INTERSPEECH 2017
August 20–24, 2017, Stockholm, Sweden

Dialect perception by older children

Ewa Jacewicz, Robert A. Fox

Department of Speech and Hearing Science, The Ohio State University, Columbus, OH, USA
jacewicz.1@osu.edu, fox.2@osu.edu

Abstract

The acquisition of regional dialect variation is an inherent part of the language learning process that takes place in the specific environments in which the child participates. This study examined dialect perception by 9-12-year-olds who grew up in two very diverse dialect regions in the United States, Western North Carolina (NC) and Southeastern Wisconsin (WI). In a dialect identification task, each group of children responded to 120 talkers from the same dialects representing three generations, ranging in age from old adults to children. There was a robust discrepancy in the children's dialect identification performance: WI children were able to identify talker dialect quite well (although still not as well as the adults) whereas NC children were at chance level. WI children were also more sensitive to cross-generational changes in both dialects as a function of diachronic sound change. It is concluded that both groups of children demonstrated their sociolinguistic awareness in very different ways, corresponding to relatively stable (WI) and changing (NC) socio-cultural environments in their respective speech communities.

Index Terms: speech perception, dialect identification, dialect acquisition, older children, sound change.

1. Introduction

Regional dialects of a language are well-recognized markers of socio-cultural identity, underlying intra- and intercultural communication. The study of the perception of dialect variation has recently emerged as an important area of research in speech communication with the overall goal of better understanding how listeners identify the dialect of a talker in situated interactions [1]. To that end, most of the dialect categorization research has been conducted with adult populations, primarily with young college-age listeners, and relatively little is known about dialect identification abilities outside this age range. There is a particularly striking gap in understanding the development of dialect perception in children. For example, when do children develop socio-cultural competence in relation to dialect variation? Relatedly, how is the awareness of social norms in pronunciation patterns shaped in the time-course of language development?

There is some evidence that early exposure to dialect variation can influence listening abilities in infancy. In particular, 6-month-old Australian infants who were exposed to American English did not show a preference for their native Australian English, whereas American infants who were not exposed to Australian English did show a preference for their native variety [2]. By 12 months, children were shown to be able to cope with dialect variation in a word segmentation task [3]. However, the ability to learn words across different dialect varieties was manifested at a later age, at the onset of the 3rd year [4]. Later in development, 4-5-year-old preschoolers demonstrated awareness of regional versus foreign accents in speech [5]. However, socio-cultural cognition in relation to regional dialects seems to develop no earlier than at pre-adolescent age. For example, 8-9-year-olds were found to approximate perceptual dialect classification patterns of adults [6], and 9-10-year-olds were able to "label" speakers based on social stereotypes (e.g., that Northern-accented American English speakers sounded smarter and Southern-accented speakers sounded nicer [7]).

The current study examined whether the perceptual abilities of older children allowed them to identify dialects in the face of extensive talker variability in the input speech. It is the case that children's dialect perception has typically been examined with fewer talkers, ranging from as few as one talker per dialect for a total of two [8] to as many as 10 talkers per dialect for a total of 40 [6]. However, presenting children with more talkers whose pronunciation varies not only in terms of the dialect itself but also with regard to a particular version of the dialect (such as when spoken by older or younger residents in the common speech community) introduces additional challenges. In particular, to interpret variation in spoken language due to changing community patterns, children must utilize their socio-indexical processing abilities. They must recognize that older and younger people speak a common regional variety although their pronunciation patterns may differ slightly in certain phonetic features. Hence, the purpose of the current study is to better understand both socio-cultural awareness of older children as well as their ability to utilize talker normalization processes, the skill that characterizes speech perception by adults.

This study is an extension of an earlier work which examined vowel identification performance of typically developing 9-12-year-olds responding to 120 talkers ranging in age from old adults to children, and representing two regional dialects and both sexes [9]. Children performed remarkably well under such demanding conditions and their overall identification rate was only 8.5% lower than that of adults responding to the same stimulus material in another study [10]. That work demonstrated that older children can successfully attend to a wide range of acoustic cues in multiple talkers' voices, which supply rich phonetic and indexical information related to dialect, age, and talker sex. Importantly, children's identification rates were significantly higher when stimuli were spoken in their native dialect. Furthermore, unlike in other studies, they made identification decisions on the basis of individual words rather than sentences or longer stretches of speech which provide comparatively more information about talker dialect.

