

Sentence context facilitation for children's and adults' recognition of native- and nonnative-accented speech

Tessa Bent ^a, Rachael Frush Holt ^b, Katherine Miller ^b, and Emma Libersky ^b

^a Department of Speech and Hearing Sciences, Indiana University

200 S. Jordan Ave.

Bloomington, Indiana 47405

^b Department of Speech and Hearing Science, The Ohio State University

110 Pressey Hall

1070 Carmack Road

Columbus, OH 43210

Date submitted: July 5, 2018

Running head: Accent influence on semantic context facilitation

Conflict of interest: No relevant conflicts of interest.

Funding Statement: This work was supported by the National Science Foundation (grant number 1461039).

^{a)} Author to whom correspondence should be addressed. Electronic mail: tbent@indiana.edu

Purpose: Supportive semantic and syntactic information can increase children's and adults' word recognition accuracy in adverse listening conditions. However, there are inconsistent findings regarding how a talker's accent or dialect modulates these context effects. Here, we compare children's and adults' abilities to capitalize on sentence context to overcome misleading acoustic-phonetic cues in nonnative-accented speech.

Method: Monolingual American English speaking 5- to 7-year-old children (n = 90) and 18- to 35-year-old adults (n = 30) were presented with full sentences or the excised final word from each of the sentences and repeated what they heard. Participants were randomly assigned to one of two conditions: native-accented (Midland American English) or nonnative-accented (Spanish- and Japanese-accented English) speech. Participants also completed the NIH Toolbox Picture Vocabulary Test.

Results: Children and adults benefited from sentence context for both native- and nonnative-accent talkers, but the benefit was greater for nonnative than native talkers. Furthermore, adults showed greater context benefit than children for nonnative talkers, but the two age groups showed similar benefit for native talkers. Children's age and vocabulary scores both correlated with context benefit.

Conclusions: The cognitive-linguistic development that occurs between the early school age years and adulthood may increase listeners' abilities to capitalize on top-down cues for lexical identification with nonnative-accented speech. These results have implications for the perception of speech with source degradation, including speech-sound disorders, hearing loss, or signal processing that does not faithfully represent the original signal.

Introduction

Accurately understanding spoken messages in adverse listening conditions involves both the interpretation of bottom-up acoustic-phonetic information and the support of top-down information. Under ideal listening conditions, listeners' reliance on top-down information (e.g., syntactic and semantic cues or real-world knowledge) can be minimal; they can accurately perceive spoken messages by heavily relying on bottom-up acoustic-phonetic information. However, speech communication frequently occurs in suboptimal conditions in which acoustic-phonetic information is masked by noise or other sounds in the environment. Talkers whose pronunciation patterns deviate from the listener's home dialect norms (e.g., unfamiliar regional dialects, nonnative accents, speech disorders) may also limit a listener's ability to rely exclusively on bottom-up information. Under these conditions, listeners may need to use top-down processing to aid in their interpretation of spoken messages. Research from the past 70 years has consistently shown that listeners are able to benefit from top-down information under a variety of suboptimal conditions (Kalikow, Stevens, & Elliott, 1977; Miller, Heise, & Lichten, 1951; Miller & Isard, 1963). In the current study, we investigate how the interaction between the talker's accent (i.e., native talker vs. nonnative talker) and the listener's age (i.e., adult vs. school-age child) impact the use of top-down cues for word identification. The investigation of this issue allows us to test how listeners can utilize top-down information, including semantic and syntactic cues, as well as information about a talker's voice when presented with word pronunciations that do not readily map onto representations of lexical items stored in memory. Even young children appear to have many of the skills that are likely necessary to accurately perceive words that deviate from previously encountered pronunciations, such as lexically-guided phonetic retuning and remapping of categorical phoneme mismatches (McQueen, Tyler, & Cutler, 2012; White &

Aslin, 2011). However, the ability to accurately perceive nonnative-accented sentences is not mature in early childhood and in fact continues to develop into adolescence (Bent, in press). Here, we explore whether one of the factors contributing to the protracted developmental trajectory for perception of unfamiliar accents is the ability to benefit from top-down cues to reinterpret the misleading acoustic-phonetic information frequently present in nonnative speech. Below we review the literature on how the use of top-down cues for word recognition is influenced by a talker's dialect or accent and a listener's age.

Semantic predictability: Talker regional dialect and nonnative accent effects

Talker native language or dialect background and accent strength can influence the extent to which listeners benefit from sentence context. In two studies assessing native- and nonnative-talker intelligibility for true-false statements (most semantic support), semantically meaningful sentences, and semantically anomalous sentences (least semantic support), the effect of semantic predictability was greater for nonnative speakers than native speakers (Behrman & Akhund, 2013; Kennedy & Trofimovich, 2008). Specifically, for the nonnative speakers, intelligibility increased with greater amounts of contextual support, whereas there was no effect for the native speakers, likely due to ceiling effects. Behrman and Akhund (2013) included nonnative talkers with three accent strengths finding that the context effect was attenuated for speakers with milder accents, but again, this result is likely due to ceiling effects (i.e., scores for the mildly-accented nonnative talkers were between 99 - 100% correct). These studies show that adult listeners can take advantage of top-down processing when the bottom-up information presents challenges for accurate word recognition due to the presence of talker pronunciation patterns that deviate from native norms. For regional dialect variation, Clopper (2012) has shown that the semantic

predictability effect is greater for a highly familiar regional dialect (i.e., Standard American English and Northern English) compared to less familiar dialects (i.e., Mid-Atlantic and Southern). In this study, sentences were mixed with noise, ensuring that intelligibility scores were not at ceiling for any of the dialects. Taken together, these studies suggest that future work on the interaction between a talker's accent and a listener's ability to benefit from context should incorporate methodologies that result in intelligibility scores that are not at ceiling (e.g., adding environmental degradation or increasing cognitive load) to better understand how these variables interact at other points on the psychometric function (Hunter & Pisoni, 2018).

