Regional dialect perception across the lifespan: Identification and discrimination

Elizabeth A. McCullough
University of Washington, USA

Cynthia G. Clopper
Ohio State University, USA

Laura Wagner
Ohio State University, USA

Abstract
Although adult listeners can often identify a talker’s region of origin based on his or her speech, young children typically fail in dialect perception tasks, and little is known about the development of regional dialect representations from childhood into adulthood. This study explored listeners’ understanding of the indexical importance of American English regional dialects across the lifespan. Listeners between 4 and 79 years old in the Midwestern United States heard talkers from the Midland, Northern, Southern, and New England regions in two regional dialect perception tasks: identification and discrimination. The results showed that listeners as young as 4–5 years old understand the identity-marking significance of some regional dialects, although adult-like performance was not achieved until adolescence. Further, the findings suggest that regional dialect perception is simultaneously impacted by the specific dialects involved and the cognitive difficulty of the task.

Keywords
Dialect perception, language development, identification, discrimination

Introduction
A voice conveys considerable personal information about the talker, including his or her age, sex, and regional dialect (Abercrombie, 1967). Adult listeners are sensitive to this information, as demonstrated by substantial previous work examining the perception of a variety of social-indexical...
dimensions. For instance, adults can accurately distinguish younger from older talkers (Ptacek & Sander, 1966), estimate talkers’ ages (Ryan & Capadano, 1978), and identify a talker’s sex (Lass, Hughes, Bowyer, Waters, & Bourne, 1976). Moreover, adult listeners can divide talkers into dialect groups (Clopper & Bradlow, 2009; Clopper & Pisoni, 2006; Clopper, Rohrbeck, & Wagner, 2012) and assign to talkers dialect labels provided by experimenters (Clopper & Pisoni, 2004; Van Bezooijen & Gooskens, 1999; Williams, Garrett, & Coupland, 1999).

Children’s sensitivity to social-indexical variation in the speech signal is less well understood. On the one hand, several recent studies have suggested that 4- to 5-year-old children are generally unsuccessful in tasks requiring explicit labeling or discrimination judgments about regional dialects (Floccia, Butler, Girard, & Goslin, 2009; Girard, Floccia, & Goslin, 2008; Wagner, Clopper, & Pate, 2014) and even adolescents do not perform as well as adults in explicit dialect identification tasks (Williams et al., 1999). On the other hand, however, 4- to 5-year-old children are able to make accurate explicit labeling and discrimination judgments about native versus foreign accents (Floccia et al., 2009; Girard et al., 2008; Wagner et al., 2014), indicating that they perceive and understand at least some kinds of social-indexical variation in speech. Children’s productions also exhibit sensitivity to social-indexical variation. For example, 9- to 14-year-old children adopt local dialect features in production when they move to a new region (Chambers, 1992) and 4-year-old children exhibit appropriate style-shifting for different listeners (Shatz & Gelman, 1973). The results of investigations involving implicit rather than explicit perceptual responses to social-indexical variation further reveal children’s sensitivity to this variation, in that children express social preferences for talkers who sound like them regardless of whether the other talkers are foreign-accented (Kinzler, Shutts, DeJesus, & Spelke, 2009) or speak with a native dialect from a different region (Kinzler & DeJesus, 2013). Thus, previous reports of children’s understanding of the significance of regional dialect variation are inconsistent, with some tasks revealing more knowledge than others. Moreover, much of this previous work has focused on preschool- and elementary-aged children, which means that the trajectory of development from early failure to adult-like success is largely unknown. In the current study, listeners’ explicit awareness of the significance of regional dialect variation was tested using two tasks with preschool-aged children through adults. The results advance the understanding of children’s developing representations of this source of social-indexical variability.

1.1 Adults’ perception of regional dialect variation

To explore the development of children’s perceptual representations of regional dialect variation, the adult-like target state must first be understood. Previous research on adults’ perception of regional dialect variation has demonstrated that adults exhibit above-chance, but below-ceiling, performance on dialect identification tasks that require fine-grained judgments about regional background. For instance, Van Bezooijen and Gooskens (1999) asked Dutch listeners to select the country, region, and province of origin for talkers of four regional varieties of Dutch. Across varieties, listeners were 90% correct about country of origin, 60% correct about region, and 40% correct about province. In a similar task with five regional varieties of American English, listeners in the United Kingdom also performed well: they were 92% correct about country, 88% correct about region, and 52% correct about area (Van Bezooijen & Gooskens, 1999). Williams et al. (1999) also tested listeners in the United Kingdom, using six regional varieties of Welsh English as well as Received Pronunciation (RP). Listeners easily identified RP (77–85% correct), but were somewhat less accurate for specific Welsh regional varieties (28–68% correct). In the United States, Clopper and Pisoni’s (2004) listeners performed forced-choice categorization with six regional varieties of American English. While they performed reliably above chance, the accuracy rates were somewhat
lower than in other studies, never exceeding 40% correct. In free classification tasks, in which listeners create, but do not necessarily label, their own groups of talkers based on regional dialect, adult listeners also perform well, but not perfectly: they both group together talkers from different regions and place talkers from the same region into different groups (Clopper & Bradlow, 2009; Clopper et al., 2012).

Imperfect dialect perception is therefore observed across languages and tasks, reflecting adult listeners’ imperfect representations of regional dialect variation in their native language. Regional varieties may be imperfectly represented due to limited familiarity with one or more relevant varieties, as when Midwestern American English listeners perceive talkers from the New England and Mid-Atlantic regions as similar, despite observable phonetic and phonological differences between them (Clopper & Pisoni, 2006), or when local listeners perceive more distinctions between neighboring varieties than non-local listeners who are less familiar with the local dialect landscape (Baker, Eddington, & Nay, 2009; Preston, 1993). Moreover, the phonetic differences between dialects may not be strongly associated with regional identities, as when Ohioans interpret perceptible Northern American English dialect features as idiosyncratic characteristics of particular talkers, rather than as indexing regional background (Campbell-Kibler, 2012), or when non-native listeners interpret two features of Southern American English as indexing different regional backgrounds because they lack knowledge about how those features co-occur in the South (Clopper & Bradlow, 2009). These latter two examples illustrate that neither the ability to simply perceive differences between dialects nor mere exposure to different varieties is sufficient to explain perceptual dialect classification behavior. Rather, adults’ representations of regional dialect variation are shaped by their familiarity with particular varieties as well as their shared cultural knowledge or stereotypes about which linguistic features are associated with those varieties. This knowledge linking variation to meaningful group categories is not always very sophisticated, such as when American English listeners distinguish primarily among non-native English varieties, non-American English varieties, and American English varieties in a free classification task (Bent, Atagi, Akbik, & Bonifield, 2016), or when Midwestern listeners exhibit perceptual “Northeastern” and “Midwestern” dialect categories that comprise multiple dialect regions identified by sociolinguists (Clopper & Pisoni, 2006). However, these representations reflect the social identity distinctions that are relevant to the adult population and it is these representations connecting phonetic variation to social groups that children must acquire as part of their sociolinguistic competence.

1.2 Children’s perception of regional dialect variation

The development of sociolinguistic competence in dialect perception is virtually untested (cf. Jones, Yan, Wagner, & Clopper, 2017; Williams et al., 1999), although the available evidence suggests a protracted developmental trajectory from early childhood through adolescence. In particular, Williams et al. (1999) reported poorer performance for 14-year-olds than adults in a six-alternative forced-choice dialect identification task, and Jones et al. (2017) did not observe adult-like accuracy in a free classification task until age 16–17 years, although adult-like dialect categories were observed by age 8–9 years. In the current study, dialect perception in 4- to 17-year-old children and in adults was examined to provide a more complete understanding of the development of regional dialect representations.

