



Research Article

The development of dialect classification across the lifespan

Zack Jones*, Qingyang Yan, Laura Wagner, Cynthia G. Clopper

*The Ohio State University, 108A Ohio Stadium East, 1961 Tuttle Park Place, Columbus, OH 43210, USA*

ARTICLE INFO

Article history:

Received 27 April 2016

Received in revised form

10 October 2016

Accepted 14 November 2016

Keywords:

Dialect perception

Free classification

Sociolinguistic development

ABSTRACT

The current study investigated the developmental trajectory of listeners' ability to perceive regional dialect variation in American English using a free classification paradigm. Listeners ranged in age from 4 years old through late adulthood. In two experiments, listeners sorted a set of talkers into groups based on where the talkers were from. Talkers from four regional dialects of American English (Midland, New England, Northern, Southern) were presented. The results showed that some of even the youngest listeners (4- to 5-year-olds) were able to separate New England talkers from talkers from other regions, providing evidence that they perceived dialect variation and exhibited some level of sociolinguistic competence. Moreover, children's dialect perception abilities showed incremental improvement through childhood and adolescence, with major developmental improvements in dialect classification accuracy observed in middle childhood (between the ages of 7 and 8 years), late childhood (between 11 and 12 years), and in the teenage years (between 15 and 16 years), when adult-like abilities were reached. The overall results have implications for the experimental methods suitable to gauge young children's sociolinguistic knowledge and for our understanding of the time course of children's development of sociolinguistic competence, particularly their acquisition of the social significance of linguistic variation.

© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

Adult listeners use acoustic-phonetic information in the speech signal to identify talkers and orient themselves in relation to those talkers along social dimensions (Giles, 1970; Purnell, Idsardi, & Baugh, 1999). They are able to identify a number of social characteristics of a talker, such as his or her social class (Labov, 1972), age (Koops, Gentry, & Pantos, 2008; Ptacek & Sander, 1966), and geographic background (Preston, 1993). Children in the process of learning their native language must learn how specific speech patterns are meaningful along these various social dimensions, in addition to acquiring the rest of their linguistic competence. In an early proposal about the development of sociolinguistic competence, Labov (1964) laid out a multi-stage developmental path that was quite protracted, with the final stages not being achieved until adolescence. Results from children's production of dialect features (Chambers, 1992; Jacewicz, Fox, & Salmons, 2011; Patterson, 1992; Roberts & Labov, 1995; Roberts, 1997) and their social evaluation of sociolinguistic variation (Giles, Harrison, Creber, Smith, & Freeman, 1983; Kinzler & DeJesus, 2013) have offered both supporting and conflicting evidence for the stages that Labov (1964) proposed. However, no single study has involved the systematic evaluation of children's sociolinguistic abilities from early childhood through adolescence, which is necessary to comprehensively examine the developmental stages proposed by Labov (1964).

The focus of the current study is the perception of one sociolinguistic variable, regional dialect. Adults are reasonably good at identifying talkers of different regions (Clopper & Pisoni, 2004a; Van Bezooijen & Gooskens, 1999; Yan, 2015), and they can use this information to place talkers into groups based on regional background (Clopper & Pisoni, 2007). However, previous work has found that children younger than 7 years old cannot reliably group talkers by regional dialect (Floccia, Butler, Girard, & Goslin, 2009; Girard, Floccia, & Goslin, 2008; Wagner, Clopper, & Pate, 2014) and little is known about the development of dialect perception through childhood and adolescence. The goal of the current study was to explore the development of listeners' abilities to group a set of unfamiliar talkers by dialect. Dialect perception by children spanning an age range from preschool through late adolescence was

* Correspondence to: Department of Linguistics, The Ohio State University, 108A Ohio Stadium East, 1961 Tuttle Park Place, Columbus, OH 43210, USA.
E-mail address: jones.5028@osu.edu (Z. Jones).

examined to assess the extent to which children's developing regional dialect perception follows the major stages suggested by Labov (1964).

1.1. Dialect perception by adult listeners

Previous research has shown that adult listeners can categorize talkers of their native language by regional dialect with above-chance accuracy. For example, Van Bezooijen and Gooskens (1999) found that native Dutch-speaking adults from the Netherlands were able to identify Dutch talkers by regional dialect and that British listeners could identify UK English talkers by local regional dialect. A study conducted in the Enshi prefecture in China showed that adult Mandarin listeners were able to accurately classify twelve talkers by county origin (Yan, 2015). Adult American English listeners can categorize fellow American English talkers with above-chance accuracy, although they do so only with broad regional category distinctions (e.g., Clopper & Pisoni, 2004b; Preston, 1993). Specifically, the dialect categories revealed by perceptual dialect categorization studies do not always correspond to the dialect divisions described in sociolinguistic or dialectology studies based on speech production. For example, Preston (1993) found that listeners categorized American English-speaking talkers from the eastern United States on the basis of a north/south distinction, despite documented evidence for distinct Northern, Midland, and Southern dialects in the target region (Labov, Ash, & Boberg, 2006; cf. Davis & Houck, 1992). Similarly, Clopper and Pisoni (2004b) found that listeners consistently categorized American English-speaking talkers from the TIMIT corpus (Fisher, Doddington, & Goudie-Marshall, 1986) into three dialect groups that did not directly correspond to the TIMIT dialect labels. Although the selected TIMIT talkers were assigned to one of six regional dialects (New England, North, North Midland, South, South Midland, or West), presumably based on contemporary American dialectology research (e.g., Carver, 1987), the listeners in Clopper and Pisoni's (2004b) study categorized talkers into three broader groups, which roughly corresponded to a New England and Northern group, a Southern group, and a Midland and Western group.

Whereas the dialect categorization studies described above (and many other such studies) presented listeners with forced-choice classification tasks, Clopper and Pisoni (2007) used a free classification paradigm to investigate how differences in individual listeners' experiences and backgrounds influenced patterns of dialect perception. Their results showed that individually, many American English listeners were able to make more fine-grained distinctions about dialect group membership than the results from forced-choice categorization tasks had suggested. However, a cluster analysis of grouping responses across listeners showed that the broad dialect groups they perceived paralleled the pattern observed by Clopper and Pisoni (2004b) for the same set of TIMIT talkers: a New England group, a Southern group, and a general Midwest/West group, composed of North Midland, Northern, and Western talkers. The results of a second free classification experiment with a set of talkers from the Nationwide Speech Project corpus (Clopper & Pisoni, 2006) showed that even in the absence of accuracy differences, patterns of dialect perception were different between listeners from the Northern dialect region and listeners from the Midland dialect region (Clopper & Pisoni, 2007). Specifically, listeners from the Northern dialect region perceived Northern talkers as more similar to Midland, New England, and Western talkers, whereas Midland listeners perceived Northern talkers as most similar to Mid-Atlantic talkers. (Both talker and listener dialect regions in Clopper and Pisoni's (2007) study were defined following Labov et al.'s (2006) dialect divisions.) These results are consistent with work showing that listener residential history also affects forced-choice dialect categorization (Baker, Eddington, & Nay, 2009; Clopper & Pisoni, 2004b; Preston, 1993).

Taken together, the previous work demonstrates that adult listeners are generally able to classify talkers of their native language by regional dialect with reasonable accuracy, although listener dialect background may influence detailed grouping patterns. Relevantly for the current study, in which the majority of participants were from the Midwestern United States, results of multiple studies have suggested that New England and Southern talkers are perceived as particularly distinctive by adult Midwestern listeners, whereas Midland, Northern, and Western talkers are perceptually more similar to each other, often grouped together to varying degrees (Clopper & Pisoni, 2004b, 2007; Preston, 1993). The current study examined the developmental trajectory of children's ability to classify talkers by regional dialect, and when children acquire adult-like dialect perception skills.

1.2. Dialect perception by children

Labov (1964) proposed that children go through multiple stages on their way to developing adult-like sociolinguistic competence. Before the age of about 5 years old, he suggested that children are primarily focused on learning the basic grammar of their language from their parents, and their productive dialect use is expected to match their parents' dialect (i.e., their primary exposure source). Once children begin to interact with their peers in the neighborhood and at school (Labov's "Vernacular" stage), their own dialect use shifts to match their peer group, although this shift was argued to happen without the child's conscious awareness. Only as children age into early adolescence (Labov's "Social Perception" stage) do they come to explicitly appreciate the social significance of dialect variation among their friends and other talkers. More advanced skills in stylistic variation and conformity to a standard are acquired even later. Social context and familiarity, however, were also proposed by Labov (1964) as cross-cutting currents in children's sociolinguistic development. For example, in his examination of children's recognition of standard forms, he found that children from higher socio-economic backgrounds typically recognized the "adult prestige norm" at an earlier age than children from lower socio-economic backgrounds.

Labov (1964) was primarily interested in linguistic variation related to social class, but he framed his argument generally enough to apply to sociolinguistic variation of other types, including regional dialect variation. There is relatively little previous research on children's categorization of regional dialects, but the existing data points to protracted development throughout childhood and

adolescence. For example, Williams, Garrett, and Coupland (1999) tested both English-speaking adolescents and adults from Wales in a forced-choice dialect categorization task. The adolescent listeners achieved 30% accuracy in categorization, and the adult listeners achieved 52% accuracy. The adolescents were presumably in Labov's Social Perception stage because they demonstrated competence with this sociolinguistic task but did not yet have fully adult-like abilities. Moreover, echoing Labov's (1964) proposal, Williams et al. (1999) argued that what develops between adolescence and adulthood is a greater familiarity with various regional dialects. In this case, the adults were argued to have greater familiarity with Welsh English than the adolescents, likely stemming from their experience and travel and a more developed understanding of the regional divisions in Wales.

Most of the other previous studies in this domain have looked at dialect perception by children before Labov's Vernacular stage or children who have just transitioned into it. Investigating regional dialect perception with children this young, however, poses special problems as the core task (i.e., grouping talkers in some way by regional dialect) presupposes an explicit awareness of the often complex relationships between the indexical dialect markers that talkers produce and the regions that talkers come from. Thus, these previous studies have used tasks that involve first building implicit social categories and then testing listeners on those categories (Floccia et al., 2009; Girard et al., 2008; Wagner et al., 2014). For example, Girard et al. (2008) trained French-speaking 5- to 6-year-old children with novel talker categories based on talker team color and talker dialect. Color categories were implicitly associated with regional dialects through training in which talkers from different colored teams spoke different regional dialects, and children were tested on their ability to categorize novel talkers into the established color-based talker groups. In one experiment, the talkers spoke with either the children's home dialect or an unfamiliar regional dialect; in this case, children were unable to categorize the novel talkers reliably. This failure is consistent with the predictions for children in Labov's Vernacular stage, who have had exposure primarily to the dialects of their parents and their newly encountered peers and have limited conscious sociolinguistic awareness. Comparable studies by Floccia et al. (2009) with British English-speaking children and Wagner et al. (2014) with American English-speaking children similarly found that 5-year-olds could not categorize talkers in a home dialect versus regional dialect condition. Moreover, Edwards et al. (2014) found that just 53% of 4- to 8-year-old African-American English (AAE)-speaking children reached above-chance levels of accuracy when asked to categorize talkers as either AAE speakers or Mainstream American English (MAE) speakers. Collectively, these results seem to support the idea that preschool and early elementary-aged children have quite limited sociolinguistic competence, that what they know is tied tightly to their home dialect, and that they are largely unable to perceive dialect variation or use it to categorize talkers.

