Volume 42 http://acousticalsociety.org/

179th Meeting of the Acoustical Society of America

Acoustics Virtually Everywhere

7-11 December 2020

Speech Communication: Paper 1pSCb10

Effects of dialect and age on the production of voiced stop closures in running speech

Lian J. Arzbecker, Ewa Jacewicz and Robert A. Fox
Department of Speech and Hearing Science, The Ohio State University, Columbus, OH, 43221; arzbecker.1@osu.edu; jacewicz.1@osu.edu; fox.2@osu.edu

Sociophonetic literature investigating consonantal variation in American English has revealed that the production of voiced stop closures in the intersonorant position differs systematically by region. This indicates that sociocultural factors may influence the phonetic realization of voiced stops. The current study examines whether the implementation of closure voicing in running speech by older children is affected by regional variation. Acoustic measurements included stop closure duration and closure voicing duration (absolute and proportional) obtained from read sentences. The prosodic prominence of the target word was varied, introducing an additional source of variation. Our sample included 48 girls and 36 women, each representing one of three US dialect regions: Midland, North, or South. We analyzed a total of 3,524 tokens. As predicted, children produced stops with less consistency than adults. Children's comparatively longer durations and greater variability—especially as a feature of lexical stress—are indicative of their more effortful productions and developing mastery of lexical stress contrastivity. Significant interactions between dialect and age group showed that children from the South dialect region deviated from their adult counterparts and implemented voicing akin to their peers from the Midland and North.

Published by the Acoustical Society of America

© 2021 Acoustical Society of America. https://doi.org/10.1121/2.0001427

1. INTRODUCTION

The production of voiced stops (i.e., /b, d, g/) in American English (AE) can vary as a function of sentential position and lexical stress, among other factors.[1] Regional dialect can introduce additional phonetic variations. For example, several studies have indicated that speakers of AE in the South produce more stops with prevoicing, nasalization, and short-lag voice onset time (VOT) relative to speakers of other regional varieties[2,3,4] (see Figure 1). Sustained voicing during a stop closure—often termed "unbroken voicing"[5]—is another hallmark of Southern AE, and this variation may be active and systematic, functioning as a marker of social identity that conveys group association.[6] One study found that speakers of Southern heritage who reside in the West dialect region—specifically, California's Central Valley—produce stops with greater voicing intensity throughout the closure.[7] This historically Southern feature (i.e., greater intensity of closure voicing) was primarily used by individuals whose livelihood depends on agriculture, suggesting an influence of social factors on realization of stop closure voicing.

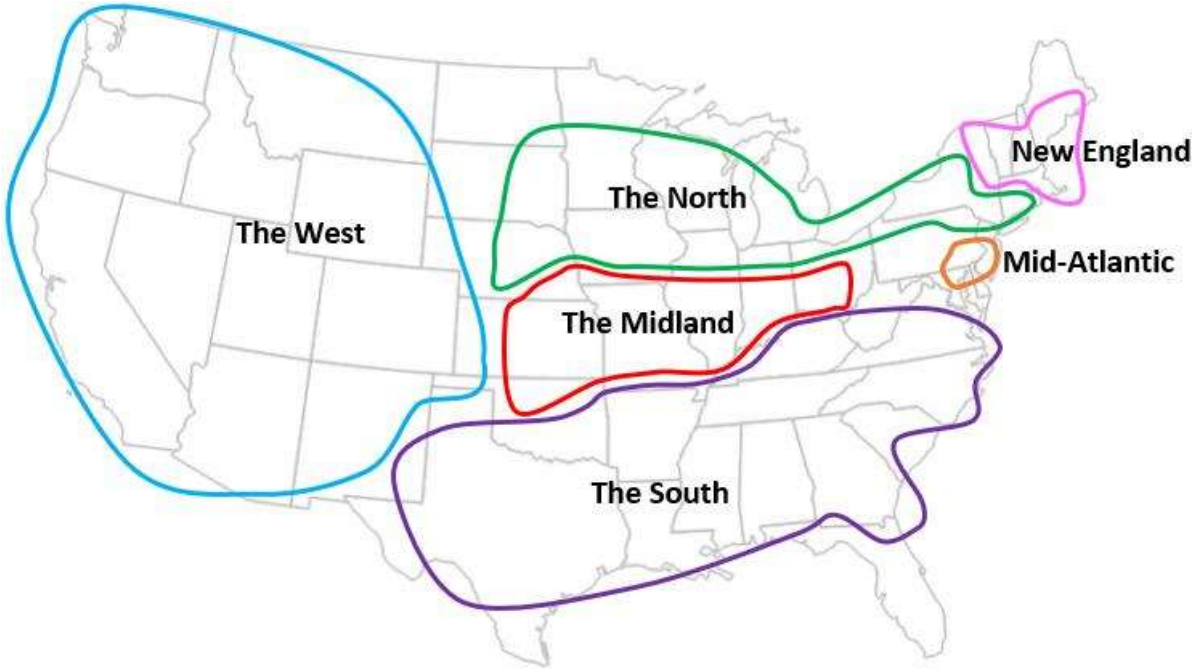


Figure 1. Map of AE dialect regions in the United States.[8,9]

