2012; Farrington *et al.*, 2018; Fox and Jacewicz, 2009; Fridland, 1999, 2001, 2012; Labov *et al.*, 2006), fronting of back vowels (e.g., Baranowski, 2008; Fridland and Bartlett, 2006; Thomas, 1989), and overall dispersion of vowels in the acoustic space (e.g., Clopper *et al.*, 2005; Jacewicz *et al.*, 2011a; Labov *et al.*, 2006). Comparatively little is known about the perception of Southern vowels by Southern speakers and, consequently, the extent to which the sub-regional background influences vowel identification remains unclear.

Evidence exists that local Southern listeners (from Birmingham, Alabama) had an advantage over the non-local listeners in identifying the vowels in their own local variety (Labov and Ash, 1997; Labov, 2010). However, in a different line of research that used a synthesized series (continua) of vowel pairs rather than naturally produced words, the local dialect advantage was less clear. In particular, it was shown that Southerners displaying local differences in vowel production identified the common synthetic vowel series similarly, irrespective of whether they grew up in Tennessee, Virginia, or North Carolina (NC; Fridland and Kendall, 2015; Kendall and Fridland, 2012). It was proposed that Southerners may, in fact, maintain more "unified" perceptual category norms across the South than what could be expected on the basis of their productions. Although their perceptual categorization patterns differed from those of listeners from Northern and Westerns regions in the United States (US), the differences among the three Southern regions were minimized, leading to the proposal that, in terms of vowel perception by Southerners, the "Southern dialect" may be more homogeneous when compared to the sub-regional variation in their vowel productions.

The aim of the current paper is to explore the "local" aspect of vowel perception and assess Southern listeners' ability to identify vowels in their own sub-regional variety in the context of sound change undergoing in their community. This focus is different from previous investigations (e.g., Labov, 2010) that reported the local dialect advantage when comparing responses of the same listeners to local versus non-local vowels produced by only a few representative talkers. In particular, we address the issue of perceptual accommodation of significant variation in vowel production across several generations of local speakers. At present, little is known about how individuals in a given speech community perceive vowels of other members of that community when there have been changes in the local vowel system related to chain shifts (e.g., SVS) or other positional variations that influence the formation of perceptual category structure. Presumably, listeners who are also users of the local variety will have accumulated their

[a]Electronic mail: jacewicz.1@osu.edu

relevant dialect experience through the lens of their own vowel production (Sumner and Samuel, 2009). Furthermore, their local identity will be influential in interpreting the variation in the community, and their long-term perceptual exposure to cross-generational vowel change will enhance their local advantage. The extent to which older and younger listeners are able to accommodate this variation has not yet been scrutinized with sufficient experimental detail, and the aim of this research is to provide new evidence to bear on this issue.

Over the past decade, we have carried out fieldwork in a small Appalachian speech community in Western NC, focusing on vowel production and sound change across several generations of the local participants. Spread over three adjacent counties, Jackson, Swain, and Haywood, this area is a part of the Inland South on the geographic map of American English (Labov *et al.*, 2006). It is the case that Appalachian dialects in the Inland South diverge from other varieties of SAE, and they can also differ substantially one from another so that the term "local" conveys not only a particular sub-regional language variety but also a rich cultural heritage of the community, dating back to Scotch-Irish immigrant settlers in the early 19th century. Situated in the Appalachian Mountains, these small and often isolated communities were naturally guarded against mobility and outside influences, which promoted distinctive living traditions, local culture, and a strong sense of belonging to the area (Clark and Hayward, 2013). Today, the region is open to the influences of mainstream American socio-cultural norms and young generations are gradually shifting toward modern lifestyles and urbanization, which is also reflected in their gradual adoption of selected mainstream American English features.

While the production of vowels and sound change in this Appalachian community was our primary interest, we also conducted perception studies with the locals to establish the correspondence between the cross-generational production patterns and listeners' ability to accommodate this variation in making vowel and dialect identification decisions. Our experimental focus was to verify listeners' success in maintaining phonetic constancy across multiple talkers whose pronunciation patterns represent not only a typical indexical variability related to talker sex and age (e.g., Harrington *et al.*, 2007; Johnson, 2006), but also dialect-specific sound-change in progress as the sex- and age-related *phonetic* changes are associated with both significant positional vowel rotations in the acoustic space and vowel-inherent time varying spectral changes (VISC; Nearey and Assmann, 1986; Morrison and Assmann, 2013); the latter variations create a substantial spectral overlap of neighboring vowels (Fox and Jacewicz, 2017).

According to recent models of speech processing, resolving the phonetic ambiguity of the variable signal involves an active perceptual process (Heald *et al.*, 2016) rather than matching the incoming sound to a prototypical representation (e.g., Syrdal and Gopal, 1986). This active process likely involves a type of intrinsic talker normalization that integrates talker-specific information in the perception of not only ambiguous phonemic contrasts (Theodore and Miller, 2010) but even when the sounds are acoustically unambiguous (Choi *et al.*, 2018). While these and other formal accounts of the interaction between the phonetic (linguistic) and talker-specific (social) cue distributions have offered new insights (e.g., Kleinschmidt and Jaeger, 2015; Kleinschmidt, 2019; Pierrehumbert, 2016), the perceptual evidence for how listeners perform under uncertainty created by talker variability is still inconclusive. Experimental results vary across studies due to differences in the design and number of talkers in perceptual tasks. For example, the processing cost involved in the perception of variable vowel productions from four talkers is significantly greater than that from a single talker (Choi *et al.*, 2018), but it is unclear whether listeners are similarly (i.e., actively) engaging in intrinsic talker normalization when additional processing cost increases to 76 different talkers (Peterson and Barney, 1952) or even to 139 different talkers (Hillenbrand *et al.*, 1995).

Possibly, when responding to *many* talkers, listeners' knowledge of a dialect-specific (and talker-specific) distribution of acoustic cues in vowels let them probabilistically infer (i.e., recover) the most likely explanation for the vowel sound they hear (Kleinschmidt and Jaeger, 2015; Kleinschmidt *et al.*, 2018). However, it is unclear which cue dimensions contribute consistent information gain for individual phonetic categories and how listeners resolve acoustic ambiguities. In particular, while the computational model may utilize vowel midpoints as cues to variability in vowel formants distributions (Kleinschmidt, 2019), it ignores the dynamic vowel structure as a robust dialect-specific acoustic cue (e.g., Jacewicz *et al.*, 2011a). On the other hand, the perceptually tracking of individual formants in naturally produced vowels is at least problematic as listeners may use information in the detailed shape of the whole spectrum rather than in the distribution of formant frequencies (Bladon, 1982; Hillenbrand and Houde, 2003; Ito *et al.*, 2001; Zahorian and Jagharghi, 1993).

Because of the complexity of the perceptual system and many unresolved issues, such as those with delineating the most relevant acoustic-phonetic cues in naturally produced vowels, talker variability remains a primary challenge for vowel perception research, and the perceptual mechanisms involved in normalization processes across vowel productions of multiple talkers are still not well understood and highly debated. In the absence of a comprehensive framework, our work with the local Appalachian listeners sought to better understand the *local* aspect of their vowel categorization, that is, their experience-based perceptual flexibility in accommodating talker variability, assuming that this flexibility reflects their knowledge of variable phonetic forms of vowels in their community. The current study builds directly on several findings from our previous research.

The first relevant finding was that the Appalachian adult listeners evidenced local dialect advantage (Jacewicz and Fox, 2012) when tested using a modified version of the classic vowel identification paradigm in Peterson and Barney (1952)

and Hillenbrand *et al.* (1995). In our study, the 12 vowel categories in hVd-context were produced by 120 talkers (3 generations of men and women, including children), 60 who spoke the local dialect and 60 who were from Wisconsin (WI), representing a Northern variety of American English. The identification accuracy for the local vowels was significantly higher (82.1%) than for the WI vowels (75.8%), although these rates were lower than in Hillenbrand *et al.* (95.8%), who did not include dialect as an additional variable. Overall, the Appalachian adults were able to perceptually separate the vowels reasonably well, suggesting that, when listening to variable productions of many talkers, a common talker-listener background facilitates perception of vowels in the listener's own variety.

The same experiment was also conducted with 9–12-yr-old local Appalachian children (Jacewicz and Fox, 2014) to better understand their abilities to cope with an extensive talker variability. On average, the children's accuracy was statistically lower when compared with the adults (the rates were 8.5% lower), but the overall pattern of identifications and confusions was quite consistent. Like the adults, the children showed local dialect advantage, identifying the Appalachian vowels with significantly greater accuracy (77.1%) than WI vowels (65.3%). We interpreted these lower rates as reflecting children's inability to accommodate extensive talker variability as efficiently as the adults, most likely due to their comparatively lesser experience with dialectal and generational variation.

The second study produced two different sets of results, however. The Appalachian listeners were presented with the same hVd-material, but their task was to identify talker dialect (local NC or WI) rather than the vowel. In these experiments, the listeners were not given the opportunity to access their dialect-specific knowledge on the basis of the wide range of socioindexical information available abundantly in longer passages of speech such as in sentences (e.g., Hay *et al.*, 2006). Rather, we aimed to establish whether they were able to access their dialect knowledge when only minimal social information was available in the phonetic content. The results showed that adult listeners were quite successful in identifying talker dialect, but that the ability to distinguish between the two dialects was dependent upon talker generation; it was greatest for the oldest talkers (66–91 years old) and diminished (although it was still above chance) for children (Fox and Jacewicz, 2011). The responses also varied with vowel category, suggesting that the changing pronunciation patterns reflecting diachronic sound change over time influenced listeners' decisions about talker dialect. These findings demonstrate the local dialect advantage in response to the oldest talkers but no comparable advantage in response to children. However, a different pattern was obtained for child listeners (Jacewicz and Fox, 2017). The children's responses were at chance, indicating their inability to identify dialects on the basis of vowel production in isolated words. This chance performance cannot be due to inexperience because the corresponding children from WI, who also participated in this experiment, were able to distinguish between the two dialects relatively well, although not yet reaching the levels of the adults. Thus, the Appalachian children showed the local dialect advantage in one task (Jacewicz and Fox, 2014) but, when explicitly asked about the talker dialect in another task (Jacewicz and Fox, 2017), they were unable to indicate which of the dialects was local and which was not. The most compelling explanation for this discrepancy is that this new generation in the community was growing up in a socio-culturally changing environment, which supplied them with a wider range of vowel pronunciation possibilities to the point that they lost the sense of what constitutes a local and a "non-local" dialect.