Explicit tests of children’s representations of regional dialect variation have focused primarily on 5- to 6-year old children, and differ substantially from studies with adults in that real-world labels like “New England” are generally avoided. When dialect identification tasks are used with children, they typically involve ad hoc category labels such as colors, and include an initial training phase to teach listeners the association between a regional variety and its ad hoc label. This training
phase is critical because the children may not have any previous experience with the varieties that are included in the task and therefore may not already have relevant associations between the phonetic features of the varieties and particular places or social groups. Using this kind of ad hoc identification task, Girard et al. (2008) found that 5- and 6-year-old children from Northern France were generally unable to make use of differences between Northern and Southern varieties of French in forced-choice identification, although they were able to accurately identify native Northern French and non-native English-accented French varieties (see also Floccia et al., 2009). Likewise, 5- and 6-year-old listeners tested by Wagner et al. (2014) in the United States could not successfully categorize talkers who spoke Midland American English versus Lancashire British English, although they were able to identify Midland American English and non-native Indian English speakers with above-chance accuracy. Similarly, Evans, Madigan, and Tome Lourido (2016) found that monolingual children in London were able to distinguish more accurately between London English and non-native Singaporean English than between London English and Yorkshire English in an ad hoc identification task, although multilingual children in London showed similar success with both tasks. Thus, as in the research with adults, the varieties that children are asked to identify have an impact on performance; 5- and 6-year-old children are able to succeed in ad hoc identification tasks involving their native variety and a non-native variety, but exhibit limited success in tasks involving their native variety and another regional variety of their native language.

This difference in perceptual performance between native and non-native varieties most likely reflects the phonetic distance between the varieties (see also Cristia et al., 2012). For example, non-native varieties may be more phonetically different from the native variety overall than are regional varieties, making them easier to distinguish from the native variety. Alternatively, or in addition, non-native varieties may differ in phonetic dimensions (e.g., consonants vs. vowels vs. prosody) that make them easier to distinguish from the native variety, even when the overall degree of accentedness is matched across non-native and native varieties (Floccia et al., 2009; Wagner et al., 2014). When studying regional dialect perception, the linguistic properties of the particular varieties involved are therefore important. Young listeners’ previous failures in identification tasks with ad hoc labels for regional dialects might reflect the use of difficult-to-distinguish regional dialects, rather than their inherent failure to understand the significance of regional dialect variation, especially given that they succeed in these tasks for native versus non-native accent comparisons. The current study considered whether more differentiable dialect choices would allow young listeners to succeed.

Overall task demands may also contribute to young listeners’ poor performance in these regional dialect identification tasks. By their very nature, identification tasks that use ad hoc labels impose substantial memory and processing demands on the listener, who must first learn and then remember which phonetic features, some of which may be unfamiliar, are associated with which of two arbitrary labels. Indeed, simpler tasks, including forced-choice localness judgment tasks without ad hoc labels, have revealed somewhat greater success for children at the same ages. In the United States, Kinzler and DeJesus (2013) found that 5- and 6-year-olds in the Northern dialect region, but not in the Southern dialect region, were able to correctly select whether a Northern or a Southern talker was more likely to live near them. Similarly, Beck (2014) reported that 5- and 6-year-old children near Philadelphia correctly chose a local Philadelphia talker, rather than a talker from the South, as sounding more like them; however, the same children responded at chance when asked directly whether a Southern talker was from the local area. Thus, whereas ad hoc identification tasks require an explicit judgment about which of two groups a particular talker belongs to (Floccia et al., 2009; Girard et al., 2008; Wagner et al. 2014), the identification tasks used by Kinzler and DeJesus (2013) and Beck (2014) required a response that was directly anchored to the listener’s personal experience by asking whether each talker sounded like them or lived near them. This kind
of local anchoring may make the task easier for young listeners because it involves only a comparison to existing representations of the in-group, rather than comparisons between two ad hoc category associations, which may or may not correspond to pre-existing representations.

Other types of tasks have also revealed young children’s sensitivity to regional dialect variation. In one study, Girard et al. (2008) found that some 5- and 6-year-olds performed above chance in a dialect discrimination task in which listeners heard speech samples from two talkers from the same regional variety or different regional varieties and decided whether they were from the same place or different places. Given that all of the trials in Girard et al.’s (2008) task involved a comparison between two different talkers, simply perceiving phonetic differences between the two talkers was insufficient to accurately identify them as coming from the same place or different places. Successful dialect discrimination required attending to the dimensions of variation that index regional background while ignoring irrelevant dimensions of variation. Although this task requires participants to determine whether two talkers are different enough along relevant phonetic dimensions to be from different places, it may be easier than an ad hoc identification task because it does not require linking talkers to particular places or groups.

More recently, in the United States, Jones et al. (2017) presented samples of Midland, Northern, Southern, and New England speech to listeners between 4 and 86 years old and asked them to group the talkers by region in a free classification task. Adult listeners placed New England and Southern talkers into distinct groups, but did not differentiate Midland from Northern talkers, consistent with Campbell-Kibler’s (2012) observations about the lack of social stereotypes associated with the Northern dialect of American English. While response patterns were not fully adult-like until listeners were in their late teenage years, 4- and 5-year-olds’ classifications offered evidence of a distinct New England group, 6- and 7-year-olds’ classifications offered evidence of a distinction between New England and Southern talkers on the one hand and Midland and Northern talkers on the other, and 8- and 9-year-olds’ classifications offered evidence of separate New England, Southern, and Midland/Northern groups, parallel to the adult pattern. As in Girard et al.’s (2008) discrimination task, these classification patterns demonstrate that children not only perceive phonetic differences between talkers, but also know which of these differences are relevant for sorting talkers by where they are from. Further, although Jones et al.’s (2017) implementation of the free classification task required explicit judgments about which of an unspecified number of groups a particular talker belonged to, the listeners defined the groups themselves and were allowed to replay the stimulus materials throughout the grouping process. These two aspects of the design may have substantially reduced the difficulty of the task by eliminating the need to remember associations between phonetic features and ad hoc category labels over the course of the entire experiment.

Together, the results of these localness judgment (Beck, 2014; Kinzler & DeJesus, 2013), dialect discrimination (Girard et al., 2008), and free classification (Jones et al., 2017) tasks provide converging evidence that the reported failures of 5- to 6-year-old children in ad hoc regional dialect identification tasks may reflect the demands of the task rather than children’s lack of knowledge about regional dialect variation, although performance in these tasks is still much lower than what would be expected of adults. In the current study, ad hoc identification was compared with dialect discrimination (following Girard et al., 2008) to explore how task demands affect children’s ability to demonstrate their representations of regional dialect variation.

1.3 The current study

The goals of the current study were to explore children’s knowledge about how variation in speech indexes talkers as being from different regions and how this knowledge develops into the adult state. The first core research question was whether young children demonstrate awareness of any
regional dialect contrasts in an identification task with ad hoc labels (Girard et al., 2008; Wagner et al., 2014). Specifically, it was predicted that regional dialect contrasts differ in their perceptual difficulty and the first goal was to determine if any regional dialect contrasts are sufficiently differentiable to offset the substantial processing demands of the task for preschool-aged children. The second core research question was how children’s representations of regional dialects develop as they age. In particular, given the prediction that regional dialect contrasts differ in their perceptual difficulty, the second goal was to determine how the development of representations of different regional dialects compares to those observed by Jones et al.’s (2017) using the free classification task. Both of the tasks in the current study provide an opportunity to examine dialect contrasts pairwise, which may provide a more nuanced perspective on the representations of individual dialects than was possible from Jones et al.’s (2017) results. The third core research question was whether an explicit regional dialect task without the ad hoc learning dimension can reveal more sensitive information about children’s representations. To address this question, a dialect discrimination task (Girard et al., 2008) was used, which mirrors the ad hoc identification task in that it is a two-alternative forced-choice task involving pairwise dialect comparisons. The critical difference between the tasks is that the dialect discrimination task does not require the participants to learn ad hoc associations between phonetic features and category labels, but instead to indicate whether two talkers are from the same place or different places. Comparing across the two tasks can reveal if removing just the requirement to learn ad hoc category associations benefits young listeners and if it does, what a dialect discrimination task can reveal about the development of regional dialect representations.

