Effects of contextual support on preschoolers' accented speech comprehension.

Permalink
https://escholarship.org/uc/item/16d4n5wt

Journal
Journal of experimental child psychology, 146

ISSN
0022-0965

Authors
Creel, Sarah C
Rojo, Dolly P
Paullada, Angelica Nicolle

Publication Date
2016-06-01

DOI
10.1016/j.jecp.2016.01.018

Peer reviewed
Effects of contextual support on preschoolers’ accented speech comprehension

Sarah C. Creel (screel@ucsd.edu)\textsuperscript{1}

Angelica Nicolle Paullada

UC San Diego

Dolly P. Rojo

UT Austin

Word count: 10480 Main text + references

\textsuperscript{1}To whom correspondence should be addressed. screel@ucsd.edu, 9500 Gilman Drive, Department of Cognitive Science, Mail Code 0515, La Jolla, CA, 92093-0515.

Acknowledgments. This study was supported by grant BCS-1230003 to SCC from the National Science Foundation.
Abstract

Young children often hear speech in unfamiliar accents, but relatively little research characterizes their comprehension capacity. The current study tested preschoolers’ comprehension of familiar-accented vs. unfamiliar-accented speech with varying levels of contextual support from sentence frames (full sentences vs. isolated words) and from visual context (four salient pictured alternatives, vs. the absence of salient visual referents). The familiar-accent advantage was more robust when visual context was absent, suggesting that previous findings of good accent comprehension in infants and young children may result from ceiling effects in easier tasks (picture fixation, picture selection) relative to the more-difficult tasks often used with older children and adults. In contrast to prior work on mispronunciations, where most errors were novel-object responses, children in the current study did not select novel-object referents above chance levels. This suggests that some property of accented speech may dissuade children from inferring that an unrecognized familiar-but-accented word has a novel referent. Finally, children showed detectable accent processing difficulty despite presumed incidental community exposure. Results suggest that preschoolers’ accented speech comprehension is still developing, consistent with theories of protracted development of speech processing.

*Keywords: accent, accented speech, speech comprehension, word recognition, adverse listening conditions*
With increasing globalization, listeners often encounter speech in regional or foreign accents different from their own. These listeners include young children. While there is a wealth of research on the effects of unfamiliar accents on speech comprehension in adults, much less is known about young children’s accented-speech comprehension, as observed by Cristia, Seidl, Vaughn, Schmale, Bradlow, and Floccia (2012) in a recent review paper. This is a significant gap in knowledge, given that accented speech provides a challenge to young children’s developing abilities in speech perception and word recognition (Bent, 2014; Nathan, Wells, & Donlan, 1998). More importantly, as Cristia et al. (2012) also note, it prevents the field from having a full developmental picture of accented speech comprehension from infancy through childhood through young and older adulthood. The current study is an attempt to decrease this gap.

Little research characterizes accented speech comprehension in young children (with, as discussed shortly, a few exceptions), and existing findings are spread across paradigms and do not always produce identical conclusions. Therefore, it is not clear how well children comprehend unfamiliar-accented speakers who they encounter in everyday environments, or what situations create greater or lesser ease in comprehension. The current study seeks to characterize accented speech processing in preschool-aged (3- to 5-year-old) children. In particular, we aim to assess the role of different types of contextual support—sentential semantic context, and visual-scene context—and the possibility that children will misidentify accented familiar words as novel words.

Accented speech may be more challenging for children in that they are generally more susceptible than adults to adverse listening conditions in speech comprehension (e.g. Fallon,
Trehub, & Schneider, 2000, 2002). Thus, perceptual difficulties introduced by speech in unfamiliar accents may be especially challenging for children. An additional reason that accented speech may be especially challenging for children is that, because preschool-aged children are rapidly acquiring vocabulary, they need to identify potentially novel word forms. This means that, if children are overly sensitive to pronunciation deviations from familiar words, they may misidentify familiar but accented words (e.g. “peeg” /pɪɡ/ for pig /pɪɡ/) as novel words. On the other hand, if children accept large variations in pronunciation, they risk failure to distinguish similar words (pick /pɪk/ vs. peek /pɪk/). The next section reviews what is known about accented speech recognition in infants, young children, and adults.

**Accented speech processing during development**

*Infants’ and toddlers’ accented speech processing.* While we do not have space for an exhaustive review, it should be noted that most developmental research on word or word-form recognition in unfamiliar accents, or accent-like variability such as deliberate mispronunciations, has focused on children up to about two years of age (accents: Best, Tyler, Gooding, Orlando, & Quann, 2009; Mulak, Best, Tyler, Kitamura, & Irwin, 2009; Schmale, Cristia, Seidl, & Johnson, 2010; Schmale & Seidl, 2009; van Heugten & Johnson, 2014, in press; Van Heugten, Krieger, & Johnson, 2015; mispronunciations: Swingley & Aslin, 2000, 2002; Swingley, 2003; White & Aslin, 2011; White & Morgan, 2008), with much less research on older children (the only exceptions being, to the best of our knowledge, the following: for accents: Bent, 2014; Bent & Atagi, 2015; Nathan, Wells, & Donlan, 1998; mispronunciations: Creel, 2012; Newton & Ridgway, in press). Cristiá et al. (2012), in reviewing accent processing studies in various age groups, note that research on children’s accent processing, compared to infants/toddlers and
PRESCHOOL ACCENT COMPREHENSION

young adults, is “less well-documented” (p. 7). (Note that there is more research with older children on recognizing or responding to the presence of accents: Floccia, Butler, Girard, & Goslin, 2009; Girard, Floccia, & Goslin, 2008; Kinzler, Dupoux, & Spelke, 2007; Kinzler, Shutts, DeJesus, & Spelke, 2009; Kinzler & DeJesus, 2013; Wagner, Clopper, & Pate, 2014.)

Infants, by some accounts, show good word or word-form recognition after 12-13 months of life. Schmale, Cristia, Seidl, and Johnson (2010) and Schmale and Seidl (2009) tested infants’ recognition of familiarized word forms, and found that infants recognize words despite unfamiliar regional accents by 12 months (Schmale et al., 2010), and foreign accents by 13.5 months (Schmale & Seidl, 2009). These word-form recognition studies fit with Swingley and Aslin’s (2000, 2002) findings from a looking-while-listening paradigm that children aged 15 months and older show some degree of recognition of non-canonically-pronounced words (in their case deliberately-mispronounced words): when hearing “Look at the tog!” children in Swingley and Aslin (2002) looked more at a pictured dog than a shoe—though they looked at the dog even more when dog was correctly pronounced. This result suggests that infants are at least partly aware of the similarity between correctly-pronounced and mispronounced forms (see White & Morgan, 2008, for further exploration). With adaptation, infants may recognize accented forms of familiar words: Van Heugten and Johnson (2014) showed accented word-form recognition in 15-month-olds following exposure to an accented reading of a familiar book, and White and Aslin (2011) found that 19-month-olds who were exposed to a vowel-shift accent (“socks” pronounced as “sacks”) visually fixated the correct pictures when hearing artificially-accented versions of familiar words like bottle (pronounced like “battle”). By age 2.0 years at the latest, children can recognize familiar words in unfamiliar accents without exposure to the accent.
PRESCHOOL ACCENT COMPREHENSION

(25 months for Canadian toddlers hearing Australian-accented English; van Heugten, Krieger, & Johnson, 2015; 19 months for Australian toddlers hearing Jamaican-accented English: Mulak et al., 2013), and newly-learned words if they have brief exposure to the accent (Schmale, Cristia, & Seidl, 2012). By 2.5 years, children no longer need accent exposure for recognizing familiar words (Van Heugten & Johnson, in press) or newly-learned words (Schmale, Hollich, & Seidl, 2011). Interestingly, the children in the Van Heugten and Johnson (in press) study were more successful at recognizing the accented words (that is, they fixated the named picture more robustly) when words were embedded in full-sentence contexts, hinting that at least one sort of context—acoustic-phonetic sentential context—may facilitate accent comprehensions.

Finally, a recent study by Schmale, Seidl, and Cristia (2015) suggests that toddlers can be induced to recognize accented tokens of newly-learned words by simply exposing them to a wide range of talkers (or even pictures of talkers) in their own accent. Two-year-old children were exposed to four American-accented talkers, either reading passages, or silently waving and gesturing. Then children were presented with a novel labeling event in a familiar accent, followed by a two-alternative choice where the same word was labeled in an unfamiliar accent. In both exposure conditions, children generalized word recognition to the unfamiliar accent. This differed from previous findings (Schmale et al., 2012), where 2-year-olds without exposure to the unfamiliar accent failed to generalize. This implies that very young children, given a minimal amount of speech variability (or even visually-implied speech variability), appear to broaden the range of pronunciations they will accept for a word. Whether this broadening overextends to accepting accent-atypical pronunciations, or confusing previously-distinct words (like /pɪk/
and /pik/), is as yet unknown. Taken together, these studies of infants and toddlers imply that very young children are well on their way to comprehending speech in unfamiliar accents.