Finally, the third study (Fox and Jacewicz, 2019) focused narrowly on the local Appalachian dialect (excluding the WI variety) and on a subset of "challenging" Southern vowels (/ɪ, e, ɛ, æ, ai/). The study sought to establish the viability of perceptual separation of neighboring vowels whose inherent time varying spectral variations (VISC) created an excessive spectral overlap. The controlled sources of this overlap included variable stress patterns in vowels, talker variability (40 local talkers), sound change (represented by the productions of adults and children), and talker sex. The experiment was conducted only with adult listeners. As in our previous studies, the listeners were only presented with isolated monosyllabic words.

In spite of the extensive variation, identification rates (IDRs) for /ɪ, e, ɛ/ were relatively high (about 80%), suggesting that perceptual separation of high and mid vowels, even in the face of mutual confusions, was not particularly challenging. However, the rates declined significantly for /æ/ (64%) and, markedly, for /ai/ (41%). Given that /æ/ was mostly confused with /ɛ/, and /ɛ/ was mostly confused with /ɪ/, it appears that, even in their own local dialect, separating spectrally overlapping mid and low vowels was far more challenging for the listeners than separating the overlapping high and mid. Importantly, talker age group was not a significant predictor for front vowels, indicating that generational differences in positions and formant dynamics did not play a dominant role in listeners' identification decisions. On the contrary, talker generation was a decisive factor for /ai/. The monophthongal productions of /ai/ in adults mostly resulted in confusion, but the increased formant movement approximating a diphthong in children resulted in higher accuracy, and more so in girls than in boys. The overall accuracy for the vowel set was higher for female talkers than for male talkers, which has also been found elsewhere (e.g., Ferguson, 2004; Hillenbrand *et al.*, 1995; Jacewicz and Fox, 2012). Also consistent with earlier studies was the finding that stressed vowels produced higher accuracy than unstressed vowels (e.g., Cutler, 2005; Fry, 1955).

The current study builds directly on the latter experiment. Given the limited and varied success in separating spectrally overlapping vowels across multiple talkers, we ask whether the obtained pattern represents the actual abilities of adult listeners to categorize vowels in their own local variety, or whether their responses are task-dependent and

the dialect-specific cues to vowel identity do not provide consistent information. Addressing this question, we included two non-local dialects in the task. The current study seeks to establish whether the addition of the non-local dialects will affect the overall identification pattern for the problematic /ɪ, e, ɛ, æ, ai/ set, and whether the Appalachian listeners will consistently demonstrate the local dialect advantage when hearing productions of the common vowel categories in local and non-local dialects. Four different outcomes seem possible. The first outcome is that the addition of the two dialects will not influence the accuracy for the local dialect, suggesting that listeners' responses represent their true abilities in identifying vowels in this particular set when no additional socioindexical information is available in speech. The second outcome is that the experience with the local dialect will be beneficial in resolving acoustic ambiguities and, thus, the local dialect advantage will be manifested in comparatively higher IDRs for the local vowels. The third outcome is that the non-local varieties will increase listeners' uncertainty, and the IDRs for the local Appalachian vowels will decrease (relative to those reported in Fox and Jacewicz, 2019). The final possibility is that the non-local productions will influence each individual vowel category differently, disturbing the overall pattern for the set reported in Fox and Jacewicz (2019).

While the current study builds on the previous results with adult listeners, our research interests are also in perceptual abilities of preadolescent children, and thus the experiment was conducted with both adult and child listeners. From the developmental perspective, children can identify clear exemplars of segmental categories well below the age of three years (e.g., Shvachkin, 1973; Werker, 2018), and the shift to an adult-like weighting of cue information seems to occur between the ages of five and seven years (Nittrouer and Miller, 1997). However, although the perceptual development may approach the adult levels when children enter the elementary school, the exact course of acquisition of sociolinguistic competence by preadolescents is largely unknown as it depends on a number of interacting social factors, such as community structure, different speaking styles across caregivers, or linguistic norms in a community as a whole, among others (e.g., Chevrot *et al.*, 2000; Foulkes *et al.*, 2005; Smith *et al.*, 2007). The acquisition of the intra- and inter-talker variation within a speech community may extend well into adolescence (McCullough *et al.*, 2019), and inconsistency in preadolescent children's categorization decisions (Hazan and Barrett, 2000) may also contribute to greater variability in their perceptual performance.

With respect to the local Appalachian children, we conjectured that they will have formed a different perceptual vowel structure than the adults because their experience with the local variation was shaped by the socio-cultural change in the community. As shown in another study, their own vowel productions differed from those of the older talkers as they represented the new generation in the community that was most influenced by the mainstream American English (Jacewicz and Fox, 2019). We thus expected the children to perform worse than the adults, not only due to their comparatively inconsistent categorization decisions across multiple talkers and lesser experience with dialect variation (Jacewicz and Fox, 2014), but also as a function of their differential perceptual organization of the vowel system, which can be inferred from systematic confusion patterns.

## II. METHODS

### A. Listeners

Two groups of listeners participated in the experiment. The first group consisted of 15 adults, 12 females and 3 males, ranging in age from 44 to 65 years old [$M = 55.7$, $SD = 6.1$]. All participants were "rooted" in Western NC, being born, raised, and spending most of their lives (other than for occasional trips) in either Jackson, Swain, or Haywood counties. They spoke the local Appalachian dialect as verified by the experimental team. The participants had never participated in any linguistic research study before. They did not participate in any of our previous perception experiments. Most of them were employed in the area in a variety of professions, and some were already retired. None reported hearing problems.

The second group consisted of 15 children, 7 boys and 8 girls, in the age range 9–12 years ($M = 10.8$, $SD = 1.2$). They were born into the local families spanning several generations, and came from the same areas as the adult listeners. The children attended local public schools and, based on the analysis of the background questionnaires completed by one of their parents, came from households of comparable socioeconomic status. None of them had participated in our previous perception studies, and all were new to linguistic research experimentation. None reported hearing problems.

### B. Stimuli

The vowels /ɪ, e, ɛ, æ, ai/ were of interest to the study because their proximity in the acoustic space and their distinctly Southern pattern of formant dynamics created a desirable condition for testing their perceptual separability. In this Appalachian variety of SAE, /ai/ is produced as the monophthong /a:/ by older talkers, the /æ/ is raised in the acoustic space, and the remaining three vowels participate in the SVS so that the positions of the lax /ɪ, ɛ/ tend to be "reversed" relative to the tense /i,e/. Furthermore, the vowels /ɪ, ɛ, æ/ exhibit an extensive formant movement known as Southern breaking (whose strength varies with talker generation). In the Southern breaking, the front vowel is partitioned (or "breaks") into two or three audible parts, and the parts are "connected" by a glide. For listeners not familiar with the Southern breaking, the minimally contrastive words such as *did* [dɪjəd], *dead* [dɛjəd], and *dad* [dæjəd] may be perceptually challenging because of the unexpected combination of the positional target(s) and the glide. All these features contribute to substantial spectral overlaps among the vowels in the /ɪ, e, ɛ, æ, ai/ set, and the nature of these overlaps is also influenced by cross-generational sound change in the community.

https://doi.org/10.1121/10.0000542

To create experimental stimuli for perceptual testing, the vowels /ɪ, e, ɛ, æ, ai/ were contained in tokens *bids*, *bades*, *beds*, *bads*, and *bides*, respectively. The tokens were produced in sentences (e.g., "Ted thinks the fall BIDS are low," where "BIDS" is in the focal position and produced with the main sentence stress; or "Ted thinks the fall bids are LOW," where "bids" occurs in a non-focal unstressed position and is produced as spectrally unreduced). The sentences were read and recorded by several generations of talkers for a large cross-dialectal study (see Jacewicz *et al.*, 2011b), and the current stimulus material for perceptual testing was selected from that corpus. The tokens *bids*, *bades*, *beds*, *bads*, and *bides* were edited out of the sentences and presented to the listeners as isolated items. Presenting these target words in isolation rather than in sentences reduced the amount of socioindexical information about each talker that is typically available in longer passages of speech. This approach was chosen to minimize the interaction of multiple linguistic and indexical features that are known to influence the social perception of language (e.g., Montgomery and Moore, 2018). Given that the focus of the study was on vowel perception across multiple talkers, our interests were in testing listeners' perceptual organization of the local vowel system and their abilities to recognize the local vowel features when only minimal information about the talkers' dialects was available in speech. This bottom-up approach to perceptual categorization assumes that listeners' acquired sensitivity to variable phonetic forms of vowels in their community will influence their decisions under uncertainty, and this phonetic knowledge will be advantageous in recognizing the local cues to vowel identity in multiple-talker productions.

To increase the spectral variation in both positions of the vowels in the acoustic space and their formant dynamics, the stimulus set included stressed and unstressed productions of the five target words. Because the experiment specifically targeted spectrally overlapping vowels, which could increase the demands of the task for untrained listeners and children, we opted for reducing the predictable physiologically based variation due to talker sex, and all talkers were male. The selected talkers represented two age groups (or generations), adults (age range 50–65 years old) and children (9–11years old). This selection was done for each dialect; the local Appalachian and the two non-local Northern varieties spoken in southeastern WI and central Ohio (OH). The WI variety was as in our prior perception experiments, and the OH variety, representing General American English, has not been used before. Together, the target items were produced by 60 talkers, 20 for each dialect (NC, WI, and OH), of which 10 were adult males and 10 were boys. Each talker contributed 10 unique exemplars of all 5 tokens, 5 stressed and 5 unstressed, for a total of 600 tokens for use in the experiment. The NC set was previously used as the "male" part of the stimulus material in Fox and Jacewicz (2019) (which also included female productions), but the WI and OH sets have not been used before.

Prior to presentation, all tokens were equalized for mean intensity and analyzed acoustically to ensure that the productions represented the variation typical of each regional variety and generation. This was done by comparing the group means for formant measurements, that is, the means for all available talkers for each generation and the means for the talkers selected for perceptual testing. The frequencies of the first two formants in vowels were sampled at five temporal locations (20%,35%,50%,65%,80%) following the procedure in Fox and Jacewicz (2009). Average dynamic formant patterns for tokens used in the experiment are shown separately for men and boys in Figs. 1 and 2, respectively. Given the narrow age ranges within each age group, formant values were not normalized for the sake of visual comparisons. As can be seen, there is a substantial variation in vowel positions and their formant dynamics across the groups and dialects, creating an extensive spectral overlap of neighboring vowels. Of interest to the current study is to establish whether NC listeners can separate these overlapping vowels as categories intended by the talker, and whether regional background and talker generation provide relevant cues to allow the separation.