The current study examined the perception of the same four American English regional dialects used by Jones et al. (2017): Midland, Northern, Southern, and New England. The study involved two tasks in a within-subject design. Experiment 1 was an identification task with ad hoc labels, as used by Girard et al. (2008) and Wagner et al. (2014). As noted above, it is not assumed that the ad hoc identification task involves merely learning to associate a new label with an existing regional dialect representation, but rather that it requires learning how phonetic variation in the speech signal is associated with place categories, which are indexed by ad hoc labels. Although some older participants may adopt the former strategy for some of the dialect contrasts, this strategy is not available for dialects that are not represented as distinct categories by adults (e.g., Midland vs. Northern American English; see Campbell-Kibler, 2012; Jones et al., 2017) or for children who do not have existing categories for the less familiar varieties. Thus, critically, none of the listeners in the ad hoc identification task needed to associate the ad hoc labels with particular regional dialects, such as “Southern” or “Northern,” and this kind of regional label was not explicitly invoked in the experiment. In the current study, young listeners’ responses on this task were not only compared to chance, as in previous work, but also interpreted in the context of adults’ responses, as even adults are not sensitive to all dialect contrasts (Clopper & Pisoni, 2006; Jones et al., 2017). This approach allows for consideration of both overall success and the trajectory of performance into adulthood.

Experiment 2 was a dialect discrimination task, like the one used with some success by Girard et al. (2008). This task required listeners to decide whether two talkers were from the same place or different places. Because dialect discrimination does not require learning ad hoc group labels, it imposes fewer processing demands than identification, which should allow us to observe children’s emerging representations of regional dialect variation. Unlike Jones et al.’s (2017) free classification task, however, it also requires overt pairwise judgments, allowing for a more direct comparison to the identification results and for a more detailed understanding of the nature of children’s developing dialect representations. Thus, the dialect discrimination task combines the strengths of the ad hoc identification task (i.e., analytical precision through pairwise comparisons) and the free classification task (i.e., dialect perception without explicit categories), providing an opportunity to obtain
converging evidence across tasks. Children’s discrimination performance was also compared to chance and interpreted in the context of adults’ performance to allow for consideration of both objective success and the developmental trajectory into adulthood for this task.

The design of both the ad hoc identification and dialect discrimination tasks required listeners to attend to the sources of variation in the speech signal that are relevant for indexing regional background. In the ad hoc identification task, listeners needed to generalize over the training materials to identify the critical features that index the dialects and then attend to those features to make their identification responses in the test phase. In the dialect discrimination task, in which listeners were always asked to compare two different talkers, listeners needed to attend to the features that index regional dialect in American English and ignore idiosyncratic differences between the two talkers. Recent work by Weatherhead, White, and Friedman (2016) demonstrates that 3- to 5-year-old children interpret a shared foreign accent as indicating shared cultural norms, such as place of origin, but not shared personal preferences for colors or games, suggesting that children at this age understand the role of language variation in marking group identity. However, for children to achieve perceptual accuracy that is comparable to adults’ performance in the dialect perception tasks in the current study, they must first acquire adult-like representations of the phonetic features that index regional background in American English and then attend to those features in these tasks.

Based on previous dialect perception research with adults (e.g., Clopper & Bradlow, 2009; Jones et al., 2017), it was predicted that for the predominantly Midwestern adult participants in this study, Southern and New England talkers would be both identifiable and discriminable from other groups, whereas Midland and Northern talkers would be easily confused with one another, leading to poor identification and discrimination. Based on Jones et al.’s (2017) results, young listeners were expected to differentiate New England talkers from the other groups first, followed later by differentiating Southern talkers from the Midwestern (Midland and Northern) talkers. Due to the additional processing demands involved in learning and remembering ad hoc labels, the dialect discrimination task, which lacked any explicit labels, was expected to reveal children’s abilities earlier than the ad hoc identification task.

2 Experiment 1: dialect identification with ad hoc labels

2.1 Methods

2.1.1 Listeners. The participants in Experiment 1 were 720 monolingual American English speakers recruited from the pool of visitors to a science center in Columbus, Ohio. They were evenly distributed among 10 age groups ranging from 4–5 years old to 50–79 years old. The age groups reflect dense sampling of children in seven two-year cohorts and more sparse sampling of adults in three cohorts of young (18–34 years), middle-aged (35–49 years), and older (50–79 years) adults. The older adult group spans a larger age range than the young and middle-aged groups, but the age distribution of these older adults is skewed young (mean = 60 years) and 70.8% of the older listeners were under 65 years old. Participants were classified as lifetime residents of the Midwest (including the Midland and Northern dialect regions), South, New England, or Other (including the Mid-Atlantic, West, or Florida) regions of the United States, following Labov, Ash, and Boberg’s (2006) proposed dialect regions. Participants who had lived in more than one of these regions were classified as Mobile. Among all participants, including relevant members of the Mobile category, 78.6% had lived in the Midwest, 21.9% had lived in the South, and 4.2% had lived in New England. While most young participants had lived only in the Midwest, older participants were more likely to have lived in multiple regions of the United States. A summary of the mean ages and residential histories for each of the 10 age groups is provided in Table 1.
Data from 214 additional individuals were excluded for the following reasons: speaking languages other than English at home \((n = 67)\); reporting no information about languages spoken \((n = 5)\); technical difficulties with the experiment software \((n = 8)\); confusion about the instructions \((n = 3)\); self-reported hearing problems, including wearing hearing aids \((n = 4)\); parental interference \((n = 1)\); and failing to complete the entire study (consisting of both this task and the task described in Experiment 2; \(n = 32\)), or failing to pass the practice session for Experiment 2 \((n = 94)\), which is described in more detail in the methods Section for Experiment 2. The 94 participants \((10.1\%)\) who failed to pass the practice session for Experiment 2 included participants in each of the 10 age groups, suggesting that this failure reflected inattention to the task in the rich science center environment where the experiment was conducted.

2.1.2 Stimulus materials. The stimulus materials for the main task were produced by 12 white female talkers from the TIMIT Acoustic–Phonetic Continuous Speech Corpus (Fisher, Doddington, & Goudie-Marshall, 1986). The talkers were all in their 20s at the time of recording. The TIMIT corpus was recorded in the mid-1980s, which means that these talkers would now be in their 50s, corresponding to the oldest listener group. Three female talkers from each of four dialect regions in the United States (Midland, Northern, Southern, and New England) were selected. The Midland region covers the lower Midwest, spanning from Ohio to Kansas and Nebraska. The Northern region refers to the upper Midwest, from western New York state to Minnesota. The Southern region includes much of the southern United States, and stretches from Virginia to Texas. New England comprises the northeastern region between Maine and eastern New York state (Labov et al., 2006). Although all of these major dialect regions are characterized by sub-regional variation, the members of each set of three talkers were selected as good representatives of their respective major dialect regions and exhibited similar dialect features in the stimulus materials.