**Accented speech processing in preschool and primary-school children.** While the work reviewed above suggests excellent accented-speech comprehension by 2-2.5 years of age, studies of older children are more equivocal. Since infant research and research on older children tend to use somewhat different paradigms, it is difficult to compare results across ages (see Creel & Quam, in press, for further discussion). On the one hand, preschool-aged children (3-5 years) hearing mispronounced words in their native-language phonological system are fairly accurate in selecting the matching item out of four pictures in a display: Creel (2012) found that when children heard a single-phonological-feature mispronunciation ("fesh" /fɛʃ/ for fish /fɪʃ/), they selected the target picture (the fish) almost 90% of the time, from an array containing pictures of a fish, two other familiar objects, and a novel object. Even when mispronunciations were more egregious (/fɑʃ/), they still selected the target picture about 60% of the time. This suggests that, similar to Swingley and Aslin’s (2000, 2002) findings with toddlers, children readily detect the similarity between the mispronounced version and their word representation. Just as interesting was what happened when children in Creel (2012) did not select the target picture (keeping in mind that target-picture selections were the predominant response overall): looking only at these non-target trials, children chose a novel object roughly 70% of the time, much higher than the chance rate of 33% (non-target choices were 1 novel picture vs. 2 familiar pictures). This latter finding implies that when children do not recognize a mispronounced word as familiar, they are most likely to assume that it is a novel form (e.g. Markman & Wachtel, 1998; see also Halberda,
This raises the possibility that unfamiliar-accented words, if unrecognized, will be treated as novel words.

While Creel’s (2012) study suggests 3- to 5-year-olds comprehend unfamiliar pronunciations fairly well, research with word or sentence repetition tasks, largely using real accents, paints a different picture (Bent, 2014; Bent & Atagi, 2015; Nathan et al., 1998; Newton & Ridgway, in press). Nathan et al. (1998) tested 4- and 7-year-old British English speakers’ abilities to repeat isolated words spoken in a familiar accent (London English) vs. an unfamiliar accent (Glasgow English). They found that both groups performed less well on the unfamiliar accent (57%, vs. 88% for the familiar accent), with improvement from the 4-year-olds (43%) to the 7-year-olds (71%). More recently, Bent (2014) extended a line of research on accented speech recognition in adults to encompass young (4-7-year-old) children. Familiar-accented (American English) or unfamiliar-accented (Korean-accented English) words were presented at a +5 dB signal-to-noise ratio with speech-shaped noise. Bent found that children were much less accurate at repeating words than adults, and were far less accurate for foreign-accented words (44% accuracy) than native-accented words (80% accuracy). Newton and Ridgway (in press), using a sentence repetition in noise task with 6- to 7-year-olds, found much higher signal-to-noise thresholds for accurate repetition of artificially-accented than for familiar-accented speech.

Thus, in contrast to studies of infants and toddlers, studies of older children (Creel, 2012 being an exception) suggest that they still have greater difficulty than adults in comprehending accented speech. These findings concur with other research suggesting that children are less adept than adults at comprehending speech in ideal (Hazan & Barrett, 2000; Ohde & Haley,
PRESCHOOL ACCENT COMPREHENSION

1997) and non-ideal listening conditions (Fallon et al., 2000, 2002). Findings of non-adultlike accent comprehension in young children are also consistent with accounts of protracted perceptual development (e.g. Creel & Jiménez, 2012; see also Bent & Atagi, 2015), which posit that fully adult-like spoken language processing requires many years of perceptual learning.

Remaining questions

Current evidence is mixed with regard to the circumstances where children succeed in recognizing accented speech. One potential explanation for differences across ages and tasks is that tasks that provide more contextual support lead to better performance. For example, hearing an accented word and selecting its referent from a small set of pictures (Creel, 2012; van Heugten et al., 2015) is likely much easier than recognizing an accented word in the absence of any visual context (e.g. Bent, 2014; Nathan et al., 1998). Similarly, it is likely easier to recognize an accented word such as “cows” in the sentence “The farmer milked the cows” than when hearing “cows” in a less predictable sentence context or as an acoustically isolated token (see Van Heugten & Johnson, in press, for evidence that 28-month-olds recognize accented words better when preceded by an accented sentence context). Thus, the current study tests the contextual support hypothesis that context—here, semantic sentence context and visual scene context—help children comprehend accented speech, potentially explaining discrepant results in the developmental literature.

A second issue concerns the types of errors children make when they misperceive accented speech. Do they tend to think the intended referent is a novel object, like children in Creel (2012)’s mispronunciation study? If they do not make novelty errors, why is this the case? While
little empirical work exists to suggest answers to this question, one possibility is that accented pronunciations may not be sufficiently different from the familiar-accent pronunciation to trigger a novelty response, such that words are slightly harder to recognize (leading to slightly more errors) but also do not register as distinctly different from any known word. That is, they may register that accented “pig” is similar to /pɪɡ/, /pig/, and /pik/, but it is not a clear instance of the novel word form /pig/. Another possibility is that children may be more likely to dismiss unfamiliar-accented speakers’ knowledge of novel object names (e.g. Corriveau, Kinzler, & Harris, 2013), dissuading them from making novel-object selections for unrecognized accented words.

The current study

These questions were addressed using a recently-developed set of sentences, the Sentences for studying Accent Understanding in Child Experiments (SAUCE). Sentences (described in more detail in Experiment 1) were recorded by ten talkers with accents familiar to monolingual English-speaking children in our area (California accents) and by ten talkers with less-familiar accents (Mexican-Spanish accents).

The stimuli were recorded and studies were run in San Diego County, California. San Diego County has a Spanish-speaking population of 25% (Ryan, 2013), though only 11% (that is, about 40% of the 25%) are estimated to speak English less than “very well” (County of San Diego, 2013), implying among other things the presence of a foreign accent. Presumably some of those who speak “very well” are English-fluent enough to have no identifiable accent, but even the 11% figure suggests that monolingual children in our area are likely to have had incidental
PRESCHOOL ACCENT COMPREHENSION

exposure to Mexican Spanish accents (see Floccia, Delle Luche, Durrant, Butler, & Goslin, 2012; Howard, Carrazza, & Woodward, 2014, for effects of community accent or language exposure), so if anything, accented-speech comprehension difficulty should be ameliorated. Such limited community exposure is common in a number of major cities or other communities with immigrant populations, and thus represents an important situation for scientific investigation. Further, Spanish is currently the most prevalent language besides English in the United States (Ryan, 2013), and thus likely the most prevalent foreign accent (or, more accurately, group of foreign accents).

Four experiments tested preschool-aged children’s comprehension of accented speech with varying degrees of contextual support. The first two used picture-pointing accuracy and visual-world eye tracking (Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995; Trueswell, Sekerina, Hill, & Logrip, 1999) to evaluate children’s comprehension of words in sentences (Experiment 1) and words in isolation (Experiment 2) when the visual scene limits possibilities to a small set of pictured referents. The next two experiments used children’s word-repetition responses to gauge recognition of words in sentences (Experiment 3) and words in isolation (Experiment 4) when there is no visual scene present to constrain children’s interpretations.

Experiment 1

The first experiment assessed unfamiliar-accented vs. familiar-accented target word recognition in the presence of both sentential and visual context: words occurred in full sentences, and children selected the named picture from a display containing only four pictures. The target words children heard were selected from a larger set of recorded sentences, where the selection
criterion was high target word transcription accuracy by adult listeners (see Stimuli, below). If children’s accent comprehension is aided by visual context, performance should be equally good (high accuracy and rapid visual fixations) for both accents. If they are aided by sentence sensicality, then sensical sentence performance should be good in both familiar and unfamiliar accents, with lower performance (lowered accuracy and slower visual fixations) for nonsensical sentences in unfamiliar vs. familiar accents.

Method

Participants. Monolingual English-speaking preschool-aged children (\(N = 32\), 16 female, age \(M = 4.61\) years, \(SD = 0.69\), range: 3.09-5.88) from local preschool and daycare facilities took part in the study. In this and following experiments, parents reported that their children received no exposure to any language besides English at home.

Stimuli. Stimuli were full sentences ending in a target word. Sentences were in either an accent familiar to monolingual English-speaking children in the region (Californian English) or less-familiar (Mexican Spanish-accented English). The full set of SAUCE sentences contains 100 sensical, 100 nonsensical, and 100 nonword-containing sentences recorded by 20 speakers (half US-accented, half Spanish-accented; half female, half male). Target words for SAUCE sentences consisted of names for color pictures that preschool-aged children labeled consistently. Adult and child cloze norming determined that target words were highly probable completions of sensical sentences (e.g., The farmer milked the cows). To create nonsensical sentences, target words were scrambled across sentence stems (e.g., The farmer milked the nose). Nonword-containing sentences matched nonsensical sentences except that a content word in the sentence had had its
onset consonant replaced with a different onset consonant, making the sentence frame itself less sensical. These sentences were not used here.