In the vowel identification task reported in [9], the children were not explicitly asked to identify talker dialect. The current study specifically asked this question in a different experiment, which tested dialect awareness in a dialect identification task.

http://dx.doi.org/10.21437/Interspeech.2017-18

2. Dialect identification experiment

2.1. Listeners

Listeners were 32 typically developing American English-speaking children ranging in age from 9-12 years. The children represented two different regional dialects. Sixteen children were born and raised in Western North Carolina (NC) and spoke the Southern variety of English, and 16 were from Southeastern Wisconsin (WI) and spoke the Northern variety [11]. The dialect spoken in the mountainous region of Western NC represents distinct features of Appalachian English, including prosody, vowel dynamics, and slower speech tempo compared to the varieties in the North. The dialect in WI shares most of its features with the mainstream General American English. All children were of comparable socio-economic status and attended local public schools.

2.2. Materials and procedure

The experiment was previously conducted with adult listeners from these two regions and the results were reported in [12]. The current study utilized the same stimulus material and followed the same experimental procedure as in [12]. The stimuli were 12 hVd-tokens recorded as isolated words: *heed, hid, hayed, head, had, hod, who'd, heard, hide, hoed, hood, hawed*. The words were recorded by 120 talkers from the same two dialect regions as the listeners (60 from NC and 60 from WI), who represented three generations of native residents in each area: 8-12-year-old children (C), 35-50-year-old young adults (P), and 66-91-year-old old adults (GP). There were 20 talkers in each age group (10 male and 10 female). For the purposes of this study, the three generations were named Children, Parents, and Grandparents (C, P, GP, respectively). Six different hVd-words were selected from each talker for a total of 720 unique exemplars used in the experiment.

Each listener was tested individually, either at Western Carolina University in NC or at University of Wisconsin-Madison in WI, using a common laboratory procedure and a common protocol at each site. The 720 words were presented in random order in three blocks of 240 items each. Signals were delivered over headphones. The listeners were told they would hear one word at a time and had to decide if the word was spoken by a NC or a WI talker. They clicked with the mouse on one of two response boxes on the computer screen, "North Carolina" and "Wisconsin." Only one repetition was allowed and children were instructed to guess if they were not sure. The experiment was under the control of a custom program in MATLAB.

2.3. Results

Listener responses were analyzed using Signal Detection Theory [13]. The identification responses were converted to d' values in order to better separate listeners' sensitivity to dialect from their response bias. In this analysis, the correct WI responses to WI tokens were "hits" and incorrect responses to NC tokens were "false alarms." As a reference, higher d' values indicate greater sensitivity and chance performance corresponds to a d' of 0.

2.3.1. Overall dialect identification pattern

In the first set of analyses, we were concerned with the overall sensitivity to dialect as a function of listener group. Also, to determine whether children's performance differed from dialect identification by adults, the current results for NC and WI children were compared with the results for the corresponding adult listeners reported in [12].

The overall pattern is shown in Figure 1, which displays mean d' values for each listener group as a function of talker generation (C, P, GP). We can observe a striking difference in the performance of NC and WI children. Essentially, NC children were unable to identify talker dialect whereas WI children were much better at this task. Still, dialect sensitivity of WI children did not reach the levels of the adults. Furthermore, all listener groups responded differentially to the versions of the dialects spoken by adults and children. The higher scores for the adult talkers (and particularly for GP-talkers in NC) indicate that listeners had less difficulty with identifying dialect of the adults than the dialect of children, who obtained comparatively lower scores.