Children's use of context cues for word identification

Children are able to take advantage of contextual cues to enhance their word recognition. For example, word recognition is facilitated in children as young as 2 years in sentence contexts compared to words in isolation (Fernald & Hurtado, 2006). Other work has shown that throughout development, children's word recognition is enhanced when supportive semantic and syntactic context is provided compared to less supportive contexts (e.g., Cole & Perfetti, 1980; Elliott, 1979; Fallon, Trehub, & Schneider, 2002). This context benefit for word recognition extends to children with hearing loss (Smiljanic & Sladen, 2013). Studies assessing whether children and adults benefit to a similar extent from semantic context have not reported consistent findings. Some evidence suggests that noise can interfere with children's ability to benefit from supportive context compared with adult listeners (Nittrouer & Boothroyd, 1990), whereas other work suggests that children and adults benefit to a similar degree from semantic context (Fallon, Trehub, & Schneider, 2000; Fallon et al., 2002).

Although a range of studies have demonstrated that, compared to adult listeners, children have more difficulty understanding talkers whose production patterns deviate from home dialect norms (Bent, 2014, in press; Bent & Atagi, 2015, 2017; Bent & Holt, in press), only two studies have investigated how semantic-contextual information influences children's perception of nonnative speech (Creel, Rojo, & Paullada, 2016; Holt & Bent, 2017). In Creel et al. (2016), children between the ages of 3 and 5 years were presented with sensible and nonsensical sentences produced by speakers of California English or Spanish-accented English in either a four-alternative forced-choice task or in a word repetition task. In both tasks, either the whole sentence or only the final word excised from the sentence was presented. Results showed that the sensible sentences only facilitated word recognition when the entire sentence was presented in the forced-choice task. When the entire sentence was presented in word repetition task, children showed equal performance on the two sentence types, and when the final excised word was presented in either task, they were more accurate in the nonsensical condition -- presumably due to talkers' more careful articulation in the nonsensical sentences than in the sensible ones. Furthermore, an effect of talker accent was absent or minimal with closed-set tasks but was relatively robust with the open-set task; however, an analysis of all experiments combined suggested that the greater accent effects in the open-set tasks may be a result of ceiling effects in the closed-set tasks rather than true differences in children's processing of accented speech across the tasks. Holt and Bent (2017) reported similar results in a sentence repetition task using high- and low-predictability sentences produced by native- and Mandarin-accented English talkers whose productions were mixed with noise. Overall, 5- to 7-year-old children were more accurate with the native talker than the Mandarin-accented talker and with the high-predictability sentences than the low-predictability sentences. In both of these studies, no interaction between

talker accent and sentence predictability was found. Taken together, along with van Heugten and Johnson (2016) who showed that toddlers can benefit from sentence context during the perception of an unfamiliar *native* variety, these studies suggest that the influence of top-down cues on children's word recognition may be similar for unfamiliar and familiar accents or dialects.

None of these studies have included comparisons with adults and the age range of the children in each of the studies has been quite narrow. The assessment of children from a narrow age range limits our ability to draw conclusions about how greater linguistic experience, cognitive development, and real-world knowledge may modulate the ability to benefit from these semantic contextual cues for nonnative-accented speech, because these skills and experiences tend to correlate strongly with age. By testing adults using the same materials and methods as the children, developmentalists are able to establish a benchmark for mature performance and can compare children to this performance level. Without knowledge of mature performance (and the range of mature performance) on a specific task, we are limited in the informed conclusions that can be made regarding observed development on that task. Although we may not be able to definitively determine the locus of age effects, this work paves the way for future investigations examining the specific mechanisms responsible for developmental differences.

In this study, we investigate the contribution of age and vocabulary size to listeners' word identification skills by testing both children and adults' perception of native and nonnative speech with and without sentence context. This study also takes a different approach from previous work by purposely selecting nonnative-accented target words that show very poor identification accuracy in isolation, thereby allowing substantial room for sentence context gains (confirmed by pilot testing; details provided below). Because we selected words in which the

incorrect candidate was almost always selected based on the acoustic information alone, the sentence context must be utilized to both inhibit the incorrect lexical item and reinterpret the "misleading" acoustic information to arrive at the target word (Aydelott & Bates, 2004). We test whether the success of this process is modulated by listener age (children vs. adults) and talker accent (native vs. nonnative). We predict that top-down cues will be beneficial regardless of listener age or talker accent. However, the size of the benefit may be modulated by both talker accent and listener age. Top-down processing may be more beneficial during the perception of nonnative- than native-accented speech because nonnative-accented speech includes bottom-up acoustic-phonetic cues that deviate from the majority of previous experienced productions, increasing the listener's need to rely on contextual cues for word identification. Furthermore, because top-down processing requires the application of linguistic and real-world knowledge, children may benefit less from top-down contextual cues than adults.

Method

Participants

Participants included 90 children (45 female, 45 male) between the ages of 5 and 7 years with 30 children in each year bracket and 30 adults (15 female, 15 male) between the ages of 18 and 35 years with an average age of 23 years. Twenty-three additional participants were tested (14 children and 9 adults), but their data were not included due to electing not to complete the experiment ($n = 7$), bilingual status ($n = 5$), extensive exposure to Spanish ($n = 5$) or Japanese ($n = 3$), or a history of speech or hearing problems ($n = 3$). Participants were recruited at the Center of Science and Industry (COSI) in Columbus, Ohio. All participants were monolingual American English speakers with no parental or self-reported history of speech, language, or hearing

problems. Participant dialect was determined based on parental-report for children or self-report for adults. The parents of child participants and adult participants were provided with a dialect map of the United States. They were asked to indicate which one of 10 regional accents, all of which were clearly identified on the map, best described the way their child or they spoke. Most children (n = 62) and adults (n = 18) were speakers of the Central Midland dialect of American English. Most other participants either were speakers of the North Central (children: n = 14; adults: n = 4) or Appalachian (children: n = 6; adults: n = 4) dialects with the few remaining participants from other dialect regions within the U.S. Adults or parents of the child participants also rated their exposure to the regional dialect and nonnative accents included in this study on a scale from 1 to 5, where 1 = no exposure or only brief casual exposure and 5 = daily home exposure. Average ratings for exposure to the Midland dialect were high for both children (M = 4.6, range = 1 - 5) and adults (M = 4.6, range = 2 - 5). Average ratings of exposure for both nonnative accents included in the study were low for both children (Japanese-accented English: M = 1.1, range = 1 - 3; Spanish-accented English: M = 1.5, range = 1 - 3) and adults (Japanese-accented English: M = 1.1, range = 1 - 2; Spanish-accented English: M = 2.1, range = 1 - 3).