It has been established that children produce stops with larger intraspeaker variability (e.g., duration of VOT, duration of voicing, shape of intraoral pressure contour) when compared to adults.[10,11] As children mature and their production consistency improves, they approach adult-like articulatory control—but not necessarily adult-like articulatory patterns. Regarding social factors, evidence exists that children's acquisition of local dialect patterns can be altered by their increased exposure to language models different from those spoken at home.[12] For example, older children growing up in a complex linguistic environment in the town of Milton Keynes, UK showed dialect leveling that trended toward dialect forms present in the community. But very young children, who were primarily exposed to the dialect of their parents, were less affected by the variation in the local speech.[13] The current study investigates developmental and regional factors that can influence older children's implementation of stop consonant voicing in running speech. Do children acquire the regional variants of stop voicing that are characteristic of their surrounding communities, or do they diverge from local patterns in favor of a more mainstream variety of AE (e.g., that which is spoken at school and in the media)? To explore the possibility that variation in stop closure voicing serves as a marker of social identity, adult and child productions of the voiced stop consonant /b/ are analyzed from three AE dialect regions. While we expect children to have acquired the between-category voicing contrast, we predict their within-category productions will still exhibit less consistency than adults. We also anticipate at least some regional influences on children's implementation of voicing during a voiced stop closure because of their developing awareness and execution of socially motivated speech variability.

2. METHODS

A. PARTICIPANTS

There were 84 female speakers in this study—36 women and 48 girls—each representing one of three AE dialect regions: Midland, North, or South. Table 1 provides a summary of participants' ages, broken down by dialect region and age group. A speaker was considered representative of her region if she were born, raised, and had lived the majority of her life in that speech community. Of the 36 adults, 16 comprised the Midland dialect region sample (central Ohio [OH]), 10 comprised the North sample (south-central Wisconsin [WI]), and 10 comprised the South sample (western North Carolina [NC]). Of the 48 children, 16 comprised each of the dialect groups. The nature of the experimental task required each to be a fluent reader. None of the participants reported a history of speech or hearing disorders.

Table 1. Average group ages in years (SD).

Dialect	Adults (N = 36)	Children (N = 48)
OH	58.5 (5.9)	10.2 (1.5)
WI	57.9 (4.1)	9.3 (1.5)
NC	58.5 (2.4)	10.3 (1.4)
Dialects combined	58.3 (4.5)	9.9 (1.5)

B. STIMULI

To elicit the voiced stops, ten target words were selected: *bits*, *bids*, *baits*, *bades*, *bets*, *beds*, *bats*, *bads*, *bites*, and *bides*. Figure 2 summarizes the vowel (syllable nucleus) and coda voicing alternations used in the creation of the stimulus set. Speakers were provided with the following definitions for the nonsense nouns: *bade*, a brand-name knife; *bad,* a blunder; and *bide*, a small animal.

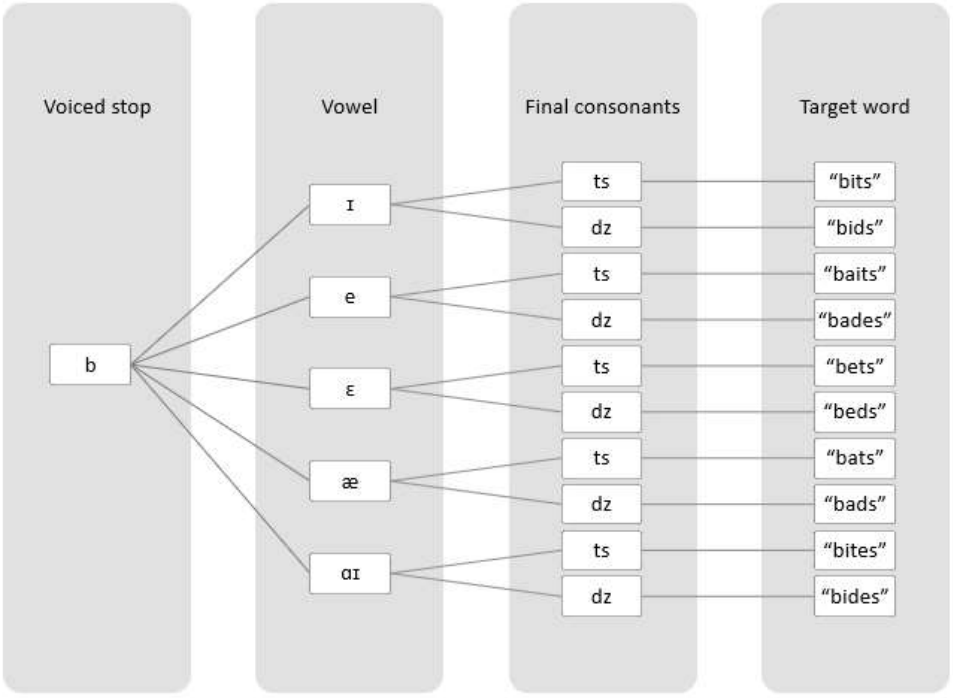


Figure 2. Target word creation.