### C. Procedure

The 600 stimuli were presented in 3 blocks of 200 tokens each, yielding 18 000 data points from all 30 listeners. The testing was done in one session lasting approximately 1 h, with short breaks between blocks. In each block, the tokens were randomized and included instances of all five vowel categories in stressed and unstressed words from all three dialects and two talker generations. We note that the stimuli were not blocked by talker dialect as it was reported in Labov (2010). Rather, our goal was to expose the listeners to a wide range of possible variations, representing a mix of local and non-local vowels. Each listener was tested individually in a quiet room. The experiment was administered by the same female research assistant, an experienced speech-language pathologist affiliated with Western Carolina University and the local schools. In terms of her linguistic background, she grew up in Northern Illinois, lived briefly in WI, and moved to the Appalachian area in Western NC 22 years prior to conducting the current study.

The stimuli were delivered over Sennheiser HD600 headphones (Sennheiser electronic GmbH & Co. KG, Wedemark, Germany) at a comfortable listening level determined for each individual listener. This was done during a 20-item practice run that was administered prior to the experiment to familiarize listeners with the procedure. The material in the practice was different than that in the experiment. On the basis of the practice, the experimenter determined whether the participant was able to do the task and match the orthographic form with the sound. If needed, the practice run was repeated one time. No response feedback was provided during the practice. No specific information about the dialect of the talkers was given to the listeners prior to the experiment, and they were only told that they would hear many talkers. The listeners were verbally instructed (and shown by the experimenter) that after hearing each word, they were to select and click with the mouse on one of the seven response boxes: beads, bids, bades, beds, bads, bides,

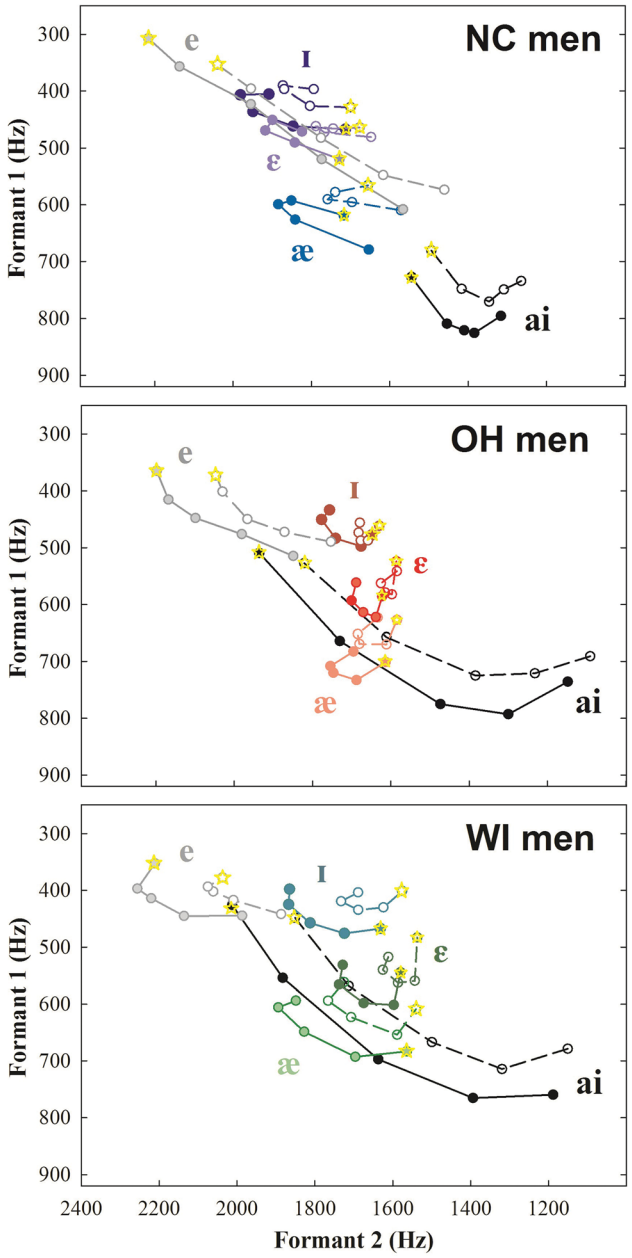


FIG. 1. (Color online) Average dynamic formant patterns for vowels used in the experimental tokens produced by men. $F1$ and $F2$ frequencies are sampled at five temporal points (20%,35%,50%,65%,80%). Filled circles, stressed vowels; open circles, unstressed. Directionality of formant movement is indicated by a yellow star superimposed on the 80% point.

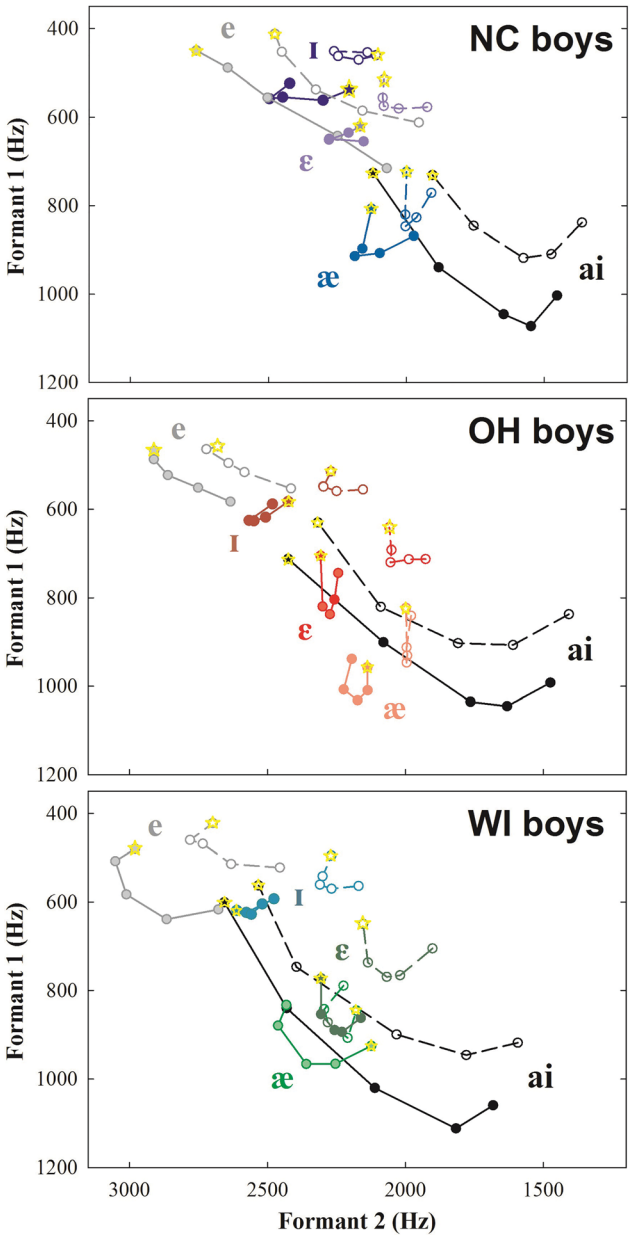


FIG. 2. (Color online) Average dynamic formant patterns for vowels used in the experimental tokens produced by boys. $F1$ and $F2$ frequencies are sampled at five temporal points (20%,35%,50%,65%,80%). Filled circles, stressed vowels; open circles, unstressed. Directionality of formant movement is indicated by a yellow star superimposed on the 80% point.

and buds. They could repeat the word one more time if uncertain, but after that they were to guess what the word was. The experiment was self-paced. A custom program in MATLAB (The MathWorks, Natick, MA) was written to control the experiment and collect the responses. The order of the blocks was counterbalanced across listeners.

## III. RESULTS

### A. Identification results for adult listeners

Average IDRs (in % correct) for the adult listeners broken down by vowel and talker dialect (local NC, and non-local OH and WI) are shown in Table I. The current IDRs for the local NC dialect could be compared directly with those in Fox and Jacewicz (2019) because the exact stimulus subset produced by NC males was used in both studies. We thus included those data in Table I for a comparison. Admittedly, the identification patterns for NC vowels were fairly consistent, even though IDRs for /e, ɛ, æ/ were slightly lower in the current experiment. However, except for the vowel /e/ in *bades*, IDRs for NC do not show evidence of local dialect advantage, and the rates for /ɛ/ and /ai/ are consistently lower when compared with both non-local dialects. The accuracy data for NC along with

https://doi.org/10.1121/10.0000542

TABLE I. Average accuracy (in %) for adult listeners.

| | NC adult listeners | | | | |
|---|---|---|---|---|---|
| Vowel intended by speaker | Local dialect (NC) | Non-local dialect (OH) | Non-local dialect (WI) | Total | Fox and Jacewicz (2019) Local dialect (NC) |
| /ɪ/ | 88.0 | 84.2 | 86.8 | 86.3 | 88.0 |
| /e/ | 74.2 | 68.0 | 55.2 | 65.8 | 77.7 |
| /ɛ/ | 67.5 | 92.3 | 82.5 | 80.8 | 74.8 |
| /æ/ | 50.2 | 61.3 | 42.2 | 51.2 | 52.5 |
| /ai/ | 50.5 | 80.8 | 77.8 | 69.7 | 46.3 |
| Total | 66.1 | 77.3 | 68.9 | 70.8 | 67.9 |

predominant and other confusions are displayed separately in Fig. 3. We observe that, as accuracy gradually declined with each vowel descending in height, the primary confusions yielded a pattern consistent with that in Fox and Jacewicz (2019) so that high and mid vowels were mutually confused with one another, /æ/ was mostly confused with /ɛ/, and the monophthongal /ai/ was most often confused with /ʌ/ (followed by /æ/). Together, the current results indicate that the inclusion of two non-local dialects in the stimulus set was neither advantageous nor detrimental to the identification of vowels in the local dialect. Rather, the pattern was consistent with earlier findings. In terms of the results for the non-local dialects, the IDRs varied with vowel category and dialect. For both OH and WI, the rates were generally high for /ɪ, ɛ, ai/ but identification of /e, æ/ was more challenging. We note that the rates for WI /e, æ/ were particularly low.