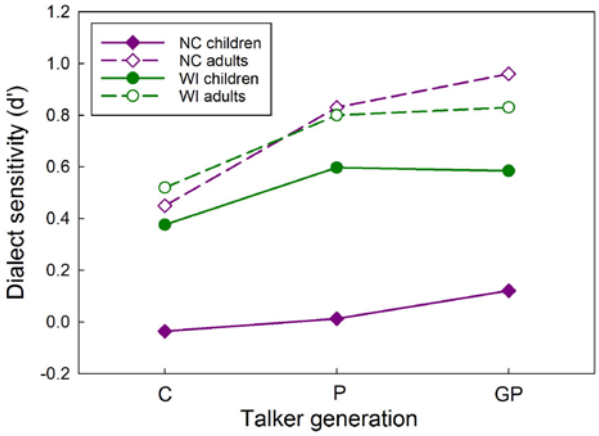

Figure 1: *Overall dialect identification pattern*

These results were analyzed using a three-way analysis of variance (ANOVA) with the between-subject factors listener age (children, adults), listener dialect (NC, WI) and talker generation (C, P, GP). Where applicable, post hoc *t*-tests were used to explore the significant results. All three main effects were significant. The children were less sensitive to talker dialect that the adults ($p<.001$) and WI listeners were more sensitive than NC listeners ($p<.001$). With respect to talker age group, sensitivity to the dialect version spoken by adults was significantly higher than sensitivity to that spoken by children ($p=.001$), but the difference between the two adult groups was not significant (Scheffé). Importantly, a significant listener age by listener dialect interaction ($p=.001$) showed that NC children were significantly less sensitive to dialect than WI children and, in turn, WI Children were less sensitive to dialect than the adults. However, NC adults and WI adults did not differ significantly one from another.

2.3.2. Dialect perception in relation to sound change

The differential performance of children revealed that WI children were able to distinguish between the two dialects relatively well but NC children performed mostly at chance level. However, the significant main effect of talker generation suggested that both groups might benefit from some pronunciation features of older talkers in assigning dialect labels whereas dialect cues in children's productions seemed to be mitigated. The second set of analyses focused only on children's performance to better understand children's awareness of the diachronic sound change in their respective

speech communities. To determine which pronunciation features might have guided dialect identification, we explored in greater detail the children's response patterns for selected words.

The greater dialect sensitivity of WI children guided our selection of the appropriate words. We selected five words which yielded the highest d' values from all words produced by GP-adults: *hid, head, had, hide, hood*. Average dialect sensitivity data for these words are displayed in Figure 2. We then compared these scores with d' values for the same words produced by children (see Figure 3).

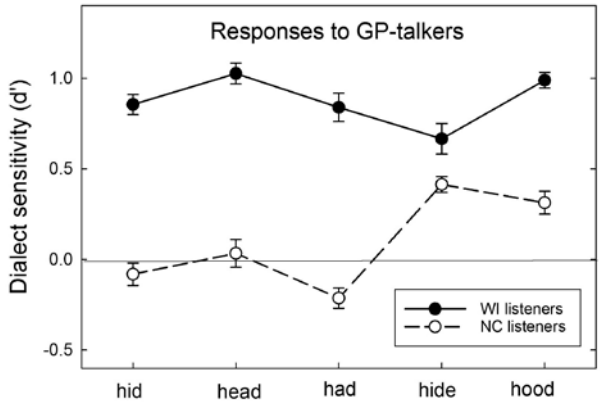

Figure 2: *Responses to the dialect of old adults (GP).*

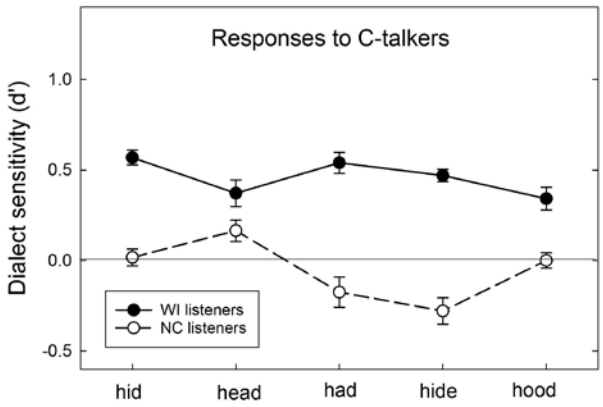

Figure 3: *Responses to the dialect of children (C).*

A repeated-measures ANOVA with the within-subject factors talker generation (C, GP) and word (5 levels) and the between-subject factor listener dialect returned significant main effects of generation (p=.003) and listener dialect (p<.001). As expected, sensitivity to dialect was greater in response to GP- than to C-groups, and WI listeners were more sensitive to dialect than NC listeners. However, a significant 3-way interaction between generation, word and listener dialect (p=.011) provided further insights into sound change.