To balance experimental control with ecological validity, the target words were inserted into sentence pairs. This two-sentence paradigm provided a more controlled environment than spontaneous speech, but nonetheless enabled production of variability that is difficult to elicit in words read as isolated items. Speaker hesitations and mistakes were more common in the first sentence of a set; thus, we chose to analyze the variable of interest—/b/ in the intersonorant position (i.e., between the liquid /l/ and a vowel)—in second sentences only. The systematic variation of prosodic prominence was achieved by instructing participants to emphasize a specific word in each sentence.[14] This was indicated for the speaker by all capital letters. A three-level stress gradient (high, intermediate, and low) was created by systematically varying the location of the word to be emphasized. Levels were defined as follows: high, when the target word was emphasized; intermediate, when the preceding word was emphasized; low, when the sentence-final word was emphasized. Specific examples are shown in Table 2.

Table 2. Systematic variation of prosodic prominence.

Stress level	Sentence pair
High	• John knows the small SCREWS are sharp. • No! John knows the small BITS are sharp.
Intermediate	• John knows the SOFT bits are sharp. • No! John knows the SMALL bits are sharp.
Low	• John knows the small bits are DULL. • No! John knows the small bits are SHARP.

C. RECORDING PROCEDURE

Recordings were obtained in a sound-attenuating booth or in a quiet room, at a sampling rate of 44.1 kHz via a head-mounted microphone. Participants were seated facing a computer monitor that displayed the sentence sets in a randomized order as controlled by a custom MATLAB program. Speakers were instructed to read each sentence set silently to themselves, then to read it aloud. Repeat recordings were allowed—and even prompted by the researcher to ensure accurate production in the case of mispronunciations, disfluency, or erroneous word emphasis. Each participant recorded 30 unique sentence sets, with most adults providing two repetitions of the set. A total of 3,540 tokens was obtained. After excluding extreme outliers (e.g., recordings with artifact or respiratory irregularities, as confirmed by visual and auditory reinspection) the final set for analysis included 3,524 tokens.

D. ACOUSTIC ANALYSIS

Prior to analysis, the audio files were downsampled to 11.025 kHz and pre-emphasized (98%). The acoustic measurements of stop closure voicing were based on four acoustic landmarks: stop closure onset, voicing offset during the closure, closure release, and voicing onset for the vowels (see Figure 3). These landmarks were identified manually from waveform displays in Adobe Audition 1.0 and cross-checked with spectrograms. Stop closure onset was defined by a decrease in waveform amplitude of the voiced vowel and shift towards a sinusoidal period of voicing during the closure. This was corroborated by a drop in formant intensity displayed on the spectrogram. Offset of voicing during the closure was determined by diminishing periodicity of the waveform. Stop closure release was defined by the first transient, often a clear spike of energy. Voicing onset for the vowel was indicated by the beginning of regular periodicity in the waveform following the release burst. All landmarks were verified by using another custom program in MATLAB that superimposed the researcher-identified measurements as lines on the waveform. If needed, final adjustments were made by simultaneously comparing the displays in MATLAB and TF32,[15] a speech analysis program.