Listeners' percent correct responses were analyzed with linear mixed-effects models in IBM SPSS Statistics v. 25 (2017). No arcsine transformation was applied prior to the analyses in light of the current recommendations for analyzing proportional data in a forced-choice task (e.g., Jaeger, 2008; Warton and Hui, 2011). The model was constructed with vowel, stress, talker age group, talker dialect, and interactions as fixed effects, adding one predictor at a time to a model that only included the intercept. Listener was a random effect. Log-likelihood comparisons were used to determine the significance of the fixed effects. The best-fitting model included all main effects and three interactions between vowel and age group, vowel and dialect, and vowel and age group and dialect.

As expected, stress had a significant effect on IDRs [$\chi^2(1) = 46.32$, $p < 0.001$] in that stressed vowels were identified with greater accuracy ($M = 75.5\%$) than unstressed vowels ($M = 66.0\%$). Talker age group also had a significant effect on IDRs [$\chi^2(1) = 10.61$, $p < 0.001$], and accuracy was higher for adults ($M = 72.9\%$) than for children ($M = 68.6\%$). The main effect of vowel was significant [$\chi^2(4) = 258.01$, $p < 0.001$]. Subsequent Bonferroni-adjusted pairwise comparisons showed that only the differences between /ɪ/ and /ɛ/ ($p = 0.133$), and /e/ and /ai/ ($p = 0.827$) were not significant, and all the other pairs were significant ($p < 0.001$). Dialect was also a significant predictor [$\chi^2(2) = 49.18$, $p < 0.001$]: IDRs for OH ($M = 77.3\%$) were significantly higher than either for NC ($M = 66.1\%$) or WI ($M = 68.9\%$; $p < 0.001$ for both), and the latter two did not differ from one another ($p = 0.255$).

Of main interest to the study was a significant interaction between vowel and dialect [$\chi^2(8) = 147.79$, $p < 0.001$]. *Post hoc* analyses revealed that dialect had differential

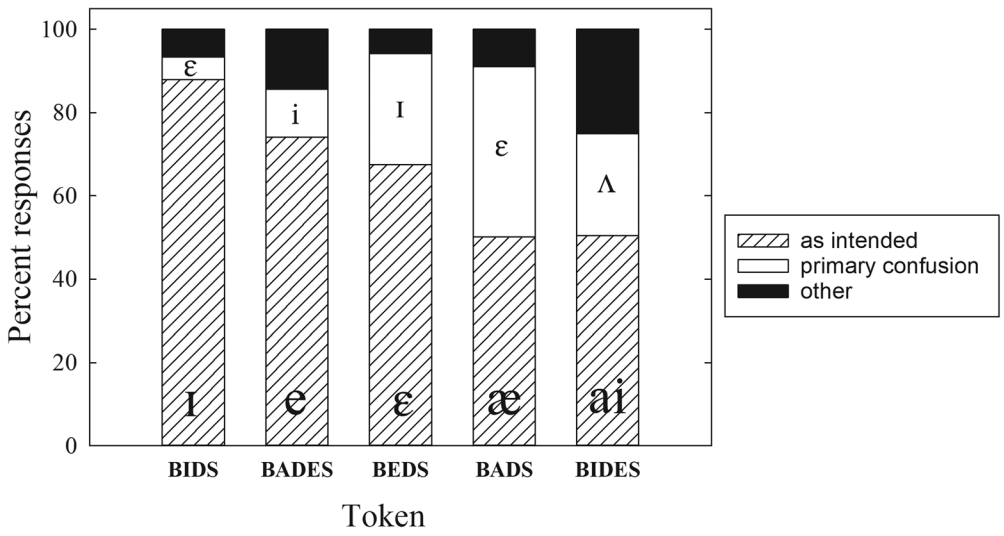


FIG. 3. Average accuracy and confusions for adult listeners for vowels in the local NC dialect.

effects on the accuracy of individual vowel categories. The IDRs for /ɪ/ were high for all three dialects, and none of the pairwise comparisons were significant. For /e/, only the difference between NC and WI reached significance ($p = 0.004$) with NC variants producing higher rates. However, the accuracy for /ɛ/ was strongly influenced by dialect with the highest accuracy for OH variants, followed by WI and NC; all comparisons were significant ($p \leq 0.005$). Although the accuracy for /æ/ was low for all three dialects, the difference between OH and WI did reach significance ($p < 0.001$) with the WI variants having the lowest IDRs. Finally, the IDRs for /ai/ in NC were significantly lower than either for OH or WI ($p < 0.001$), and the latter two did not differ from one another. This interaction showed that accuracy for individual vowel categories (except for /ɪ/) was differentially influenced by talker dialect.

However, there was also a significant three-way interaction between vowel, dialect, and age group [$\chi^2(10) = 23.27$, $p = 0.010$]. The locus of this interaction, visualized in Fig. 4, was in the differential response to the vowels /ɛ/ (*beds*) and /æ/ (*bads*) of adults (left panel) and children (right panel) in the local NC dialect. Subsequent *post hoc* *t*-tests revealed no significant difference in the IDRs for /ɛ/ ($M = 60.7\%$) and /æ/ ($M = 57.7\%$) in adults ($p = 0.650$), but the difference was significant for children ($M = 74.3\%$ and $M = 42.7\%$, respectively, $p < 0.001$). Comparing the age groups, there was a significant difference in the IDRs for adults and children for /ɛ/ ($p = 0.021$) but not for /æ/ ($p = 0.082$). These results indicate that listeners identified the vowel /ɛ/ in their local NC dialect more accurately in children's speech than in adults' speech, whereas identification of /æ/ was equally challenging. However, the accuracy for the two vowels in the adults' productions was comparable (about 60% correct) so that both vowels were confused with other vowels with equal proportions. This was not the case for the children, as the accuracy for /æ/ was disproportionally low, and the accuracy for /ɛ/ was high. We will return to this finding in Sec. IV.

A visual summary of the identification patterns for adult listeners is presented in Fig. 5. The results are broken down by talker age group, dialect, stress, and vowel. Shown are percent correct identifications as intended by talker (color coded for each dialect), predominant confusions, and confusions with other vowel categories. The primary confusions were remarkably consistent across dialects and age groups in spite of extensive variability and spectral overlap of vowels as shown in Figs. 1 and 2. Without exception, /ɪ/ was confused with /ɛ/, /e/ was confused with /i/, /ɛ/ was confused with /ɪ/, and /æ/ was confused with /ɛ/. The confusions for /ai/ were more dialect-dependent, so that NC variants were mostly confused with /ʌ/, and OH and WI variants were confused with /æ/ and sometimes with /ɛ,ɪ/. This overall pattern of confusions matches that found in Fox and Jacewicz (2019) for NC talkers, indicating that inclusion of additional variation in OH and WI speech did not influence listeners' categorization decisions with respect to vowels in their own dialect.

### B. Identification results for child listeners

Average IDRs for the child listeners are shown in Table II. There is a striking difference between the two listener groups in that for all three dialects, the accuracy for /ɪ/ was lower in the children than in the adults, while the reverse was true for /æ/. Apparently, there was a global trade-off in children's vowel identification across the dialects. Compared with the adults, they were able to utilize more cues to the identity of the low vowel /æ/ and fewer cues to the high vowel /ɪ/. It is also apparent that accuracy for the local NC dialect did not gradually decline with each vowel descending in height as found in the adults (see Table I and Fig. 3). Rather, the high and mid vowels were identified with a comparable accuracy, and the IDRs for /æ,ai/ were higher than for the adults. Shown in Fig. 6 are the children's accuracy data for the local NC dialect along with confusions. There is one notable exception to the pattern of the primary confusions found in the adults. The vowel /ɛ/ was mostly confused with /æ/ and not with /ɪ/, which further supports the observation that children perceived the relations among vowel height differently than did the adults.

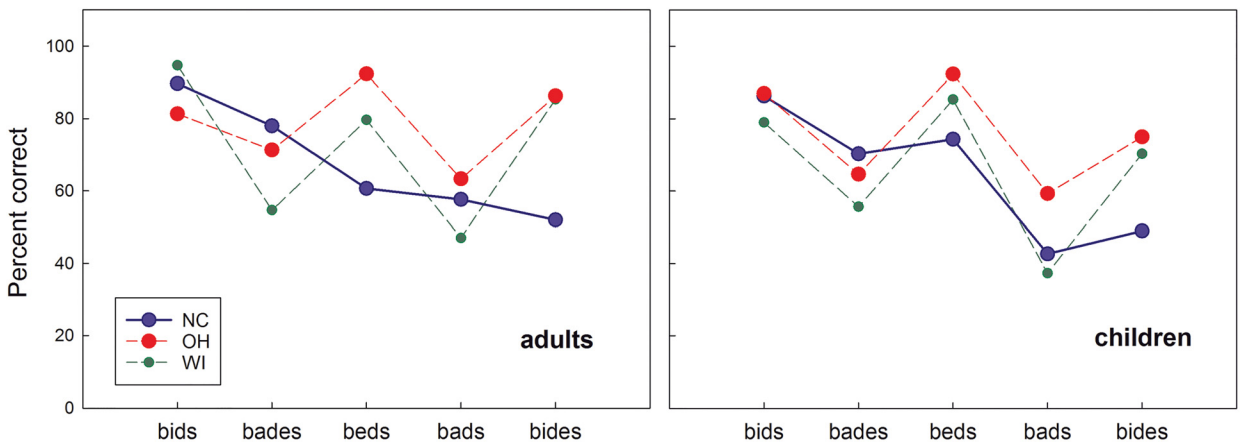


FIG. 4. (Color online) Average accuracy for adult listeners broken down by vowel token, talker dialect, and age group (adults, left panel; children, right panel).