However, even with these young children, the picture is not completely clear. In a second experiment conducted by Girard et al. (2008), the talkers spoke with either the child's home dialect or a foreign-accented variety; in this case, the 5- and 6-year-old children successfully categorized the novel talkers (see also Wagner et al., 2014). This result raises the possibility that the extent to which children's sociolinguistic skills can be revealed in a categorization task may depend on the nature of the variation being examined, and that children may have begun to acquire the social meaning of linguistic forms even in the early Vernacular stage. Moreover, as children get just slightly older, there is some evidence that their abilities improve: 7-year-old listeners in Floccia et al.'s (2009) study had some success in categorizing talkers using a home dialect versus regional dialect contrast, and some of the older children in Edwards et al.'s (2014) study were able to reliably distinguish AAE talkers from MAE talkers. Furthermore, the work on children's development of social evaluations has shown that children as young as 5 and 6 years old show preferences for talkers from their local community, implicitly suggesting that children of this age can already distinguish talkers of different dialect backgrounds (Kinzler & DeJesus, 2013). These results are unexpected if children at these young ages truly lack an appreciation of dialect variation, as Labov (1964) suggested.

Thus, before concluding that children in early and middle childhood lack basic sociolinguistic competence, and specifically that children at these ages are unable to perceive regional dialect variation, it is worth considering whether the children in the previous studies were tested with optimal methods for their sociolinguistic knowledge and cognitive abilities. One reason that young children may have difficulty with dialect categorization is their limited phonological working memory abilities, in particular, their inability to retain unfamiliar speech sequences (Gathercole, Willis, Baddeley, & Emslie, 1994). The forced-choice categorization tasks that have been used previously have typically required children not only to remember each talker's speech, but also to link that speech to newly built talker categories (e.g., Floccia et al., 2009; Girard et al., 2008; Wagner et al., 2014; cf. Edwards et al., 2014). For adults, socio-indexical information is processed in parallel with linguistic information (Mullennix & Pisoni, 1990), and thus the two types of information share working memory capacity. If young children have difficulty remembering the details of novel phonological strings, it is perhaps unsurprising that they do not have the processing resources to spare for connecting those strings to novel social categories. The failure of 5-year-olds to succeed in previous dialect categorization tasks therefore may not reflect children's inability to perceive the relevant sociolinguistic information in the speech signal, but rather their inability to make use of the information in the speech signal while also processing the relevant social dimensions.

1.3. The current study

The current study explored the development of regional dialect classification from childhood through adulthood with four white American English regional dialects using the free classification paradigm developed by Clopper and Pisoni (2007; see also Clopper (2008)). By using a wide listener age range, we were able to examine children's competence across a long developmental time period and determine when listeners' abilities become adult-like in this particular domain of sociolinguistic competence and whether children's abilities improve along the trajectory laid out by Labov (1964). By using a range of dialects, moreover, the study allowed for stronger comparisons with the existing adult and child data, which involve a number of different varieties. Finally, by using the free

classification paradigm, the youngest children included in the study were given a better chance to succeed in a task gauging dialect perception. The free classification task offers listeners the opportunity to group talkers with more individual interpretation both by allowing them to make as many or as few groups as they wish and by allowing them to group individual talkers with any other talker(s); thus, it does not require children to learn any ad hoc social categories. It also allows participants to listen to the talkers multiple times and to compare them in a side-by-side manner, lessening the working memory burden of the task relative to forced-choice tasks. Thus, the free classification task allows both children and adults to display their knowledge of regional dialect variation to a fuller extent than would be possible in a forced-choice classification task.

Dialect classification by children and adults was examined in two experiments. In Experiment 1, listeners ranging from 8 to 86 years old were presented with read speech samples of either 20 female or 20 male talkers from four broad dialect regions in the United States (Midland, New England, North, and South) and were asked to sort the talkers by their regional background. The results of Experiment 1 showed that even the 8- and 9-year-olds exhibited similar classification patterns to the patterns observed for the adult listeners. To further explore when classification abilities first emerge in childhood, Experiment 2 involved a modified free classification task with a smaller set of more child-friendly materials, suitable for use with 4- to 11-year-olds. The speech samples in Experiment 2 were a subset of those used in Experiment 1 and represented the same four U.S. dialect regions. The results of Experiment 2 showed that children first begin to classify talkers like adults at around 8- to 9-years old. However, most of the younger children in Experiment 2, the 4- to 7-year-olds, were also able to classify talkers in terms of dialect, showing that even in early childhood, children can perceive dialect differences and use those differences in an explicit dialect classification task, suggesting that they already have some level of sociolinguistic competence.

2. Experiment 1

2.1. Methods

2.1.1. Participants

The data included in Experiment 1 were collected at the Buckeye Language Network's Language Sciences Research Lab in the Labs in Life exhibit at the Center of Science and Industry (COSI), a science museum located in Columbus, Ohio, USA. The participants in Experiment 1 included 400 children and adults who were recruited from among the visitors to the museum. The data from eight additional participants were excluded because they did not complete the task and the data from seven additional participants were excluded because they were not native English speakers. Listeners in Experiment 1 were all monolingual native speakers of American English and reflected the typical ethnic breakdown of the visitor population to the museum, with over 90% of listeners being white. The listeners were assigned to one of six listener age groups: 8- to 9-year-olds, 10- to 11-year-olds, 12- to 13-year-olds, 14- to 15-year-olds, 16- to 17-year-olds, and adults, who ranged from 18 to 86 years old.

Given that individual listeners' residential history has been shown to affect dialect classification performance (Baker et al., 2009; Clopper & Pisoni, 2004b; Preston, 1993), some limited demographic data were collected to identify the broad dialect background of each listener. To determine region of origin of each listener, the regional distinctions described by Labov et al. (2006) were used as criteria for categorizing listener dialect. Using these criteria, the state of Ohio is home to two dialect regions, the Midland and Northern regions. Listeners who reported being exclusively from either of these two regions were included in a broad Midwest group to represent the region local to the state of Ohio. Listeners from either the Eastern New England or Western New England regions were included in a broad New England group. Listeners included in the South or the West categories were from the corresponding regions described by Labov et al. (2006). Listeners who had lived in two or more dialect regions during their lifetime, or who had lived in regions outside of the U.S., were considered to be from multiple dialect regions. The number of listeners in each age group from each dialect region is shown in Table 1. As shown in the table, the majority of the listeners in all age groups were from the Midwest region, and a substantial number of additional listeners had lived in multiple dialect regions.

2.1.2. Stimulus materials

Speech samples from 20 white female and 20 white male talkers, between 20 and 29 years old at the time of recording, were selected from the TIMIT Acoustic-Phonetic Continuous Speech Corpus (Fisher et al., 1986). Five female talkers from each of four

Table 1
Number of listeners in Experiment 1 by listener age group and listener dialect region.

Listener age	Listener dialect region					Total
	Midwest	New England	South	West	Multiple	
8–9 yrs. (<i>M</i> =8.34 yrs.)	38	1	6	1	2	48
10–11 yrs. (<i>M</i> =10.47 yrs.)	29	3	13	3	4	52
12–13 yrs. (<i>M</i> =12.42 yrs.)	30	1	3	0	9	43
14–15 yrs. (<i>M</i> =14.29 yrs.)	21	0	3	0	10	34
16–17 yrs. (<i>M</i> =16.48 yrs.)	17	0	2	0	4	23
18–86 yrs. (<i>M</i> =38.10 yrs.)	109	0	8	0	83	200
Total	244 (61.0%)	5 (1.3%)	35 (8.8%)	4 (1.0%)	112 (28.0%)	400



Fig. 1. A map of the continental United States showing the four dialect regions included in the current study. Areas in white indicate regions that were not presented to listeners in the current study.

Table 2

Number of listeners in each age group in each of the two talker gender conditions in Experiment 1.

Listener age	Talker gender condition	
	Female	Male
8–9 yrs.	23	25
10–11 yrs.	28	24
12–13 yrs.	19	24
14–15 yrs.	19	15
16–17 yrs.	12	11
18–86 yrs.	110	90
Total	211	189

dialect regions in the United States (Midland, New England, Northern, Southern) were selected for the female talker condition, and five male talkers from each of the same four dialect regions were selected for the male talker condition. The four talker dialect regions are shown on the map in Fig. 1. Although these dialects have changed over time since the TIMIT recordings were made in the mid-1980s, the 40 stimulus talkers in this study were selected because they exhibit the expected phonetic and phonological features of their respective dialect, as discussed in more detail below. The current ages of the selected TIMIT talkers would place them in the same age cohort as the adult listeners (although they are slightly older than the average age for that listener group), and thus, in the cohort of parents, or perhaps grandparents, of the youngest listeners. The dialect-specific features of the stimulus materials were therefore expected to be relatively familiar to all but possibly the youngest participants, even if the participants themselves speak a more advanced variety of a particular dialect.

Talker gender has been shown to be a salient perceptual dimension that listeners are unable to ignore in the free classification task (Clopper & Pisoni, 2007). Thus, individual listeners were randomly assigned to one of two talker gender conditions so that each participant heard talkers of only one gender. The number of listeners in each age group assigned to each talker gender condition is shown in Table 2.

The auditory stimuli consisted of one token of each talker reading the same sentence: “She had your dark suit in greasy wash water all year.” This sentence was used because it contains a number of dialect-specific phonological variables and phonetic characteristics that distinguish talkers from the four dialect regions. Clopper and Pisoni (2004a) analyzed the acoustic attributes of tokens of this sentence produced by a larger set of male talkers in the TIMIT corpus, and identified the acoustic-phonetic properties that served as perceptual cues to talker dialect. Evidence of r-lessness for the New England talkers was seen in tokens of the word ‘dark.’ The vowel /æ/ in the word ‘had’ was also observed to be backed in tokens produced by New England talkers, but fronted in tokens produced by Northern talkers. A high degree of voicing of the fricative in the word ‘greasy’ was a significant predictor of Southern talkers’ speech and Southerners also produced /u/-fronting in the word ‘suit’ relative to the New England talkers. None of the Southern talkers were r-less. Although the talkers displayed variability within each dialect, the previous acoustic analysis of the stimulus sentence tokens confirmed that dialect-specific properties were correlated with the male talkers’ regional dialect background (Clopper & Pisoni, 2004a). Further, as noted above, the dialect-specific variation observed in these materials reflects current properties of these varieties, despite changes in the dialects over time since the TIMIT corpus was recorded.

2.1.3. Procedure

Listeners were seated in front of a personal computer equipped with a mouse and headphones. Twenty black rectangles, each with a talker’s initials in white text were shown on the left of the computer screen and a 16 × 16-cell grid was presented on the right. An example of the visual display as presented to the listeners is shown in the left panel of Fig. 2. Each black rectangle was

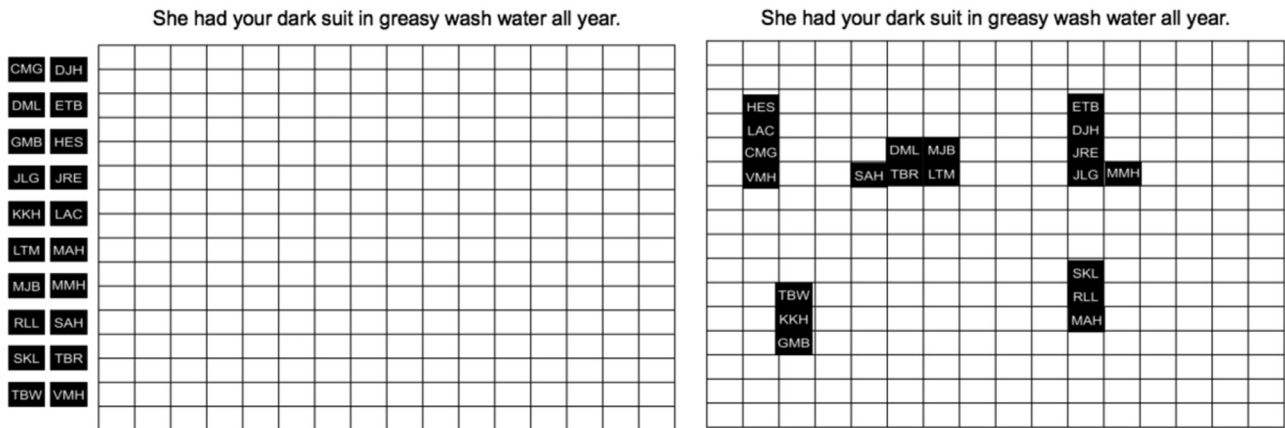


Fig. 2. The graphical configuration of stimuli that listeners were presented with at the beginning of the free classification task in Experiment 1 (left) and the talker grouping made by one listener at the end of the free classification task (right).