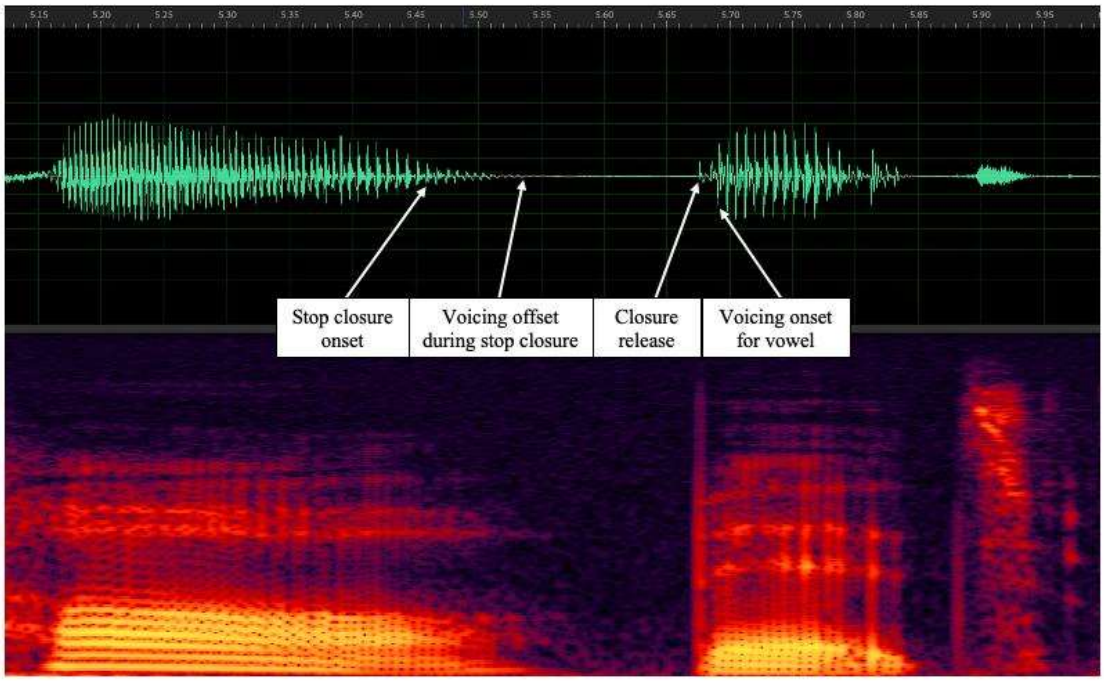


Figure 3. Adobe Audition 1.0 display of a stop closure /b/ occurring in the intersonorant position in "small bats". Shown are the acoustic landmarks used in the analysis of stop voicing.

3. RESULTS

Based on the landmarks, the acoustic variables of interest included closure duration and voicing duration during the closure, expressed in absolute (milliseconds) and proportional terms (proportion of closure duration to word duration, and proportion of voicing duration to closure duration). These four acoustic measures were included as dependent variables in the statistical analyses. VOT was not analyzed in the current study. Preliminary statistical modeling informed the decision to analyze the effects of stress level (high, intermediate, low) and coda consonants ([t], [d]) in separate linear mixed models. All data analyses and visualizations were run in R. For the effects of stress level, the models were fitted for each dependent variable separately by adding age group, dialect, stress level, and their interactions as fixed-effects predictors. For speaker-level effects, models were fitted with random intercepts and random slopes, assessing the addition of by-speaker slopes on final model fit. Closure duration and voicing duration during the closure were modeled using the lme4 package.[16] In these models, significance was evaluated by applying the Satterthwaite approximation of degrees of freedom for *t*-tests and *F*-tests, implemented using the lmerTest package.[17] Decisions about significance were based on Type III ANOVA and backward elimination of non-significant effects. Proportion variables (i.e., proportion of closure to word and proportion of voicing during the closure) were modeled using the glmmTMB package.[18] The beta regression models were fitted using maximum likelihood estimation via the TMB package,[19] using *beta_family* function and logit link for the conditional mean. Significance was evaluated by applying the *z* distribution to the Wald *t*-values and from a Type III ANOVA table (*Anova.glmmTMB* function).

The four separate models for the effects of coda consonants were fitted exactly as for the effects of stress level, with coda consonants instead of stress level as a fixed-effects predictor. In sections A–D, we present the main results for the effect of stress level separately for each dependent variable. In section E, we highlight the key findings for the effects of coda consonants and present preliminary results of additional analyses.

A. CLOSURE DURATION

Compared to adults, children produced stop closures with longer absolute durations and greater interspeaker variability (Figure 4). The main effect of dialect was not significant, but there was a significant interaction between dialect and age group [$F(2, 68) = 4.32$, $p = 0.017$]. The interaction arose because the difference between adults and children was significant for WI and NC, but not OH. Furthermore, the main effect of stress level was significant [$F(2, 74) = 177.48$, $p < 0.001$] and pairwise comparisons showed that all three SWPs differed significantly from one another ($p < 0.001$), with longer closure durations found in higher emphasis positions. There was also a significant interaction between age group and stress level [$F(2, 74) = 13.77$, $p < 0.001$]. The interaction revealed greater differences between adults and children in the high stress position as compared to the intermediate and low positions.

B. PROPORTION OF CLOSURE DURATION TO WORD DURATION

Addressing individual and group differences in speaking rate,[20] closure duration was also analyzed as a proportion in the word. The results showed that even when closure duration was assessed in relative terms, a significant interaction still arose between dialect and age group [$\chi^2(2) = 11.97$, $p = 0.003$]. Adults produced closures with shorter durations than children for all dialects, and the difference between NC adults and NC children was the most striking (i.e., NC adults' closures were significantly shorter than NC children's closures). These findings were consistent with the results for absolute duration. The main effect of stress level was also significant—but only the high SWP differed significantly from any other position ($p < 0.001$), as the differences between the intermediate and low SWPs were reduced and not significant.(Figure 5). Overall, across the three stress levels, children's proportions exhibited greater variability than adults' proportions.