The interaction was first explored with independent t-tests. For GP-productions, the differences between WI and NC listeners were significant for all but one word (*hide*). For C-productions, the differences were not significant for two words, *head* and *hood*. These results indicate that certain acoustic characteristics of the vowels in these three words were particularly salient resulting in either enhanced dialect identification in NC listeners or reduced dialect identification in WI listeners.

However, before we attempt to explain these patterns, we need to consider if the significant differences in the perceptual behavior of children were meaningful with respect to chance performance. This possibility was explored with one-sample t-tests (one-tailed), which examined if the d' score for each word was significantly greater than $d' = 0$ (chance level). The results for WI listeners showed that all d' scores were significantly greater than 0, indicating that even the reduced dialect sensitivity for *hide, head* and *hood* was still above chance. On the contrary, none of the d' scores except for *hide* (GP) for NC listeners were significantly greater than 0. This indicates that their better dialect identification score for *hide* in GP-productions reflected an enhanced dialect sensitivity rather than a chance performance. Their sensitivity for both *head* and *hood* was still at chance, however.

2.3.3. Acoustic characteristics of perceptually salient vowels

We now relate the dialect identification patterns to the acoustic characteristics of vowels in *hid, head, had, hide, hood* (ɪ, ɛ, æ, aɪ, ʊ, respectively) produced by GP-adults in NC (Figure 4) and WI (Figure 5). To approximate the dynamic formant change in a vowel, the F1 an F2 frequencies were measured at 20-35-50-65-80% time points. In the plots, vowel symbols were placed next to the 80% point. The plots represent average normalized male and female productions, first converted to z-scores [14] and then rescaled to hertz values [15]. Further details of these procedures can be found in [16, 17].

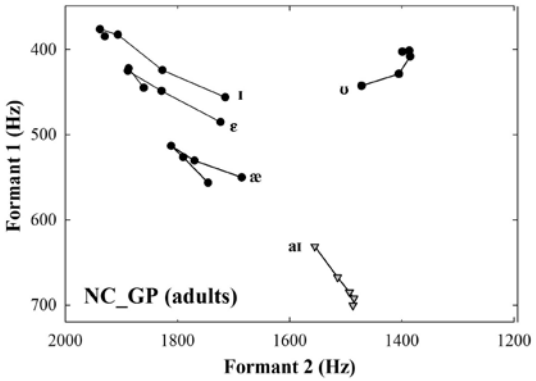

Figure 4: *Vowel trajectories of old adults in NC.*

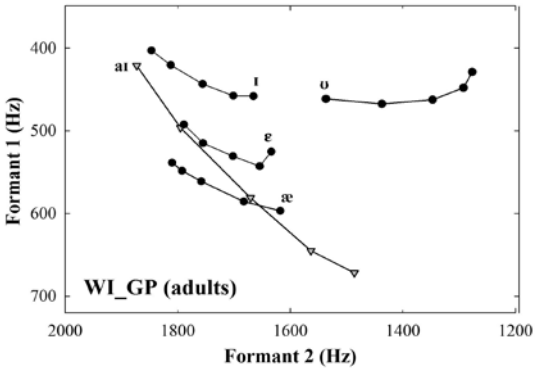

Figure 5: *Vowel trajectories of old adults in WI.*

As can be seen, there are notable dialectal differences in both the positions of the vowels in the F1 and F2 plane, and in the amount and nature of their formant dynamics. The dynamics of the front vowels in *hid, head* and *had* in NC show a characteristic mark of Southern breaking (the vowel "breaks" into two parts) and this feature is absent in WI vowels [11]. A robust difference can also be seen in the dynamics of the vowel in *hide*, which is produced as a monophthong [a] in NC and as a full diphthong [aɪ] in WI. The monophthongal variant is considered a hallmark of Southern American English and this feature was salient even for the current NC listeners who otherwise performed at chance level. On the other hand, the lower scores for WI listeners may reflect their confusion of the vowel with the true monophthongal [ɑ] in *hod*, and their relative unfamiliarity with this Southern feature.