Table 3

Mean grouping data from all listeners in Experiment 1, by listener age group and talker gender condition. Standard deviations of each measure are shown in parentheses. Average measures across both talker gender conditions are shown in bold directly below the two corresponding condition means.

Age group	Gender condition	Talker groups	Hit rate	False alarm rate	Difference score
8–9 yrs. (N=48)	Female	7.96 (3.95)	0.25 (0.23)	0.14 (0.15)	0.11 (0.16)
	Male	6.88 (3.80)	0.33 (0.23)	0.17 (0.12)	0.15 (0.14)
10–11 yrs. (N=52)	Female	6.50 (3.05)	0.36 (0.19)	0.15 (0.14)	0.20 (0.14)
	Male	6.83 (3.32)	0.32 (0.22)	0.17 (0.16)	0.15 (0.15)
12–13 yrs. (N=43)	Female	6.16 (2.32)	0.36 (0.19)	0.15 (0.09)	0.21 (0.19)
	Male	5.83 (2.68)	0.39 (0.18)	0.16 (0.09)	0.23 (0.16)
14–15 yrs. (N=34)	Female	6.74 (3.05)	0.37 (0.22)	0.10 (0.06)	0.27 (0.22)
	Male	6.20 (2.21)	0.35 (0.17)	0.14 (0.09)	0.20 (0.13)
16–17 yrs. (N=23)	Female	6.83 (2.44)	0.41 (0.25)	0.08 (0.04)	0.33 (0.23)
	Male	5.82 (1.78)	0.46 (0.13)	0.09 (0.04)	0.36 (0.11)
18–86 yrs. (N=200)	Female	6.44 (2.31)	0.41 (0.17)	0.11 (0.09)	0.29 (0.17)
	Male	5.92 (2.22)	0.45 (0.21)	0.13 (0.08)	0.32 (0.19)
		6.21 (2.28)	0.43 (0.19)	0.12 (0.09)	0.31 (0.18)

associated with an audio file containing a token of the stimulus sentence produced by one of the talkers. The stimulus sentence was printed above the grid for the listeners' reference.

The listeners were informed that each black rectangle represented a different talker and that the talkers came from different parts of the United States. They were asked to listen to each talker by clicking on the rectangles, and move the rectangles into the grid by dragging them. They were instructed to group the talkers based on where they thought the talkers came from. Talkers from the same place or region were to be placed in the same group by positioning talker rectangles within contiguous cells of the grid. Listeners were permitted to make as many groups as they wanted, with any number of talkers in each group. They could also listen to the talkers and alter the groupings as many times as they wanted. The complete instruction script used in Experiment 1 is provided in the [Appendix](#). A sample talker grouping is shown in the right panel of [Fig. 2](#).

2.2. Results

2.2.1. Talker grouping

Listeners varied in the number of talker groups they made, making as few as 2 groups and as many as 18 groups, with a median of 6 groups. The number of talkers per group ranged from 1 to 10, with a median of 3 talkers per group. The third column of [Table 3](#) shows the mean number of talker groups by listener age and talker gender condition for all listeners. An ANOVA on the number of talker groups that listeners created revealed no significant main effects of listener age or talker gender condition. The interaction was also not significant. These results suggest that listeners in the different age groups did not significantly differ in the number of talker groups they formed.

A hit rate of talker pairings was then calculated for each listener. A correct talker pairing, or hit, was achieved when a listener accurately placed two talkers from the same dialect region in the same group. Given five talkers from each of four different dialect regions, a total of 40 correct pairings were possible in an individual grouping solution ($4 \times (4+3+2+1) = 40$). A hit rate was then

calculated for each listener as the proportion of the number of correct pairings achieved out of the total number of possible correct pairings. The fourth column of Table 3 shows the mean hit rate for each listener age group in each talker gender condition. Along with hit rate, a false alarm rate was calculated. An incorrect talker pairing, or false alarm, was made when a listener placed two talkers from different dialect regions together in the same talker group. A total of 150 false alarms were possible for any given listener's grouping solution $((5 \times 15) + (5 \times 10) + (5 \times 5) = 150)$. False alarm rates were calculated for each listener as the proportion of the number of incorrect talker pairings made out of the total number of possible incorrect talker pairings. False alarm rates are shown in the fifth column of Table 3. Because the hit rate and false alarm rate are only interpretable relative to the number of talker groups formed, a difference score was also calculated for each listener. This measure consisted of the difference between an individual listener's hit rate and his or her false alarm rate. These difference scores are conceptually similar to d' scores, a common measure of sensitivity used in signal detection theory (Macmillan & Creelman, 1991; Stanislaw & Todorov, 1999). The rightmost column of Table 3 shows the mean difference scores by listener age and talker gender condition.

An ANOVA was conducted to examine whether the difference scores were significantly different by listener age group and talker gender condition. The results revealed a significant effect of listener age group ($F(5, 394) = 12.27$; $p < 0.001$). Post-hoc t -tests comparing difference scores across age groups showed that the 8- to 9-year-olds were significantly less accurate than the 12- to 13-year-olds and older groups ($p < 0.01$ for all age comparisons). The four youngest groups, ranging from 8- to 15-year-olds, were also significantly less accurate than 16- to 17-year-olds and 18- to 86-year-olds ($p < 0.05$ for all age comparisons). None of the other pairwise comparisons were significant. There was no main effect of talker gender condition, nor was the interaction between listener age and talker gender condition significant. Taken together, the results show that while listeners of different ages did not differ in the number of talker groups they produced in the free classification task, the listeners' difference scores revealed developmental changes: 8- to 9-year-olds were less accurate than listeners 12 years old and above and adult-like levels of performance were not achieved until listeners were 16 years old.

As shown in Table 1, 61% of the listeners in Experiment 1 had lived exclusively in the Midwest, which includes both the Midland and Northern dialect regions. The remaining listeners had lived in other dialect regions or in multiple dialect regions, but the proportion of the non-Midwestern group to the total sample of listeners was small, especially for the youngest listeners. To reduce noise in the full analysis of all listeners' responses and to control for potential effects of listener residential history on performance, the subset of Midwestern listeners' responses was analyzed independently from the responses from the entire listener set. Results of this subset analysis showed very similar patterns in talker grouping and difference scores by listener age group and talker gender condition. The only difference in the two analyses was that the main effect of listener age ($F(5, 238) = 3.15$; $p < 0.01$) was significant in the analysis of the number of talker groups for the Midwestern listeners. Post-hoc t -tests comparing the number of talker groups across age groups revealed that the 12- to 13-year-olds and 18- to 86-year-olds made significantly fewer talker groups than the 8- to 9-year-olds and 10- to 11-year-olds ($p < 0.05$ for all age group comparisons). None of the other pairwise comparisons were significant. Compared to the results of the analysis based on all of the listeners, in which age did not significantly impact the number of talker groups produced, the Midwestern 8- to 11-year-olds made more talker groups than the Midwestern 12- to 13-year-olds and the adults, suggesting that they perceived more differences between the talkers, leading to more detailed perceptual categories of talkers. Differences between listener age groups in the difference scores were the same among the Midwestern listeners as for the full listener set.

2.2.2. Perceptual talker similarity

Based on the clear disjunctions in grouping accuracy revealed by the statistical analysis of the difference scores achieved by different listener age groups, the data from the 8- to 9-year-olds and the 10- to 11-year-olds were combined into an 8- to 11-year-old group, the data from the 12- to 13-year-olds and the 14- to 15-year-olds were combined into a 12- to 15-year-old group, and the data from the 16- to 17-year-olds and the 18- to 86-year-olds were combined into a 16- to 86-year-old group for cluster analyses of perceptual talker similarity. The goal of the cluster analyses was to examine the developing fine-grained patterns of talker grouping across the three combined listener age groups: 8- to 11-year-olds ($N = 100$), 12- to 15-year-olds ($N = 77$), and 16- to 86-year-olds ($N = 223$).¹ This analysis complements the analyses presented in Section 2.2.1 because although measures of talker groups and difference scores can reveal listeners' developing accuracy to group talkers by regional dialect, they cannot show how listeners of various ages differ in their specific grouping patterns for individual talkers within and across dialects.

To conduct the analysis of talker similarity, a similarity matrix of the twenty talkers was first created for each individual listener from his/her grouping solution. If a listener placed two talkers into the same group, the matrix cell for that relationship was filled with a value of 1, specifying a pairing. Cells filled with a value of 0 indicated a pair of talkers who were not grouped together. The individual similarity matrices were then summed over listeners in each of the six listener groups (3 listener age groups \times 2 talker gender conditions) to obtain six separate group similarity matrices. The value of each cell of the group matrix showed the total number of times that each pair of talkers was grouped together across listeners. As the value of each cell represented the sum of all listeners' matrices, larger cell values reflected more frequent grouping of the two corresponding talkers together. Greater cell values then, can be interpreted as greater perceived similarity between the two talkers, whereas smaller values indicate less perceived similarity between the two talkers. The similarity matrices were then submitted to the additive similarity tree program ADDTREE, which is used for efficiently representing proximity data (Cortier, 1982). The output for each input similarity matrix was an additive tree solution that

¹ The cluster solutions for the 16- to 17-year-olds are qualitatively similar to the cluster solutions for the adults, further justifying combining these two age groups into a large 16–86-year-old group.

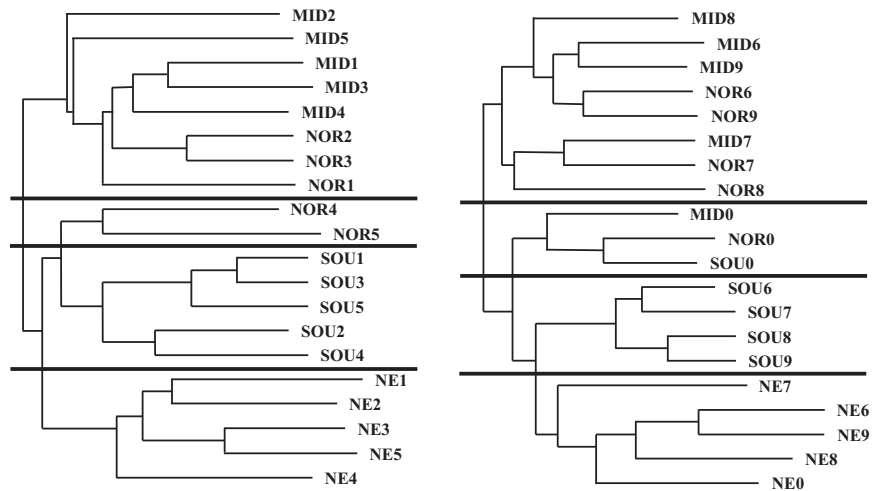


Fig. 3. Cluster solutions for the 8- to 11-year-old listeners in Experiment 1 for the female talker condition (left) and the male talker condition (right). Each talker is denoted with the label of his or her dialect region (MID for Midland, NE for New England, NOR for Northern, and SOU for Southern), followed by a numeral unique to that individual talker for that dialect region.