Stimuli

Stimuli included 32 sentences selected from the Bamford–Kowal–Bench (BKB) Standard Sentence Test (Bamford & Wilson, 1979), the Hearing in Noise Test - Children's Version (HINT-C; Nilsson, Soli, & Gelnett, 1996), the Pediatric Speech Intelligibility-Sentences (PSI-Sentences; Jerger & Jerger, 1984), and the Lexical Neighborhood Sentence Test (LNST; Eisenberg, Martinez, Holowecky, & Pogorelsky, 2002). Recordings of the sentences were taken from the Hoosier Database of Native and Non-native Speech for Children (Atagi & Bent, 2013;

Bent, 2014) as produced by 2 native speakers of Spanish, 2 native speakers of Japanese, and 4 native speakers of English with a Midland dialect. Talker gender was balanced within accents. The nonnative speakers of English had lived in the United States for one year or less at the time of recording, started learning English between the ages of 13 - 21 years, and were between the ages of 25 - 29 years at the time of recording. Their accents were moderately strong (i.e., ranged from 3.8 - 6.2) based on a rating scale from 1 - 9, where 1 = no foreign accent and 9 = very strong foreign accent (Atagi & Bent, 2013).

To select the experimental stimuli, 64 sentences produced only by the nonnative speakers (16 sentences for each talker) were chosen based on the authors' judgments that the final word in the sentence would likely be misperceived when played in isolation. Final words were excised from the sentences. In pilot testing, the final excised word and sentence in its entirety were presented to a group of 20 monolingual American English-speaking adults for identification (5 listeners for each talker). All words were presented in random order before all sentences, which were also randomized. From this larger set of 64 sentences, eight sentences from each nonnative talker were selected that resulted in the greatest accuracy differences between the isolated word presentation and the word-in-sentence presentation. It should be noted here that this method for selecting the stimuli may have inflated effect sizes for the adult nonnative condition. Using a different method for stimulus selection, such as selecting stimuli based on word-in-isolation accuracy alone, native talker pilot results, or pilot data from child listeners, may result in different outcomes or smaller effect sizes than reported here.

The final stimulus set included 32 unique sentences and 32 words that were the excised final words from the sentences. The set of native stimuli included the same sentences and

corresponding excised words. The sentences and excised words were equated for total root-mean-square amplitude using Adobe Audition (Adobe System Corporation, San Jose, CA).

Procedures

Testing occurred in the Language Sciences Research Laboratory at the Center of Science and Industry. Stimuli were presented binaurally at a comfortable level under Audiotechnica headphones (model 8TH-770COM) using E-Prime, Version 2.0.8.22 (Psychology Software Tools, 2007) on a Dell Optiplex 790 desktop computer. After the consent process, adult participants and the parents of child participants completed a demographic and language background questionnaire.

Participants were randomly assigned to either the native or the nonnative condition with 80 listeners in the nonnative condition (60 children with 20 for each year, and 20 adults) and 40 listeners in the native condition (30 children with 10 for each year, and 10 adults). Participants were presented with two blocks of 32 trials each in the following order: (a) words in isolation and (b) sentences. Talkers and items were randomized within each block for each participant, but all participants received the word-in-isolation block before the sentence block. Participants were either instructed to repeat the word for the word-in-isolation block or to repeat as much of the sentence as possible for the sentence block. No feedback was provided, but participants received general encouragement.

Between the word-in-isolation and sentence blocks, the NIH Toolbox Picture Vocabulary Test (iPad format) (Gershon et al., 2013; National Institutes of Health and Northwestern University) was administered. On each trial of this adaptive test, an audio recording of a word is

played, and the participant must choose the picture that best matches the word by touching one of four pictures on the screen. Testing time is approximately five minutes.

For the speech perception component, responses were recorded by hand by the experimenter at the time of testing. Scoring was completed off-line. Any questionable articulations on the part of the younger children were clarified by asking follow-up questions, asking the child to point, or to describe or repeat the word they said. The NIH Toolbox Picture Vocabulary Test is scored within the app based on participant performance to the adaptive test. For each participant, an age-corrected standard score is derived using normative mean of 100 and standard deviation of 15. At the conclusion of testing, the purpose of the study was explained to the participant and/or the participant's parent.

Results

Word identification accuracy was calculated for all words in the word-in-isolation condition and for the sentence-final word in the sentence condition. Words were scored with a strict scoring criterion. Words with added or deleted morphemes were counted as incorrect. For example, if a participant responded "play" for the target word "playing", their response would be counted as incorrect due to the deletion of the progressive morpheme (i.e., "-ing"). Similarly, if a participant responded "balls" for "ball" their response would be incorrect due to the addition of the plural "-s" morpheme. Accuracy scores were converted to rationalize arcsine units (RAU) to facilitate meaningful comparisons across the entire range of the scale (Studebaker, 1985). Note that analyses were carried out on transformed data, but figures are presented in percent correct for readability.

Word recognition accuracy

The first analysis focused on the comparison between the adults and the children (Figure 1). Data were analyzed with a repeated-measures ANOVA with two between-subjects variables (talker accent: native or nonnative; listener age: adult or child) and one within-subject variable (context condition: word or sentence). All three main effects were significant in the expected directions. Listeners were more accurate in the sentence condition than in the isolated word condition, $F(1, 116) = 85.30, p < .001, \eta_p^2 = .869$. Adults were more accurate than children, $F(1, 116) = 771.31, p < .001, \eta_p^2 = .424$. Listeners were more accurate in the native condition than in the nonnative condition, $F(1, 116) = 930.34, p < .001, \eta_p^2 = .889$. All two-way interactions were also significant. The context condition by listener age interaction arose because adults benefited more from sentence context relative to the words in isolation (55 RAU difference) than children (42 RAU difference), $F(1, 116) = 10.96, p = .001, \eta_p^2 = .086$. The condition by talker accent interaction arose because listeners showed a greater benefit from sentence context for the nonnative condition (51 RAU difference) than the native condition (34 RAU difference), $F(1, 116) = 45.14, p < .001, \eta_p^2 = .280$. The listener age by talker accent interaction arose because the performance difference between the children and adults was larger for the nonnative talker (25 RAU difference) than the native talker (14 RAU), $F(1, 116) = 7.07, p = .009, \eta_p^2 = .057$. Finally, the three-way interaction was significant, $F(1, 116) = 6.93, p = .01, \eta_p^2 = .056$. To facilitate the interpretation of the three-way interaction, two follow-up ANOVAs were conducted, one for the native talker condition and one for the nonnative talker condition. For both of these ANOVAs, there were significant main effects of context and listener age (all $p < .001$). In the nonnative condition, there was an interaction between context condition and listener age, $F(1, 78) = 7.07, p < .001, \eta_p^2 = .250$, with children showing smaller gains between the sentence and the word-in-