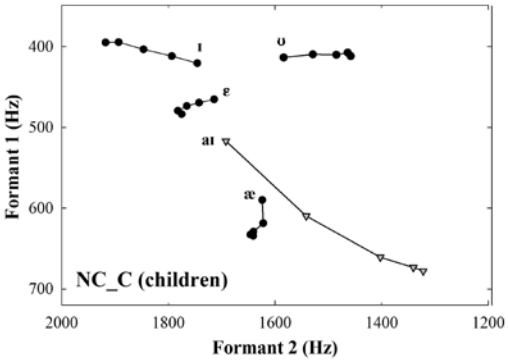

Figure 6: *Vowel trajectories of NC children.*

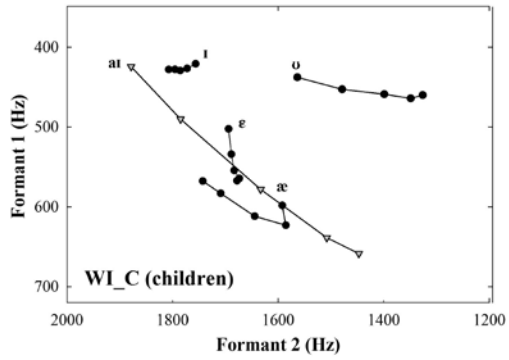

Figure 7: *Vowel trajectories of WI children.*

Figures 6 and 7 show the corresponding vowel plots for children in the C-groups. The generational changes in both the positions of vowels in the vowel space and in their formant dynamics are apparent, signifying the operation of a diachronic sound change in each speech community. Although, overall, children's productions in both dialects are comparatively more "monophthongal," it is apparent that the Southern variant of the vowel in *hide* in NC children undergoes a diphthongization and approximates the formant movement of that spoken in WI.

3. General discussion

The main finding of this study was that WI children were able to distinguish between the two dialects whereas NC children were mostly unable to do so. Furthermore, WI children, and to much lesser extent NC children, were sensitive to changes in the pronunciation patterns in their respective speech communities as a function of a diachronic sound change. In particular, the older versions of both dialects spoken by adult generations supplied more of the dialect-inherent cues that facilitated dialect identification than the newer versions spoken by children.

How can we explain the discrepancy in the perceptual performance of NC and WI children? The most compelling explanation is that WI children were relatively unfamiliar with the Southern dialect and thus could more readily detect the differences between the characteristic features of both varieties. Importantly, they were able to categorize dialects while coping with extensive talker variability in the input. However, NC children have been exposed to a comparatively greater dialect variation over the course of their development. It is the case that the stereotypic Southern features are gradually shifting toward mainstream American English forms, and this is also true for the relatively isolated speech community in Western North Carolina studied here. The cultural change is particularly apparent in younger adults and children, whose increased mobility, exposure to mainstream media, modern lifestyle and education brought about significant changes in pronunciation patterns. Growing up in a socio-culturally changing environment, the current NC children might have lost a sense of what constitutes a "local" versus a "non-local" dialect. Consequently, after being exposed to a great variety of pronunciation features in the speech of older and younger generations, it might be difficult for them to decide which features "belong" to a particular dialect.

One of the aims of the current study was to better understand socio-cultural awareness of older children. Both groups of children demonstrated this knowledge in very different ways. WI children were more readily responsive to the acoustic markers of the two dialects and were able to associate particular acoustic cues with dialect and talker generation. NC children might have implicit knowledge of dialect and generational variations, given the previous findings that children's identification rates were significantly higher for vowels in their native dialect [9]. However, when explicitly asked to distinguish between two dialects, the same children may find the task overly difficult. This is because their socio-cultural awareness admits a wider range of pronunciation possibilities, which may be associated with context-dependent situated interactions rather than with a particular dialect.