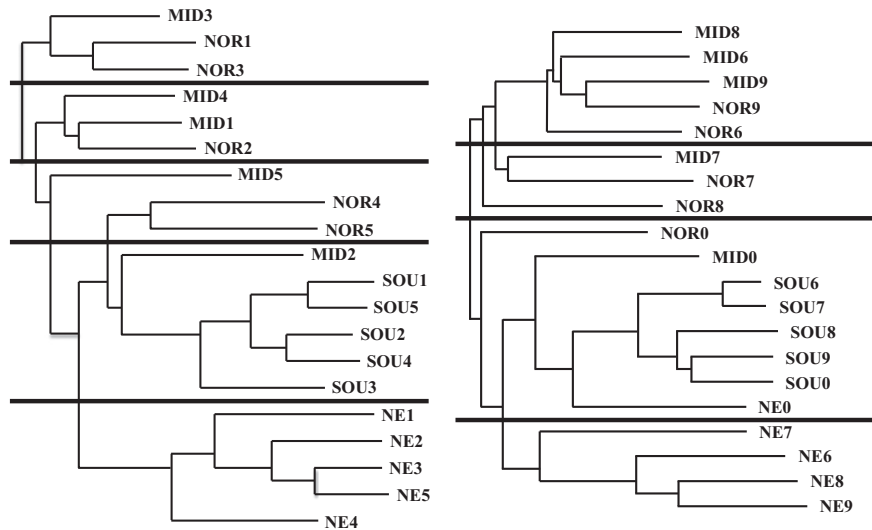


Fig. 4. Cluster solutions for the 12- to 15-year-old listeners in Experiment 1 for the female talker condition (left) and the male talker condition (right). Each talker is denoted with the label of his or her dialect region (MID for Midland, NE for New England, NOR for Northern, and SOU for Southern), followed by a numeral unique to that individual talker for that dialect region.

graphically represents the relative distances between individual talkers and clusters of talkers. Unlike other hierarchical cluster solutions, which are interpreted in terms of the node height connecting two items, relative perceptual distance in ADDTREE solutions is represented by horizontal branch length, and the model distance between any two objects equals the total horizontal distance of the shortest path traversed to connect them within the diagram. Thus, distances from both the nodes and the root of the tree are relevant for interpreting these solutions. Figs. 3–5 show the cluster solutions for the 8- to 11-year-olds, 12- to 15-year-olds, and 16- to 86-year-olds, respectively. In each of these figures, the left panel shows the cluster solution for listeners in the female talker condition, and the right panel shows the cluster solution for listeners in the male talker condition. Hand-drawn solid horizontal lines were added to indicate interpreted cluster divisions.

As shown in the left panel of Fig. 3, for 8- to 11-year-old listeners in the female talker condition, four clusters are apparent. The top cluster, a combined Midland/Northern cluster, includes all five of the Midland talkers and three of the Northern talkers. The second cluster includes the remaining two Northern talkers, who are further joined more loosely to the third major cluster, made up of all five Southern talkers. All five of the New England talkers appear in the final major cluster at the bottom of the solution. For the 8- to 11-year-old listeners in the male talker condition, four clusters are also present, as shown in the tree in the right panel of Fig. 3. The combined Midland/Northern cluster at the top includes four Midland talkers and four Northern talkers. One Midland talker, one Northern talker, and one Southern talker are grouped into the second cluster. The two bottom clusters include four Southern talkers, and all five of the New England talkers, respectively. Thus, when classifying the entire set of talkers, the 8- to 11-year-olds perceived distinct perceptual dialect clusters composed of New England talkers, Southern talkers, and a combination of Midland and Northern talkers.

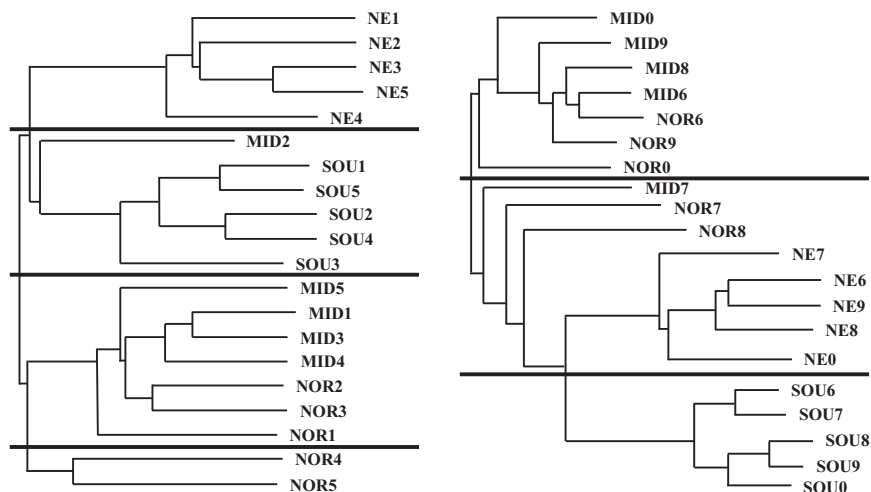


Fig. 5. Cluster solutions for the 16- to 86-year-old listeners in Experiment 1 for the female talker condition (left) and the male talker condition (right). Each talker is denoted with the label of his or her dialect region (MID for Midland, NE for New England, NOR for Northern, and SOU for Southern), followed by a numeral unique to that individual talker for that dialect region.

Turning to the 12- to 15-year-olds, for those listeners in the female talker condition, five clusters appear in the solution in the left panel of Fig. 4. The top cluster is composed of two Northern talkers and one Midland talker. The second cluster includes two Midland talkers and one Northern talker. The third cluster is made up of two Northern talkers, who, together with MID2 and MID5 are joined loosely to the fourth major cluster, composed of all five Southern talkers. While the horizontal distances between any two of the five Southern talkers are small, indicating relatively high perceptual similarity among the Southern talkers, the horizontal distances that connect the Southern talkers to the third cluster and to the individual MID2 and MID5 talkers are quite large, indicating a relatively weak level of similarity between these Midwestern talkers and the talkers in the Southern cluster. Further, the relatively short branches connecting the Midland/Northern clusters to the root of the tree reflect relatively short perceptual distances between these two clusters and the third cluster and the individual MID2 and MID5 talkers. Thus, the details of the cluster solution suggest a perceptual separation of the Southern talkers from the Midland and Northern talkers for this listener group. The fifth major cluster features all five of the New England talkers. For 12- to 15-year-old listeners in the male talker condition, four clusters are apparent in the solution shown in the right panel of Fig. 4. Three Midland talkers and two Northern talkers are grouped together in the first cluster and are joined with a second cluster including one Midland talker and one Northern talker to form a combined Midland/Northern major cluster. The third cluster includes all five Southern talkers in a tight configuration, with one New England talker and the remaining Midland talker more distantly joined to this cluster. Four New England talkers formed the fourth cluster. Two Northern talkers (NOR8 and NOR0) were not included in any of the major clusters. Thus, the cluster solutions for the 12- to 15-year-old listeners in both talker gender conditions show that they perceived three main dialect clusters: a New England cluster, a Southern cluster, and a combined Midland/Northern cluster. This pattern is similar to the solutions for the younger children shown in Fig. 3, although the cluster solutions for the 12- to 15-year-old listeners are somewhat noisier than those for the younger children.

The left panel of Fig. 5 shows that the five New England talkers formed a distinct major cluster in the analysis of the female talker condition for the 16- to 86-year-olds, as did the set of five Southern talkers. The large third cluster includes four Midland talkers along with three Northern talkers. A final cluster, which is joined to the third cluster, includes the other two Northern talkers. The remaining Midland talker (MID2) was perceived as not belonging definitively to any major cluster, but is closest in horizontal distance to the talkers within the Southern cluster. The cluster analysis for the 16- to 86-year-old listeners in the male talker condition, shown in the right panel of Fig. 5, shows three major perceptual clusters: a New England cluster, a Southern cluster, and a combined Midland/Northern cluster. Two of the Northern talkers and one of the Midland talkers are connected very distantly to the major cluster comprising the New England cluster and the Southern cluster, but the relatively long horizontal branches that connect those clusters with the Midland and Northern talkers suggest weak inclusion in that major cluster. The grouping patterns for both talker gender conditions suggest that the 16- to 86-year-olds largely perceived three dialect groups overall; a New England group, a Southern group, and a combined Midland/Northern group.

Overall, listeners of different age groups showed comparable perceptual dialect spaces that included three major dialect clusters: a New England cluster, a Southern cluster, and a combined Midland/Northern cluster. The talkers from the New England and Southern dialect regions, who exhibit speech patterns that are perceptually and culturally salient within the United States (Preston, 1986, 1989), appeared in their own clusters and were, for the most part, easily distinguished from the other talkers. In general, the New England talkers were the most perceptually distinct from the other talkers, as suggested by the relatively long perceptual distances (i.e., long branches in the trees) that separate them from the other talkers and the relatively short distances that join the New England talkers together within their clusters. Talkers from the Midland and the North were clustered together and typically formed a large, combined Midland/Northern cluster, although some talkers from these regions were occasionally connected loosely (i.e., by long branches) to talkers from the South.

To parallel the analysis conducted on the data from the Midwestern subset of listeners for grouping and difference scores, cluster analyses for the same subset were conducted to explore potential differences from the cluster solutions for the full dataset as a result

of residential history. For all listener age groups and talker conditions, the solutions for the Midwestern listeners showed relatively clear and separate New England and Southern clusters of talkers, along with a single, combined cluster of mostly Midland and Northern talkers. The cluster analyses of responses from both the full set of listeners and the Midwestern subset therefore suggest that while New England talkers and Southern talkers are perceptually distinct from talkers from other regions, Midland and Northern talkers are not very distinct from each other and are typically perceived as belonging to the same dialect group.

2.3. Discussion

The results of Experiment 1 showed that although listener age did not affect the number of talker groups formed in the free classification task, it affected the overall accuracy with which listeners grouped talkers into categories by regional dialect. The increases in difference scores in relation to listener age showed that from middle childhood through adolescence, and from the teenage years into adulthood, listeners improve in their ability to classify talkers on the basis of regional dialect. Significant increases in classification accuracy were observed at the beginning of adolescence (at 12 years) and in the teenage years (at 16 years). Similar patterns were also observed for the difference scores of the Midwestern listener subset. These improvements in classification accuracy correspond well to the ages suggested by Labov (1964) for children's entry into the Social Perception stage of sociolinguistic competence and for their transition into the even more advanced Stylistic Variation stage, respectively. The accuracy results therefore lend strong support to the idea that the changes in schooling and social exposure noted by Labov (1964) for these stages are accompanied by changes in sociolinguistic competence.

The cluster analyses of the listeners' grouping solutions revealed consistent patterns of perceptual talker similarity across listener age groups and talker gender conditions. The analyses for all listener age groups in both talker gender conditions showed a distinct cluster of most, if not all five, of the New England talkers. Also, these New England talker clusters featured no talkers from other dialect regions. A second distinct cluster that emerged in the cluster solutions for all listeners was a cluster of Southern talkers. Clusters of Southern talkers also did not typically include close connections with talkers from other regions. Lastly, the solutions for listeners from all age groups and in both talker gender conditions featured a combined Midland/Northern cluster of primarily Midland and Northern talkers. Despite grouping patterns in which some individual Midland or Northern talkers were not tightly included in any cluster, even the solutions for the youngest listeners, the 8- to 11-year-olds, featured a cluster composed entirely of Midland and Northern talkers.