isolation condition (46 RAU) compared to the adults (66 RAU). However, in the native condition the interaction between context condition and age was not significant, $F(1, 38) = 0.18$, $p = .673$, $\eta_p^2 = .005$; adults and children had similar differences between the sentence and word-in-isolation conditions (35 and 33 RAU, respectively). These results suggest that the children and adults could equally capitalize on sentence context in the native talker condition and thus showed similar gains from the addition of sentence context compared to the word-in-isolation condition. For the nonnative talker condition, both the children and adults were more accurate in the sentence condition compared to the word-in-isolation condition, but the adults showed larger gains than the children with the addition of sentence context.

Part of the reason that adults might be more accurate on the final word in the sentences compared to children is that they are accurately identifying more of the words in the sentences before the final words. To assess this possibility, another ANOVA was conducted focusing on the sentences only. In this analysis, there were two between-subjects variables (talker accent: native or nonnative; listener age: adult or child) and one within-subject variable (word position: final target word or words before final target word). In this analysis, all main effects were significant. Listeners were more accurate on the words before the final target words than on the target words, $F(1, 116) = 23.91$, $p < .001$, $\eta_p^2 = .171$. Adults were more accurate than children, $F(1, 116) = 60.34$, $p < .001$, $\eta_p^2 = .342$. Accuracy was higher for the native talkers than the nonnative talkers, $F(1, 116) = 250.84$, $p < .001$, $\eta_p^2 = .684$. All two-way interactions were also significant. The word position by listener age interaction arose because children were more accurate on the words before the final word (86 RAU) than on the final word (76 RAU), but the adults showed similar accuracy on target words and words before the target word (101 and 100 RAU, respectively), $F(1, 116) = 10.96$, $p = .001$, $\eta_p^2 = .086$. The word position by talker accent

interaction arose because listeners were more accurate on the preceding words than the target word for the nonnative condition (17 RAU better for preceding words than target words), but for the native condition, the result was in the opposite direction with slightly more accurate performance on the final word than the preceding words in the sentence (6 RAU difference), $F(1, 116) = 17.20, p < .001, \eta_p^2 = .129$. The listener age by talker accent interaction arose because there was a greater adult-child difference in accuracy for the nonnative conditions (85 vs. 55 RAU) than for the native condition (116 vs. 106 RAU), $F(1, 116) = 14.82, p < .001, \eta_p^2 = .113$. The three-way interaction was not significant.

In addition to the comparisons across adults and children, the data from the children were compared across the three age brackets: 5-year-olds, 6-year-olds, and 7-year-olds (Figure 2). As in the first analysis, talker accent was a between-subjects variable, and the context condition was a within-subject variable. This analysis showed main effects of context condition with more accurate performance in the sentence than in the isolated word condition, $F(1, 84) = 549.11, p < .001, \eta_p^2 = .867$. A significant main effect of talker accent was also observed, $F(1, 84) = 1079.23, p < .001, \eta_p^2 = .928$, with more accurate performance for the native than the nonnative talker. The effect of age group was also significant, $F(2, 84) = 3.82, p = .026, \eta_p^2 = .083$. Follow-up post-hoc Bonferroni tests with adjusted p -values showed that the 6- and 7-year-olds were significantly more accurate than the 5-year-olds ($p = .013$ and $.026$, respectively), but that the 6- and 7-year-olds were not significantly different from one another. Two of the two-way interactions were significant: context condition by talker accent, $F(1, 84) = 15.33, p < .001, \eta_p^2 = .154$, and context condition by age group, $F(2, 84) = 6.41, p = .003, \eta_p^2 = .132$. For the context condition by talker accent interaction, as in the first analysis, the improvement from the words in isolation to the sentence context was greater for the nonnative talkers than the native talkers. The

context condition by age group interaction showed that older children showed greater facilitation from sentence context than younger children. The remaining two-way interaction between age group and talker accent and the three-way interaction were not significant.

Relation between word recognition accuracy and vocabulary scores

The final accuracy analysis investigated whether the scores on the vocabulary measure were related to a child's ability to benefit from sentence context. For this analysis, the children's vocabulary scores were correlated with the difference between the word identification accuracy in the sentence context and word identification accuracy for the words in isolation. The correlations between these difference scores and vocabulary scores were significant for both the nonnative condition, $r = .324$, $n = 60$, $p = .011$, and the native condition, $r = .415$, $n = 30$, $p = .023$, suggesting that children with larger vocabularies are better able to take advantage of sentence context regardless of the talker's accent. However, when partial correlations were conducted, which controlled for age, the correlations were no longer significant, suggesting that there is not an effect of vocabulary size independent of age when listening to either native or nonnative talkers. Another possibility is that because vocabulary and age are so highly correlated partialing out age removes the shared variance between the two factors. Indeed, for our child participants, there was a highly significant correlation between age and vocabulary, $r = .558$, $n = 90$, $p < .001$. Furthermore, there was a significant correlation between age and difference scores for the nonnative condition, $r = .384$, $n = 60$, $p = .002$, but only a trend for the native condition, $r = .321$, $n = 30$, $p = .083$. In fact, when controlling for vocabulary, there was still a significant partial correlation between age and difference scores in the nonnative condition, $r = .268$, $n = 60$, $p = .040$, but this partial correlation was not significant for the native condition.