Specifically, the stimulus materials comprised a recording of the sentence “She had your dark suit in greasy wash water all year” produced by each of the 12 talkers. This sentence was recorded by each talker in the TIMIT corpus because it exhibits a variety of regional pronunciation differences in American English. For instance, the three talkers from the South said [ɡɹiizi] rather than [ɡɹisi] for “greasy” and the three talkers from New England omitted the postvocalic /ɹ/ in the words “dark” and “year” (none of the Southern talkers were non-rhotic). To quantify dialect-based differences and ensure that these talkers reflect current patterns of variation in American English, a selection of acoustic measures was extracted from the target sentence; these measures

### Table 1. Ages and residential histories of participants in Experiments 1 and 2.

<table>
<thead>
<tr>
<th>Age range (years)</th>
<th>Midwest</th>
<th>South</th>
<th>New England</th>
<th>Other</th>
<th>Mobile</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>4–5 (mean (M = 4.6))</td>
<td>58</td>
<td>3</td>
<td>1</td>
<td>4</td>
<td>6</td>
<td>72</td>
</tr>
<tr>
<td>6–7 ((M = 6.6))</td>
<td>46</td>
<td>15</td>
<td>0</td>
<td>5</td>
<td>6</td>
<td>72</td>
</tr>
<tr>
<td>8–9 ((M = 8.5))</td>
<td>46</td>
<td>8</td>
<td>0</td>
<td>9</td>
<td>9</td>
<td>72</td>
</tr>
<tr>
<td>10–11 ((M = 10.6))</td>
<td>41</td>
<td>11</td>
<td>0</td>
<td>6</td>
<td>14</td>
<td>72</td>
</tr>
<tr>
<td>12–13 ((M = 12.5))</td>
<td>44</td>
<td>5</td>
<td>1</td>
<td>9</td>
<td>13</td>
<td>72</td>
</tr>
<tr>
<td>14–15 ((M = 14.5))</td>
<td>47</td>
<td>5</td>
<td>0</td>
<td>10</td>
<td>10</td>
<td>72</td>
</tr>
<tr>
<td>16–17 ((M = 16.4))</td>
<td>43</td>
<td>2</td>
<td>0</td>
<td>10</td>
<td>17</td>
<td>72</td>
</tr>
<tr>
<td>18–34 ((M = 24.6))</td>
<td>40</td>
<td>4</td>
<td>1</td>
<td>2</td>
<td>25</td>
<td>72</td>
</tr>
<tr>
<td>35–49 ((M = 40.6))</td>
<td>25</td>
<td>2</td>
<td>0</td>
<td>5</td>
<td>40</td>
<td>72</td>
</tr>
<tr>
<td>50–79 ((M = 60.1))</td>
<td>18</td>
<td>0</td>
<td>1</td>
<td>6</td>
<td>47</td>
<td>72</td>
</tr>
<tr>
<td>Total</td>
<td>408</td>
<td>55</td>
<td>4</td>
<td>66</td>
<td>187</td>
<td>720</td>
</tr>
</tbody>
</table>

(56.7%) (7.6%) (0.6%) (9.2%) (26.0%)
Northern talkers were expected to have raised and fronted \(/æ/ in “had” and lowered and fronted \(/\alpha/ in “wash” as part of the Northern Cities Shift (Labov et al., 2006). The \(/\alpha/ in “suit” was expected to be fronted for Midland and Southern talkers, and Southern talkers were predicted to have high proportions of voicing of the fricative in “greasy.” Southern talkers were also predicted to have relatively slow speaking rates, while New England talkers were expected to have faster speaking rates (Byrd, 1994), as well as high values for the r-lessness measure, indicating the lack of rhoticity.

Values for each of the acoustic measures were obtained via monitored automatic measurements on events hand-marked in Praat TextGrids, and the means of these measures are presented in Table 3. The patterns were generally as predicted; the only exceptions to the expected patterns were that \(/\alpha/ was not lowered for Northern talkers and that Northern talkers, rather than New England talkers, had the fastest speaking rates. For each measure, a one-way analysis of variance was performed to determine whether the acoustic–phonetic property differed significantly across talker regions. The main effect of region was significant for fricative voicing, \(F(3, 8) = 17.4, p < 0.001\), and for r-lessness, \(F(3, 8) = 17.4, p < 0.001\). Post-hoc Tukey tests revealed that Southern talkers had higher proportions of voicing of the fricative in “greasy.” Southern talkers were also predicted to have relatively slow speaking rates, while New England talkers were expected to have faster speaking rates (Byrd, 1994), as well as high values for the r-lessness measure, indicating the lack of rhoticity.

Values for each of the acoustic measures were obtained via monitored automatic measurements on events hand-marked in Praat TextGrids, and the means of these measures are presented in Table 3. The patterns were generally as predicted; the only exceptions to the expected patterns were that \(/\alpha/ was not lowered for Northern talkers and that Northern talkers, rather than New England talkers, had the fastest speaking rates. For each measure, a one-way analysis of variance was performed to determine whether the acoustic–phonetic property differed significantly across talker regions. The main effect of region was significant for fricative voicing, \(F(3, 8) = 17.4, p < 0.001\), and for r-lessness, \(F(3, 8) = 17.4, p < 0.001\). Post-hoc Tukey tests revealed that Southern talkers had higher proportions of voicing of the fricative in “greasy.” Southern talkers were also predicted to have relatively slow speaking rates, while New England talkers were expected to have faster speaking rates (Byrd, 1994), as well as high values for the r-lessness measure, indicating the lack of rhoticity.

Values for each of the acoustic measures were obtained via monitored automatic measurements on events hand-marked in Praat TextGrids, and the means of these measures are presented in Table 3. The patterns were generally as predicted; the only exceptions to the expected patterns were that \(/\alpha/ was not lowered for Northern talkers and that Northern talkers, rather than New England talkers, had the fastest speaking rates. For each measure, a one-way analysis of variance was performed to determine whether the acoustic–phonetic property differed significantly across talker regions. The main effect of region was significant for fricative voicing, \(F(3, 8) = 17.4, p < 0.001\), and for r-lessness, \(F(3, 8) = 17.4, p < 0.001\). Post-hoc Tukey tests revealed that Southern talkers had higher proportions of voicing of the fricative in “greasy.” Southern talkers were also predicted to have relatively slow speaking rates, while New England talkers were expected to have faster speaking rates (Byrd, 1994), as well as high values for the r-lessness measure, indicating the lack of rhoticity.

Values for each of the acoustic measures were obtained via monitored automatic measurements on events hand-marked in Praat TextGrids, and the means of these measures are presented in Table 3. The patterns were generally as predicted; the only exceptions to the expected patterns were that \(/\alpha/ was not lowered for Northern talkers and that Northern talkers, rather than New England talkers, had the fastest speaking rates. For each measure, a one-way analysis of variance was performed to determine whether the acoustic–phonetic property differed significantly across talker regions. The main effect of region was significant for fricative voicing, \(F(3, 8) = 17.4, p < 0.001\), and for r-lessness, \(F(3, 8) = 17.4, p < 0.001\). Post-hoc Tukey tests revealed that Southern talkers had higher proportions of voicing of the fricative in “greasy.” Southern talkers were also predicted to have relatively slow speaking rates, while New England talkers were expected to have faster speaking rates (Byrd, 1994), as well as high values for the r-lessness measure, indicating the lack of rhoticity.

Values for each of the acoustic measures were obtained via monitored automatic measurements on events hand-marked in Praat TextGrids, and the means of these measures are presented in Table 3. The patterns were generally as predicted; the only exceptions to the expected patterns were that \(/\alpha/ was not lowered for Northern talkers and that Northern talkers, rather than New England talkers, had the fastest speaking rates. For each measure, a one-way analysis of variance was performed to determine whether the acoustic–phonetic property differed significantly across talker regions. The main effect of region was significant for fricative voicing, \(F(3, 8) = 17.4, p < 0.001\), and for r-lessness, \(F(3, 8) = 17.4, p < 0.001\). Post-hoc Tukey tests revealed that Southern talkers had higher proportions of voicing of the fricative in “greasy.” Southern talkers were also predicted to have relatively slow speaking rates, while New England talkers were expected to have faster speaking rates (Byrd, 1994), as well as high values for the r-lessness measure, indicating the lack of rhoticity.

Values for each of the acoustic measures were obtained via monitored automatic measurements on events hand-marked in Praat TextGrids, and the means of these measures are presented in Table 3. The patterns were generally as predicted; the only exceptions to the expected patterns were that \(/\alpha/ was not lowered for Northern talkers and that Northern talkers, rather than New England talkers, had the fastest speaking rates. For each measure, a one-way analysis of variance was performed to determine whether the acoustic–phonetic property differed significantly across talker regions. The main effect of region was significant for fricative voicing, \(F(3, 8) = 17.4, p < 0.001\), and for r-lessness, \(F(3, 8) = 17.4, p < 0.001\). Post-hoc Tukey tests revealed that Southern talkers had higher proportions of voicing of the fricative in “greasy.” Southern talkers were also predicted to have relatively slow speaking rates, while New England talkers were expected to have faster speaking rates (Byrd, 1994), as well as high values for the r-lessness measure, indicating the lack of rhoticity.