https://doi.org/10.1121/10.0000542

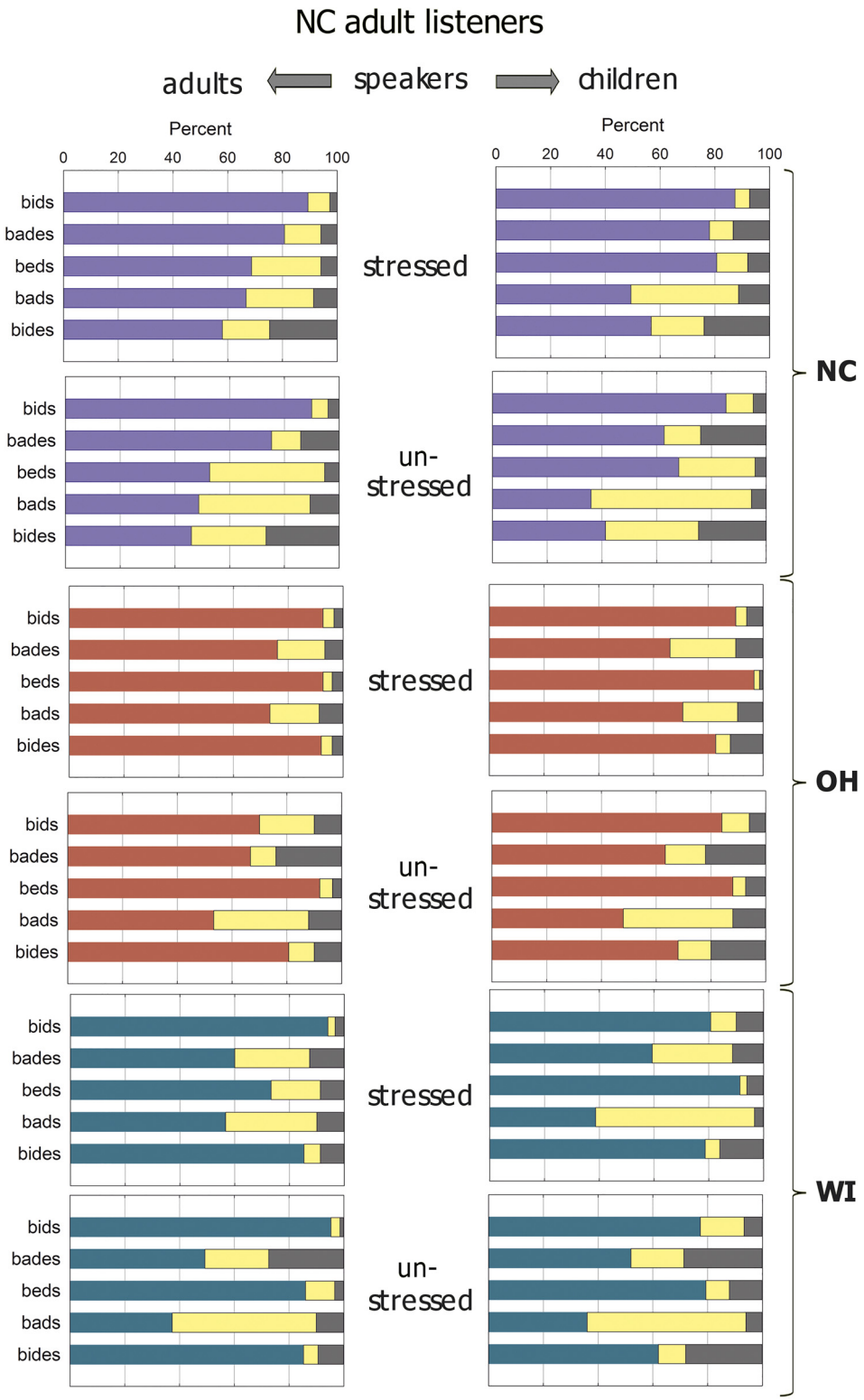


FIG. 5. (Color online) Response pattern for adult listeners. Identification as intended by NC (purple), OH (red), and WI (green) talkers; primary confusion are in yellow, and other confusions are in gray.

The children's accuracy data were analyzed statistically exactly like those of the adult listeners. The best-fitting linear mixed-effects model for children included the main effects of stress, dialect, and vowel, and three interactions between vowel and dialect, vowel and stress, and vowel and dialect and stress. Consistent with the adults, stress had a significant effect on IDRs [$\chi^2(1) = 28.062$, $p < 0.001$] so that stressed vowels were identified with greater accuracy ($M = 73.2\%$) than unstressed vowels ($M = 66.0\%$). The significant main effect of dialect [$\chi^2(2) = 11.14$, $p < 0.001$] revealed that IDRs for OH were significantly higher ($p = 0.003$, $M = 72.4\%$) than for WI ($M = 67.0\%$); this effect was consistent for the adults as well. However, unlike for the adults, neither OH ($p = 0.196$) nor WI ($p = 0.405$) differed significantly from NC ($M = 69.4\%$), indicating that, on average, children identified vowels in their local

TABLE II. Average accuracy (in %) for child listeners.

| | NC child listeners | | | |
|---|---|---|---|---|
| Vowel intended by speaker | Local dialect (NC) | Non-local dialect (OH) | Non-local dialect (WI) | Total |
| /ɪ/ | 76.2 | 60.5 | 67.3 | 68.0 |
| /e/ | 73.7 | 73.5 | 61.7 | 69.6 |
| /ɛ/ | 73.2 | 86.8 | 80.5 | 80.2 |
| /æ/ | 58.5 | 70.5 | 56.0 | 61.7 |
| /ai/ | 65.5 | 70.7 | 69.3 | 68.5 |
| Total | 69.4 | 72.4 | 67.0 | 69.6 |

NC dialect at rates comparable with the two other dialects. The main effect of vowel was significant [$\chi^2(4) = 74.3$, $p < 0.001$]. Subsequent Bonferroni-adjusted pairwise comparisons showed that the differences between /ɪ/-/e/, /ɪ/-/æ/, /ɪ/-/ai/, and /æ/-/ai/ were not significant, and all the other differences were significant ($p \leq 0.002$). The main effect of vowel is thus manifested differently for children than for the adults.

The results for talker age group for children were also not consistent with the results for the adults. In particular, there was no main effect of age group [$\chi^2(1) = 10.6$, $p = 0.390$], and there were no meaningful interactions with age group. Rather, children seem to have been more influenced by the variation in stress. There was a significant interaction between vowel and stress [$\chi^2(4) = 16.4$, $p < 0.001$]. *Post hoc* analyses revealed that stress had a significant effect on IDRs for /ɪ/ ($p < 0.001$) and /æ/ ($p < 0.001$) but not for any other vowel. This indicates that even if IDRs were higher across all stressed vowels when compared with unstressed, the identification of /ɪ/ and /æ/ was more challenging when the vowels were unstressed.

Of interest is a significant three-way interaction between vowel and stress and dialect [$\chi^2(10) = 20.7$, $p = 0.020$]. This interaction is displayed in Fig. 7 and shown separately for stressed (left panel) and unstressed (right panel) vowels. The locus of the interaction was in differential responses to /ɪ/ and /æ/ as a function of dialect. *Post hoc* analyses showed that the accuracy for the OH /ɪ/ was significantly affected by stress ($p < 0.001$), and IDRs were substantially lower in unstressed positions ($M = 47.3\%$) than in stressed ($M = 73.7\%$). The accuracy for /æ/ was affected by stress for both NC and OH; the difference between unstressed ($M = 49.3\%$) and stressed ($M = 67.7\%$) was significant for NC ($p = 0.003$) and, respectively ($M = 63.0\%$ versus 78.0%), for OH ($p = 0.004$). No such significant effects were found for the remaining three vowels, and no significant effects of stress were found for WI. Overall, this three-way interaction revealed that identification of the two vowels was challenging for children, and their IDRs were significantly affected by variation due to dialect and stress.

A corresponding summary of the identification patterns for child listeners is presented in Fig. 8. In terms of the predominant confusions, there was a consistency with the adults for the vowels /ɪ/ (confused with /ɛ/), /e/ (confused with /i/), and /æ/ (confused with /ɛ/). A very different pattern was found for /ɛ/, which was confused with /ɪ/ for all NC variants and /æ/ for both OH and WI. This indicates that NC children perceived the variants of /ɛ/ in their local dialect as adults did, that is, as "belonging" more to the high vowels group, whereas those in the two other dialects were perceptually "closer" to the low /æ/. There was also comparatively more variability in the confusions for /ai/. For all three dialects, the variants of *bides* produced by children tended to be confused with *bids*, although this was also true for those spoken by the OH adults. The NC adults' productions were confused with *buds* when stressed and with *bads* when unstressed, and WI adults' *bides* were only confused with *bades*. Clearly, this variability in child listeners does not correspond to the more regular pattern in adult listeners, indicating that at least some of children's decisions may be related to their inconsistent judgments of an ambiguous vowel quality. This inconsistency is likely reflective of their developmental immaturity when dealing with an

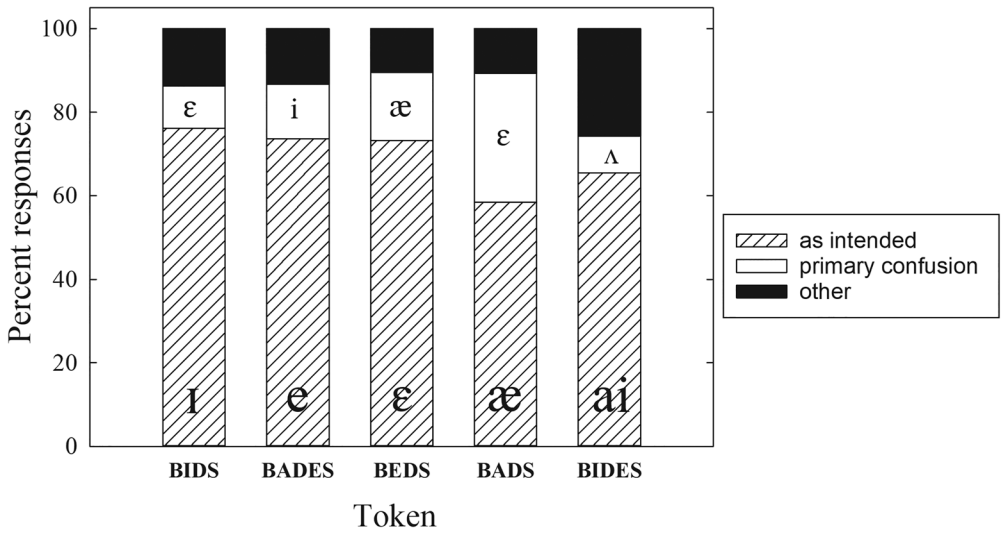


FIG. 6. Average accuracy and confusions for child listeners for vowels in the local NC dialect.

https://doi.org/10.1121/10.0000542

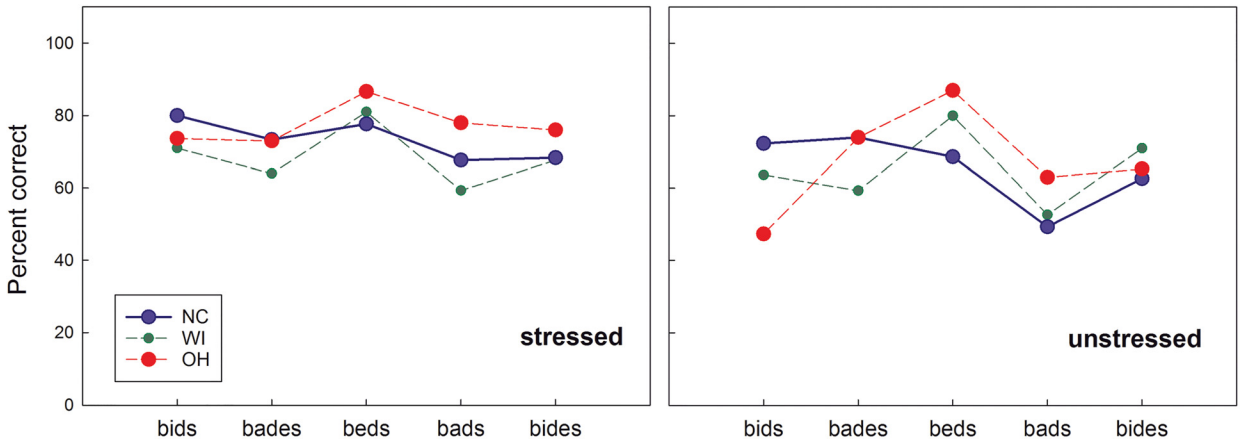


FIG. 7. (Color online) Average accuracy for child listeners broken down by vowel token, talker dialect, and stress (stressed vowels, left panel; unstressed vowels, right panel).

extensive talker variability as this skill characterizes perceptual abilities of adults.