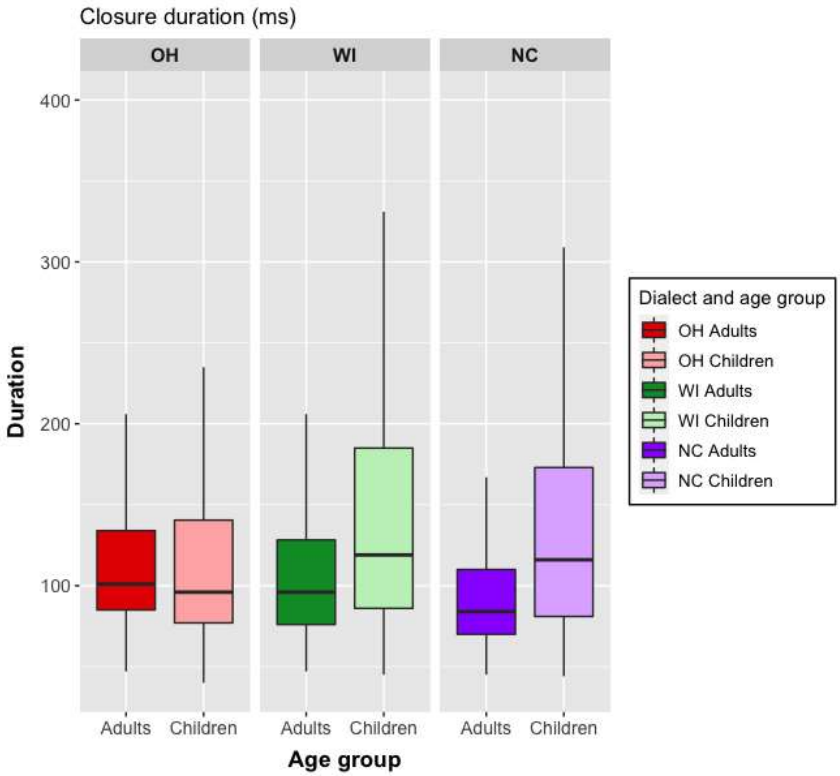


Figure 4. Closure duration (in ms) by dialect and age group.

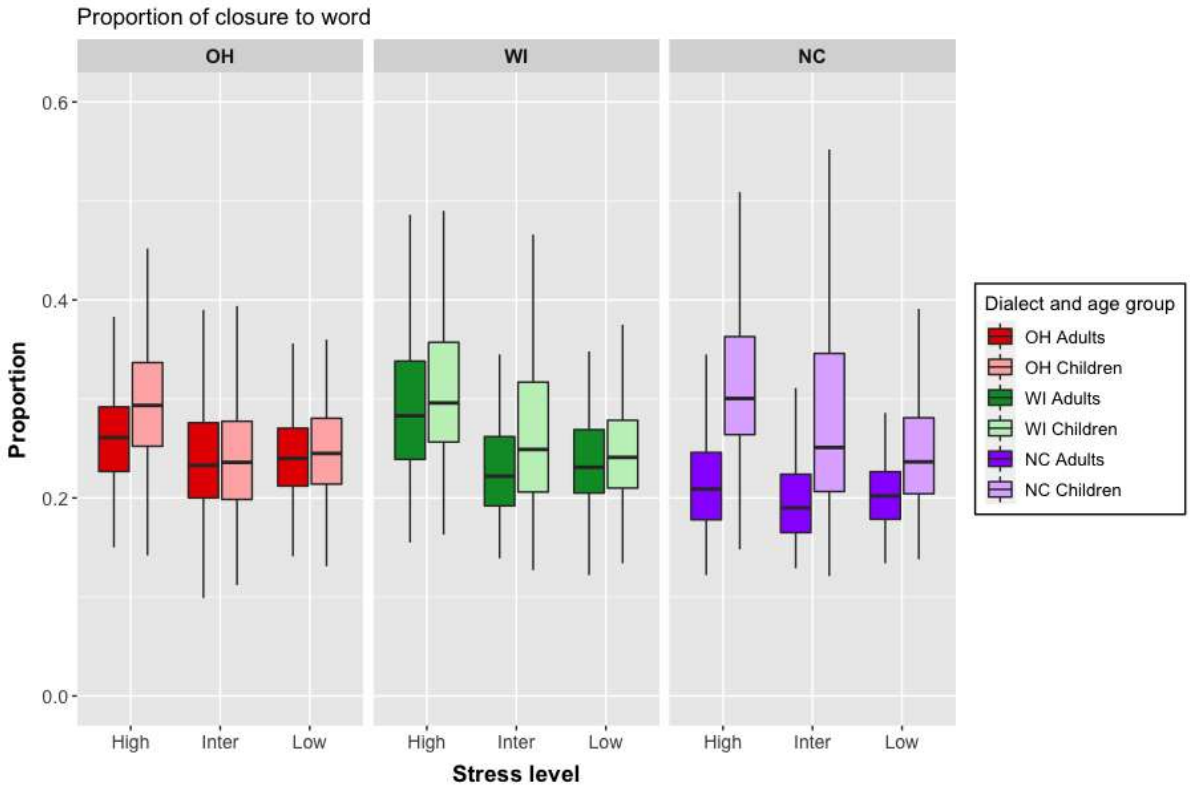


Figure 5. Proportion of closure to word duration by dialect, age group, and stress level.