Values for each of the acoustic measures were obtained via monitored automatic measurements on events hand-marked in Praat TextGrids, and the means of these measures are presented in Table 3. The patterns were generally as predicted; the only exceptions to the expected patterns were that \(/\alpha/ was not lowered for Northern talkers and that Northern talkers, rather than New England talkers, had the fastest speaking rates. For each measure, a one-way analysis of variance was performed to determine whether the acoustic–phonetic property differed significantly across talker regions. The main effect of region was significant for fricative voicing, \(F(3, 8) = 17.4, p < 0.001\), and for r-lessness, \(F(3, 8) = 17.4, p < 0.001\). Post-hoc Tukey tests revealed that Southern talkers had higher proportions of voicing of the fricative in “greasy.” Southern talkers were also predicted to have relatively slow speaking rates, while New England talkers were expected to have faster speaking rates (Byrd, 1994), as well as high values for the r-lessness measure, indicating the lack of rhoticity.
The stimulus materials for the training and practice sessions consisted of the same sentence produced by 12 white male talkers from the TIMIT corpus, including three talkers from each of the same four dialect regions. As in the test materials, all of the talkers were in their 20s at the time of recording, were selected to be representative of their major dialect regions, and within each dialect group, produced similar dialect features in the stimulus sentence.

2.1.3 Procedure. Each participant was seated at a personal computer station equipped with headphones in a quiet research space in the science center. Experiment 1 was conducted immediately after Experiment 2. All participants completed the two tasks in the same order so as not to encourage direct associations between the ad hoc color categories and regional dialect in the dialect discrimination task (for further detail, see the methods Section for Experiment 2). Participants completed both tasks in approximately 15–20 minutes.

Experiment 1 began with a brief training session (four trials) in which the listener heard two male talkers from each of two dialect regions associated with smiley faces of different bright colors. That is, each talker from one dialect region was introduced as, for instance, one of “the purple people,” and a purple smiley face was displayed prominently on the screen while his voice played. For the two talkers from the other dialect region, the same approach was used, but with a different color. To limit the duration of the experiment, each participant learned color associations for only two of the four regional dialects. Each of the six possible combinations of dialects was presented to 12 participants in each age group. Assignment of color to talker dialect was randomized across participants. Color pairings were piloted with colorblind individuals to ensure that the two colors were sufficiently distinguishable.

A brief practice session (two trials) followed the training session. During the practice session, the computer screen displayed two brightly colored smiley faces with the name of the color written under each one. The question “Who is talking?” appeared at the top of the screen and was also asked verbally by the experimenter for young participants. On each practice trial, the participant heard a new male talker from one of the two dialect regions and indicated which color group he belonged with. The correct answer was given after each response.

The test session was identical to the practice session, except that participants heard female voices and the correct answers were not revealed. A sample response screen from the test phase is shown in Figure 1. The test phase consisted of six randomly ordered trials (all three talkers from each of the two regions). Participants aged 8 years and older used the mouse to click on a response on the computer screen, and the response was logged by the experiment presentation software. Participants aged 7 years and younger, as well as a small number of older children who were not
comfortable with the computer interface, answered verbally or pointed to a smiley face, and the experimenter then clicked on the corresponding response on the screen. This task is described as “ad hoc identification” to emphasize that participants were not asked to provide real-world regional dialect labels like “Southern” or “New England”; rather, as in previous work with children (Floccia et al., 2009; Girard et al., 2008; Wagner et al., 2014), they learned associations between colors and the phonetic features of the dialects during the training session and were tested on these learned associations with new talkers.

2.1.4 Analysis. Given that listeners in this task chose from two fixed response options to identify the dialect of each talker, identification accuracy rates were compared against chance using two-tailed one-sample \( t \)-tests. While this approach resulted in a total of 60 tests (6 pairwise dialect combinations \( \times \) 10 listener age groups), the data in each test were entirely independent due to the between-subject design. Nonetheless, a Bonferroni correction was applied within each age group (\( \alpha = 0.0083 \)) to control for the large number of tests. In the context of this analysis, “success” is defined as performance that reliably exceeded chance.

2.2 Results

A summary of performance in the ad hoc identification task is shown in Figure 2, with asterisks indicating which accuracy rates significantly exceeded chance. The full \( t \)-test results are reported in Table A1 in the Appendix. Listeners younger than 10 years old were generally unsuccessful at identifying regional dialects, except when 6- and 7-year-olds heard talkers from New England and the Midland (“New England–Midland”). In pre-adolescence, however, dialect identification abilities quickly fell into place: 10- and 11-year-olds’ accuracy rates exceeded chance for all dialect combinations involving Southern talkers, as well as for the New England–Midland contrast. While 12- and 13-year-old listeners performed at chance when hearing talkers from New England and the South (“New England–South”), by age 14–15 years listeners matched the adult-like pattern of above-chance performance on five of the six dialect combinations. For the sixth combination, which involved Midland and Northern talkers (“Midland–North”), the dialects were not successfully identified by listeners in any age group. Finally, 50- to 79-year-old listeners declined in performance for many dialect contrasts, exceeding chance only when Midland talkers were combined with talkers from New England or the South. As shown in Figure 2, this lack of success by the oldest listener group partially reflects greater variability for some dialect contrasts (e.g., “New England–North” and “North–South”) within this listener group than within the young and middle-aged adult groups.

2.3 Discussion

The results of the ad hoc identification task clearly indicate that different regional dialects have an impact on listeners’ performance. For instance, as in Jones et al.’s (2017) study, even adult listeners in this task were unable to differentiate between Midland and Northern talkers. This result is not surprising, given that the acoustic analysis described above revealed that no acoustic–phonetic property significantly set apart Midland or Northern talkers. Further, the Northern dialect is not strongly stigmatized within Ohio, and its pronunciation patterns are interpreted by listeners in the central Ohio area as idiosyncratic details rather than as dialect features (Campbell-Kibler, 2012). However, this finding highlights the importance of considering children’s performance in the context of adults’ skills. In particular, children’s failure to accurately identify Midland and Northern talkers in this task should not be interpreted as a failure in dialect identification in general, because adults also fail with this dialect contrast.
Figure 2. Ad hoc identification accuracy (M = Midland, N = North, S = South, and NE = New England) for each age group. Error bars represent 99% confidence intervals over subject means. Asterisks represent significant differences from chance performance.

Developmental patterns offer another piece of evidence that the regional varieties under consideration are important, as the dialect contrasts in this experiment were mastered at different times. On the whole, listeners seemed to have representations of variation in speech associated with New
England and the South by adolescence, as expected based on Jones et al.’s (2017) results. However, the ad hoc identification task in the current experiment revealed more detailed information about the perception of specific pairwise combinations of regional dialects than the previous free classification task, in which listeners’ responses may have been motivated by comparisons among any or all varieties simultaneously. On the one hand, for this listener population, the New England–Midland dialect combination was easiest to differentiate: even 6- and 7-year-olds succeeded with this contrast, although this success was not consistent until age 10–11 years. On the other hand, the most difficult combination for this listener population, apart from the Midland–North contrast for which performance never exceeded chance, was New England–North, for which performance did not exceed chance until age 14–15 years.