## IV. GENERAL DISCUSSION

The current study produced the following new evidence regarding the issue of the local dialect advantage in the perception of sub-regional variation in Southern vowels. Importantly, we were able to reproduce the general pattern of identifications and confusions of the /ɪ, e, ɛ, æ, ai/ set reported for adult listeners in Fox and Jacewicz (2019). The current experiment used a different group of local adult listeners and increased the task demands by adding variable productions of these vowels from two different dialects. Thus, irrespective of whether listeners only heard the vowels spoken in their local dialect (Fox and Jacewicz, 2019) or whether their local variants were mixed with other regional varieties as in the current experiment, the overall IDRs and the patterns for individual vowels were consistent and highly comparable. This indicates that when responding to monosyllabic tokens listeners' identification choices reflected their ability to utilize a specific set of cues to the identity of each vowel in their own dialect, and the robustness of these cues was not influenced by the context of non-local dialects. However, the cues were not equally robust across the vowels in the set.

In particular, accuracy was highest for the high vowel /ɪ/ and declined for the low vowels, resulting in the overall accuracy rate of 69.4% for the local NC set. This rate corresponds to the 70.0% rate in Fox and Jacewicz (2019). However, the local adult listeners were significantly more accurate when identifying OH vowels ($M = 77.3\%$), representing a mainstream variety of American English, suggesting that, irrespective of their experience with the local Appalachian dialect, the cues in the corresponding OH vowels were comparatively less ambiguous. This finding does not demonstrate that the local listeners had an advantage in the identification of vowels in their own dialect, casting some doubts on the stance that local listeners should always perform better when identifying vowels in their own dialect than vowels in a non-local variety.

The current results do provide evidence, however, that the adult listeners were able to perceive generational differences in vowel production in their own dialect. The significant three-way interaction between vowel and dialect and age group revealed that talker generation influenced the identifications of /ɛ/. While we cannot be certain that listeners indeed utilized the acoustic cues in talkers' productions displayed in Figs. 1 and 2, there is an indication that both positional differences in the $F1$ by $F2$ plane and the differences in formant dynamics can be associated with the obtained identification patterns. In particular, the acoustic proximity of /ɛ/ and /ɪ/ in NC men and their extensive formant movement minimized the contrast between the two vowels, which was manifested in comparatively lower accuracy for /ɛ/ and a greater number of confusions with /ɪ/. However, as the /ɛ/-vowel became more "monophthongal" in NC boys and increased its acoustic distance from /ɪ/, listeners' accuracy significantly improved and confusions with /ɪ/ declined. We note that the accuracy for /ɛ/ in both OH and WI varieties was high, and this success may be attributed to their greater acoustic separation from /ɪ/, as well as its reduced formant movement.

By the same reasoning, the raised position of /æ/ in NC and WI has likely caused extensive confusions with /ɛ/, which were less frequent for the comparatively lower (and more "monophthongal") OH variant. The identification pattern for /ɛ/ and /æ/ suggests that, even in listeners' own dialect, perceptual separation of spectrally overlapping vowels can be challenging, and listeners may not be able to utilize additional "secondary" cues to vowel identity (Klatt, 1982) when spectral cues are dominant. In those situations, the experience with the local dialect has limited advantage, and decisions with respect to vowel quality are likely made at the lexical-semantic level of speech processing. The case of /ai/ further supports this interpretation. The accuracy for the NC variant of /ai/ was again low, and there were extensive confusions with other monophthongs. Apparently, the increased formant movement in NC boys still did not provide the adequate amount of a diphthongal change that would disambiguate the signal. However, the listeners were more successful in identifying the

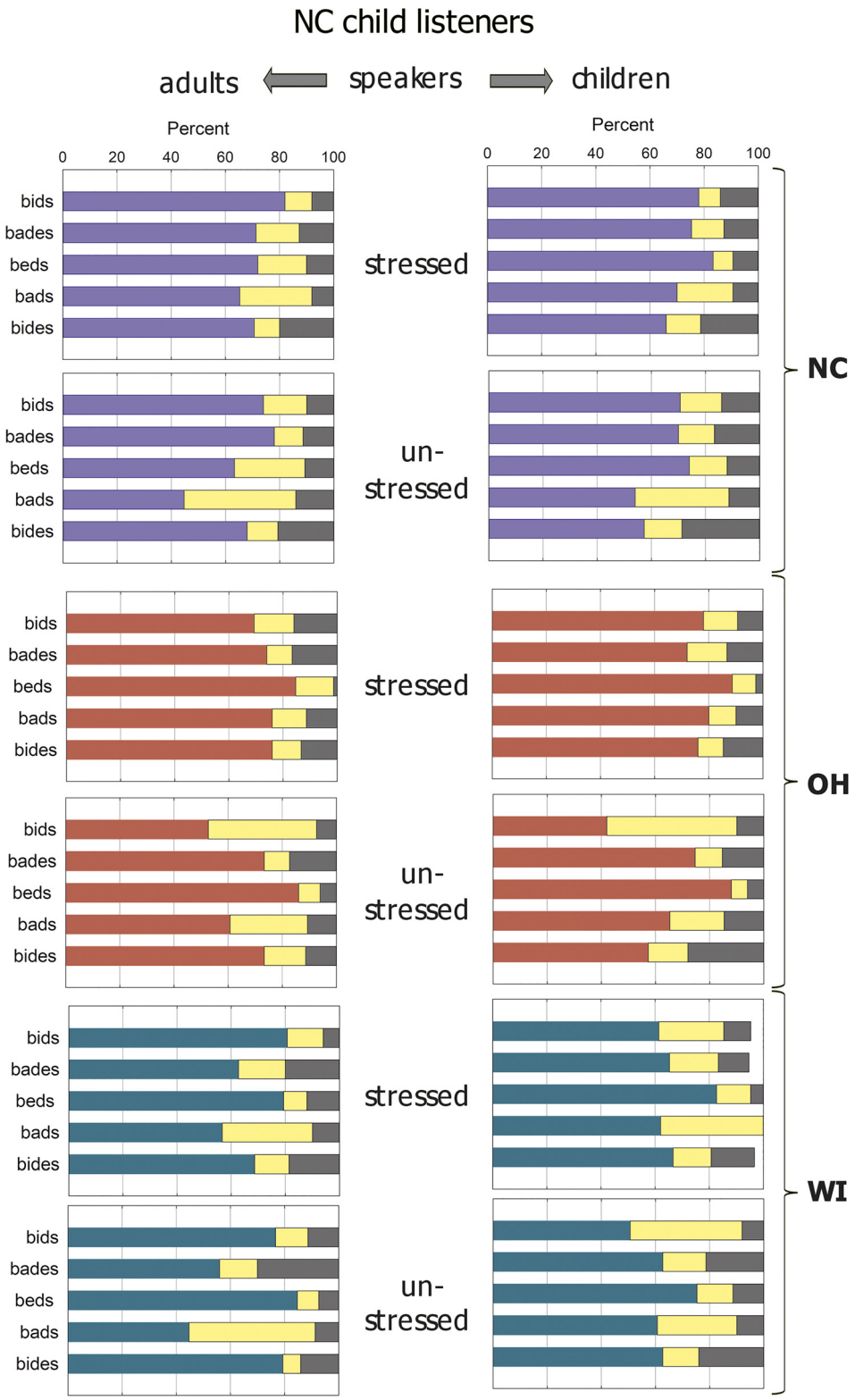


FIG. 8. (Color online) Response pattern for child listeners. Identification as intended by NC (purple), OH (red), and WI (green) talkers; primary confusion are in yellow, and other confusions are in gray.

diphthong in both OH and WI varieties, indicating that the increased dynamics of the full diphthong and the fact that the second diphthongal target was actually produced by the talkers (compare Figs. 1 and 2), were more important for the listeners than the experience with the local dialect.

On the other hand, the case of /e/ does demonstrate the local dialect advantage. The accuracy was relatively high for the NC variant (and for the OH variant), but it was significantly lower for WI. It is the case that the WI /e/ is located closer to /i/ in the formant space and has less formant movement than either NC or OH. Confusions with /i/ can be extensive, even for local WI listeners, as it was documented in our previous work (Jacewicz and Fox, 2012). The current study provides additional evidence for the extensive confusion with /i/ from NC listeners.

https://doi.org/10.1121/10.0000542

Turning to the results for NC child listeners, our prediction that the accuracy for children would be lower when compared with the adults was not supported. Overall accuracy rates for the two listener groups were comparable. However, the differences between the groups were apparent, and they provide important new insights with respect to vowel identification by older children. The major difference was that, for all three dialects, the children were able to utilize comparatively more cues to the identity of /æ/ and comparatively fewer cues to /ɪ/. Speculating on possible sources of this pattern, it could be that children perceived the low vowel /æ/ through the lens of their own production (Kendall and Fridland, 2012; Sumner and Samuel, 2009). These younger listeners were of the same age as the children who produced the speech materials and, as we showed in our previous work, the local NC /æ/ that is raised in older talkers has lowered its position in children, approximating the norms of mainstream American English (Jacewicz *et al.*, 2011a,c). Thus, the children could have been more sensitive to the contrast between /æ/ and /ɛ/, being able to perceive more distinctive cues in either vowel.