Of the ten talkers per accent that we originally recorded, six talkers per accent (3 male, 3 female in each accent; total 12) were used in the current study. This allowed good counterbalancing of talkers across participants and conditions. Mexican-accented speakers were older on average ($M = 32.0$, $SD = 13.6$, range: 20-55) than Californian-accented speakers ($M = 22.8$, $SD = 6.6$, range: 18-36), in part because we had to recruit from the general community in order to find speakers we judged to be sufficiently accented in English. Mexican-accented speakers had begun learning English at age 11 or later ($M = 16.2$, $SD = 6.9$, range: 11-30), while Californian-accented speakers had all learned English from birth. Multiple recordings were made of each sentence for each talker so that we could screen out gross errors and disfluencies and select the most artifact-free recording. The six Spanish-accented speakers had slower average speech rates (measured by sentence duration) than the six California-accented speakers (2632 milliseconds ± 313 ms vs. 2234 ms ± 204 ms; $t(10) = 3.78$, $p = .004$).

Adult speakers of American English ($N = 49$)$^1$ transcribed sentences and words excised from sentences. Target word recognition accuracy was calculated for each utterance, and—for the current study—recordings were chosen to have high adult transcription accuracy. The logic was that this would provide a stronger test of children’s susceptibility to accent unfamiliarity; showing that children have difficulty comprehending words that adults also have difficulty comprehending would be less striking.

$^1$ For expediency, some listeners at our linguistically-diverse university were early acquirers of American English, rather than native speakers. Preliminary analyses indicated no group differences, so results are collapsed across groups.
Overall, adult listeners were more accurate in transcribing familiar-accented (94.6%, $SD = 22.5\%$) than unfamiliar-accented target words (90.7%, $SD = 29.0\%$; $\chi^2 = 27.64, p < .0001$), consistent with previous reports of adult difficulty in comprehending accented speech (e.g. Bradlow & Bent, 2008). Each accented word token used in Experiment 1 was transcribed correctly both in isolation and in its full-sentence context by at least 3 out of 4-5 adults each (Table 1). To guard against the possibility that adults had learned or adapted from previous hearings of the word (each word or sentence was heard five times per participant, from different speakers), accuracies for selected sound files were also calculated on their first hearing, showing accuracy that was still high (Table 1, lower). While all of the Spanish-accented recordings were included in the adult transcription study, we included only a subset (about 20%) of the familiar-accented recordings as a familiar-accent baseline measure, in order to maximize participants’ time with accented recordings. Therefore, transcription accuracy is not available for familiar-accented materials.

Sentences for the child study were 48 matched pairs with identical sentence stems (The farmer milked the cows/nose). Twenty-five words occurred as both a sensical target and a nonsensical target, but never for the same child. Children only heard one version of each sentence (sensical or nonsensical), and only heard a given word as target once. Six of the original ten speakers of each accent were selected for inclusion in the experiment (3 female, 3 male in each group), for ease of counterbalancing with 48 trials. Importantly, the possible sensical target did not appear as an option on nonsensical sentence trials (e.g., the cows picture on “milked the cows” trials was replaced with a picture of a nose on “milked the nose” trials), and vice versa. This means that
PRESCHOOL ACCENT COMPREHENSION

there were no highly-plausible lures amongst the pictures on nonsensical trials—just one novel picture, and two pictures with labels phonologically-unrelated to the target and semantically unrelated to the sentence.

**Table 1.** Means (SDs) of proportion correct transcriptions of experimental sentences by adult English-speaking listeners.

<table>
<thead>
<tr>
<th>Adult transcriptions</th>
<th>Sensicality</th>
<th>in sentence</th>
<th>isolated word</th>
</tr>
</thead>
<tbody>
<tr>
<td>All instances</td>
<td>Sensical</td>
<td>0.990 (0.053)</td>
<td>0.910 (0.138)</td>
</tr>
<tr>
<td></td>
<td>Nonsensical</td>
<td>0.956 (0.089)</td>
<td>0.931 (0.108)</td>
</tr>
<tr>
<td>First instance</td>
<td>Sensical</td>
<td>0.989 (0.104)</td>
<td>0.882 (0.325)</td>
</tr>
<tr>
<td></td>
<td>Nonsensical</td>
<td>0.906 (0.293)</td>
<td>0.885 (0.320)</td>
</tr>
</tbody>
</table>

*Procedure.* Children were tested in a quiet room or area in their daycare or preschool. Sentences were played over child-sized KidzGear headphones (http://www.gearforkidz.com). Pictures and sound files were presented on a Mac Mini running Matlab R2008a, using Psychtoolbox3 (Brainard, 1997; Pelli, 1997) and the Eyelink Toolbox (Cornelissen, Peters, & Palmer, 2002). On each trial, four pictures appeared: the target picture, one novel object picture, and two familiar-object pictures that did not overlap substantially in sound or meaning with the target word. After 500 milliseconds, the sound file began playing. Children were asked to point to the picture that the person was talking about. An experimenter recorded children’s pointing responses by clicking the mouse on the indicated picture. When necessary, the experimenter prompted children to point or to clarify their points. Visual fixations were recorded by an Eyelink 1000 eye tracker (SR Research, Mississauga, ON) operating in remote mode.

Each child heard 48 sentences total, equally distributed across four conditions: sensical/familiar accent, sensical/unfamiliar accent, nonsensical/familiar accent, nonsensical/unfamiliar accent.
Across children, each sentence was heard equally often in each of the four conditions. There were 8 different lists so that each word in each condition (sensical or senseless) was heard from 4 different talkers, 2 per accent. Each list was presented in a fixed order, which was random with the constraint that no picture that was a target on trial $n$ could recur (as a non-target) any earlier than trial $n+5$. This was intended to prevent children from paying undue attention to distractor pictures as a result of recent occurrence as a target.

Children also completed the Peabody Picture Vocabulary Test, Fourth Edition (PPVT-IV; Dunn & Dunn, 2007). This allowed us to assess whether language knowledge facilitated accented speech comprehension. In most cases, both PPVT scores and chronological age correlated with overall performance. For conciseness, we present vocabulary size findings from all four experiments in a separate section prior to the General Discussion.

**Results**

There were substantial differences in both accuracy and visual fixation measures as a function of sentence sensicality. However, the data suggested no accent-related differences in accuracy or visual fixations to targets.

*Pointing responses.* Data were analyzed in R (R Core Team, 2008) using the package *lme4* version 1.1-7 (Bates, Maechler, Bolker, & Walker, 2015), using logistic mixed-effects models. Mixed-effects models have the advantage of accounting for subjects and items variability in a single statistic. Jaeger (2008) has argued that logistic regression is more appropriate than
transformed or untransformed accuracy data because accuracy is fundamentally binomially distributed.

**Figure 1.** Experiment 1, children’s accuracy in pointing to the target picture. Throughout, error bars are standard errors.

A logistic regression model was computed with target selection as the binary dependent variable, and Accent (familiar, unfamiliar; within-subjects and items) and Sensicality (sensical sentence, nonsensical sentences; within-subjects and items) as sum-coded predictors. The starting model always contained maximal random effects structure, that is, subject and item (target-word) intercepts, and subject and item slopes for both the main effects and the interaction, as recommended by Barr, Levy, Scheepers, and Tily (2013). In the current experiment, the model with the maximal random effects structure did not converge, so correlations between random effects were dropped from the model as suggested by Barr et al. (2013). The model without random effects correlations converged successfully. Throughout, effects are reported via model
comparison using the `anova()` function in R. That is, for each reported significance statistic, two models were compared: both had the same random effects structure, but one omitted the fixed factor of interest (for example, Accent). A significant difference in variance accounted for indicates that that factor (here, Accent) improves the fit of the model.

Overall, children were more accurate for words heard in sensical sentences (93.2%) than nonsensical sentences (86.5%; $\chi^2(1) = 9.72, p = .002$). However, there was no effect of Accent ($\chi^2 = 0.87, p = .35$; unfamiliar accented, 88.9% vs. familiar accented, 90.8%), nor was there an Accent x Sensicality interaction ($\chi^2 = 0.00, p = .95$).

**Table 2.** Experiments 1 and 2, proportions of novelty responding (SDs) and proportion of error trials that were novelty responses (chance = .333).