C. CLOSURE VOICING DURATION

For all three dialects, adults produced closures with significantly greater voicing durations compared to children (Figure 6). There was a significant interaction between age group and stress [$F(2, 72) = 1.53$, $p < 0.001$], with adults producing longer closure voicing in the high stress position compared to the intermediate and low positions, respectively. Children's voicing closure durations did not show this effect, and unlike closure duration, their closure voicing durations were not significantly affected by stress level or dialect.

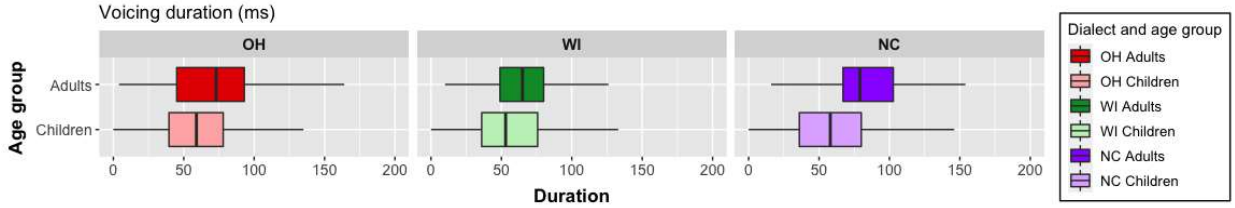


Figure 6. Voicing during closure (in ms) by dialect and age group.

D. PROPORTION OF VOICING DURATION TO CLOSURE DURATION

After adjusting for closure duration, the relative measure of voicing duration (i.e., the proportion of voicing during the closure) yielded markedly different results from its raw values in milliseconds. A significant interaction between age group and stress level [$\chi^2(2) = 36.83$, $p < 0.001$] predicted a gradient increase in the amount of voicing with each lower SWP for children (Welch's t-test, $p < 0.001$ for all), but the difference between high and intermediate SWPs only approached significance for adults ($p = 0.013$, Bonferroni-corrected at $\alpha = 0.013$). The model fit was further improved by a significant dialect by age group interaction [$\chi^2(2) = 17.44$, $p < 0.001$], predicting more voicing for NC adults than for any other group. Since NC adults produced almost fully voiced closures, there was little room for effects of stress position. Therefore, they differed significantly from OH and WI (Figure 7). Importantly, there were no significant differences between any of the child groups, suggesting that NC children may have diverged from the voicing patterns of NC adults.

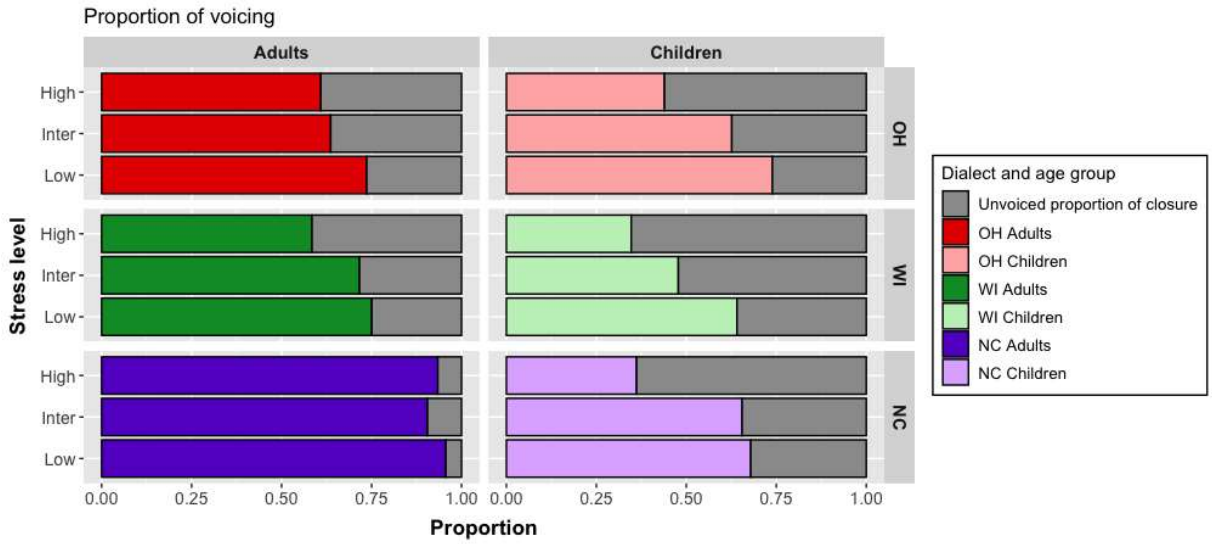


Figure 7. Proportion of voicing during closure by dialect, age group, and stress level.

E. FURTHER ANALYSES OF CLOSURE VOICING

Regarding the effects of coda voicing, both age groups adhered to the same trend: closure durations were longer for the voiceless consonant (i.e., [t]) than they were for the voiced consonant (i.e., [d]). However, stop closure durations (in ms) for the children were greater than those for the adults. The model for the proportional measure of closure voicing included two significant interactions, dialect by age group, and coda by age group voicing. These two interactions revealed that the proportion of voicing during the closure was smaller for the voiceless coda and larger for the voiced coda ($p = 0.005$), and that this pattern was consistent for all dialect and age groups—except for NC adults who differed significantly from any other group ($p < 0.001$). Again, NC adults produced almost entirely fully voiced closures, and were therefore minimally affected by variation as a function of coda voicing (Figure 8).