Response type

In addition to the analyses focused on accuracy, we investigated participants' response types for the nonnative talkers when they responded incorrectly (Table 1). Although children were less accurate overall compared to adults, their response strategies appeared to be aligned with the adults. That is, regardless of age, in cases in which the participants did not provide the correct answer, they tended to give another real word as their response. For example, for the items "ball" (from *The puppy played with the ball*) and "hat" (from *Yesterday he lost his hat*), the most common incorrect responses were *bowl* (26 / 80 listeners) and *head* (25 / 80 listeners), respectively. Nonword responses were much less common, although they accounted for about 14% of children's responses in the word-in-isolation condition. For example, for the items "legs" (from *The table has three legs*) and "blue" (from *The sky was very blue*), the most common incorrect responses were *lex* (21 / 80 listeners) and *voo* (11 / 80 listeners), respectively. Words from other languages were also counted as nonwords. For example, one of the most common "nonword" responses from adults was a response of *gracias* for the stimulus "glasses" (accounting for 25% of all adult nonword responses). Trials in which the listeners did not provide a response were also relatively rare overall, although did account for approximately 15% of children's trials in the sentence condition. In addition to comparing response types for the nonnative condition across the children and adults, we examined whether the accuracy for specific items was correlated for the adults and children. Average accuracy scores for each target word presented in isolation ($n = 32$) and in sentence context ($n = 32$) were calculated for the adults and children. We then correlated these accuracy scores between the children and adults. Accuracy scores were significantly correlated for both context conditions: word-in-isolation, $r = .832$, $n = 32$, $p < .001$, and sentence context, $r = .653$, $n = 32$, $p < .001$. This analysis suggests that

the children and adults found similar items more or less difficult to accurately identify. Lastly, we compared the most common responses by the adults and children, whether the response was correct or incorrect, to determine whether the most common responses were similar across the two age groups. For the word-in-isolation condition, the most common response for each item accounted for 40% of the children's responses and 47% of the adults' responses. For the word-in-isolation condition, the adults and children had the same most common response for 20 of the 32 items (63%). For the sentences, the most common response for each item accounted for a larger percentage of the data because there were many more correct answers, 72% and 75% of children's and adults' responses, respectively. Again, there was substantial overlap for the most common responses by children and adults, although much of this finding can be accounted for by the correct responses. The most common response was the same for adults and children for 26 of the 32 items (81%). Taken together, these results suggest that although adults' word recognition accuracy is higher than children's, they might use similar response strategies when confronting unfamiliar pronunciations of known words.

Across-talker effects

The data were divided by talker to determine if there was a substantial influence of language background (Spanish vs. Japanese) or specific talker. As is shown in Table 2, the accuracy patterns varied to some extent across talkers, but the overall pattern of responses was consistent for both the children and adults across the nonnative talkers.

Discussion

This study investigated 5- to 7-year-old children and adults' perception of native and nonnative talkers' productions of full sentences and words excised from sentences. The nonnative talkers' word productions were selected to be difficult to identify when presented in isolation relative to the same items in sentence context. This design allowed for the assessment of how children and adults can utilize top-down cues including syntactic and semantic information as well as familiarity with the talker's voice to facilitate the lexical re-interpretation of misleading acoustic-phonetic information. By testing both children and adults, we could assess whether listeners' abilities to benefit from top-down contextual information are influenced by the linguistic and cognitive development and real-world knowledge acquisition that occurs between the early school age years and adulthood.

Developmental comparisons

The findings from the native talker condition support previous work, which has found that children and adults benefit to a similar extent from top-down semantic cues (Fallon et al., 2000, 2002). However, ceiling performance in the native sentence condition may have underestimated adults' abilities to benefit from context. Future work should continue to assess children's and adults' abilities to benefit from top-down cues at different points along the psychometric function, particularly as previous work with speech in noise suggests that the slopes of the psychometric functions are different for words in isolation and words in sentences (Miller et al., 1951). In the nonnative condition, adults were better able than children to take advantage of sentence context, which was much more difficult overall due to the deviations in the productions relative to native norms. This differential context benefit between the children and adults accords with earlier work showing that adverse conditions - due to the addition of

noise - can make it difficult for children to capitalize on top-down cues to word recognition (Nittrouer & Boothroyd, 1990). The current study adds to the previous speech-in-noise findings by showing that children may also have more difficulty than adults with integration of syntactic and semantic information when listening adversity stems from unfamiliar pronunciation patterns due to the presence of a nonnative accent rather than to adversity stemming from environmental degradation.

The adult-child difference in the ability to benefit from syntactic and semantic context may partially arise from the fact that children understood fewer words preceding the final target word in the sentence. Children's difficulty in understanding the initial part of the sentence very likely accounts for some of the difference between children and adults. In fact, children identified significantly fewer of the words preceding the final target word than adults for both the native and the nonnative conditions. However, the interaction between word position and listener age suggests that the final words posed a particular challenge for the children. That is, although adults showed similar accuracy rates for the context and the target words, children were less accurate on the final words than on the preceding context. Adults' enhanced executive functions, particularly abilities that are drawn upon during test-taking situations (e.g., following directions, maintaining attention, and working memory) may allow them to more adeptly take advantage of contextual cues than children who have less test-taking experience and robust executive functioning (Huizinga, Dolan, & van der Molen, 2006). In the sentence condition, listeners must hold the entire sentence in memory. It is possible that the requirement to repeat back the entire sentence taxed the children's working memory to a greater extent than adults leading to decreases in available processing resources when encoding and retrieving the final word in the sentence. A design in which listeners only had to repeat the final word of the sentence could lessen the

memory load, but this method would not allow for the assessment of word recognition accuracy for the sentence context. This design also does not ensure that listeners in fact listen to the sentence, which is important for a study examining the effects of sentence context; they may instead simply wait for the final word and just repeat that item back. Children's difficulty in encoding and processing the unfamiliar accent could also lead to a decrease in the ability to efficiently integrate the top-down information with the following signal. This potential processing difficulty may also lead to decreases in children's ability to inhibit the activation of the previously accessed, incorrect lexical candidate. Even in adults, increases in cognitive processing demands or increases in listening effort have been observed for nonnative-accented speech relative to native-accented speech (Van Engen & Peelle, 2014; Yi, Smiljanic, & Chandrasekaran, 2014), but adults' increased linguistic and conceptual knowledge may allow them to benefit from sentence context to a greater degree than children.