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Unfamiliar accent</th>
<th>Familiar accent</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sensical</td>
<td>Senseless</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Novel</td>
<td>.021 (.042)</td>
<td>.044 (.056)</td>
</tr>
<tr>
<td>Familiar</td>
<td>.052 (.112)</td>
<td>.104 (.120)</td>
</tr>
<tr>
<td><strong>Proportion</strong></td>
<td><strong>.286</strong></td>
<td><strong>.298</strong></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Novel</td>
<td>.047 (.056)</td>
<td>.034 (.047)</td>
</tr>
<tr>
<td>Familiar</td>
<td>.078 (.099)</td>
<td>.083 (.116)</td>
</tr>
<tr>
<td><strong>Proportion</strong></td>
<td><strong>.375</strong></td>
<td><strong>.289</strong></td>
</tr>
</tbody>
</table>

Here and in other experiments, we ran models including Trial Half (first half of trials, second half of trials) as an additional predictor, to assess whether children adapted to the accent across the course of the experiment. Since no significant effects or interactions with Trial Half were found in any experiments, we do not discuss them further.
**Novelty errors.** Our second question was whether children would misrecognize familiar but accented words as novel word forms. If so, more than one-third of “errors” (non-target responses) should be novel-picture selections. If a child did not choose the target, there were three remaining possibilities, one of which was the novel object. Thus, the proportion of novelty errors predicted by chance is 1/3 (~.33). Proportions of errors that were novel errors showed little visible evidence of novelty responding (Table 2). To verify this statistically, we computed a logistic mixed-effects regression, including only non-target responses, and we assessed the probability that an error was a selection of the novel picture (1; otherwise 0). To ensure that the intercept term assessed deviation from chance, rather than deviation from .50, an offset—the log odds of 1/3—was included in the model specification. Also, because of the relatively sparse number of errors, which precludes inclusion of large numbers of predictors, only the random slopes for Accent were included, to test effects of accent on novelty responding (following Barr et al., 2013). Neither the intercept term ($\chi^2 = 0.89, p = .34$) nor the effect of Accent ($\chi^2 = 0.67, p = .41$) approached significance. This suggests that novelty errors were no more than would be predicted by chance either overall (proportion novel = .295) or on accented trials (proportion novel = .294).

**Visual fixations.** Looks were analyzed in a one-second (1000-ms) time window ranging from 200 ms to 1200 ms. This more than encompasses the full duration of target words (accented: 498 ms ± 120 ms; familiar-accented: 476 ms ± 103 ms; a small-but-significant difference, $p < .05$) and includes a 200-ms positive shift to allow for the amount of time (roughly 200 ms) needed to execute a speech-signal-based saccade (Matin, Shao, & Boff, 1993; see also Salverda, Kleinschmidt, & Tanenhaus, 2014). Following Barr (2008), linear regression models were...
PRESCHOOL ACCENT COMPREHENSION

computed on empirical-logit-transformed target looks by participants and by items, with Accent and Sensicality as within-subjects and within-items factors. Factors were mean-centered and sum-coded. Due to the small number of data points for both participants and items analyses, random effects correlations were removed, as well as the random effect of the interaction. To test interaction significance, the base model and the model without the interaction were computed with only the intercept and interaction random effect term (again, without random effects correlations), as recommended by Barr et al. (2013).

An effect of Sensicality ($\chi^2 = 27.40, p < .0001; \chi^2 = 20.87, p < .0001$) revealed more target looks overall when targets were embedded in sensical sentences. However, there was no effect of Accent ($\chi^2 = 2.23, p = .14; \chi^2 = 2.14, p = .14$) and no interaction ($\chi^2 = 0.12, p = .73; \chi^2 = 0.05, p = .82$). Thus, mirroring the accuracy data, there does not appear to be a deficit in recognition speed for accented speech.

Discussion

Children showed equivalently high comprehension accuracy for familiar-accented (Californian) and less familiar-accented (Mexican Spanish) spoken words in sentence contexts. Our findings are somewhat consistent with the hypothesis that a very small set of response possibilities (four pictures) facilitates comprehension so that the effects of accented speech on comprehension are eliminated. However, the role of sentence context in accented speech comprehension was not supported: while responses were more accurate for sensical sentences overall, this effect did not interact with accent. Thus, at least for visual context, our findings are consistent with the contextual support hypothesis. A second account of this finding, though, is that because the
Figure 2. Experiment 1, looks to pictures, top: sensical sentences; bottom: nonsensical sentences. Fam. = familiar accent, Unfam. = unfamiliar accent.
sentences were chosen based on high adult transcription accuracy, they were already easy to comprehend either because they were among the least accented in the set of recordings, or because their accent characteristics were less deleterious to recognition.

To dissociate these two explanations, the next experiment tested comprehension of target words alone, with the pre-target-word portion of each sentence removed. These explanations produce different predictions: If words are weakly accented and do not elicit comprehension difficulty, then children will again show no effects of accent familiarity. However, if comprehension is bolstered by semantic context (or hindered, in the case of nonsensical sentences), an accent familiarity effect will appear.

We also found no tendency for children to think that unrecognized accented words were novel words. This is perhaps not surprising given that accented speech comprehension was not detectably weaker than unaccented speech comprehension. Thus, in Experiment 2, we again assessed whether children showed novelty responding for the unfamiliar accent.

Experiment 2

Method

Participants. Monolingual English-speaking children \( N = 32, 16 \) female, age \( M = 4.72 \) years, \( SD = 0.63 \), range: 3.24-5.58) from the same population as in Experiment 1 took part.

Stimuli and Procedure. These were identical to Experiment 1, except that children heard only the target words instead of full sentences.
Results

Accuracy. As in Experiment 1, a logistic regression model was computed with target selection as the binary dependent variable, with Accent and Sensicality as contrast-coded predictors. The maximal random effects structure was used, that is, subject and item (target-word) intercepts, and subject and item slopes for both the main effects and the interaction. Because the maximal model with random-effects correlations did not converge, the random effects correlations were dropped. The model then converged.

Overall, children were slightly more accurate for words from nonsensical sentences (91.5%) than those drawn from sensical sentences (88.5%; $\chi^2 = 3.74, p = .05$). While it is not immediately intuitive that there should be any difference between words drawn from the two contexts, it accords with reports that speakers tend to reduce phonetic information in semantically predictable contexts (e.g. Lieberman, 1963). The effect of Accent was significant ($\chi^2 = 5.44, p = .02$), with lower accuracy overall for unfamiliar-accented words (87.9% vs. 92.2%). The interaction missed significance ($\chi^2 = 2.82, p = .09$). The marginal difference seen in Figure 3, where the accent familiarity effect appears smaller for sensical sentences, may suggest that the unfamiliar-accented speakers, as nonnative speakers, were speaking sentences more carefully overall, leading to weaker effects of predictability on care of articulation (see Warner & Tucker, 2011, on milder reduction in read speech vs. conversational speech). That is, the nonnative speakers were less likely to hypoarticulate in predictable contexts. This accords with our impressionistic judgment that unfamiliar-accented speakers were articulating quite carefully. The marginal nature of the interaction, however, cautions against any interpretation.
To assess whether removing sentential context in the current experiment generated substantively larger accent effects than in Experiment 1, we ran a second model with Accent, Sensicality, and Experiment (1 or 2) and their interactions as predictors. There was a marginal effect of Accent ($\chi^2 = 2.73, p < .10$), suggesting slightly weaker performance for unfamiliar-accented speech overall. There was also a Sensicality x Experiment interaction ($\chi^2 = 9.03, p = .003$), consistent with the effect of Sensicality reversing between the two experiments. However, there was no Accent x Experiment interaction ($\chi^2 = 1.66, p = .20$), suggesting that removal of sentential context in Experiment 1 did not actually increase accent effects. No other effects approached significance. This finding implies that the accent effect in the current study, while significant, was not detectably larger in magnitude than in Experiment 1.

![Figure 3](image_url)

**Figure 3.** Experiment 2, children’s accuracy in pointing to target pictures.

**Novelty errors.** As in Experiment 1, there was no tendency to select the novel objects more than familiar non-targets. In a logistic regression model on errors with an offset as in Experiment 1,
and with random intercepts and slopes of Accent, neither the intercept term ($\chi^2 = 0.51, p = .48$) nor the effect of Accent ($\chi^2 = 0.01, p = .93$) approached significance. This suggests that novelty errors were no more than would be predicted by chance either overall (novel responses, .036; familiar distractor responses, .064; proportion novel, .359) or on accented trials (novel responses, .040; familiar distractor responses, .081; proportion novel, .333).

Visual fixations. Linear regression models were computed on empirical-logit-transformed target looks by participants and by items, with Accent and Sensicality (of the original sentence) as within-subjects and within-items factors. As in Experiment 1, due to the small number of data points for participants and items analyses, random effects correlations were removed, as well as the random effect of the interaction. To test interaction significance, the base model and the model without the interaction were computed with only the intercept and interaction random effect term (again, without random effects correlations), as recommended by Barr et al. (2013).