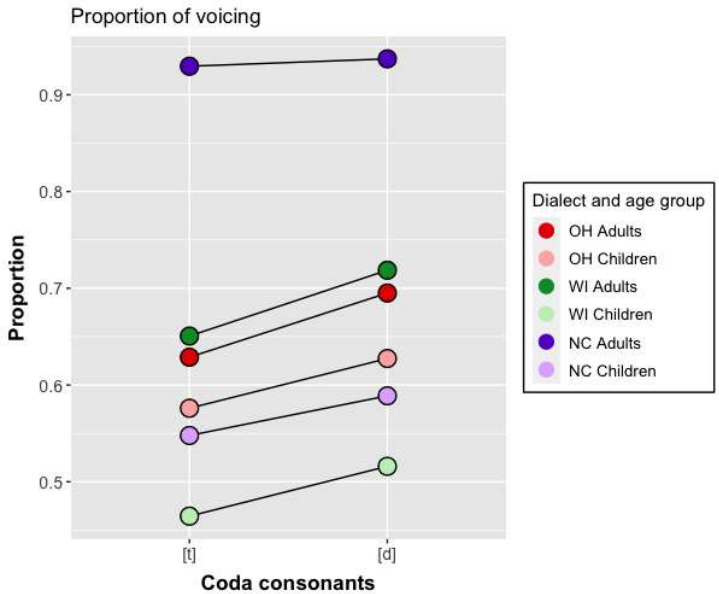


Figure 8. Proportion of voicing in closure by coda consonant, dialect, and age group.

The frequency of fully voiced closures (i.e., continuous voicing) was calculated for each dialect and age group. NC adults had the greatest number of fully voiced closures (Figure 9, left panel). NC adults also had the highest rate of nasalized closures (Figure 9, right panel).

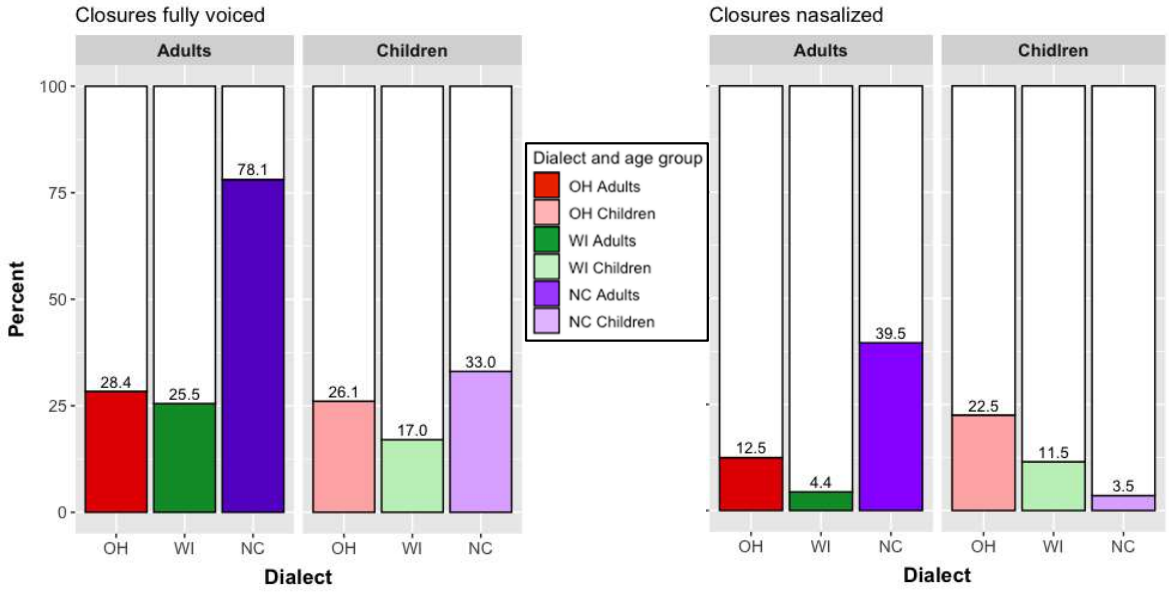


Figure 9. Left panel: percentage of fully voiced closures by dialect and age group. Right panel: percentage of closures with voicing present that were also nasalized.

4. DISCUSSION

As expected, children produced closures with significantly longer durations than adults—both in absolute terms and as a proportion in a word. The raw closure durations varied systematically with stress level across all dialect and age groups; the longest closures were associated with high prosodic prominence, followed by intermediate and low, respectively. The proportional measure of closure duration best reflected the differences between adults and children, with NC adults and NC children showing the greatest age group differences compared to the other two dialect-age group pairings. This result indicates that NC children may be diverging from the local variety in favor of a mainstream AE (e.g., the dialect spoken in the Midland and North). The proportional measure revealed that the contrast between intermediate and low stress positions was not significant for adults of any dialect; however, it was significant for WI and NC children—but not for OH children. OH children's prosodic contrasts were similar to the adult groups, suggesting that they may follow a different developmental trajectory than their Northern and Southern AE-speaking peers.