This explanation also seems plausible with respect to the vowel /ɪ/ because the ongoing sound change in children due to the influences of the mainstream variety is manifested not only in lowering of /æ/ but also of /ɪ/ and /ɛ/ (Jacewicz *et al.*, 2011a). This lowering of the front lax vowel group could explain the increased confusions of /ɪ/ with /ɛ/, and could also contribute to the differential pattern of confusions for the vowel /ɛ/. If this speculative interpretation can be supported in future work, we would have experimental evidence that children perceive the front vowels differently than do adults on the basis of their own experience with vowel production. Relatedly, the sound change in children would explain their differential identification responses to the vowels /æ/ and /ɪ/ as a function of stress, as indicated by significant two- and three-way interactions with stress. That is, the increased confusions for the unstressed variants of both vowels could result from children's differential perceptual organization along the height dimension for front vowels, and some of the identification uncertainties could have been resolved as more cues became available when the vowels were stressed.

However, we cannot rule out that some of the identification decisions in children were related to their still maturing perceptual skills, especially when faced with extensive talker variability. The existing literature converges in findings that both linguistic perceptual skills and those related to the interpretation of indexical information, such as classification of regional dialects, are not fully adult-like until well into adolescence (e.g., Hazan and Barrett, 2000; Jacewicz and Fox, 2014; Jones *et al.*, 2017; McCullough *et al.*, 2019). Thus, it could be that the children's uncertainty and some of the confusions reflected their still inefficient processing of the variation in the vowel production of multiple talkers.

Since our perception experiment allowed for a second repetition of the token, we inquired into the frequency of the second stimulus repetition in both adults and children. Table III summarizes the number of tokens heard twice broken down by vowel, listener group, and stress position. Out of a total of 1.09% tokens that were heard twice by the adults, 0.32% was for stressed vowels and 0.77% was for unstressed vowels. The total for children was more than twice as much, 2.54%, of which 0.97% was for stressed vowels and 1.57% was for unstressed vowels. The obvious difference between the groups is that children's repetitions were far more frequent when the vowels were unstressed, but it is of note that they also repeated more of the stressed vowels. These differences suggest that children were comparatively more uncertain about vowel identity, which corresponds to their greater reliance on stress cues in making identification decisions. This interpretation is consistent with the finding that English has a large perceptual benefit of stress, and the perception of English unstressed vowels, especially short unstressed vowels, can be quite poor (Warner and Cutler, 2017). The children's repetitions seemed unrelated to the dialect and age of the talker because they were distributed with almost equal proportions across the dialects and age groups. In the absence of a clear pattern in the repetitions, we are left with the interpretation of these inconsistencies that children chose to repeat a given token due to uncertainties related to the task demands, insufficient attention paid to the stimulus, or some other factors that inhibited their identification abilities, including perceptual asymmetries related to sound change as discussed above.

The lingering question stemming from the current study is whether the identification results would differ had the target stimuli contained more contextual information about talker dialect. Had the listeners heard the whole sentences (and not only individual words extracted from these sentences), they would have been able to utilize more of the talker-specific and dialect-specific indexical cues to resolve at least some of the phonetic ambiguities. The rich sociophonetic literature, perhaps beginning with Niedzielski (1999), provides converging evidence that social information about the talkers has a great influence on phonetic categorization. However, the redundancy of sociophonetic information in longer passages of speech necessitates at least some forms of a top-down processing as it would be impossible to attend to and process every phonetic detail. Since top-down processing is highly influenced by context,

TABLE III. Number of tokens heard twice.

| | NC Adults | | NC children | |
|---|---|---|---|---|
| Vowel/token | Stressed | Unstressed | Stressed | Unstressed |
| /ɪ/ *bids* | 1 | 11 | 13 | 33 |
| /e/ *bades* | 8 | 19 | 14 | 21 |
| /ɛ/ *beds* | 6 | 14 | 20 | 30 |
| /æ/ *bads* | 7 | 15 | 17 | 26 |
| /ai/ *bides* | 7 | 10 | 23 | 32 |
| Total | 29 (0.32%) | 69 (0.77%) | 87 (0.97%) | 142 (1.57%) |

experience, beliefs, and expectations (e.g., Gilbert and Sigman, 2007), it is certainly plausible that the adult Appalachian listeners' knowledge of the local manifestation of the sociophonetic variation would influence their vowel categorization decisions to a greater extent than what was possible in the current study.

However, the current results are informative with respect to the influences of dialect variation and talker variability on bottom-up processes involved in vowel perception in isolated context-free monosyllabic words. In particular, while listeners were primarily guided by their phonetic knowledge of vowel identity, we also have evidence that their identification decisions were influenced by top-down modulation of acoustic information in resolving ambiguities. This indicates that, at the single-syllable level, perceptual bottom-up and top-down processes can interact in making vowel identification decisions. Additional support for this interpretation comes from our previous studies that demonstrated listeners' sensitivity to talker dialect in hVd-words (Fox and Jacewicz, 2011; Jacewicz and Fox, 2017).

It was also the case that, in the vowel set used in this study, the phonetic cues to vowel identify were not equally robust, such that the vowel /ɪ/ was not as easily confused with its neighbors, whereas the identification of /æ/ was challenging. This tendency was also reported in other studies (e.g., Hillenbrand *et al.*, 1995; Ladefoged and Broadbent, 1957), indicating that the raised variant of /æ/, such as in the Appalachian and WI varieties, fell well into the "area" of /ɛ/, and thus experience with the local dialect was of limited utility when no additional context was available at the single-syllable level. Possibly, the perceptual category for /ɛ/ is much broader than that for /ɪ/ as suggested in Ladefoged and Broadbent (1957, p. 101): "the relative position of the vowel /ɛ/ as in *bet* can be anywhere in a comparatively large area." The current study supports this observation as the confusions with /ɛ/ were prevalent for /æ/, and particularly for the unstressed variants.

## V. CONCLUSIONS

The aim of the current paper was to assess the ability of Southern listeners to identify vowels in their own subregional variety in the context of sound change undergoing in their Appalachian speech community in Western NC. Building on prior work, the current study sought to establish whether adult and child listeners would demonstrate the local dialect advantage when identifying a subset of spectrally overlapping vowels in local and two non-local varieties. The results indicate that accuracy levels are highly dependent upon the robustness of cues in individual vowel categories irrespective of their local versus non-local status rather than on the sole experience with the local dialect. The local dialect advantage was not demonstrated for most vowel categories when the performance was compared with the non-local mainstream variety spoken in Ohio. However, adult listeners did show sensitivity to generational changes in vowel production, indicating their familiarity with the local norms.

The overall accuracy for children was comparable with that for the adults. However, there were notable differences in the identification and confusion patterns, suggesting that children perceive some of the front vowels differently than do adults. This differential pattern seems to be related to the sound change in the community, and children's own experience with vowel production may underlie their perceptual organization of front vowels differing in height. Also, compared with adults, children's accuracy can be influenced by the variation in linguistic stress to a greater extent, with increased confusion when the vowels are unstressed. Although duration cues are known to be the best predictors of vowel identification scores (Neel, 2008), children seem to rely more on the longer durations of stressed vowels than do adults. The results also indicate that the children's uncertainty and some of the confusion reflects their still inefficient processing of the variation in the vowel production of multiple talkers.

In conclusion, the current study provides evidence that the experience and familiarity with the local dialect can be of limited advantage when listeners are asked to identify vowels in monosyllabic words across multiple talkers. Rather, the accuracy is dependent upon the robustness of cues in individual vowel categories, and these cues may well be less ambiguous in non-local variants. However, the bottom-up processes underlying vowel categorization in isolated monosyllabic words do seem to interact with the top-down processing of dialect-specific information, indicating that socioindexical knowledge can be accessed not only in listening to longer passages of speech, but it can also be activated at the single-syllable level.

## ACKNOWLEDGMENTS

Research reported in this paper was supported by The National Institute on Deafness and Other Communication Disorders (NIDCD) under Award No. R01DC006871. The content is solely the responsibility of the authors and does not necessarily represent the official views of The National Institutes of Health (NIH). We thank Joseph Salmons for his contributions to this research and Janaye Houghton for data collection in NC. We also thank Erica Huston and Lian Arzbecker for help with data displays.

Baranowski, M. (**2008**). "The fronting of the back upgliding vowels in Charleston, South Carolina," Lang. Var. Change **20**, 527–551.

Bladon, A. (**1982**). "Arguments against formants in the auditory representation of speech," in *The Representation of Speech in the Peripheral Auditory System*, edited by R. Carlson and B. Granstrom (Elsevier Biomedical, Amsterdam), pp. 95–102.

Chevrot, J.-P., Beaud, L., and Varga, R. (**2000**). "Developmental data on a French sociolinguistic variable: The word-final post-consonantal /R/," Lang. Var. Change **12**, 295–319.

Choi, J. Y., Hu, E. R., and Perrachione, T. K. (**2018**). "Varying acoustic-phonemic ambiguity reveals that talker normalization is obligatory in speech processing," Atten. Percept. Psychophys. **80**, 784–797.

Clark, A. D., and Hayward, N. M. (eds.) (**2013**). *Talking Appalachian: Voice, Identity, and Community* (The University Press of Kentucky, Lexington, KY).