An effect of Sensicality ($\chi^2 = 22.42, p < .0001; \chi^2 = 6.29, p = .01$) revealed fewer target looks overall when targets were drawn from sensical sentences. Somewhat similar to the accuracy effects, this suggests weaker recognition of words drawn from sensical contexts. There was no effect of Accent ($\chi^2 = 1.42, p = .23, \chi^2 = 0.81, p = .37$), nor was there an interaction ($\chi^2 = 2.64, p = .10, \chi^2 = 1.26, p = .26$). Thus, despite modest-but-significant accuracy differences, visual fixations appeared unaffected by accent, much like Experiment 1.
Figure 4. Experiment 2, looks to pictures, top: words from sensical sentences, bottom: words from nonsensical sentences. Fam. = familiar accent, Unfam. = unfamiliar accent.
Discussion

The current experiment yielded two findings. First, the effect of accent was more evident here—where there were no preceding sentence contexts—than in Experiment 1, where sentence contexts preceded target words. This may indicate that sentence contexts either semantically or phonologically facilitate accented word recognition, consistent with the contextual support hypothesis. However, the non-significant Experiment x Accent interaction in the cross-experiment analysis cautions that the two experiments may both represent relatively good performance in accented speech comprehension, with only a mild overall effect of accent.

The second finding is that this experiment replicated Experiment 1 in showing that children do not make novelty responses when they misperceive or fail to recognize accented words. The current experiment is a stronger demonstration of this finding in that unfamiliar-accented words showed lower accuracy relative to familiar-accented words. This stands in stark contrast to Creel’s (2012) study, where children who did not recognize mispronounced words (e.g., hearing /fɔʃ/ and pointing to a fish) overwhelmingly selected novel-object pictures as their referents. This may indicate that unfamiliar-accented speech somehow cues young children to dismiss novel-word interpretations. This might happen because the accented word is not distinctly different enough from familiar-accented pronunciations (that word might be “pig” or “peeg” or “peek,” but it is not clearly the novel form “peeg”). It might instead happen based on children’s inferences about accented speakers’ vocabulary knowledge—if the child thinks accented speakers are less likely to know names for things (Corriveau et al., 2013), then the child may not believe that accented speaker will know the name for a novel object.
However, there is another explanation of the absence of novelty responding that has nothing to do with accentedness. Specifically, children in Experiments 1 and 2 completed the PPVT-IV vocabulary assessment, many of them before the main experiment for reasons of expediency. It may be that completing a four-alternative vocabulary test in which they were asked to select familiar referents prior to the main experiment steered children into a pragmatic set of selecting familiar pictures. To test this, we ran an additional experiment on 48 children, all of whom received the PPVT post-experimentally (accented materials only; 50% nonsensical full sentences, 50% isolated words; presented in separate blocks, order counterbalanced across children). Novelty responding did not emerge in this control experiment, suggesting that the PPVT is not the reason for children’s missing novelty responses.

Interim summary

Experiment 1 suggested that 3-5-year-old children are fairly good at recognizing words in unfamiliar accents. Experiment 2 suggested greater difficulty for accented words in the absence of sentential context, though a cross-experiment analysis did not suggest strong differences in the effects in the two experiments, but rather a mild effect of accent on comprehension in both experiments. Further suggesting that the effects of accent familiarity are quite small, visual fixations did not show a familiar-accent advantage in either experiment. A second finding is that, unlike cases where children hear mispronunciations in their native phonology (Creel, 2012), children did not tend to select novel objects when they failed to recognize accented words.

The relatively mild effects of accented speech on comprehension suggest one of two things. One possibility raised already is that these may be very easy-to-comprehend accented utterances.
Alternatively, visual context—which narrows down one’s selections to four possibilities—may be especially facilitative. Thus, the next two experiments removed visual context cues, and asked children simply to repeat target words. Recordings were transcribed offline by coders blind to accent condition. If unfamiliar accents become more difficult to process in the absence of highly-constraining visual context, then effects of accent should be more evident in Experiments 3 and 4 than they were in Experiments 1 and 2.

Experiment 3

Method

Participants. \( N = 32 \) monolingual English speaking preschoolers (15 female, age = 4.35 years, SD = 0.61, range: 3.33-5.77).

Stimuli. The auditory stimuli were the same spoken sentences as presented in Experiment 1.

Procedure. Children wore a Califone 2810PA-AV microphone headset (Califone, San Fernando, CA, US), which was connected via a Griffin iMic USB adapter (Griffin Technology, Nashville, TN) to a Mac Mini running Matlab. The Matlab PsychPortAudio engine within PsychToolbox3 was used to record sound files. Children were asked to repeat the last word in the sentence. The experimenter read the following directions, including two guided examples:

> You will hear some people talking. Your job is to repeat the last word they say, as fast as you can. Like, if I said, “The chicken crossed the road,” you would say… [child should
say ROAD; if not, try again]. Or if I said, “I like to eat chocolate,” you would say…

[child should say CHOCOLATE; if not, try again]. Are you ready? Here we go!

Then the child heard 48 sentences. Once the sentence ended, recording began immediately. Recording continued until the experimenter clicked the mouse (indicating that the child had responded) or for a maximum of 10 seconds, whichever happened first.

Coding of responses. Each utterance was saved as a .wav file with a generic name (such as SubjectAA01_item1.wav) so that responses could be transcribed without knowledge of which condition its spoken stimulus was in. Two coders (SCC and ANP) transcribed responses independently and blind to accent condition (familiar or unfamiliar). One coder (SCC) rechecked their own responses with knowledge of sentence content (sensical or nonsense), but remained blind to accent condition. Note that some children responded primarily with full sentences or sentence fragments, rather than single words, which unblinded the sensicality status of the sentences. When children did produce multiple words, the response was counted as correct as long as the target word was among those words. If children made certain sound substitutions characteristic of young children’s speech productions (interchanges between l/r/w; changing /f/ to /s/; changing /s/ to /θ/ or /z/ to /∂/; reducing clusters, such as producing glasses as “gases”; simplifying words of 3 or more syllables by dropping a reduced vowel or entire reduced syllable; rarely, dropping or adding a plural (-s) or diminutive (-y) marker) they were also counted as correct (see Table 3). R-colored and l-colored vowels were also scored leniently given children’s erratic liquid productions. Other sound substitutions (different vowels; changes in consonant place of articulation, voicing, or manner), deletions, or insertions were counted as incorrect.
PRESCHOOL ACCENT COMPREHENSION

Agreement on target word correctness was 89.2%. For the remaining trials, a third coder, and in a few cases a fourth, transcribed utterances and the predominant response was selected (in cases where no coders produced the same response, the predominant response for each segment or phonetic feature was selected; for example, for transcriptions /ɡɑɪp/, /tɑɪp/ and /tɛɪp/, the resolution was /tɑɪp/). Children failed to respond, or failed to respond completely, within the 10-second time period only 4.4% of the time (exactly 34/768 trials in each accent condition). Such trials were excluded from analyses, along with trials where children stated they did not know, and trials containing interruptions, excessive noise, or distortion, and responses indicating inattention (unfamiliar accent: 16/768, or 2.1%; familiar accent, 13/768, or 1.7%). In total, 6.3% of trials were excluded from analyses, with roughly similar numbers per accent condition.

Errors (Table 4) were further classified as phonetically related to the target; production of another word or words from the sentence; a word with apparent semantic relation to the sentence; and unclassifiable. We chose not to break down phonetically-related errors into lexical errors (producing a real word) and phonetic errors (producing a nonsense word) as we believe that they are not easily dissociated for two reasons: (1) lexical errors could occur unintentionally via a phonetic misinterpretation; (2) lexical errors could be the result of coders’ lexical biases. One feature of the recordings in Experiments 3 and 4 was that, by and large, it was not clear which utterances were repetitions of unfamiliar-accented words; this surprised us given Nathan et al.’s (1998) report suggesting that London 4-year-olds often imitated characteristics of the unfamiliar accent (Scottish English). With the exception of one child who trilled two r’s, children’s responses did not appear to closely follow accented patterns subphonemically. Nonetheless, the majority of children’s errors were featurally similar to the original word.
## Table 3. Experiments 3 and 4, examples of coded target words.