The patterns seen in the cluster solutions offer insight into the relative perceptual salience of the speech from the four dialect regions, especially for Midwestern listeners. First, the separation of New England and Southern talkers from the combined cluster of Midland and Northern talkers suggests that listeners perceive differences among the New England talkers, the Southern talkers, and "Midwestern" talkers from the Midland and Northern dialect regions. Further, none of the analyses for any listener age group showed a distinctive Midland cluster apart from a distinctive Northern cluster. Thus, listeners did not seem to perceive a regionally significant difference between the Midland and Northern talkers, despite the many acoustic-phonetic differences in the speech of talkers from these two regions. Previous work in perceptual dialectology has revealed corroborating evidence of Ohioans' failure to fully recognize Midland and Northern talkers as speakers of two different regional varieties. From interviews with lifelong residents of Ohio, Campbell-Kibler (2012) found that the speech of Northern talkers is neither stigmatized nor fully enregistered in the minds of Ohioans. That is, respondents did not recognize Northern speech as particularly distinct from a "general American" variety, which is also associated with Midland speech. By contrast, they were much more aware of a Midland-Southern distinction, making strongly negative evaluations of both Southern speech and speakers of Southern American English. Similar perceptions by the listeners in the current study are reflected in the separate clusters of Southern talkers. However, the clusters of New England talkers were relatively tighter and more separate from the other talkers than were the clusters of Southern talkers. This pattern suggests that despite the widely recognized division between "Midwestern" (Midland/Northern) speech and Southern speech, the talkers from the New England region were the most perceptually different from the Midland and Northern talkers, perhaps reflecting a combination of the linguistic and geographical distance of New England speech and speech from Ohio.

The youngest children in Experiment 1 were of the right age to be in Labov's (1964) Vernacular stage. However, the divisions among the dialect varieties perceived by these youngest children corresponded to sociolinguistically relevant divisions for the adults, in contrast to Labov's (1964) suggestion that children in this stage are too young to be aware of the social significance of this type of variation. Previous work has not found evidence that children under the age of 7 years old can categorize talkers on the basis of regional dialect (Flocchia et al., 2009; Wagner et al., 2014), which may reflect either their being at an earlier stage of sociolinguistic development, as suggested by Labov (1964), or that they have not been tested with sufficiently sensitive methods, as suggested in Section 1.3. Experiment 1 used a free classification task as a sensitive method to gauge dialect perception that does not require explicitly connecting dialect variation to specific places. The results of this experiment suggest that children in the Vernacular stage have already begun to acquire the social meaning of regional dialect variation. Experiment 2 used a simplified free classification task to probe dialect perception abilities in children as young as 4 years old. Children of this age would presumably be in Labov's (1964) pre-Vernacular stage and were predicted to be too young to exhibit awareness of sociolinguistic variation. Thus, we expected to observe developmental changes in dialect grouping behavior across age groups in Experiment 2. To ensure that the free classification task in Experiment 2 was not too taxing on the attention of these younger listeners, the task was simplified both by decreasing the number of stimuli presented and by manipulating characteristics of the visual representation of the auditory stimulus tokens. The task included stimuli that were a subset of the stimuli used in Experiment 1, and the ages of the listeners in Experiment

Table 4
Number of listeners in Experiment 2 by listener age group and listener dialect region.

Listener age	Listener dialect region					Total
	Midwest	New England	South	West	Multiple	
4–5 yrs. ($M=4.52$ yrs.)	33	0	3	1	3	40
6–7 yrs. ($M=6.55$ yrs.)	31	0	3	2	8	44
8–9 yrs. ($M=8.40$ yrs.)	31	0	3	0	11	45
10–11 yrs. ($M=10.39$ yrs.)	25	0	0	0	16	41
Total	120 (70.6%)	0 (0.0%)	9 (5.3%)	3 (1.8%)	38 (22.4%)	170

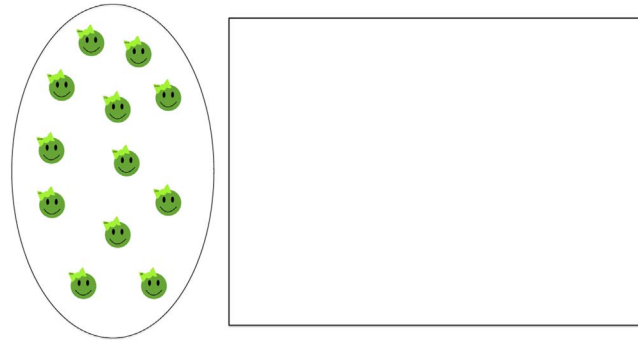


Fig. 6. The graphical configuration of stimuli that listeners were presented with at the beginning of the free classification task in Experiment 2.

2 ranged from 4 to 11 years old. The inclusion of 8- to 11-year-olds allowed the performance of these older children in Experiment 2 to be compared to the performance of the 8- to 11-year-old listeners in Experiment 1.

3. Experiment 2

3.1. Methods

3.1.1. Participants

The participants in Experiment 2 included 170 listeners who were tested in the same lab where Experiment 1 was conducted. All participants were monolingual native speakers of American English, and as in Experiment 1, over 90% of listeners were ethnically white. The data from an additional 39 children were excluded from the analysis due to failure to complete the classification task as instructed. This failure resulted when all 12 talkers were classified into one group or when each of the 12 talkers was classified as separate from all other talkers and placed in her own group. Twenty-three 4- to 5-year-olds, 12 6- to 7-year-olds, 3 8- to 9-year-olds, and 1 10- to 11-year-old were excluded on this basis. Data from an additional 8 4- to 5-year-olds, 1 6- to 7-year-old, and 1 8- to 9-year-old were excluded from the analysis due to failure in the control task (see below). The remaining data came from 40 4- to 5-year-olds, 44 6- to 7-year-olds, 45 8- to 9-year-olds, and 41 10- to 11-year-olds.

A summary of the listeners by age group and residential history is shown in Table 4. As shown in the table, the majority of the listeners were from the Midwest, although there was a sizable minority of listeners from other regions.

3.1.2. Stimulus materials

The sentence tokens from a subset of twelve of the female talkers used in Experiment 1 were presented. Listeners heard the twelve talkers read the same sentence as in Experiment 1, “She had your dark suit in greasy wash water all year.” There were three talkers from each of the four American English dialect regions. The talkers from Experiment 1 who were included in Experiment 2 were: MID1, MID3, and MID4 from the Midland dialect region; NE2, NE3, and NE5 from the New England dialect region; NOR1, NOR2, and NOR3 from the Northern dialect region; and SOU1, SOU3, and SOU5 from the Southern dialect region. In the cluster analyses of the listeners in the female talker condition in Experiment 1, these talkers consistently appeared within tight, inner clusters that represented their respective dialect regions, which was the basis for their selection and inclusion as stimuli in Experiment 2.

3.1.3. Procedure

Children began the experiment by completing a short training phase in which they practiced sorting six individual smiley face icons on the computer screen by either each smiley face’s hat (baseball cap or top hat) or smiley face color (red, yellow, blue). Each smiley face was one of three colors and wore one of two different types of hat, making six unique icons. The smiley faces were presented to listeners simultaneously and in a random arrangement on the screen. Listeners learned through sorting that if grouped according to hat, the objects should be placed into two groups: one with smiley faces wearing top hats and one with smiley faces wearing baseball

Table 5

Mean grouping data in Experiment 2 by listener age group. Standard deviations of each measure are shown in parentheses.

Age group	Talker groups	Hit rate	False alarm rate	Difference score
4–5 yrs. (N=40)	3.40 (1.92)	0.45 (0.25)	0.39 (0.21)	0.05 (0.14)
6–7 yrs. (N=44)	4.77 (1.87)	0.31 (0.21)	0.22 (0.17)	0.09 (0.14)
8–9 yrs. (N=45)	4.67 (2.15)	0.42 (0.25)	0.19 (0.15)	0.23 (0.21)
10–11 yrs. (N=41)	4.56 (1.80)	0.39 (0.27)	0.16 (0.10)	0.23 (0.26)

caps. Alternatively, if grouped by color, the faces should be arranged in three groups: blue smiley faces, yellow smiley faces, and red smiley faces. Listeners were told that they would be grouping more faces and objects through the course of the experiment, and that any grouping solutions they produced were acceptable.

Listeners continued to the experimental task once the experimenter was certain that they understood the task objective. Although the auditory stimuli in Experiment 2 were a subset of the stimuli used in Experiment 1, the visual presentation of the talkers was different and more child-friendly. Twelve visually identical, green smiley face icons were presented simultaneously to listeners, analogous to the labeled black rectangles presented in Experiment 1. The smiley face icons appeared within an oval in one large, randomly assorted group on the left side of the computer screen. Each icon was linked to an audio file of one of the twelve unique tokens of the stimulus sentence that played through the headphones when clicked on by the listener. Fig. 6 shows the display that each listener saw during the experimental trial. The main task instructions given to the listeners were similar to the instructions in Experiment 1. The complete instruction script used in Experiment 2 is provided in the Appendix. A grid for talker placement was not provided to simplify the visual display and the task requirements for the younger participants. Therefore, smiley face icons needed only to be closely positioned to constitute a group. For the youngest listeners with less-developed proficiencies at using a computer mouse, the experimenter managed the moving of smiley face icons around the computer screen, but since the listeners wore headphones for the entirety of the experimental trial, the experimenter was unable to hear the stimuli and simply moved the icons in response to the child's pointing.

Once listeners completed the experimental task to their individual satisfaction, the experimenter inspected the grouping solution on the computer screen to ensure that group membership of each smiley face icon was apparent. If the experimenter was unclear about group membership, he asked the listener to provide clarification.

Following the experimental task, listeners were given a control classification task to insure that they understood how to complete the grouping task. Listeners were instructed to sort objects of varying dimensions (e.g., tall and skinny vs. short and fat) by object shape (ovals, triangles, stars, and hearts). The control task solution was coded for accuracy for each listener. If listeners failed to accurately classify each shape into a group with shapes similar to it, their data were excluded from analysis due to an inferred lack of understanding of the grouping task. As described above, data from 10 children were excluded from the analysis due to failure in the control task.

3.2. Results

3.2.1. Talker grouping

As in Experiment 1, the first measure calculated from the data was the number of talker groups for each individual listener's grouping solution. The mean number of talker groups produced by each listener age group is shown in Table 5. As mentioned above, the data from children who formed only one talker group or who formed 12 groups of 1 talker each were excluded. An ANOVA was conducted to test for effects of listener age group on number of talker groups formed and revealed a significant main effect of listener age group ($F(3, 166)=4.43$; $p<0.05$). Post-hoc t -tests revealed that with a mean of 3.4 talker groups, the youngest age group, the 4- to 5-year-olds, formed significantly fewer talker groups when compared pairwise against each of the three older listener age groups ($p<0.01$ for all comparisons). The three oldest age groups each formed just over 4.5 talker groups on average, and were not significantly different from each other in the number of talker groups produced.