Although the adults were more accurate in all conditions, the response strategies of the adults and children appeared to be similar. That is, for trials in which they did not accurately identify the target word, both children and adults typically provided a real word rather than phonetically imitating the talker's production as a nonword. Nonword responses were most common in the nonnative word-in-isolation condition for both children and adults, but still relatively rare (14% and 7% of responses, respectively). When sentence context was added, nonword response rates dropped substantially for both age groups (4% for children and 0.5% for adults). Thus, although children are more likely to provide a nonword response than adults, they appear to be biased towards reporting known lexical items rather than assuming that the production is an unknown, novel word. This interpretation accords with prior work showing that children do not select novel items when presented with nonnative-accented speech (Creel et al.,

2016), unlike when they are presented with native-accented words with mispronunciations (Creel, 2012). Support for the idea that adults and children employ similar response strategies also comes from the findings that adults' and children's accuracy scores across items were significantly correlated and that the specific responses given by the two age groups showed substantial overlap, even when accuracy rates differed across the age groups.

Children's context use: Vocabulary and age effects

In addition to the comparison between adults and children, we also assessed how age and vocabulary size influenced children's abilities to capitalize on sentence context. We found that older children showed greater benefit from sentence context than younger children, when compared across year brackets (5-, 6-, and 7-year-olds) or when age was analyzed as a continuous variable. Further, children with higher vocabulary scores on the NIH Toolbox Picture Vocabulary Test showed greater sentence context benefit in both the native and nonnative conditions. However, there was not a significant effect of vocabulary scores when the effect of age was partialled out. Prior work on the influence of vocabulary knowledge on the perception of native and nonnative speech and the use of semantic context has also failed to observe an effect of vocabulary size separate from the age effect (Bent, 2014; Creel et al., 2016). These findings suggest that although vocabulary knowledge may certainly play a role in a child's ability to utilize semantic cues during word recognition, there may be other linguistic or cognitive changes during development that underlie the increase in semantic predictability benefit in the early school age years. This idea was supported by the finding that age significantly correlated with context benefit in the nonnative condition, even when partialing out vocabulary scores. Therefore, it remains possible that there are other cognitive-linguistic skills beyond vocabulary

that increase with age and are related to the ability to benefit from sentence context with nonnative talkers. There are many substantial developmental changes in the 5- to 7-year-old age range, including entry into school, which will likely result in acquisition of greater amounts of real-world knowledge and exposure to greater number of speakers, possibly including those with varied accents and dialects. As children learn more about the world, their ability to benefit from sentence context may increase because utilization of top-down cues requires an understanding of the relationship between frequent events, as well as the corresponding linguistic labels that code these events. More exposure to these types of sentences and relationships may facilitate the ease with which the information is accessed and utilized during the perception of speech under challenging conditions.

Native and nonnative talker comparisons

Both children and adults benefited more from sentence context for the nonnative condition than the native condition. These results align with previous work (Behrman & Akhund, 2013; Kennedy & Trofimovich, 2008) demonstrating that adult listeners receive more benefit from sentence context with nonnative than native speech. However, in these earlier studies, adults were at or near ceiling for all context conditions for the native speakers (Behrman & Akhund, 2013; Kennedy & Trofimovich, 2008), which did not allow for a semantic context benefit in the native talker conditions. Here, listeners reached ceiling level performance in the native-talker sentence condition (99% and 95% correct for adults and children, respectively), although unlike in earlier work, identification accuracy was not at ceiling for the words in isolation, which provided room for the use of context. Nonetheless, the greater advantage of sentence context in the nonnative condition may partially stem from ceiling effects in the native

sentence condition. Future studies could attempt to equate performance in the native and nonnative conditions through the use of environmental degradation or by increasing the cognitive demands of the task. However, adding noise to speech so that scores are not at ceiling and a sentence context benefit can be observed may introduce other factors that will need to be considered: the perception of speech in noise or babble may involve the recruitment of different cognitive resources or processing strategies relative to processing of speech in quiet (McLaughlin, Baese-Berk, Bent, Borrie, & Van Engen, 2018).

In contrast to earlier findings with adults showing the greater context benefit for nonnative- than native-accented speech, children's semantic predictability benefit has been reported to be similar for native- and nonnative-accented speech (Creel et al., 2016; Holt & Bent, 2017). Yet, we found that children benefited more from context in the nonnative than the native condition. There are several methodological differences between this study and the previous two that may account for the difference in findings. First, in Holt and Bent (2017), the sentences in both the native and nonnative condition were presented in noise whereas in this study all materials were presented in quiet. Second, the stimuli used previously were sentences differing in predictability (Holt & Bent, 2017) or sentences differing in sensicality and the final excised word from these sentences (Creel et al., 2016) rather than excised words compared to sensical sentences. Third, the children in Creel et al. (2016) were substantially younger than those in the current study (3 to 5 years vs. 5 to 7 years of age). Lastly, there was greater talker variability in the current study compared to Holt and Bent (2017), who only had one talker for each accent, although there was a relatively similar extent of variability compared to Creel et al. (2016), who had six talkers (three male and three female) per accent (although only one nonnative accent was included). Here, each listener was presented with four talkers of both genders. Listeners in the

nonnative condition also received two nonnative accents (Japanese-accented and Spanish-accented). The extent of talker variability in each study may account for some of the differences in results. Although in the opposite direction, an effect of talker variability has been shown previously by Clopper (2012) who found that adult listeners showed greater semantic facilitation with a familiar dialect when the talker's dialect randomly varied from trial to trial, but did not show an effect of dialect familiarity on semantic predictability when dialects were blocked. Due to the number of methodological differences across all of these studies, it is not possible to definitively determine how each of these factors influences children's and adults' abilities to benefit from top-down cues across nonnative accents and regional dialects. Future studies should continue to investigate these issues by manipulating these variables within the study (e.g., testing a wider age range, assessing talker variability effects in blocked and randomized conditions, comparing quiet vs. noise-added conditions).