<table>
<thead>
<tr>
<th>Target</th>
<th>Example correct responses</th>
<th>Example incorrect responses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Car</td>
<td>car /kɑɹ/, caw /kʌw/</td>
<td>dar /dɑɹ/, tar /tɑɹ/</td>
</tr>
<tr>
<td>Glasses</td>
<td>glasses /glæstʃ/, gwatheth, /gwæθθ/, gases /gæstʃ/, gas /gæs/</td>
<td>glosses /glosɛtʃ/, nasses /næstʃ/</td>
</tr>
<tr>
<td>Dinosaur</td>
<td>dinosaur /dɹɑnəsɔɹ/, din′saur /dɹɑnəsɔɹ/, dai-saur /dɹoɪsəɹ/, daisow /dɹoɪsəw/</td>
<td>tinosaur /tɹoɪnæsɔɹ/</td>
</tr>
<tr>
<td>Shark</td>
<td>/ʃɔrk/, /ʃawk/, /sɔwk/</td>
<td>Chark /tʃɔrk/, chawk /tʃɔwk/</td>
</tr>
</tbody>
</table>

## Table 4. Experiment 3, classifications of production errors. Familiar = familiar accent, Unfamiliar = unfamiliar accent.

<table>
<thead>
<tr>
<th>Error type</th>
<th>Accent</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Familiar</td>
</tr>
<tr>
<td>phonetically related to target</td>
<td>.096</td>
</tr>
<tr>
<td>another word(s) in sentence</td>
<td>.083</td>
</tr>
<tr>
<td>semantically related to sentence</td>
<td>.003</td>
</tr>
<tr>
<td>incorrect (no apparent reason)</td>
<td>.035</td>
</tr>
<tr>
<td><strong>Total errors</strong></td>
<td>.216</td>
</tr>
</tbody>
</table>

**Figure 5.** Experiment 3, children’s target-word repetition accuracy. Familiar = familiar accent, Unfamiliar = unfamiliar accent.
PRESCHOOL ACCENT COMPREHENSION

Results

As evident from Figure 5, unfamiliar-accented words showed lower repetition accuracy. A logistic regression model was computed with target word repetition accuracy as the binary dependent variable, and Accent (familiar, unfamiliar) and Sensicality (sensical sentence, nonsensical sentence) as sum-coded within-subject and within-item (target-word) predictors. The maximal random effects structure was used (subject and item intercepts as well as subject and item slopes for both the main effects and their interaction). Children’s repetition accuracy was higher overall for familiar-accented (77.8%) than unfamiliar-accented targets (67.5%; $\chi^2 = 2.00, p = .16$) and no interaction ($\chi^2 = 0.58, p = .45$).

Discussion

There was a large effect of unfamiliar accent here: children repeated unfamiliar-accented words less accurately than familiar-accented ones. Recall that Experiment 1, which also presented words in sentence contexts, but additionally provided a fixed set of visual representations to select from, did not show effects of accent. This suggests that removing visual context hinders unfamiliar-accented speech processing more than familiar-accented speech processing.

Interestingly, there was no facilitation from sentence context cues. The absence of a sensicality effect differs from Experiment 1, and might result from children listening with more attention focused on the ends of the sentences, such that preceding context was less facilitative than in

2 While the full model converged with random effects correlations, the model with the Sensicality effect removed did not converge. Therefore the base model and the no-Sensicality model were recalculated without random effects correlations, both of which converged. These models were then compared to derive the reported p-value.
Experiment 1. Another possible, but less-parsimonious, account is that facilitative effects of sensical sentence context (as seen in Experiment 1) combined with effects of hypoarticulation of predictable words (as hinted at in Experiment 2), which worked in the opposite direction, such that effects cancel out. However, both of these explanations are speculative and do not explain the major finding, that is, poorer repetition of words in the less-familiar accent.

The final experiment replicated Experiment 3, but removed both sources of context: sentence context and visual context. If these sources of context combined improve children’s accent processing, then accented speech processing should be especially difficult here. On the other hand, if the major factor is the presence or absence of visual context, then results of the final experiment should look similar to Experiment 3.

Experiment 4

Method

Participants. A new sample of 32 monolingual English speaking preschoolers took part (15 female, $M = 4.22$, $SD = 0.66$, range: 3.05-5.72).

Stimuli and Procedure. Auditory stimuli were the same isolated-word tokens as presented in Experiment 2. Procedure was largely the same as Experiment 3, except that children were simply asked to repeat isolated words, rather than to repeat the last word of the sentence.

Coding of responses. This followed Experiment 3 (see Table 5). As in Experiment 3, a small proportion of trials (37/1536, or 2.4%) elicited responses of “I don’t know”, non-responses, or

---

3 Thanks to Gaja Jarosz for suggesting this at the 2014 Boston University Conference on Language Development.
PRESCHOOL ACCENT COMPREHENSION

off-topic responses (e.g. fits of laughter), or suffered from recording issues (child too soft, sound clipping, excessive external noise). Such response failures, while rare, were more common for unfamiliar-accented words (25, or 3.3%) than familiar-accented words (12, or 1.6%). There were fewer failures to respond than in Experiment 3, perhaps because the task here, to repeat a single word, is cognitively simpler than Experiment 3’s task, to segment out the final word and repeat it.

Table 5. Experiment 4, classifications of production errors. Familiar = familiar accent, Unfamiliar = unfamiliar accent.

<table>
<thead>
<tr>
<th>Error type</th>
<th>Accent</th>
<th>raw N</th>
<th>Accent</th>
<th>raw N</th>
</tr>
</thead>
<tbody>
<tr>
<td>phonetically related to target</td>
<td>0.245</td>
<td>185</td>
<td>0.322</td>
<td>239</td>
</tr>
<tr>
<td>semantically related to target</td>
<td>0.001</td>
<td>1</td>
<td>0.000</td>
<td>0</td>
</tr>
<tr>
<td>incorrect (no apparent reason)</td>
<td>0.034</td>
<td>26</td>
<td>0.032</td>
<td>24</td>
</tr>
<tr>
<td>Total errors</td>
<td>0.280</td>
<td>212</td>
<td>0.354</td>
<td>263</td>
</tr>
</tbody>
</table>

Results

A logistic regression model formally identical to that in Experiment 3 was computed on target word repetition accuracy. Children’s repetition accuracy was higher overall for familiar-accented (71.9%) than unfamiliar accented words (64.6%; $\chi^2 = 7.12, p = .008$). There was also an effect of sensicality ($\chi^2 = 5.29, p = .02$), such that words drawn from nonsensical sentences (72.5%) were repeated more accurately than those from sensical sentences (64.0%). This repeats the inverted effect of sensicality seen in Experiment 2, and presumably reflects speakers’ more casual articulation in expected sentence contexts. The interaction of these terms missed statistical significance ($\chi^2 = 2.36, p = .12$).
To assess whether removing both visual and sentential context generated even larger accent effects, a second model with Accent, Sensicality, and Experiment and their interactions as predictors compared the results of the current experiment to that of Experiment 3. There was an overall effect of Accent ($\chi^2 = 14.98, p = .0001$), suggesting weaker performance for unfamiliar-accented speech overall. There was also a Sensicality x Experiment interaction ($\chi^2 = 6.94, p = .008$), such that the effect of Sensicality reversed between the two experiments. However, there was no Accent x Experiment interaction ($\chi^2 = 1.75, p = .19$), suggesting that removal of sentential context in Experiment 4 did not lead to worse performance—and also that the presence of sentential context did not lead to worse accented-speech performance. No other effects approached significance.
Discussion

These results concur with the findings of Experiment 3 that, in the absence of picture context, children are less accurate at repeating unfamiliar-accented words than familiar-accented ones. Interestingly, there does not appear to be a substantial change in the size of the accent effect when words are excised from sentence contexts. This may mean that sentential context on the whole is relatively less impactful than visual context.

In the next two sections, we discuss findings from the set of experiments as a whole. First, we consider effects of vocabulary knowledge on performance in all experiments. Then, we compare findings across all experiments, to assess the degree to which each type of context (visual and sentential) impacts comprehension.