A hit rate and a false alarm rate were calculated by individual listener using the same methods as in Experiment 1. The mean hit rates and false alarm rates for each listener age group are shown in Table 5. In considering hit rate, the 4- to 5-year-old age group had the highest hit rate of all the age groups, while they also had the highest false alarm rate of all the age groups. These trends arise from the significantly fewer number of talker groups that the youngest age group formed, which increased the likelihood of producing both hits and false alarms. Therefore, as in Experiment 1, difference scores by listener were calculated from individual listener hit and false alarm rates. The mean difference score for each listener age group is shown in Table 5. An ANOVA on overall difference scores revealed a significant main effect of listener age group ($F(3, 166)=9.36$; $p<0.05$). Comparing difference scores pairwise across age groups with post-hoc t -tests, the two youngest age groups, the 4- to 5-year-olds and the 6- to 7-year-olds, were both significantly less accurate in their grouping of talkers than the two oldest age groups, the 8- to 9-year-olds (compared to 4- to 5-year-olds, $t=-4.67$, $p<0.01$; compared to 6- to 7-year-olds, $t=-3.78$, $p<0.01$) and the 10- to 11-year-olds (compared to 4- to 5-year-olds, $t=-3.67$, $p<0.01$; compared to 6- to 7-year-olds, $t=-2.91$, $p<0.01$). That is, a definitive shift in the ability to accurately classify talkers was observed between the 4- to 7-year-olds and the 8- to 11-year-olds.

As in Experiment 1, an analysis of the data from the subset of Midwestern listeners was conducted to control for listener residential history. As shown in Table 4, 120 listeners in the current experiment were from the Midwestern dialect regions (i.e.,

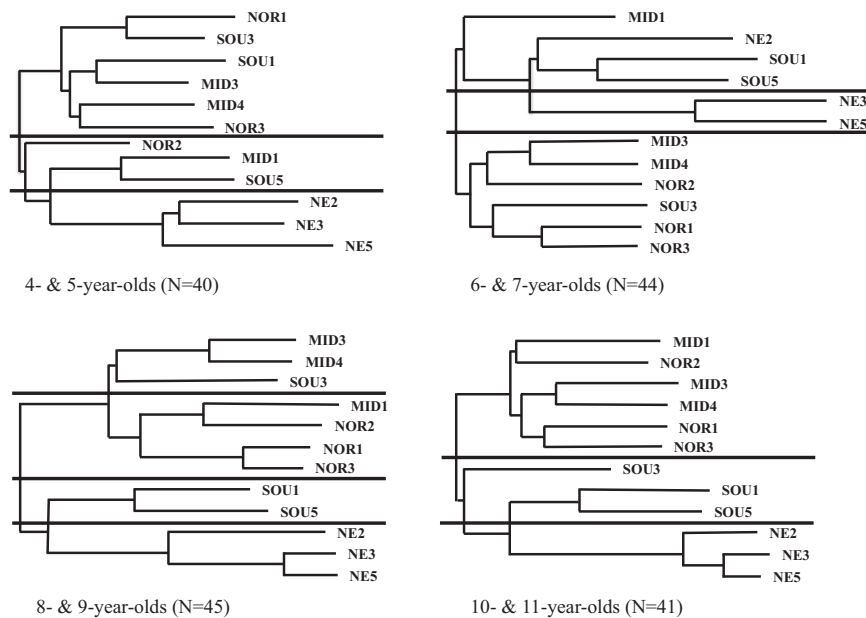


Fig. 7. Cluster solutions for each listener age group in Experiment 2. Each talker is denoted with the label of her dialect region (MID for Midland, NE for New England, NOR for Northern, and SOU for Southern), followed by a numeral unique to that individual talker for that dialect region.

Midland and North). The results of these analyses showed the same qualitative patterns as were found in the analysis of the full listener set.

3.2.2. Perceptual talker similarity

A cluster analysis was carried out using listeners' grouping solutions to investigate how listeners classified individual talkers and which dialect groups they formed. Each listener's grouping solution was transformed into a 12×12 talker similarity matrix in which each matrix cell represented the pairwise comparison of any two given talkers. As with the individual similarity matrices obtained in Experiment 1, the matrices were then summed by listener age group and the summed similarity matrices were submitted to the ADDTREE program (Corter, 1982) to visualize perceptual talker similarity. Fig. 7 shows the output additive tree solutions for each listener age group.

In the case of the 4- to 5-year-old listeners, the cluster solution shown in the top left panel of Fig. 7 reveals three major clusters. The first major cluster is composed of two talkers from each of the Midland, Northern, and Southern dialect regions. The second cluster includes one Midland talker and one Southern talker, and the third cluster, at the bottom of the solution, includes all three of the New England talkers. One Northern talker (NOR2) is not strongly linked to any cluster, which suggests that she was perceived as equally similar to, or equally distinct from, the other eight talkers in the top two clusters, who were all approximately the same distance from the root of the tree. The New England talkers are all further from the root of the tree, suggesting that they were perceived as more distinct than the other talkers. This clustering pattern therefore suggests that the 4- to 5-year-olds who successfully completed the task perceived a distinction between a New England group and a group of talkers that did not belong to the New England group.

The solution for the 6- to 7-year-olds in the top right panel of Fig. 7 also shows three major perceptual clusters of talkers. The first cluster is made up of two Southern talkers, and it has a loose connection to one of the New England talkers. The first cluster is further joined to a second major cluster of two New England talkers. The third cluster is composed of two Midland talkers, all three Northern talkers, and a Southern talker. As in the cluster solution for the 4- to 5-year-olds, one talker does not fit well into any of the three clusters. Midland talker MID1 is joined to the two clusters with the New England and Southern talkers, but is still fairly distant from these two clusters, given the relatively large distances of the talkers in these clusters from the root of the tree. Further, the two clusters that contain the three New England talkers and two of the Southern talkers have mostly longer branches adjoining the talkers within the clusters than does the bottom third cluster. This clustering pattern suggests that the talkers in the bottom cluster were perceived as mutually more similar than were the five talkers in the top two clusters, possibly reflecting listeners' greater degree of familiarity with the speech of talkers from the Midland and Northern dialect regions, given that about 70% of the listeners in that age group were from the Midwest.

The cluster solution for the 8- to 9-year-old listeners in the bottom left panel of Fig. 7 shows two major perceptual clusters, each made up of two minor clusters. The top minor cluster includes two Midland talkers and a Southern talker and is joined to the second minor cluster, which includes the third Midland talker and all three of the Northern talkers to form a larger Midland/Northern cluster. The third minor cluster is composed of two Southern talkers, and the fourth minor cluster, at the bottom of the solution, includes all of the New England talkers. The Southern and New England minor clusters are joined into a major cluster, but with a more distant connection than the two minor clusters in the Midland/Northern cluster at the top. The minor clusters suggest that four groups roughly corresponding to the talkers' actual dialect backgrounds were perceived by the 8- to 9-year-old listeners.

The solution for the oldest age group, the 10- to 11-year-olds, in the bottom right panel of Fig. 7, is comparable to the solution for the 8- to 9-year-old listeners. The first cluster includes all three Midland talkers and all three Northern talkers. The second cluster includes the same two Southern talkers from the minor Southern cluster in the 8- to 9-year-olds' solution. The Southern cluster is loosely joined to a very tight New England cluster containing all three New England talkers. Southern talker SOU3 is more distantly connected to this combined Southern/New England cluster.

To control for potential effects of residential history on dialect classification performance, a second cluster analysis was carried out on the data from only the Midwestern listeners. Very similar results were observed in the subset cluster analyses as in the full analyses shown in Fig. 7. The youngest listener age group formed a distinct New England talker group and a combined Midland, Northern, and Southern group, the 6- to 7-year-olds formed a combined New England and Southern group and a combined Midland and Northern group, and listeners 8 years old and above formed the three clusters seen in Experiment 1: a New England cluster, a Southern cluster, and a combined Midland/Northern cluster.

3.2.3. Comparing across experiments

To make viable inferences that span from the youngest children (in Experiment 2) to the adolescents and adults (in Experiment 1), it is critical to establish that the slight changes in methodology across the two experiments did not change the core effects. Children aged 8–11 years old participated in both of the experiments and so a direct examination of the effect of the experimental differences was possible for this age range. The results of a series of independent-sample *t*-tests showed that the 8- to 9-year-olds in Experiment 1 formed more talker groups on average than the 8- to 9-year-olds in Experiment 2 ($t=4.24, p<0.01$), and that the 10- to 11-year-olds in Experiment 1 formed more talker groups on average than the 10- to 11-year-olds in Experiment 2 ($t=4.03, p<0.01$). Thus, there was a significant difference in the number of groups that the 8- to 11-year-olds made across the experiments, which is not particularly surprising given the reduction in the number of talker tokens from 20 in Experiment 1 to 12 in Experiment 2. To compare accuracy across experiments, further independent-sample *t*-tests showed that the 8- to 9-year-olds in Experiment 1 obtained lower difference scores than the 8- to 9-year-olds in Experiment 2 ($t=-2.63, p<0.05$). In contrast, the difference scores for the 10- and 11-year-old listeners were not different across experiments ($t=-1.04, p=0.30$). Thus, 10- to 11-year-olds performed with similar accuracy in both experiments, but the 8- to 9-year-olds performed significantly better in Experiment 2 than in Experiment 1, in part because their hit rates were much higher when there were fewer talkers. With respect to the specific groups formed, children aged 8–11 years old performed roughly equivalently across the two experiments, creating separate New England and Southern talker clusters and a combined Midland/Northern cluster. Overall, it appears that the methodological differences across experiments made the task somewhat easier to perform, particularly for younger children, but the methodological differences did not change the qualitative picture of children's dialect classification ability.

3.3. Discussion

The results that emerged from Experiment 2 provide limited support for Labov's (1964) proposal regarding sociolinguistic competence in young children. The youngest listeners, 4- to 5-year-olds, would presumably be on the cusp between Labov's Pre-Vernacular and Vernacular stages. On the one hand, their accuracy scores were significantly lower than those of the two oldest groups of children in this experiment, consistent with an earlier developmental stage. In addition, they made the fewest talker groups overall, suggesting that they do not yet have the level of fine-grained dialect perception that older children and adults have. Further, a comparatively high proportion of these young children did not complete the free classification task with multiple, distinctive talker groups (37%) and their data were therefore excluded from further analysis. Given that these children were able to appropriately group objects in the control (non-dialect-related) task, it is unlikely that they were cognitively incapable of the task demands. It appears instead that they may have lacked a relevant aspect of sociolinguistic knowledge. Specifically, these children may have failed to understand that regional dialect variation can be a meaningful dimension with which to classify talkers into groups, or alternatively, they may have been unaware of how the specific acoustic-phonetic properties of the stimulus materials that differentiated the talkers could be markers of regional groupings. These failures are consistent with Labov's (1964) predictions for children in the Pre-Vernacular and early Vernacular stages.

On the other hand, the cluster analysis of the 4- to 5-year-old listeners who succeeded in forming distinctive talker groups suggests that they perceived the New England talkers as different from the talkers from the other dialect regions, despite their relatively low accuracy in grouping overall. Their ability to perceive the New England talkers as distinct from the talkers from the other regions suggests that these young listeners are sensitive to the sociolinguistic importance of regional dialect variation, even if they could apply that sensitivity to only one group of talkers in this set of speech samples. Thus, despite the high failure rate of this age group, many 4- to 5-year-olds demonstrated a certain level of sociolinguistic competence that allowed them to group talkers based on regional dialect.

All of the remaining children in this experiment would presumably be in Labov's (1964) Vernacular stage, but each group showed a different profile of sociolinguistic competence. The 6- and 7-year old children were similar to the younger children in terms of accuracy and had a lower, but still relatively high, failure rate (21%). However, they patterned with the two older age groups in terms of the number of groups they created, suggesting an increased ability in fine-grained dialect perception compared to the 4- to 5-year-olds. More strikingly, these children – unlike the children in all of the other age groups – perceived two mixed groups of talkers: one made up of mostly New England and Southern talkers and a second group composed of mostly Midland and Northern talkers. The structure of the clusters for the 6- to 7-year-olds therefore may reflect a distinction in perception that is based on Midwestern versus other

dialects. That is, these listeners appear to have separated the Midwestern (Midland and Northern) talkers from the non-Midwestern (Southern and New England) talkers, but they did not strongly distinguish between the non-Midwestern talkers, unlike older children.