Clopper, C. G., Pisoni, D., and de Jong, K. (**2005**). "Acoustic characteristics of the vowel systems of six regional varieties of American English," J. Acoust. Soc. Am. **118**, 1661–1676.
Cutler, A. (**2005**). "Lexical stress," in *The Handbook of Speech Perception*, edited by D. Pisoni, and R. Remez (Wiley-Blackwell, Hoboken, NJ), pp. 264–289.
Dodsworth, R., and Kohn, M. (**2012**). "Urban rejection of the vernacular: The SVS undone," Lang. Var. Change **24**, 221–245.
Farrington, C., Kendall, T., and Fridland, V. (**2018**). "Vowel dynamics in the Southern Vowel Shift," Am. Speech **93**, 186–222.
Ferguson, S. H. (**2004**). "Talker differences in clear and conversational speech: Vowel intelligibility for normal-hearing listeners," J. Acoust. Soc. Am. **116**, 2365–2373.
Foulkes, P., Docherty, G. J., and Watt, D. (**2005**). "Phonological variation in child-directed speech," Language **81**, 177–206.
Fox, R. A., and Jacewicz, E. (**2009**). "Cross-dialectal variation in formant dynamics of American English vowels," J. Acoust. Soc. Am. **126**, 2603–2618.
Fox, R., A., and Jacewicz, E. (**2011**). "Perceptual sensitivity to dialectal and generational variations in vowels," in *INTERSPEECH 2011: Proceedings of the 12th Annual Conference of the International Speech Communication Association*, Florence, Italy, pp. 2921–2924.
Fox, R. A., and Jacewicz, E. (**2017**). "Reconceptualizing the vowel space in analyzing regional dialect variation and sound change in American English," J. Acoust. Soc. Am. **142**, 444–459.
Fox, R. A., and Jacewicz, E. (**2019**). "Perceptual separation of spectrally overlapping vowels," in *Proceedings of the 19th International Congress of Phonetic Sciences*, edited by S. Calhoun, P. Escudero, M. Tabain, and P. Warren (Australasian Speech Science and Technology Association Inc., Canberra, Australia), pp. 874–878.
Fridland, V. (**1999**). "The Southern Shift in Memphis, Tennessee," Lang. Var. Change **11**, 267–285.
Fridland, V. (**2001**). "The social dimension of the Southern Vowel Shift: Gender, age and class," J. Sociolinguist **5**, 233–253.
Fridland, V. (**2012**). "Rebel vowels: Southern vowel shift and the N/S speech divide," Lang. Linguist. 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., and Kendall, T. (**2015**). "Within-region diversity in the Southern Vowel Shift: Production and perception," in *Proceedings of the 18th International Congress of Phonetic Sciences*, edited by M. Wolters, J. Livingstone, B. Beattie, R. Smith, M. MacMahon, J. Stuart-Smith, and J. Scobbie (University of Glasgow, Glasgow, UK).
Fry, D. B. (**1955**). "Duration and intensity as physical correlates of linguistic stress," J. Acoust. Soc. Am. **27**, 765–768.
Gilbert, C. D., and Sigman, M. (**2007**). "Brain states: Top-down influences in sensory processing," Neuron **54**, 677–696.
Harrington, J., Palethorpe, S., and Watson C. I. (**2007**). "Age-related changes in fundamental frequency and formants: A longitudinal study of four speakers," in *INTERSPEECH 2007: Proceedings of the 8th Annual Conference of the International Speech Communication Association*, Baixas, France, pp. 2353–2376.
Hay, J., Nolan, A., and Drager, K. (**2006**). "From *fush* to *feesh*: Exemplar priming in speech perception," Linguist. Rev. **23**, 351–379.
Hazan, V., and Barrett, S. (**2000**). "The development of phonemic categorization in children aged 6–12," J. Phon. **28**, 377–396.
Heald, S., Klos, S., and Nusbaum, H. C. (**2016**). "Understanding speech in the context of variability," in *Neurobiology of Language*, edited by G. Hickok and S. Small (Academic, San Diego, CA), pp. 195–208.
Hillenbrand, J. M., Getty, L. A., Clark, M. J., and Wheeler, K. (**1995**). "Acoustic characteristics of American English vowels," J. Acoust. Soc. Am. **97**, 3099–3111.
Hillenbrand, J. M., and Houde, R. A. (**2003**). "A narrow band pattern-matching model of vowel perception," J. Acoust. Soc. Am. **113**, 1044–1055.
IBM. (**2017**). *IBM SPSS Statistics, v. 25* (International Business Machines Corp., Armonk, NY).
Ito, M., Tsuchida, J., and Yano, M. (**2001**). "On the effectiveness of whole spectral shape for vowel perception," J. Acoust. Soc. Am. **110**, 1141–1149.
Jacewicz, E., and Fox, R. A. (**2012**). "The effects of cross-generational and cross-dialectal variation on vowel identification and classification," J. Acoust. Soc. Am. **131**, 1413–1433.
Jacewicz, E., and Fox, R. A. (**2014**). "The effects of indexical and phonetic variation on vowel perception in typically developing 9- to 12-year-old children," J. Speech Lang. Hear. Res. **57**, 389–405.
Jacewicz, E., and Fox, R. A. (**2017**). "Dialect perception by older children," in *INTERSPEECH 2017: Proceedings of the 18th Annual Conference of the International Speech Communication Association*, Stockholm, Sweden, pp. 354–358.
Jacewicz, E., and Fox, R. A. (**2019**). "The old, the new, and the in-between: Preadolescents' use of stylistic variation in speech in projecting their own identity in a culturally changing environment," Dev. Sci. **22**, e12722.
Jacewicz, E., Fox, R. A., and Salmons, J. (**2011a**). "Cross-generational vowel change in American English," Lang. Var. Change **23**, 45–86.
Jacewicz, E., Fox, R. A., and Salmons, J. (**2011b**). "Vowel change across three age groups of speakers in three regional varieties of American English," J. Phon. **39**, 683–693.
Jacewicz, E., Fox, R. A., and Salmons, J. (**2011c**). "Regional dialect variation in the vowel systems of typically developing children," J. Speech Lang. Hear. Res. **54**, 448–470.
Jaeger, T. F. (**2008**). "Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models," J. Mem. Lang. **59**, 434–446.
Johnson, K. (**2006**). "Resonance in an exemplar-based lexicon: The emergence of social identity and phonology," J. Phon. **34**, 485–99.
Jones, Z., Yan, Q., Wagner, L., and Clopper, C. G. (**2017**). "The development of dialect classification across the lifespan," J. Phon. **60**, 20–37.
Kendall, T., and Fridland, V. (**2012**). "Variation in perception and production of mid front vowels in the U.S. Southern Vowel Shift," J. Phon. **40**, 289–306.
Klatt, D. H. (**1982**). "Prediction of perceived phonetic distance from critical-band spectra: A first step," in *Proc. IEEE Int. Conf. Speech, Acoust. Signal Process.*, pp. 1278–1281.
Kleinschmidt, D. F. (**2019**). "Structure in talker variability: How much is there and now much can it help?," Lang. Cogn. Neurosci. **34**, 43–68.
Kleinschmidt, D. F., and Jaeger, T. F. (**2015**). "Robust speech perception: Recognize the familiar, generalize to the similar, and adapt to the novel," Psychol. Rev. **122**(2), 148–203.
Kleinschmidt, D. F., Weatherholtz, K., and Jaeger, T. F. (**2018**). "Sociolinguistic perception as inference under uncertainty," Top. Cogn. Sci. **10**, 818–834.
Labov, W. (**2010**). *Principles of Linguistic Change. III: Cognitive and Cultural Factors* (Blackwell, Oxford), pp. 1–419.
Labov, W., and Ash, S. (**1997**). "Understanding Birmingham," in *Language Variety in the South Revisited*, edited by C. Bernstein, T. Nunnally, and R. Sabino (University of Alabama Press, Tuscaloosa, AL), pp. 508–573.
Labov, W., Ash, S., and Boberg, C. (**2006**). *Atlas of North American English: Phonetics, Phonology, and Sound Change* (Mouton de Gruyter, Berlin), pp. 1–318.
Ladefoged, P., and Broadbent, D. E. (**1957**). "Information conveyed by vowels," J. Acoust. Soc. Am. **29**, 98–104.
McCullough, E. A., Clopper, C. G., and Wagner, L. (**2019**). "Regional dialect perception across the lifespan: Identification and discrimination," Lang. Speech **62**, 115–136.
Montgomery, C., and Moore, E. (**2018**). "Evaluating S(c)illy voices: The effects of salience, stereotypes, and co-present language variables on real-time reactions to regional speech," Language **94**, 629–661.
Morrison, G. S., and Assmann, P. F. (**2013**). *Vowel Inherent Spectral Change* (Springer, Berlin, Germany), pp. 1–286.
Nearey, T. M., and Assmann, P. F. (**1986**). "Modeling the role of vowel inherent spectral change in vowel identification," J. Acoust. Soc. Am. **80**, 1297–1308.
Neel, A. (**2008**). "Vowel space characteristics and vowel identification accuracy," J. Speech Lang. Hear. Res. **51**, 574–585.
Niedzielski, N. (**1999**). "The effect of social information on the perception of sociolinguistic variables," J. Lang. Soc. Psychol. **18**, 62–85.
Nittrouer, S., and Miller, M. E. (**1997**). "Predicting developmental shifts in perceptual weighting schemes," J. Acoust. Soc. Am. **101**, 2253–2266.
Peterson, G. E., and Barney, H. L. (**1952**). "Control methods used in a study of the vowels," J. Acoust. Soc. Am. **24**, 175–184.

Pierrehumbert, J. B. (**2016**). "Phonological representation: Beyond abstract versus episodic," Annu. Rev. Linguist. **2**, 33–52.
Shvachkin, N. K. (**1973**). "The development of phonemic perception in early childhood," in *Studies of Child Language Development*, edited by C. A. Ferguson and D. I. Slobin (Holt, Rinehart and Winston, New York), pp. 92–127.
Smith, J., Durham, M., and Fortune, L. (**2007**). "Community, caregiver and child in the acquisition of variation in a Scottish dialect," Lang. Var. Change **19**, 63–99.
Sumner, M., and Samuel, A. (**2009**). "The effect of experience on the perception and representation of dialect variants," J. Mem. Lang. **60**, 487–501.
Syrdal, A. K., and Gopal, H. S. (**1986**). "A perceptual model of vowel recognition based on the auditory representation of American English vowels," J. Acoust. Soc. Am. **79**, 1086–1100.
Theodore, R. M., and Miller, J. L. (**2010**). "Characteristics of listener sensitivity to talker-specific phonetic detail," J. Acoust. Soc. Am. **128**, 2090–2099.
Thomas, E. R. (**1989**). "The implications of /o/ fronting in Wilmington, North Carolina," Am. Speech **64**, 327–333.
Warner, N., and Cutler, A. (**2017**). "Stress effects in vowel perception as a function of language-specific vocabulary patterns," Phonetica **74**, 81–106.
Warton, D. I., and Hui, F. K. (**2011**). "The arcsine is asinine: The analysis of proportion in ecology," Ecology **92**, 3–10.
Werker, J. F. (**2018**). "Perceptual beginnings to language acquisition," Appl. Psycholinguist. **39**, 703–728.
Zahorian, S., and Jagharghi, A. (**1993**). "Spectral shape features versus formants as acoustic correlates for vowels," J. Acoust. Soc. Am. **94**, 1966–1982.