Effects of vocabulary size

Because there is substantial interest in the role of vocabulary size in accented speech comprehension (Bent, 2014; Best et al., 2009; Mulak et al., 2009; Van Heugten et al., 2015) and comprehension generally (e.g. Borovsky, Elman, & Fernald, 2012), we briefly overview the effects of vocabulary size on performance across experiments (summarized in Table 6). Since vocabulary size correlates strongly with age, we also consider the effects of chronological age. Age correlated with accuracy in Experiments 1 and 3, and vocabulary correlated with accuracy in Experiments 1, 3, and 4. In all experiments, age and vocabulary size (that is, PPVT-IV raw score) correlated positively with each other. To assess whether vocabulary size had effects independently of age, models were computed with age and vocabulary as predictors, including
the interaction of the two and their interactions with other predictors. Results are summarized in Table 6. In only one case did the effect of vocabulary size explain additional variance once age was accounted for: accuracy in Experiment 1. Of course, this reflected simply that children with larger vocabularies were more accurate overall, rather than being selectively more accurate on accented trials (which did not, in Experiment 1, differ from familiar-accented trials). In other cases, age appeared to account for independent variance. This pattern suggests that age, vocabulary, or both are related to better comprehension overall, but that this benefit is spread across both familiar-accented and unfamiliar-accented speech (see Bent, 2014, for a similar pattern). One possibility is that stronger representations of familiar words provides better top-down support for recognizing less-familiar (accented) realizations of those words, though other accounts emphasize phonological constancy as the operative mechanism (Best et al., 2009;

**Table 6.** Effects of age and vocabulary size on performance across experiments.

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Correlations</th>
<th>Effects present in model$^1$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Age &amp; Acc</td>
<td>Vocab &amp; Acc</td>
</tr>
<tr>
<td>1</td>
<td>***0.630</td>
<td>***0.739</td>
</tr>
<tr>
<td>2</td>
<td>0.288</td>
<td>0.125</td>
</tr>
<tr>
<td>3</td>
<td>**0.492</td>
<td>*0.393</td>
</tr>
<tr>
<td>4</td>
<td>0.295</td>
<td>*0.422</td>
</tr>
</tbody>
</table>

* $p < .05$, ** $p < .01$, *** $p < .001$  

Note. Acc = overall accuracy, Vocab = PPVT-4 raw vocabulary score, ET = eye-tracked visual fixations to target.  
$^1$ Models included Sensicality, Accent, Age and Vocabulary, and all interactions. Because items analyses for eye tracking had to be calculated separately, and items themselves do not have ages or vocabulary scores, only subjects random effects were included in eye tracking analyses (intercepts and all slopes).  
$^2$ There was, however, an interaction of Age x Vocabulary, which appeared to cancel the main effects of each factor. The nature of the interaction was that children with low values on both measures showed especially low target looks.
There was also an interaction of Sensicality x Accent x Age. This appeared to result from increases of looks with age in all Sensicality x Accent cells except for sensical unfamiliar-accented.

Mulak et al., 2009). Another account is that with age comes not only vocabulary knowledge, but real-world knowledge, which can serve as a source of constraint on comprehension.

Cross-experiment comparisons

A major question of interest was whether the task itself increased the difficulty of accented speech processing in children. It should be borne in mind that the two tasks being compared are quite different, and thus a direct comparison should be interpreted with caution. We compared the four experiments in a single model, with Accent and Sensicality as within-subjects and within-items factors, and Visual Context (present, absent) and Sentence Context (present, absent) as between-subjects, within-items factors. Intercepts and all random slopes were included; the full model with random correlation parameters failed to converge, so the model was rerun without random correlations. There was an overall effect of Accent ($\chi^2 = 111.99, p = .0005$), as well as an effect of Visual Context ($\chi^2 = 70.97, p < .0001$). Sensicality and Sentence Context (sentence vs. word) interacted ($\chi^2 = 24.17, p = .0007$), consistent with higher accuracy for sensical targets with full sentence contexts, but the reverse for isolated words. Interestingly, there was no Accent x Visual Context interaction ($\chi^2 = 0.83, p = .36$), suggesting that the numerically larger accent effects in the absence of visual context (Experiments 3 and 4) did not statistically differ from accent effects in the presence of visual context (Experiments 1 and 2), at least in the log-odds space of the logistic regression model. An Accent x Visual Context x Sentence Context interaction ($\chi^2 = 3.84, p = .05$) appeared to result from a nonsignificantly larger accent effect for individual words than for sentences in the presence of visual context ($\chi^2 = 0.18, p = .67$), and a
nonsignificant effect in the opposite direction when visual context was absent ($\chi^2 = 1.78, p = .18$). No other effects reached significance. This analysis suggests that, for the current materials, changing the task did not change the relative difficulty of familiar and unfamiliar-accented materials—it simply changed the overall accuracy level, affecting both familiar and unfamiliar accents.

What does this mean with regard to context effects (or, if one prefers, task effects)? It may indicate that accented speech imposes a fairly consistent increment in difficulty regardless of task, but if a given task is more difficult overall, overall performance pulls away from ceiling and accent effects are easier to detect. This ceiling effect is muted in the log odds space of logistic regression, but not in more traditional analytic frameworks, including those used in many speech comprehension experiments. It is worth noting that ceiling effects cannot logically be eliminated by logistic models (or any data transformation), given that ceiling effects are a property of the measure itself and not the presumed underlying cognitive faculty.

**General Discussion**

We sought to answer two questions: in what circumstances do children show good recognition of unfamiliar-accented speech, and what kinds of comprehension errors do they make? The answer to the first question seems to be that when visual contextual support is removed children’s comprehension may falter. With supportive visual context and words in full sentences (Experiment 1), children were equally accurate at recognizing target words in unfamiliar and familiar accents—and, interestingly, equally inaccurate for nonsensical sentences. When sentential context was removed (Experiment 2), unfamiliar-accent accuracy undershot familiar-
accent accuracy, yet the accent effect did not differ in magnitude from the non-significant effect in Experiment 1. For both full sentences and isolated words (Experiments 3 and 4 respectively), removing visual context resulted in unfamiliar-accented word recognition that was lower than familiar-accented word recognition.

Looking across all studies, logistic regression analyses did not suggest differences in the size of the accent effect, even though the accent effect was numerically larger for the production task (Experiments 3 and 4) than for the recognition task (Experiments 1 and 2). The most cautious interpretation of this finding is that accent effects are not larger but are more robust in the face of decreased visual contextual support. This further suggests a reason for seemingly variable effects of accented speech comprehension across different age groups: tasks typically used with infants tend to provide a high degree of contextual constraint, often in the form of providing a small set of pictures as response options (Creel, 2012; Schmale et al., 2015; Swingley & Aslin, 2000, 2002; White & Morgan, 2008; White & Aslin, 2011). Studies showing less success in accent comprehension tend to provide less constraint, typically asking children to repeat words in the absence of pictured referents or highly-constraining sentences (Bent, 2014, 2015; Nathan et al., 1998). Findings here suggest that seemingly-different effect sizes may be quite comparable between tasks, with the difference being that easier tasks are at a functional ceiling and this makes the same effect size more difficult to detect.

Second, we asked whether children make erroneous novel-object attributions for accented familiar words, as they do for mispronounced words in their native accent (“fesh” for fish; Creel, 2012). Here, the answer is no. Both Experiment 1 and Experiment 2 presented novel pictures as
response alternatives, but in neither experiment were there even trends toward selecting novel objects. These data suggest that children do not automatically assume that an unfamiliar accented word is a novel word. Though more research is needed, these data are consistent with the hypothesis that accented speech elicits alternative comprehension strategies from young children. That is, children may have learned, perhaps from incidental contact in their daily lives, not to invoke novelty responding when hearing an accented speaker. That is, monolingual English-speaking children may have encountered Spanish-accented speakers in the community and may have observed that they sometimes provide atypical labels for familiar things. Previous studies suggest that 24-month-olds are less likely to trust someone who has mislabeled familiar objects to know the name of a novel object (e.g. Koenig & Woodward, 2010), and in particular, Corriveau et al. (2013) and Hoicka and Akhtar (2011) have found that young children are less likely to credit foreign-accented or foreign-language speakers with knowing words for objects. Previous researchers have suggested that children are less likely to deploy mutual exclusivity when hearing a different language for both bilingual (Byers-Heinlein & Werker, 2009, 2013; Houston-Price, Caloghiris, & Raviglione, 2010) and monolingual children (Au & Glusman, 1990), but this is the first account of diminished novelty responding for different accents. This account of course presumes that young children can detect (explicitly or implicitly) that a speaker has an accent, a task that other studies suggest is difficult for 5-year-olds (Floccia, Butler, Girard, & Goslin, 2009; Girard, Floccia, & Goslin, 2008; Wagner, Clopper, & Pate, 2014; though see Kinzler, Dupoux, & Spelke, 2007, for a different account).

Other explanations for the absence of novelty responding are more perceptually-based. One is that children are uncertain of the status of the accented words. Consider a form such as p?g,
PRESCHOOL ACCENT COMPREHENSION

where ? represents a sound ambiguous between /i/ and /i:/ it is not clearly the familiar-accented form /pig/, nor is it clearly a novel form /pig/. That is, the accented word forms presented to children in the current study may not have been sufficiently phonetically distant from familiar forms to mark them clearly as novel words. This predicts that more strongly-accented words that differ more starkly from their familiar-accent counterparts should lead to novelty responding, in contrast to the mildly-accented target words presented in the current study. A second possibility is preexisting perceptual adaptation to the accent. This would predict that mild mispronunciations that violate the accent (hearing a Spanish-accented speaker say “pag” for pig) would be more likely to generate novelty responding than mispronunciations that are consistent with the accent (asking the same speaker to say “peeg” for pig). A final possibility is that, following Schmale et al.’s (2015) general expansion hypothesis, our use of multiple speakers led to more lax speech sound boundaries in children, which would drop their detection rates of (potential) novel words. General expansion might be taken to predict that only highly-novel words in an accent would generate novelty responding (e.g. blicket, when pictured familiar objects have names that are all very different from blicket). Distinguishing amongst these various explanations exceeds the scope of the current paper, but is being pursued in our lab.