The two older age groups showed similar categorization performance in several respects, including the number of talker groups produced, accuracy scores, and cluster solutions. The accuracy and cluster solutions for the two oldest age groups also show developmental improvements relative to the two younger age groups and the cluster solutions resemble the solutions for the listeners in Experiment 1: these children formed three talker groups, by separating the New England talkers and the Southern talkers from each other, and creating a third group of Midwestern talkers. But even within this general pattern of similarity, there is evidence for differences between these age groups. The first difference was observed in overall accuracy across experiments for each age group. For the 8- to 9-year-olds, accuracy was higher in Experiment 2, which involved more child-friendly methods, than in Experiment 1, suggesting these children may have found the overall task easier in Experiment 2. By contrast, the accuracy of the 10- to 11-year-olds was equivalent across the two experiments, despite the change in methods. In addition, there were subtle differences in the clustering patterns between the two older age groups. Although the 8- to 9-year-olds showed a similar clustering pattern to the 10- to 11-year-olds, less mixing was observed within the Midland/Northern cluster for the 8- to 9-year-olds' solution than for the 10- to 11-year-olds' solution. While more work would be needed to explore this pattern, it suggests that 8- to 9-year-olds may be more sensitive to the place-based acoustic-phonetic differences between talkers from the Midland and Northern dialect regions than 10- to 11-year-olds, perhaps because they are not yet aware that these acoustic-phonetic differences characterize a dialect distinction that is not particularly socially meaningful for adults (see [Campbell-Kibler \(2012\)](#)), and thus are not relevant parameters for grouping talkers by region. The distinctive patterns of performance across these age groups therefore demonstrate greater sociolinguistic skills than would be predicted based on [Labov's \(1964\)](#) proposal, and they suggest variation and continued development of sociolinguistic competence within [Labov's \(1964\)](#) Vernacular stage.

4. General discussion

In the current study, a free classification task was used to explore how the ability to classify talkers by regional dialect develops through childhood and adolescence and when it may emerge in early childhood. Three main results were obtained in this study. First, children as young as 4 to 5 years old can perceive regional dialect differences and use them as a basis for grouping talkers. Although more than one third of the youngest listeners in Experiment 2 did not succeed in classifying talkers into multiple groups and although the youngest listeners were not as adept as older children in the task, these data are the first to document that many children as young as 4 years old can succeed in this kind of dialect perception task. Second, children's abilities to perceive dialect differences show ongoing, incremental improvement through childhood and do not reach adult-like levels until children are in late adolescence, consistent with [Labov's \(1964\)](#) proposal for a protracted process of development of sociolinguistic competence. Third, listeners of all ages exhibit qualitatively comparable patterns of perceptual dialect similarity, which may stem from the acoustic-phonetic properties of the speech stimuli, sociolinguistic knowledge about regional variation, or a combination of both.

The youngest listeners in Experiment 2 were 4- to 5-year-olds, and the previous literature has consistently shown that children in this age range fail to use regional dialect information to place talkers into groups ([Flocchia et al., 2009](#); [Girard et al., 2008](#); [Wagner et al., 2014](#)). Nevertheless, the majority of the youngest children in the current study were able to perceive regional dialect differences in the free classification task and consistently made a coherent group of New England talkers that was distinct from the other talkers. Although these children were only able to extract the New England talkers as a separate group, and they were significantly less accurate overall than their older peers, their success strongly suggests that they have the core cognitive abilities and social awareness in place to make this type of sociolinguistic judgment. Moreover, this early success is not consistent with the stages suggested by [Labov \(1964\)](#), in which children of this age should exhibit limited social knowledge about variation. Below we discuss some reasons that the New England dialect may have been comparatively easy to perceive, but an additional factor is likely the specific task that we used. Free classification, unlike forced choice grouping methods, may ease the processing difficulties inherent in grouping people by regional dialect and allow younger children to demonstrate their competence in this area.

In terms of developmental changes, listeners make significant improvements in their ability to accurately classify talkers according to regional dialect around the age of 8 years, again around the age of 12 years, and yet again around the age of 16 years. The latter two improvements in accuracy roughly correspond to [Labov's \(1964\)](#) transitions into the Social Perception and Stylistic Variation stages, respectively. They also correspond, again, roughly, to the ages at which many American children change schools, from entry into middle school/junior high and into high school, respectively. Previous research, also partially motivated by [Labov's \(1964\)](#) work, has shown that social networks and school environments have a large impact on listeners' conceptions of socially salient language varieties and styles, as well as the linguistic markers that characterize them ([Eckert, 1989](#); [Lambert, Giles, & Picard, 1975](#)). While we do not have detailed information about the school environment of our participants, school changes seem like a promising explanation for why these particular ages are important for observing developmental changes in sociolinguistic competence.

We consider next how the different dialects included in this study influenced children's ability to group the talkers by region. Talkers from the New England region were consistently grouped together by all of the listeners, including the youngest ones who were able to successfully complete the task. Conversely, the Midland and Northern talkers were treated as part of a single unified group by all listeners, including adults. The classification of the Southern talkers underwent the most change across development: the 4- to 5-year-olds grouped them loosely with the Midland/Northern talkers; the 6- to 7-year-olds grouped the Southern talkers with the

New England talkers as a distinct group separate from the Midland/Northern talkers; and beginning with 8- to 9-year-olds, participants created a distinct group for the Southern talkers separate from both the New England talkers and the Midland/Northern talkers.

These changes in grouping patterns bear little relationship to the stages suggested by Labov (1964). On the one hand, the youngest children, presumably in the Pre-Vernacular stage, are too precocious inasmuch as they can succeed at distinguishing even one regional dialect from the set. On the other hand, children undergo several shifts in grouping ability within what should be a single stage, the Vernacular. Given that Labov (1964) suggested that children in the Vernacular stage do not yet have awareness of the social significance of speech variation, the adult-like clusters formed by children as young as 8 years old are somewhat surprising. However, Labov (1964) did not rely exclusively on ages or concomitant schooling to account for sociolinguistic ability and was explicit about the role of social context in influencing dialect exposure and awareness. We consider now the kinds of additional information and contexts that might be important for understanding the changes in children's dialect perception abilities within the Vernacular stage.

One possible explanation for the observed trajectory lies in the acoustic-phonetic differences across the dialects. The regional dialects in this study contained distinctive acoustic-phonetic properties, for example r-lessness in the word 'dark' and /æ/ backing in 'had' in New England speech; /u/-fronting in the word 'suit' and voicing of the fricative in the word 'greasy' for Southern speech; and /æ/ fronting in 'had' in Northern speech. Perhaps particular features in the New England talkers' speech, such as r-lessness, allow young listeners to distinguish their tokens of the stimulus sentence from tokens produced by talkers from the other three regions. For example, r-lessness involves the absence of a segment, which may make it perceptually salient for young children. Similarly, the alternation between [s] and [z] in 'greasy' for the Southern talkers may be more salient to young children than more subtle vowel variation. The cues that adult listeners attend to when categorizing talkers by regional dialect have been explored in previous research (e.g., Clopper & Pisoni, 2004a), but little is known about the cues that children and adolescents attend to when identifying talkers by dialect or which cues they acquire awareness of first. Understanding which cues, or which types of cues, are especially relevant for children would allow us to make more general predictions about which speech varieties children should be able to distinguish most easily. Future research on children's abilities to perceive various kinds of phonetic and phonological variation that is linked to dialect variation is therefore a critical step in addressing questions relating to the role of acoustic-phonetic salience in dialect perception during childhood.

An alternative, but not mutually exclusive, explanation for the different trajectories for the dialects may lie in children's developing social awareness. It is unlikely that the younger listeners in this study had the relevant knowledge to link acoustic-phonetic cues to specific regions of the U.S., but they may have the ability to distinguish between local and non-local talkers based on their relative familiarity with certain speech patterns. Children living in the Midwest have likely had the least experience with talkers from the New England dialect region due to the sheer distance between the Midwest and New England, and are therefore likely to be the least familiar with the speech patterns that New England talkers produce, and more likely to perceive New England talkers as different from talkers in their own local communities. There are likely more substantial levels of exposure for many Midwestern children to talkers from the Southern dialect region, as it directly borders their home region to the south. This geographic reality may explain why 4- to 5-year-olds grouped the Southern talkers with the Midland/Northern talkers in the larger group of 'local' talkers. However, as children get older, their differential levels of exposure to Southern speech compared to the truly local Midland and Northern dialects appear to be leveraged to create a distinct Southern group. The results for the listeners in the 6- to 7-year-old age group further suggest the existence of an intermediate stage during which Southern talkers are grouped with New England talkers as part of a potentially generalized 'non-local' talker group. If this interpretation is correct, it predicts that how listeners from other regions of the United States – or indeed, other parts of the world – classify talkers will depend on their familiarity with and social connectedness to the varieties in question.

Further, it is important to interpret children's performance in the context of that of adults, who presumably model locally relevant social knowledge. For example, the fact that all of the listeners – adults included – grouped Midland and Northern talkers together likely stems from the similar sociolinguistic status of talkers from the Midland and Northern dialect regions. As noted above, these two dialects are characterized by different acoustic-phonetic features, but it has been shown that Northern dialect features are comparatively less socially meaningful for adult listeners from Ohio than features that characterize the Southern dialect (Campbell-Kibler, 2012). This lack of regional salience demonstrates that speech from the Northern dialect region has not been fully enregistered by the local community as a separate variety. Therefore, the talkers from the Midland and the Northern dialect regions are not reliably differentiated as representing different groups based on their speech. The fact that the 8- to 9-year-olds in Experiment 2 show modest evidence of separate Midland and Northern groups, but that the 10- to 11-year-olds have lost that potential distinction provides some evidence that this lack of enregisterment of the Northern dialect is learned. Thus, the local social status of particular varieties may also be a critical component to understanding the development of dialect perception in childhood.

Beyond the social understanding required to link dialects to speech communities, adult-like abilities to classify talkers according to regional dialect also depend on a certain degree of geographic knowledge. Specifically, the acquisition and refinement of geographic knowledge may support listeners' developing improvements in grouping talkers according to dialect through childhood and adolescence. Children begin acquiring geographic and sociocultural knowledge regarding their own nation and other nations from about 5 to 6 years old, but this knowledge is further refined well into adolescence and likely even into adulthood (Barrett & Short, 1992; Jahoda, 1962; Piaget & Weil, 1951). Further research is needed to gauge the developmental course of children's explicit geographical awareness from early childhood to adulthood and how that development corresponds to changes in their dialect classification abilities. We expect that children's acquisition of finer distinctions between geographic regions would allow them to construct more accurate representations of regional dialect variation and facilitate the ability to link regional dialects to identifiable geographic locations or entities. We further expect the continued refinement of geographic knowledge and increased exposure to variation in adulthood to lead to variation in dialect classification performance in adults. Although we obtained data from a large

number of adult participants in this study, including a relatively large percentage (41%) who had lived in multiple dialect regions, we leave an exploration of this potential variation within the adult population for future research.

Taken together, the results of the two experiments support Labov's (1964) general proposal that sociolinguistic competence develops over a protracted period from early childhood through adolescence. Echoing Labov's (1964) discussion, we have highlighted a number of factors that likely shape this developmental trajectory, including changes in peer groups associated with changes from elementary to middle to high school, in children's perceptual sensitivity to dialect-specific variants of different kinds, in familiarity with particular varieties, in sociolinguistic knowledge acquired from the local community, and in geographic knowledge. With the exception of geographic knowledge, these factors should similarly contribute to the trajectory of social classification along other dimensions, such as age, gender, ethnicity, race, or social class. In particular, to the extent that children are sensitive to the phonetic and phonological characteristics of the speech of talkers from these groups and that they have sufficient exposure to the relevant social categories, we predict that they will succeed with explicit classification. Thus, children would be expected to exhibit adult-like gender classification much earlier than adult-like regional dialect classification, given large acoustic differences in f_0 and vowel space across genders and a prominent role of gender in the local society. Across all social domains, however, these factors should be teased apart in future research to understand their separate and combined contributions to children's developing skills in perceptual social classification.