These results therefore provide additional evidence that regional dialect identification with ad hoc category labels is very challenging for children younger than approximately 10 years old, consistent with the previous studies using this type of task (Floccia et al., 2009; Girard et al., 2008; Wagner et al., 2014). The development of children’s representations of regional dialect cannot be tracked well by a task that is simply too difficult for many young listeners. Therefore, Experiment 2 involved a different task in an attempt to better target development prior to age 10 years. The dialect discrimination task used by Girard et al. (2008) required participants to say whether two talkers were from the same place or different places. While it involved more listening per trial than the ad hoc identification task—two talkers rather than one—there was no need for listeners to learn or remember how phonetic variation is associated with ad hoc category labels. A similar dialect discrimination task was used in Experiment 2 to reduce overall task demands and allow a clearer trajectory of perceptual development to emerge.

3 Experiment 2: dialect discrimination

3.1 Methods

3.1.1 Listeners. The same 720 monolingual American English listeners who participated in Experiment 1 also completed Experiment 2 in the same experimental session.

3.1.2 Stimulus materials. The stimulus materials in the main task were the same as those used in the main task in Experiment 1. Additional stimulus materials were used for the practice task, which tested discrimination abilities unrelated to speech. To maintain the auditory modality of the task, these stimuli consisted of four ascending musical scales: one in each of two musical keys played by each of two synthesized musical instruments (piano and tuba).

3.1.3 Procedure. Experiment 2 was conducted immediately before Experiment 1 in the same quiet research space in the science center, and involved the two regional dialects that the participant would not hear in Experiment 1. A brief practice session preceded the task. In the practice session, the computer screen displayed a pair of smiley faces of the same bright color accompanied by the word “same,” and a separate pair of smiley faces, each of a different bright color, accompanied by the word “different.” The question “Did you hear two songs played on the SAME instrument or on DIFFERENT instruments?” appeared at the top of the screen and was also asked verbally by the experimenter for young participants. For each practice trial \((n = 4)\), the participant heard two musical scales in different keys and judged whether they were played on the same musical instrument or on different musical instruments. As in Experiment 1, participants aged 8 years and older responded by clicking on a response on the computer screen, while participants aged 7 years and younger responded verbally or by pointing, and the experimenter clicked on the response on the
During the test session, the computer screen was similar to the screen displayed during the practice session, except that the question at the top was “Did you hear two people from the SAME place or from DIFFERENT places?” This question was also asked verbally by the experimenter for young participants. A sample of this screen is shown in Figure 3. For each test trial, the participant heard two talkers produce the stimulus sentence and were asked to indicate if the talkers were from the same place or different places. Each of the six pairwise combinations of dialects involved 15 test trials (i.e., all 15 possible pairings of the six talkers), six of which were same-dialect pairings and nine of which were different-dialect pairings. Same-talker pairings were not included because the question the participants were asked to answer involved comparing two people. Trial order was randomized separately for each participant, as was talker order within each trial.

Brightly colored smiley faces were shown in both the ad hoc identification and dialect discrimination tasks, but only the ad hoc identification task in Experiment 1 imposed a direct association between color and region; in the dialect discrimination task in Experiment 2, two smiley faces of the same bright color represented a “same” response regardless of which region the talkers were believed to be from. As noted above, all participants completed Experiment 2 before Experiment 1 so as not to encourage direct associations between color categories and regional dialect in the dialect discrimination task. To further differentiate the two tasks, for each participant, the colors used in the two tasks were different, with purple and yellow faces in the ad hoc identification task and blue and orange faces in the dialect discrimination task, or vice versa, by random assignment.

3.1.4 Analysis. The analysis for Experiment 2 was identical to the analysis for Experiment 1.

3.2 Results

A summary of performance in the dialect discrimination task is shown in Figure 4, with asterisks indicating which accuracy rates significantly exceeded chance. The full $t$-test results are reported in Table A2 in the Appendix. Overall, success was evident earlier in the dialect discrimination task than in the ad hoc identification task. By 4–5 years old, listeners successfully discriminated New England talkers from both Midland and Northern talkers, and continued to do so consistently
through adulthood. Moreover, at age 6–7 years, performance exceeded chance for the New England–South and North–South contrasts, although neither result persisted; for the former, consistent success began at 10–11 years old, and for the latter, at 12–13 years old. In fact, at 12–13

![Dialect discrimination accuracy](image)

**Figure 4.** Dialect discrimination accuracy (M = Midland, N = North, S = South, and NE = New England) for each age group. Error bars represent 99% confidence intervals over subject means. Asterisks represent significant differences from chance performance.
years old, listeners exhibited adult-like accuracy for all dialect contrasts, including successful performance on the Midland–South contrast, but generally not on the Midland–North contrast. Two isolated age groups (10- to 11-year-olds and 16- to 17-year-olds) exceeded chance performance for the Midland–North contrast, however. Finally, in this task, listeners in the oldest age group did not differ from other adults in their response patterns.

3.3 Discussion

Overall, listeners were successful in this dialect discrimination task at younger ages than listeners in the ad hoc identification task in Experiment 1, despite the fact that the two experiments used the same dialects—indeed, the exact same speech tokens—and shared superficial task demands (e.g., both involved a two-alternative forced choice between colored smiley faces). In particular, 4- and 5-year-olds successfully discriminated New England talkers from Midland and Northern talkers, although children in Experiment 1 could not consistently perform ad hoc identification of these contrasts until 10–11 years old. Moreover, listeners’ dialect discrimination responses were adult-like across the range of dialect combinations by 12–13 years old, which is younger than in the ad hoc identification task, where this milestone was achieved at 14–15 years. Considering the dialect combinations individually, in most cases listeners were able to discriminate talkers from two dialects before they could assign ad hoc color labels to the talkers from those dialects, as expected. The only exceptions to this trend were the Midland–South and North–South combinations, for which listeners succeeded in ad hoc identification at 10–11 years old, but performed at chance in the dialect discrimination task until 12–13 years old. However, 10- and 11-year-old listeners were nearly successful at discriminating Northern from Southern talkers ($p < 0.05$ before Bonferroni correction), suggesting overlap in the timing of success across tasks for this dialect combination.

The developmental pattern of performance in the dialect discrimination task was broadly consistent with the trajectory observed by Jones et al. (2017) using a free classification task with the same regional varieties. The 4- and 5-year-old children in the current study successfully discriminated New England talkers from Midland and Northern talkers, and children of the same age in Jones et al.’s (2017) study considered New England talkers to be perceptually dissimilar from other talkers. Subsequent advancements later in development were also similar, although the developmental time frames diverged somewhat across studies. Jones et al. (2017) observed a combined group of New England and Southern talkers at 6–7 years old and separate groups of New England and Southern talkers by 8–9 years old. In contrast, dialect discrimination involving Southern talkers was not consistently successful until 12–13 years old. Together, the results of both studies suggest that Midwestern children discriminate New England talkers from Midwestern (Midland and Northern) talkers earlier than they discriminate Southern talkers from Midwestern talkers.

4 General discussion

The results from the two experiments showed how children gradually develop adult-like competence in the domain of regional dialect perception, as well as the ways in which children’s performance is constrained by the dialects themselves and by task demands. With respect to differences across dialects, the primarily Midwestern listeners tested in the current study showed the earliest success with contrasts involving the New England variety and limited success at any age with the Midland–North combination. With respect to differences across tasks, the results suggest that a cognitively difficult task, such as the ad hoc identification task in Experiment 1, might mask a listener’s competence in dialect perception, especially at younger and older ages.
4.1 Differences across regional dialect contrasts

The first core research question was whether regional dialect contrasts are differentially difficult, and the results of the two experiments clearly demonstrate that they are. With the exception of the Midland–North combination (which is discussed separately below), adults were successful at both dialect discrimination and, at least until age 50 years, ad hoc identification for all dialect contrasts. Adult-like abilities, however, emerged gradually over the course of childhood, and were highly influenced by the specific dialects being examined. The New England–Midland contrast showed the earliest success in Experiment 1, and in Experiment 2, the youngest listeners succeeded with both this contrast and with the New England–North contrast. Success with contrasts involving the Southern dialect emerged after these successes with contrasts involving the New England dialect. These results are consistent with Jones et al.’s (2017) findings that for listeners in the Midwestern United States, New England talkers are the first to form a cohesive, separate group in children’s free classification responses, followed by the separation of Southern talkers from Midwestern and then New England talkers.