Contributions

The current study makes several novel contributions to the literature. First, this study tests young children, an age group whose accented-speech comprehension is somewhat understudied, thus adding to the developmental picture of accented speech comprehension (see Cristia et al., 2012). Critically, this is the first study to directly compare two tasks, one which is used frequently with infants and toddlers and provides higher contextual support (visual fixations to named pictures)
and another that is used more with young children and provides less contextual support (word repetition; Bent, 2014; Bent & Atagi, 2015; Nathan et al., 1998; Newton & Ridgway, in press), and we find that accent processing difficulty is more robust in the task with less contextual support. Further, confirmation of difficulty in accent processing by young children provides further support for theories of protracted speech sound development (e.g. Creel & Jiménez, 2012). A second contribution is that we find detectable processing difficulty for an accent that is present in a sizable minority of our child participants’ community. This suggests that findings of better comprehension of community accents (Floccia et al., 2012) may depend on that accent being exhibited by the majority of speakers. A third contribution is that, as mentioned, we find no tendencies to misinterpret accented familiar words as referring to novel objects, a finding that remains to be explored more fully. These contributions together substantially enrich our picture of accented speech processing during development.

Comparison to previous studies

Role of adaptation. As we noted briefly earlier, there was no evidence of accent adaptation in any of the experiments. This contrasts with studies where adults have shown adaptation to real or fictitious accents (Bradlow & Bent, 2008; Clarke & Garrett, 2004; Kraljic & Samuel, 2005, 2006; Norris, McQueen, & Cutler, 2003), and with studies of infants and toddlers suggesting acceptance of accented variants after brief exposure (Schmale et al., 2010; Schmale & Seidl, 2009; White & Aslin, 2011). This might be because adaptation happened within a single trial or a small number of trials. Alternatively, children may be to some extent pre-adapted to the Spanish accent via (limited) community exposure (see Floccia et al., 2012, and Howard et al., 2014). Of course, even complete accent adaptation cannot yield perfect comprehension in every
PRESCHOOL ACCENT COMPREHENSION

circumstance. For example, if an accented speaker collapses the distinction between the English vowels /i/ and /ɪ/, adaptation to this vowel merger will not help the listener distinguish the words *peek* and *pick*, because those words would be homophonous in such an accent.

*Role of language knowledge.* As noted earlier, vocabulary size and chronological age correlated with overall accuracy, though not for accented speech specifically. This fits with existing findings in infants and toddlers that vocabulary size predicts better accented speech comprehension (Mulak et al., 2013; van Heugten et al., 2015), and with studies which suggest that age increases from childhood to adulthood are correlated with accented speech comprehension (Bent 2014; Bent & Atagi, 2015; as well as Nathan et al., 1998). In a group of 4- to 7-year-olds, Bent (2014) found correlations of children’s speech-in-noise recognition with both vocabulary size and age, with no clear differences in the strength of the correlations for familiar and unfamiliar accents. In Bent’s (2014) data, vocabulary appeared to explain independent variance once age was controlled for. In the studies presented here, age accounted for independent variance in some experiments, while in others, vocabulary did. This makes it somewhat difficult to say whether one exerts a stronger influence. In any case, the general lack of interactions with accent suggests that vocabulary knowledge may aid accented speech comprehension because a larger and presumably more robust knowledge of words provides greater top-down word-form knowledge that supports word recognition in any accent (see van Heugten et al., 2015, for a very similar argument). Chronological age presumably also correlates with increases in accented-speech experience, which might aid accented-speech comprehension. However, if greater accent exposure with age should not also increase *familiar*-accented speech comprehension, a pattern that is more consistent with increases in vocabulary size.
Limitations and alternative explanations

Poor production indicates poor comprehension. The assumption in the latter two experiments is that if children mis-imitated a word, they did not understand what word was being said, and that this failure to understand resulted from weaker visual contextual support. An alternative interpretation is that comprehension was unaffected: children understood accented words perfectly well, but were trying to imitate tokens faithfully, and this faithful imitation gave the spurious appearance of failure to recognize an accented word. Similarly, they might have comprehended well, but had heightened difficulty in imitating the unfamiliar-accented word forms. While Nathan et al. (1998) asked children to define words to assess understanding, we opted not to do so as we were concerned that the task was too metalinguistic for our youngest participants. Thus, it is unknown to what extent our children understood apparently mis-imitated words. While this explanation cannot be dismissed entirely, there are several factors that cast doubt on it. First, Nathan et al. (1998) report that 4-year-old children, the most comparable in age to our own sample, provided the correct definition for such “phonetic responses” only 16% of the time, with incorrect, factitious, or no definitions on 84% of phonetic-response trials, suggesting that children largely misunderstood words when they mis-imitated them. Second, in Bent’s (2014) study of children’s comprehension of accented words in noise, the author reported that when children produced non-words, they were very unlikely to define them (only five instances in the entire sample). Nonetheless, future work pursuing the contextual support hypothesis might consider a comprehension task that is more cognitively challenging.
PRESCHOOL ACCENT COMPREHENSION

The critical task difference is amount of context. We argue here that the major difference between the comprehension and production tasks is a difference in amount of supportive visual context. However, the tasks differ in other ways that have nothing to do with contextual support. In particular, comprehension is easier than production and can be observed at earlier developmental time points (e.g. Bergelson & Swingley, 2012). Future studies might consider a comprehension task with more alternatives (making the context less supportive), or a production task with pictures present (lending greater contextual support to the production task).

Generalizability to real-world accented speech perception. Here we found a relatively modest decrement in children’s comprehension when they were presented with accented speech. It is significant that this difficulty appears even though the accent may be somewhat familiar to children in San Diego: while all children’s parents reported that their child had had no exposure to languages besides English at home, children in this region are likely to encounter Spanish accents in many public settings given the high proportion of Spanish-speaking residents. Recall that Floccia et al. (2012) found that bidialectal English children understood the rhotic (r-containing) community accent’s pronunciations of words like *bird* while they did not understand non-rhotic (r-less) pronunciations of *bird*. This is especially striking in that at least one of the child’s parents had a non-rhotic accent, suggesting a very strong role for community language input. Nonetheless, our study suggests that there are limits to community exposure. This may mean that effects of community (vs. home) exposure occur mainly when the community accent is the majority accent (as in Floccia et al.’s study), rather than a minority of accents (from 11%-25%, as in our study).
PRESCHOOL ACCENT COMPREHENSION

It is important to emphasize that children’s performance here may represent best-case accented speech comprehension in children for several reasons. First, recordings were screened for high adult comprehension accuracy. While this was a principled decision, it may have yielded the most-intelligible subset of the stimuli, reducing apparent accented-speech effects. Second, the accented speech used here consisted of scripted sentences, which may be more carefully articulated than is typical of conversational speech (Warner & Tucker, 2011). More fluent, conversational accented speech might present greater comprehension difficulty. We have some hints of this in two senses: impressionistically, our accented speakers were quite careful in their articulations, to a greater extent than our familiar-accented speakers, and were still less comprehensible to children; and further, children’s response patterns implied that mild phonological reduction of words due to contextual predictability decreased intelligibility. In sum, a number of considerations suggest that, for many realistic listening environments, children’s accented speech processing may be weaker on average than the current study indicates. A complementary interpretation is that children perform best at comprehending accented speech in highly contextualized settings.

Conclusion

We provide the first study to compare two paradigms assessing young children’s comprehension of accented speech, bridging between literatures on infants and toddlers on the one hand and young children on the other. Children’s comprehension difficulty in accented speech was more evident when facilitative sources of context were removed. We detected accent processing difficulty despite the presence of the relatively less-familiar accent in our monolingual participants’ broader community, suggesting that low-frequency community accents may
Preschool Accent Comprehension

generate more processing difficulty than high-frequency ones (as in Floccia et al., 2012), and supporting theories of protracted development of speech processing. We also found that, when children failed to recognize accented words, they did not tend to interpret them as novel word-forms, apparently discarding known tendencies toward disambiguation or mutual exclusivity. This lack of novelty responses might have a perceptual basis, or instead might reflect a high-level social-cognitive tendency to infer that accented speakers have weaker word knowledge.
References


PRESCHOOL ACCENT COMPREHENSION


doi:10.1111/j.1467-7687.2012.01173.x


doi:10.1016/j.tics.2015.09.006


doi:10.1121/1.1323233


PRESCHOOL ACCENT COMPREHENSION