Although the results of the current study are broadly consistent with Labov's (1964) proposal, the results do not fully support the specific developmental stages that he suggested. In particular, it appears that children in the Pre-Vernacular stage already show some level of dialect awareness, however rudimentary. Moreover, the results observed in the perceptual cluster analyses suggest that children pass through a greater number of intermediate stages of ability in the Vernacular stage on their way to full adult-like competence. However, this study examined only one aspect of sociolinguistic competence: regional dialect classification. It is therefore critical to explore other dimensions of sociolinguistic competence to fully flesh out a model of the developmental trajectory of skills in this area. As suggested above, this expansion should include research on other social dimensions. However, this expansion should also consider other skills in the perception domain, including social attitude judgments (e.g., Giles et al., 1983; Kinzler & DeJesus, 2013), cross-variety intelligibility (e.g., Nathan, Wells, & Donlan, 1998), and explicit identification and discrimination between varieties (e.g., Floccia et al., 2009; Girard et al., 2008; Wagner et al., 2014) throughout childhood to determine how the trajectories of these skills align within and across children.

5. Conclusions

The results of the current study provide evidence for both the emerging ability to perceive dialect variation in children as young as 4 to 5 years old and gradual developmental improvements in perceptual dialect classification throughout childhood and adolescence. The youngest listeners had lower accuracy rates and a tendency to form comparatively fewer talker groups than older listeners, but cluster analyses of their talker grouping solutions showed that they were able to classify New England talkers separately from the other dialect groups. This latter finding provides the first evidence that some children as young as 4–5 years old can successfully group talkers by regional dialect. With increasing age, Southern talkers were gradually separated from Midland and Northern talkers, and by the age of 8–9 years, children were significantly more accurate than their younger peers and consistently formed three dialect groups that qualitatively match those produced by adults: a New England, a Southern, and a combined Midland/Northern group. Classification accuracy continued to improve with age, showing significant gains for 12- to 13-year-olds and for 16- to 17-year-olds, when adult-like accuracy was achieved. Thus, these results demonstrate both the early emergence of children's ability to group talkers by dialect as well as a protracted trajectory of development in dialect perception that does not become adult-like until well into adolescence. This trajectory of development reflects ongoing changes throughout the lifespan in sociolinguistic awareness, including familiarity and exposure to linguistic and social aspects of dialect variation. The overall shape of the trajectory observed in this study is therefore expected to be found in explorations of other social categories and other social perception tasks, although the exact timing of the trajectory will vary with the sociolinguistic awareness of the relevant varieties and their associated social categories.

Acknowledgments

This work was supported by a seed grant and a graduate summer research fellowship from the Center for Cognitive and Brain Sciences at Ohio State University.

Appendix

The instructions used for the free classification task in Experiments 1 and 2 are provided in (1) and (2), respectively.

(1) Instructions for participants (children and adults) in Experiment 1

In this experiment, you will be asked to make groups of talkers based on their regional accent. On the screen, you will see a set of icons, representing different talkers. The talkers come from different parts of the United States. If you double-click on an icon, you will hear that talker reading a sentence. Your task is to listen to each of the talkers and then make groups of talkers based on their accents. You can listen to the talkers as many times as you want and you can rearrange the icons until you are satisfied that you have

grouped all of the talkers with the same accent into one group. You may make as many groups as you want. Each group may have as many talkers as you want. You do not need to have the same number of talkers in each group.

(2) Instructions for children in Experiment 2

In this experiment, you will play a computer game. On the screen, you will see a set of shapes and a big rectangle. The shapes are different people from different parts of the country and you can listen to the people talk by double-clicking on the shapes. Your job is to put all of the people into groups on the big rectangle. You should put people who you think are from the same part of the country in a group together. You should put people who you think are from different parts of the country in different groups. You can make as many groups as you want and you can have different numbers of people in each group.

References

- Baker, W., Eddington, D., & Nay, L. (2009). Dialect identification: The effects of region of origin and amount of experience. *American Speech*, 21, 34–56.
- Barrett, M., & Short, J. (1992). Images of European people in a group of 5-10-year-old English schoolchildren. *British Journal of Developmental Psychology*, 10(4), 339–363.
- Campbell-Kibler, K. (2012). Contestation and enregisterment in Ohio's imagined dialects. *Journal of English Linguistics*, 40(3), 281–305.
- Carver, C. M. (1987). *American regional dialects: a word geography*. Ann Arbor: University of Michigan Press.
- Chambers, J. K. (1992). Dialect acquisition. *Language*, 68, 673–705.
- Clopper, C. G. (2008). Auditory free classification: Methods and analysis. *Behavior Research Methods*, 40, 575–581.
- Clopper, C. G., & Pisoni, D. B. (2004a). Some acoustic cues for the perceptual categorization of American English regional dialects. *Journal of Phonetics*, 32, 111–140.
- Clopper, C. G., & Pisoni, D. B. (2004b). Homebodies and army brats: Some effects of early linguistic experience and residential history on dialect categorization. *Language Variation and Change*, 16, 31–48.
- Clopper, C. G., & Pisoni, D. B. (2006). The Nationwide Speech Project: A new corpus of American English dialects. *Speech Communication*, 48, 633–644.
- Clopper, C. G., & Pisoni, D. B. (2007). Free classification of regional dialects of American English. *Journal of Phonetics*, 35, 421–438.
- Corter, J. E. (1982). ADDTREE/P: A Pascal program for fitting additive trees based on Sattah and Tversky's ADDTREE algorithm. *Behavior Research Methods & Instrumentation*, 14, 353–354.
- Davis, L. M., & Houck, C. L. (1992). Is there a Midland dialect? – again. *American Speech*, 67, 61–70.
- Eckert, P. (1989). *Jocks and burnouts: social categories and identity in the high school*. New York: Teachers College Press.
- Edwards, J., Gross, M., Chen, J., MacDonald, M. C., Kaplan, D., Brown, M., & Seidenberg, M. S. (2014). Dialect awareness and lexical comprehension of mainstream American English in African American English-speaking children. *Journal of Speech, Language, and Hearing Research*, 57, 1883–1895.
- Fisher, W. M., Doddington, G. R., & Goudie-Marshall, K. M. (1986). The DARPA speech recognition research database: specification and status. In *Proceedings of the DARPA speech recognition workshop* (pp. 93–99).
- Floccia, C., Butler, J., Girard, F., & Goslin, J. (2009). Categorization of regional and foreign accent in 5- to 7-year-old British children. *International Journal of Behavioral Development*, 33, 366–375.
- Gathercole, S. E., Willis, C. S., Baddeley, A. D., & Emslie, H. (1994). The children's test of nonword repetition: A test of phonological working memory. *Memory*, 2, 103–127.
- Giles, H. (1970). Evaluative reactions to accents. *Educational Review*, 22, 211–227.
- Giles, H., Harrison, C., Creber, C., Smith, P., & Freeman, N. (1983). Developmental and contextual aspects of children's language attitudes. *Language and Communication*, 3, 141–146.
- Girard, F., Floccia, C., & Goslin, J. (2008). Perception and awareness of accents in young children. *British Journal of Developmental Psychology*, 26, 409–433.
- Jacewicz, E., Fox, R. A., & Salmons, J. (2011). Regional dialect variation in the vowel systems of typically developing children. *Journal of Speech, Language, and Hearing Research*, 54, 448–470.
- Jahoda, G. (1962). Development of Scottish children's ideas and attitudes about other countries. *The Journal of Social Psychology*, 58, 91–108.
- Kinzler, K. D., & DeJesus, J. M. (2013). Northern=smart and Southern=nice: The development of accent attitudes in the United States. *The Quarterly Journal of Experimental Psychology*, 66, 1146–1158.
- Koops, C., Gentry, E., & Pantos, A. (2008). The effect of perceived speaker age on the perception of PIN and PEN vowels in Houston, Texas. *University of Pennsylvania Working Papers in Linguistics: Selected papers from NAWAV*, 36(14), 91–101.
- Labov, W. (1964). Stages in the acquisition of standard English. In R. Shuy (Ed.), *Social dialects and language learning* (pp. 77–104). Champaign, IL: National Council of Teachers of English.
- Labov, W. (1972). *The social stratification of (r) in New York City department stores*. In: *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania Press 43–69.
- Labov, W., Ash, S., & Boberg, C. (2006). *The atlas of North American English: Phonetics, phonology and sound change*. Berlin: Mouton de Gruyter.
- Lambert, W. E., Giles, H., & Picard, O. (1975). Language attitudes in a French-American community. *International Journal of the Sociology of Language*, 1975.4, 127–152.
- Macmillan, N.A., & Creelman, C.D. (1991). *Detection theory: A user's guide* (pp. 147–149). Cambridge: Cambridge University Press.
- Mullennix, J. W., & Pisoni, D. B. (1990). Stimulus variability and processing dependencies in speech perception. *Perception & Psychophysics*, 47, 379–390.
- Nathan, L., Wells, B., & Donlan, C. (1998). Children's comprehension of unfamiliar regional accents: A preliminary investigation. *Journal of Child Language*, 25, 343–365.
- Patterson, J. L. (1992). *The development of sociolinguistic phonological variation patterns for (ing) in young children*. Doctoral dissertation. University of New Mexico.
- Ptacek, P. H., & Sander, E. K. (1966). Age recognition from voice. *Journal of Speech, Language, and Hearing Research*, 9, 273–277.
- Piaget, J., & Weil, A. M. (1951). The development in children of the idea of the homeland and of relations with other countries. *International Social Science Bulletin*, 3, 561–578.
- Preston, D. R. (1986). Five visions of America. *Language in Society*, 15, 221–240.
- Preston, D. R. (1989). *Perceptual dialectology: Nonlinguists' views of areal linguistics*. Providence, RI: Foris.
- Preston, D. R. (1993). Folk dialectology. In D. R. Preston (Ed.), *American dialect research* (pp. 333–377). Philadelphia: John Benjamins.
- Purnell, T., Idsardi, W., & Baugh, J. (1999). Perceptual and phonetic experiments of American English dialect identification. *Journal of Language and Social Psychology*, 18, 10–30.
- Roberts, J., & Labov, W. (1995). Learning to talk Philadelphia: Acquisition of short a by preschool children. *Language Variation and Change*, 7, 101–112.
- Roberts, J. (1997). Acquisition of variable rules: a study of (–t, d) deletion in preschool children. *Journal of Child Language*, 24, 351–372.
- Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers*, 31, 137–149.
- Van Bezooijen, R., & Gooskens, C. (1999). Identification of language varieties: The contribution of different linguistic levels. *Journal of Language and Social Psychology*, 18, 31–48.
- Wagner, L., Clopper, C. G., & Pate, J. K. (2014). Children's perception of dialect variation. *Journal of Child Language*, 41, 1062–1084.
- Williams, A., Garrett, P., & Coupland, N. (1999). Dialect recognition. In D. R. Preston (Ed.), *Handbook of perceptual dialectology* (pp. 345–358). Philadelphia: John Benjamins.
- Yan, Q. (2015). The perceptual categorization of Enshi Mandarin regional varieties. *Journal of Linguistic Geography*, 3, 1–19.