Clinical implications

Future investigations in this area should also be extended to children with communication disorders. Evidence is mixed about whether children with speech disorders have particular difficulty processing unfamiliar speech varieties. Compared to children with typical speech and language development, children with speech disorders show a deficit with an unfamiliar regional dialect in a lexical decision task (Nathan & Wells, 2001), but not for an unfamiliar regional dialect or nonnative accent when carrying out instructions in the Token Test (e.g., "Touch the white circle") (Harte, Frizelle, & Gibbon, 2017). Although comparing these studies is difficult due to the differences in both method and speech varieties included, the results suggest that the difficulties observed for the children with speech disorders in the lexical decision task may stem

from bottom-up processing deficits (e.g., phonetic discrimination) (Nathan & Wells, 2001). These children may be able to compensate for these processing deficits through the use of top-down knowledge. That is, the findings from the token test, in which children with speech disorders did not show a deficit on the unfamiliar speech varieties, included much more contextual support (full sentences as well as a closed set of shapes, colors, and possible actions) (Harte et al., 2017). Future studies should explicitly evaluate whether children with speech disorders can benefit from top-down information during the perception of unfamiliar speech varieties. Ideally, these investigations would manipulate the availability of contextual support along with both nonnative accents and regional dialects that vary in their proximity to home dialect production norms.

Conclusion

Bottom-up acoustic-phonetic cues to a target word's identity do not always sufficiently support accurate word recognition. These challenging word recognition situations can arise due to competing sounds in the environment or speakers with unfamiliar pronunciation patterns. In these cases, listeners must recruit top-down information to support the identification of the intended lexical items. We assessed how both 5- to 7-year-old children and adults capitalized on sentence cues to identify nonnative-accented words that showed very low accuracy rates in isolation ($\leq 15\%$ correct). The results showed that both child and adult listeners could utilize the cues present in the sentences to support word identification, but adults showed greater facility at taking advantage of the top-down cues. A similar effect was seen within the group of children, where older children derived more benefit from the sentence cues than younger children. The context benefit correlated with vocabulary size, but the effect of vocabulary was not independent

of age. These results suggest that cognitive-linguistic development and the application of real-world knowledge, both within the early school age years and from childhood to adulthood, supports the integration of semantic information with partial or misleading acoustic-phonetic cues to word identity. Children's less adept use of top-down information during the perception of nonnative speech may partially explain the protracted developmental trajectory of word recognition abilities for unfamiliar speech varieties (Bent, in press).

Acknowledgements

We are grateful to the families who participated in this research and the Labs in Life Language Pod at the Center of Science and Industry (Columbus, OH).

References

- Atagi, E., & Bent, T. (2013). Auditory free classification of nonnative speech. *Journal of Phonetics*, 41(6), 509-519. doi:10.1016/j.wocn.2013.09.003
- Aydelott, J., & Bates, E. (2004). Effects of acoustic distortion and semantic context on lexical access. *Language and Cognitive Processes*, 19(1), 29-56.
- Bamford, J., & Wilson, I. (1979). Methodological considerations and practical aspects of the BKB sentence lists. In J. Bench & J. Bamford (Eds.), *Speech-hearing Tests and the Spoken Language of Hearing-impaired Children* (pp. 148-187). London: Academic.
- Behrman, A., & Akhund, A. (2013). The influence of semantic context on the perception of Spanish-accented American English. *Journal of Speech, Language, and Hearing Research*, 56(5), 1567-1578.
- Bent, T. (2014). Children's perception of foreign-accented words. *Journal of Child Language*, 41(6), 1334-1355. doi:10.1017/S0305000913000457
- Bent, T. (in press). Development of unfamiliar accent comprehension continues through adolescence. *Journal of Child Language*, 1-12.
- Bent, T., & Atagi, E. (2015). Children's perception of nonnative-accented sentences in noise and quiet. *Journal of the Acoustical Society of America*, 138(6), 3985-3993.
doi:10.1121/1.4938228
- Bent, T., & Atagi, E. (2017). Perception of nonnative-accented sentences by 5- to 8-year-olds and adults: The role of phonological processing. *Language and Speech*, 60(1), 110-122.
doi:10.1177/0023830916645374

- Bent, T., & Holt, R. F. (in press). Shhh... I need quiet! Children's understanding of American, British, and Japanese-accented English speakers. *Language and Speech*.
doi:10.1177/0023830918754598
- Clopper, C. G. (2012). Effects of dialect variation on the semantic predictability benefit. *Language and Cognitive Processes, 27*(7-8), 1002-1020.
- Cole, R. A., & Perfetti, C. A. (1980). Listening for mispronunciations in a children's story: The use of context by children and adults. *Journal of Verbal Learning and Verbal Behavior, 19*(3), 297-315.
- Creel, S. C. (2012). Phonological similarity and mutual exclusivity: on-line recognition of atypical pronunciations in 3-5-year-olds. *Developmental Science, 15*(5), 697-713.
doi:10.1111/j.1467-7687.2012.01173.x
- Creel, S. C., Rojo, D. P., & Paullada, A. N. (2016). Effects of contextual support on preschoolers' accented speech comprehension. *Journal of Experimental Child Psychology, 146*, 156-180. doi:10.1016/j.jecp.2016.01.018
- Eisenberg, L. S., Martinez, A. S., Holowecky, S. R., & Pogorelsky, S. (2002). Recognition of lexically controlled words and sentences by children with normal hearing and children with cochlear implants. *Ear and Hearing, 23*(5), 450-462.
doi:10.1097/01.aud.0000034736.42644.be
- Elliott, L. L. (1979). Performance of children aged 9 to 17 years on a test of speech-intelligibility in noise using sentence material with controlled word predictability. *Journal of the Acoustical Society of America, 66*(3), 651-653. doi:10.1121/1.383691
- Fallon, M., Trehub, S. E., & Schneider, B. A. (2000). Children's perception of speech in multitalker babble. *Journal of the Acoustical Society of America, 108*(6), 3023-3029.