While voicing duration was also analyzed in absolute and relative terms, the proportional measure best displayed the relationship between closure voicing and dialects and age groups. For all participants, prosodic prominence yielded the same pattern: the greatest proportion of voicing in the closure was in the low stress positions (when closures were shortest) than the intermediate and high positions, respectively. Akin to closure duration, the smaller proportions of closure voicing duration in children may be indicative of a more effortful production relative to the adults, as suggested by their comparatively longer closure gap. Only adult closure voicing was significantly affected by dialect. Specifically, NC adults produced largely fully voiced closures, leaving little room for the effects of prosodic prominence or variations as a function of coda voicing. This voicing pattern was markedly different from either OH or WI adults, whose partially voiced closures were consistent with the underlying breath-stream control mechanism proposed for mainstream AE.[21] The results for children showed that this mechanism was also effective in all child groups, so that NC children produced general voicing patterns similar to those of OH and WI children. However, it is noteworthy that NC children exhibited the greatest frequency of unbroken voicing for all child groups. This suggests that NC children may be diverging from regionally marked patterns, but some characteristics of their stop closure productions retain aspects of their local dialect.

5. CONCLUSIONS

Compared to adults, older children produced the voiced stop /b/ with greater variability and longer closure duration. Children also differed from adults in their execution of varying levels of prosodic prominence, suggesting that their productions are still not adult-like. NC children appeared to deviate from two locally marked NC adult patterns of closure voicing (i.e., unbroken voicing and nasalization), following the voicing patterns of mainstream AE. However, traces of the regional NC dialect could still be found in their speech.

REFERENCES

[1]L. Davidson, "Variability in the implementation of voicing in American English obstruents," J. Phon. **54**, 35–50 (2016). https://doi.org/10.1016/j.wocn.2015.09.003

[2]W. Herd, "Sociophonetic voice onset time variation in Mississippi English," J. Acoust. Soc. Am. **147**(1), 596–605 (2020). https://doi.org/10.1121/10.0000545

[3]L. Hunnicutt and P. A. Morris, "Prevoicing and aspiration in Southern American English," Penn. Working Papers in Ling. **22**(1), 215–224 (2016). https://repository.upenn.edu/pwpl/vol22/iss1/24/ (Last viewed May 18, 2021).

[4]A. Walker, "Voiced stops in the command performance of Southern US English," J. Acoust. Soc. Am. **147**(1), 606–615 (2020). https://doi.org/10.1121/10.0000552

[5]L. Lisker and A. Abramson, "A cross-language study of voicing in initial stops: Acoustical measurements," WORD **20**(3), 384–422 (1964). https://doi.org/10.1080/00437956.1964.11659830

[6]E. Jacewicz, R. A. Fox, and S. Lyle, "Variation in stop consonant voicing in two regional varieties of American English," J. Int. Phon. Assoc. **39**(3), 313–334 (2009). https://doi.org/10.1017/S0025100309990156

[7]R. J. Podesva, P. Eckert, J. Fine, K. Hilton, and S. Jeong, "Social influences on the degree of stop voicing in inland California," Penn. Working Papers in Ling. **21**(2), 167–176 (2015). https://repository.upenn.edu/pwpl/vol21/iss2/19/ (Last viewed May 18, 2021).

[8]W. Labov, S. Ash, and C. Boberg, "The dialects of North American English," in *Atlas of North American English: Phonetics, phonology and sound change* (Mouton de Gruyter, 2006), Chap. 11, pp. 148. https://doi.org/10.1515/9783110206838

[9]C. G. Clopper, D. B. Pisoni, and K. de Jong, "Acoustic characteristics of the vowel systems of six regional varieties of American English," J. Acoust. Soc. Am. **118**(3), 1661–1676 (2005). https://doi.org/10.1121/1.2000774

[10]J. H. Lowenstein and S. Nittrouer, "Patterns of acquisition of native voice onset time in English learning children," J. Acoust. Soc. Am. **124**(2), 1180–1119 (2008). https://doi.org/10.1121/1.2945118

[11]L. L. Koenig and J. C. Lucero, "Stop consonant voicing and intraoral pressure contours in women and children," J. Acoust. Soc. Am. **123**(2), 1077–1088 (2008). https://doi.org/10.1121/1.2828065

[12]P. Foulkes and G. Docherty, "The social life of phonetics and phonology," J. Phon. **34**(4), 409–438 (2006). https://doi.org/10.1016/j.wocn.2005.08.002

[13]P. Kerswill and A. Williams, "Creating a new town koine: Children and language change in Milton Keynes," Lang. in Soc. **29**(1), 65–115 (2000). https://doi.org/10.1017/s0047404500001020