Given that regional labels were not explicitly invoked in either of the current experiments, the early acquisition of the perceptual contrast between New England talkers and Midwestern talkers, followed by the acquisition of the perceptual contrast between Southern talkers and other talkers, likely reflects a combination of the phonetic distances among the varieties (Floccia et al., 2009; Wagner et al., 2014) and the listeners’ experience with the varieties (Baker et al., 2009; Clopper & Pisoni, 2006). First, the pronunciations of the New England and Southern talkers in the present investigation were set apart by the acoustic measures of r-lessness and fricative voicing, respectively. These large and consistent deviations from other talkers were potentially easy to perceive and represent as socially meaningful even for young listeners, and r-lessness may have been an especially good index of regional dialect, leading to the earliest success with contrasts involving New England talkers. Second, although detailed information about the listeners’ exposure to different American English dialects was not obtained, from the residential histories reported in Table 1 it seems likely that many listeners, and especially many young listeners, were most familiar with the dialects of the Midwest (Midland and North) and less familiar with New England talkers and possibly even Southern talkers. Hearing a familiar dialect (i.e., Midland or Northern) alongside an unfamiliar one with clearly perceptible phonetic differences (i.e., New England or Southern) might have facilitated listeners’ discrimination and identification performance. As noted above, this local anchoring hypothesis has been advocated by previous researchers using ad hoc identification tasks (Floccia et al., 2009; Girard et al., 2008; Wagner et al., 2014), who have suggested that children’s success depends on the inclusion of a very phonetically different unfamiliar speech variety—in those studies, a foreign-accented variety—as well as the presence of the child’s native variety (see especially Wagner et al., 2014). The current results lend support to the idea that children benefit from the presence of a high versus low familiarity contrast that is supported by substantial phonetic distance; however, they also demonstrate that the less familiar variety can be a native one. The specific regional varieties used in previous studies may simply have not been sufficiently unfamiliar or phonetically different to facilitate children’s perception. The previously observed distinction in performance between regional and non-native varieties is therefore likely accidental, and a sufficiently familiar and/or phonetically similar non-native variety (e.g., the speech of a highly proficient native Spanish speaker of English in many parts of the United States) would be predicted to be difficult for children to distinguish from their native variety of American English.

Beyond specific exposure to particular varieties, overall exposure to variation may also contribute to performance in dialect perception tasks. For example, Evans et al. (2016) observed higher accuracy in the regional dialect condition of their ad hoc identification task for
multilingual children than for monolingual children. They interpreted this result in terms of exposure to variation: multilingual children have more exposure to language variation in their environment and may therefore exhibit better performance in tasks that require them to associate phonetic variation with social groups. To explore the potential role of general exposure to variation on dialect perception in the current study, a series of analyses on the subset of participants who had lived in multiple different dialect regions (i.e., the Mobile participants in Table 1) was conducted. The between-subject manipulation of the pairwise dialect comparisons in these experiments did not permit a full analysis of the effect of dialect exposure on performance, so in this analysis the Mobile participants were collapsed into three age groups: 54 10- to 17-year-olds, 65 18- to 49-year-olds, and 47 50- to 79-year-olds (see Table 1), with 5–14 listeners per pairwise dialect comparison in each task. Too few children under the age of 10 years had lived in multiple different regions for inclusion in this analysis. The results did not reveal evidence of better ad hoc identification or dialect discrimination performance for the Mobile subset than for the full set of participants, at least partially due to the relatively small numbers of participants in some analyses. An exploration of the potential role of overall dialect exposure on the development of dialect perception is therefore left for future work.

Unlike the results involving the New England and Southern dialects, phonetic distance and dialect familiarity do not provide a straightforward explanation for the Midland–North contrast. Adults as well as children failed on this contrast in both experiments, suggesting that the difficulty with this dialect combination results from the representations of the dialects themselves rather than from developmental factors or task demands. Specifically, adults who have substantial exposure to both of these varieties interpret the phonetic differences between them as idiosyncratic pronunciation patterns rather than as patterns reflecting regional affiliation (Campbell-Kibler, 2012), suggesting the lack of association between their respective phonetic features and regional categories. Despite this general failure to explicitly perceptually distinguish the Midland and Northern dialects, however, in Experiment 1, identification of the New England–Midland contrast was consistent by 10–11 years, while the New England–North contrast was the last to be acquired, with performance not exceeding chance until 14–15 years. Similarly, performance in Experiment 1 on contrasts involving Northern talkers, but not Midland talkers, fell below chance for the 50- to 79-year-old group. Thus, although Midland and Northern talkers were not identifiably different, they were also not perceived identically and facilitated pairwise identification with other dialects in different ways. That is, these results reveal that listeners can perceive and represent phonetic differences between the Midland and Northern varieties, even if those differences generally cannot be harnessed in explicit identification and discrimination tasks involving the two varieties themselves. Indeed, representational differences between these two dialects must exist for the 10- to 11-year-olds and 16- to 17-year-olds who successfully discriminated Midland from Northern talkers in Experiment 2 (see also Jones et al., 2017). To the extent that perception of the Midland–North contrast depends on associating their distinct phonetic features with distinct social groups, perception of dialect combinations involving one or both of these dialects might be expected to shift over time with changes in the social meaning of these varieties.

4.2 The development of representations of regional dialects

The second core research question was how children’s representations of how speech indexes people as being from different places develops over time. The results of Experiment 2 showed that for some dialect contrasts, 4- and 5-year-olds can use information in the speech signal that indexes a talker’s region of origin to successfully discriminate talkers as being from different places. Together
with Jones et al.’s (2017) findings, this result provides evidence from the youngest children examined attesting to successful regional dialect perception in a task that is explicitly about place: children as young as 4 years old understand that language variation can in principle index group identity. Although similar results may be obtained by asking children about aliens (Floccia et al., 2009), families (Wagner et al., 2014), or shared cultural norms (Weatherhead et al., 2016), the current results further reveal young children’s ability to attend to relevant dimensions of variation to identify groups of talkers, in this case defined in terms of place, while ignoring irrelevant dimensions related to idiosyncratic talker variability or other group identities.

Despite early success, however, regional dialect perception skills continue to develop throughout childhood, and do not look fully adult-like until well into adolescence. The results from this study and from Jones et al.’s (2017) study suggest that the development of regional dialect representations takes a decade or more, beginning at 4–5 years old (or perhaps earlier) and continuing through the teenage years. Children do not exhibit adult-like success in dialect discrimination until age 12–13 years, in ad hoc identification until age 14–15 years, and in free classification until age 16–17 years. That is, understanding phonetic variation and how it indexes information about a talker’s region of origin is a protracted and complex process. Given that regional dialect perception requires detailed knowledge about both geography and language, the long time-course of its development is perhaps not surprising. Indeed, even abstract geographical knowledge is not well-developed until age 10–11 years (Jahoda, 1963, 1964; Piaget & Weil, 1951), and it takes time for children to learn about the specific geographical constructs that are culturally relevant to them. Similarly, Labov (1964) proposed that the development of sociolinguistic competence continues throughout childhood and adolescence as children learn about the social constructs that are relevant to them and how they are marked linguistically in their speech community. At the same time, personal experience listening to talkers from different places, and even traveling to or moving to new locations, serves to help children develop robust representations of regional linguistic variation (see also Williams et al., 1999). Thus, although children have a basic understanding of the link between language variation and place as early as 4–5 years old, they continue to build representations of the culturally relevant dialect categories and the linguistic features that index them over the next decade into late adolescence.