- Fallon, M., Trehub, S. E., & Schneider, B. A. (2002). Children's use of semantic cues in degraded listening environments. *Journal of the Acoustical Society of America*, *111*(5 Pt 1), 2242-2249.
- Fernald, A., & Hurtado, N. (2006). Names in frames: infants interpret words in sentence frames faster than words in isolation. *Developmental Science*, *9*(3), F33-40. doi:10.1111/j.1467-7687.2006.00482.x
- Gershon, R. C., Wagster, M. V., Hendrie, H. C., Fox, N. A., Cook, K. F., & Nowinski, C. J. (2013). NIH toolbox for assessment of neurological and behavioral function. *Neurology*, *80*(11 Supplement 3), S2-S6.
- Harte, J., Frizelle, P., & Gibbon, F. (2017). The effect of different speaker accents on sentence comprehension in children with speech sound disorder. *Clinical Linguistics & Phonetics*, *32*(8), 689-705. doi:/10.1080/02699206.2017.1419379
- Holt, R. F., & Bent, T. (2017). Children's use of semantic context in perception of foreign-accented speech. *Journal of Speech, Language, and Hearing Research*, *60*, 223-230.
- Huizinga, M., Dolan, C. V., & van der Molen, M. W. (2006). Age-related change in executive function: Developmental trends and a latent variable analysis. *Neuropsychologia*, *44*(11), 2017-2036.
- Hunter, C. R., & Pisoni, D. B. (2018). Extrinsic Cognitive Load Impairs Spoken Word Recognition in High- and Low-Predictability Sentences. *Ear and Hearing*, *39*(2), 378-389. doi:10.1097/AUD.0000000000000493
- Jerger, S., & Jerger, J. (1984). *Pediatric speech intelligibility test*. St. Louis, MO: Auditec

- Kalikow, D. N., Stevens, K. N., & Elliott, L. L. (1977). Development of a test of speech-intelligibility in noise using sentence materials with controlled word predictability. *Journal of the Acoustical Society of America*, *61*(5), 1337-1351. doi:10.1121/1.381436
- Kennedy, S., & Trofimovich, P. (2008). Intelligibility, comprehensibility, and accentedness of L2 speech: The role of listener experience and semantic context. *Canadian Modern Language Review*, *64*(3), 459-489.
- McLaughlin, D. J., Baese-Berk, M. M., Bent, T., Borrie, S. A., & Van Engen, K. J. (2018). Coping with adversity: Individual differences in the perception of noisy and accented speech. *Attention, Perception, & Psychophysics*, *80*(6), 1559-1570.
doi:doi:10.3758/s13414-018-1537-4
- McQueen, J. M., Tyler, M. D., & Cutler, A. (2012). Lexical retuning of children's speech perception: Evidence for knowledge about words' component sounds. *Language Learning and Development*, *8*(4), 317-339. doi:10.1080/15475441.2011.641887
- Miller, G. A., Heise, G. A., & Lichten, W. (1951). The intelligibility of speech as a function of the context of the test materials. *Journal of Experimental Psychology*, *41*(5), 329-335.
- Miller, G. A., & Isard, S. (1963). Some perceptual consequences of linguistic rules. *Journal of Verbal Learning and Verbal Behavior*, *2*(3), 217-228.
- Nathan, L., & Wells, B. (2001). Can children with speech difficulties process an unfamiliar accent? *Applied Psycholinguistics*, *22*(3), 343-361.
- Nilsson, M., Soli, S. D., & Gelnett, D. J. (1996). *Development of the Hearing in Noise Test for Children (HINT-C)*. Los Angeles, CA: House Ear Institute.

- Nittrouer, S., & Boothroyd, A. (1990). Context effects in phoneme and word recognition by young children and older adults. *Journal of the Acoustical Society of America*, 87(6), 2705-2715.
- Psychology Software Tools. (2007). E-Prime 2.0. Pittsburgh, PA.: Psychology Software Tools.
- Smiljanic, R., & Sladen, D. (2013). Acoustic and semantic enhancements for children with cochlear implants. *Journal of Speech, Language, and Hearing Research*, 56(4), 1085-1096.
- Studebaker, G. A. (1985). A rational arcsine transform. *Journal of Speech and Hearing Research*, 28(3), 455-462.
- Van Engen, K. J., & Peelle, J. E. (2014). Listening effort and accented speech. *Frontiers in Human Neuroscience*, 8, 577. doi:10.3389/Fnhum.2014.00577
- van Heugten, M., & Johnson, E. K. (2016). Toddlers' word recognition in an unfamiliar regional accent: The role of local sentence context and prior accent exposure. *Language and Speech*, 59(3), 353-363. doi:10.1177/0023830915600471
- White, K. S., & Aslin, R. N. (2011). Adaptation to novel accents by toddlers. *Developmental Science*, 14(2), 372-384. doi:Doi 10.1111/J.1467-7687.2010.00986.X
- Yi, H. G., Smiljanic, R., & Chandrasekaran, B. (2014). The neural processing of foreign-accented speech and its relationship to listener bias. *Frontiers in Human Neuroscience*, 8, 768. doi:10.3389/Fnhum.2014.00768

Table 1: Response type (in percent) separated by listener age and listening context for the nonnative talkers

	Context	Correct response	Incorrect real word response	Nonword response	No response
Children	Word	6.6	74.9	14.3	4.2
	Sentence	44.7	36.3	4.3	14.7
Adults	Word	15.0	76.4	6.9	1.7
	Sentence	78.4	18.0	0.5	3.1

Table 2: Mean accuracy scores for the four talkers separated by listener age and listening context

Talker	Percent correct (range)			
	Children		Adults	
	Sentence	Word	Sentence	Word
Spanish male	47 (12 - 88)	10 (0 - 30)	68 (25 - 100)	25 (0 - 80)
Spanish female	37 (10 - 53)	4 (0 - 10)	81 (50 - 100)	12 (0 - 45)
Japanese male	37 (8 - 65)	4 (0 - 12)	78 (35 - 100)	8 (0 - 25)
Japanese female	58 (10 - 87)	8 (0 - 22)	87 (55 - 100)	16 (0 - 70)

Figure captions

Figure 1: Word identification accuracy scores in percent correct for the children (black bars) and adults (striped bars) for the word-in-isolation condition and the full sentence condition. Scores for the nonnative speakers are on the left and the native speakers are on the right. Error bars represent 1 standard deviation.

Figure 2: Word identification accuracy scores in percent correct for the 5-year-old children (black bars), 6-year-old children (gray bars), and the 7-year-old children (striped bars) for the word-in-isolation condition and the full sentence condition. Scores for the nonnative speakers are on the left and the native speakers are on the right. Error bars represent 1 standard deviation.