The study also found that the performance of WI children was still not adult-like when actually compared with adult listeners responding to the same stimulus material [12]. This suggests that perceptual abilities of older children are still maturing and developmental improvements are still to be expected, perhaps well into adolescence, as has been suggested elsewhere [6, 9].

4. Acknowledgements

This work was supported by NIDCD/NIH research grant R01DC006871.

5. References

[1] C. Clopper and D. Pisoni, "Perception of dialect variation," in D. Pisoni and R. Remez (eds.), *The Handbook of Speech Perception*, pp. 313-337. Oxford: Blackwell Publishing, 2005.

[2] C. Kitamura, R. Panneton, M. Deihl, and A. Notley, "Attuning to the native dialect: When more means less," in *Proceedings of the*

Eleventh Australian International Conference on Speech Science & Technology, Auckland, New Zealand, 2006.

[3] R. Schmale, A. Cristia, A. Seidl, and E. K. Johnson. "Developmental changes in infants' ability to cope with dialect variation in word recognition," *Infancy*, 15, pp. 650-662, 2010.

[4] C. Best, M. Tyler, T. Gooding, C. Orlando, and C. Quann, "Development of phonological constancy: Toddlers' perception of native- and Jamaican-accented words," *Psychological Science,* 20, pp. 539-542, 2009.

[5] K. Kinzler, K. Corriveau, and P. Harris. "Children's selective trust in native-accented speakers," *Developmental Science*, 14, pp. 106-111, 2011.

[6] Z. Jones, Q. Yan, L. Wagner, and C. Clopper. "The development of dialect classification across the lifespan," *Journal of Phonetics,* 60, pp. 20-37, 2017.

[7] K. Kinzler, and J. DeJesus, "Northern=smart and Southern=nice: The development of accent attitudes in the United States," *The Quarterly Journal of Experimental Psychology*, 66, pp. 1146-1158, 2013.

[8] C. O'Connor, and F. Gibbon. "Familiarity of speaker accent on Irish children's performance on a sentence comprehension task," *Journal of Clinical Speech and Language Studies*, vol. 18, pp. 1-17, 2010-2011.

[9] E. Jacewicz and R. A. Fox. "The effects of indexical and phonetic variation on vowel perception in typically developing 9- to 12-year old children," *Journal of Speech, Language and Hearing Research*, 57, pp. 389-405, 2014.

[10] E. Jacewicz and R. A. Fox. "The effects of cross-generational variation on vowel identification and classification," *Journal of the Acoustical Society of America*, 131, pp. 1413-1433, 2012.

[11] W. Labov, S. Ash, and C. Boberg, *Atlas of North American English: Phonetics, Phonology, and Sound Change*. Berlin: Mouton de Gruyter, 2006.

[12] R. A. Fox and E. Jacewicz. "Perceptual sensitivity to dialectal and generational variations in vowels," in *INTERSPEECH 2011 – Annual Conference of the International Speech Communication Association*, *Florence, Italy, Proceedings*, 2011, pp. 2921-2924.

[13] N. A. Macmillan and C.D. Creelman. *Detection Theory: A User's Guide, 2nd ed.* Mahwah, NJ: Erlbaum, 2005.

[14] B. Lobanov. "Classification of Russian vowels spoken by different speakers," *Journal of the Acoustical Society of America*, 49, pp. 606-608, 1971.

[15] E. Thomas and T. Kendall. "NORM: The vowel normalization and plotting suite." 2007. Available at: http://ncslaap.lib.ncsu.edu/tools/norm/.

[16] E. Jacewicz, R. A. Fox, and J. Salmons. "Cross-generational vowel change in American English," *Language Variation and Change,* 23, pp. 45-86, 2011.

[17] R.A. Fox and E. Jacewicz. "Cross-dialectal variation in formant dynamics of American English vowels," *Journal of the Acoustical Society of America*, 126, pp. 2603-2618, 2009.