4.3 Approaches to testing regional dialect perception

The final core research question was methodological, and asked how children’s performance was influenced by task demands. Previous work using ad hoc identification tasks suggested that 5- to 6-year-old children could not successfully identify regional varieties (Floccia et al., 2009; Girard et al., 2008; Wagner et al., 2014; cf. Evans et al., 2016). The results of the current study support those findings: children in Experiment 1 did not succeed with any contrasts until 6–7 years old, and were not consistently successful until 10–11 years old. However, this task may be too difficult to be sensitive to these young children’s actual competence in this domain because 4- to 5-year-old children demonstrated some success in the dialect discrimination task in Experiment 2, as well as in Jones et al.’s (2017) free classification task. Given the similarity in the developmental patterns between dialect discrimination (Experiment 2) and free classification (Jones et al., 2017) in contrast to ad hoc identification (Experiment 1), the critical factor that appears to increase task difficulty is the assignment of an explicit ad hoc category label to the talkers of each dialect.

The substantial cognitive load involved in remembering the color/dialect associations may also account for the oldest listeners’ decline in performance, given that age-related cognitive decline in working memory begins in young adults and continues throughout the age range examined in the
current study (Park & Reuter-Lorenz, 2009; Salthouse, 2009; Verhaeghen & Salthouse, 1997). It is also possible, however, that the oldest listeners’ poorer performance reflects unreported hearing loss. An examination of the subset of older listeners under age 65 years (n = 51) revealed qualitatively similar results to the analysis of all 72 listeners in the oldest group, suggesting that the variability and decline in performance observed for this oldest listener group in the ad hoc identification task was not due to especially poor performance by listeners who were 65 years old and older. Thus, questions remain about what makes the ad hoc identification task so difficult, but given that older adults also appear to find it challenging, it is likely that the source of the difficulty lies in some aspect of general cognitive processing, such as working memory.

Although consistent success in the ad hoc identification task was not observed until 10 years old, it is important to note that in principle, the task is not too difficult for 5- and 6-year-olds. In previous studies using this task, although children generally failed to distinguish regional dialect contrasts, they consistently succeeded with native versus foreign accents (Floccia et al., 2009; Girard et al., 2008; Wagner et al., 2014). Thus, children’s past and present failures in ad hoc identification can be attributed to the use of difficult contrasts (which can nonetheless be distinguished in easier tasks) combined with a difficult task (which can nonetheless lead to success with easier contrasts).

5 Conclusions

The goal of this investigation was to explore children’s developing representations of phonetic variation in speech and how it indexes talkers’ regions of origin. The results indicate that children begin to understand the connection between language variation and group identity, such as regional background, in their preschool years. That is, by 4–5 years old, children know in principle that phonetic variation can reflect a person’s place of origin, and they have the cognitive capacity to discriminate some dialect contrasts. However, it takes an additional decade for them to master the details of the culturally relevant groups and the attendant phonetic variation in their native language, leading to adult-like patterns of ad hoc identification and dialect discrimination performance in adolescence.

While the knowledge exhibited by the listeners in the current study was specific to the American Midwest, the implications of these results extend beyond this particular set of American English dialects. Fundamentally, these findings demonstrate that given the right task and the right language varieties, even preschool-aged children can demonstrate their knowledge about how phonetic variation indexes regional background. It appears that the right task is one that requires little to no explicit labeling, even ad hoc labels, and that the right varieties are those that involve substantial differences in phonetic realization and/or familiarity. Finally, a protracted trajectory of development in dialect perception might be expected across languages and cultures, although the precise timing of the trajectory may vary as a function of children’s exposure to linguistic variation and its social meaning in the local speech community.

Acknowledgements

The authors thank Eryn Ahlers, Emily Behm, and Liz Nugent for assistance with data collection.

Funding

This work was supported by a seed grant from the Center for Cognitive and Brain Sciences at Ohio State University and by an Undergraduate Research Award from the Buckeye Language Network at Ohio State University.
References


### Appendix

**Table A1.** t-values from tests against chance for ad hoc identification accuracy by listener age and regional dialect contrast condition (M = Midland, N = North, S = South, and NE = New England).

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>NE–M</td>
<td>3.08</td>
<td>3.95</td>
<td>2.68</td>
<td>3.78</td>
<td>7.24</td>
<td>4.41</td>
<td>17.38</td>
<td>6.20</td>
<td>5.06</td>
<td>4.30</td>
</tr>
<tr>
<td>NE–N</td>
<td>0.22</td>
<td>0.83</td>
<td>0.94</td>
<td>2.24</td>
<td>2.11</td>
<td>5.63</td>
<td>8.02</td>
<td>4.08</td>
<td>4.51</td>
<td>2.51</td>
</tr>
<tr>
<td>NE–S</td>
<td>0.71</td>
<td>0.67</td>
<td>1.15</td>
<td>4.42</td>
<td>1.13</td>
<td>5.23</td>
<td>4.00</td>
<td>8.07</td>
<td>3.26</td>
<td>2.24</td>
</tr>
<tr>
<td>M–S</td>
<td>0.52</td>
<td>1.15</td>
<td>1.57</td>
<td>4.16</td>
<td>5.33</td>
<td>16.58</td>
<td>7.39</td>
<td>10.34</td>
<td>7.24</td>
<td>10.38</td>
</tr>
<tr>
<td>N–S</td>
<td>−1.48</td>
<td>1.11</td>
<td>2.92</td>
<td>3.22</td>
<td>3.59</td>
<td>7.24</td>
<td>10.56</td>
<td>10.38</td>
<td>16.26</td>
<td>1.71</td>
</tr>
<tr>
<td>M–N</td>
<td>0.46</td>
<td>0.27</td>
<td>0.00</td>
<td>1.00</td>
<td>0.46</td>
<td>1.59</td>
<td>−0.71</td>
<td>0.00</td>
<td>2.14</td>
<td>0.00</td>
</tr>
</tbody>
</table>

All degrees of freedom = 11; boldface type t-values are significant after Bonferroni correction (p < 0.0083); italicized t-values are significant before correction (p < 0.05).

**Table A2.** t-values from tests against chance for dialect discrimination accuracy by listener age and regional dialect contrast condition (M = Midland, N = North, S = South, and NE = New England).

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>NE–M</td>
<td>3.77</td>
<td>4.53</td>
<td>9.03</td>
<td>6.03</td>
<td>5.52</td>
<td>10.16</td>
<td>7.92</td>
<td>10.19</td>
<td>16.42</td>
<td>9.56</td>
</tr>
<tr>
<td>NE–N</td>
<td>4.93</td>
<td>5.82</td>
<td>5.16</td>
<td>4.59</td>
<td>6.98</td>
<td>7.06</td>
<td>6.60</td>
<td>6.53</td>
<td>6.92</td>
<td>4.62</td>
</tr>
<tr>
<td>NE–S</td>
<td>2.27</td>
<td>6.38</td>
<td>2.54</td>
<td>4.39</td>
<td>7.36</td>
<td>5.27</td>
<td>4.00</td>
<td>5.65</td>
<td>6.07</td>
<td>8.51</td>
</tr>
<tr>
<td>M–S</td>
<td>1.35</td>
<td>2.50</td>
<td>1.86</td>
<td>2.00</td>
<td>4.29</td>
<td>7.36</td>
<td>7.36</td>
<td>6.19</td>
<td>5.57</td>
<td>4.87</td>
</tr>
<tr>
<td>N–S</td>
<td>−0.41</td>
<td>4.14</td>
<td>2.40</td>
<td>3.04</td>
<td>5.69</td>
<td>4.07</td>
<td>6.86</td>
<td>5.83</td>
<td>3.88</td>
<td>3.68</td>
</tr>
<tr>
<td>M–N</td>
<td>−1.07</td>
<td>2.55</td>
<td>0.78</td>
<td>3.27</td>
<td>0.93</td>
<td>1.30</td>
<td>3.25</td>
<td>0.16</td>
<td>0.81</td>
<td>−0.91</td>
</tr>
</tbody>
</table>

All degrees of freedom = 11; boldface type t-values are significant after Bonferroni correction (p < 0.0083); italicized t-values are significant before correction (p < 0.05).