Developmental Science

Developmental Science (2012), pp 1–18 DOI: 10.1111/j.1467-7687.2012.01173.x

PAPER

Phonological similarity and mutual exclusivity: on-line recognition of atypical pronunciations in 3–5-year-olds

Sarah C. Creel

Department of Cognitive Science, University of California San Diego, USA

Abstract

Recent research has considered the phonological specificity of children's word representations, but few studies have examined the flexibility of those representations. Tolerating acoustic–phonetic deviations has been viewed as a negative in terms of discriminating minimally different word forms, but may be a positive in an increasingly multicultural society where children encounter speakers with variable accents. To explore children's on-line processing of accented speech, preschoolers heard atypically pronounced words (e. g. 'fesh', from fish) and selected pictures from a four-item display as eye movements were tracked. Children recognized similarity between typical and accented variants, selecting the fish overwhelmingly when hearing 'fesh' (Experiment 1), even when a novel-picture alternative was present (Experiment 2). However, eye movements indicated slowed on-line recognition of accented relative to typical variants. Novel-picture selections increased with feature distance from familiar forms, but were similarly sensitive to vowel, onset, and coda changes (Experiment 3). Implications for child accent processing and mutual exclusivity are discussed.

Introduction

A crucial aspect of development is vocabulary acquisition: by learning words for things, children can send and receive messages to and from those around them. This code critically depends on recognizing familiar words despite changes in surface form. 'Cat', for instance, still means a four-legged furry bewhiskered meowing entity when it is spoken with high or low pitch, quickly or slowly. While some evidence suggests that children less than a year old have trouble recognizing word forms across marked acoustic differences (Houston & Jusczyk, 2000; Schmale, Cristià, Seidl & Johnson, 2010; Schmale & Seidl, 2009; Singh, 2008), with generalization across these factors occurring roughly around age 1 year, it is largely unknown how unfamiliar speech variability affects word recognition later in childhood.

One prevalent source of variation in word forms is *accents*. Accents stem from differences in speakers' phonological realizations of the same verbal material, due to regional variation or native-language background. Accent variability is widespread in children's environments: many children in the United States grow up hearing and speaking sociolects (African American English: Green, 2002; Chicano English: Eckert, 2008; Fought, 1999), which may differ from the accent they are later schooled in. Children growing up in bilingual environments are also likely to hear foreign-accented speech in each language. For such bilinguals, the developmental course of sound categorization differs from that observed in monolinguals (Werker & Tees, 1984), such that certain sounds in a language may not be distinguished in early infancy (Bosch & Sebastián-Gallés, 2003; Sebastián-Gallés & Bosch, 2009). Later on, bilingual toddlers and preschoolers accept accent-like mispronunciations unless dominant in the language containing the contrast (Ramon-Casas, Swingley, Sebastián-Gallés & Bosch, 2009). Native-speaking adults who easily discriminate a sound distinction appear to represent accented 'mispronunciations' as separate lexical items (Sebastián-Gallés, Echeverria & Bosch, 2005; Sebastián-Gallés, Vera-Constán, Larsson, Costa & Deco, 2009). Accents are often associated with social class (Eckert, 2008; Fought, 1999; Green, 2002). Even very young children can use accent or language familiarity to make decisions about in-group vs. out-group status (Hirschfeld & Gelman, 1997; Kinzler, Dupoux & Spelke, 2007). Thus, understanding accent variation is relevant for social as well as linguistic development.

Accents differ in phonetic realizations of consonants and vowels, deletion or insertion of phonemes, and even prosody. Residents of Rochester and Detroit may say 'log' more like other US speakers say 'lag' (Labov, Ash & Boberg, 2006); African-American-English speakers may say 'log' similarly to 'lock' via the phonological process of word-final devoicing (Kohler, Bahr, Silliman, Bryant, Apel & Wilkinson, 2007). As a result, unfamiliar accents impede familiar-word recognition in adults (Bradlow &

Address for correspondence: Sarah C. Creel, University of California San Diego, Cognitive Science, 9500 Gilman Drive Mail Code 0515, La Jolla, CA 92093-0515, USA; e-mail: creel@cogsci.ucsd.edu

Bent, 2008), children (Nathan, Wells & Donlan, 1998), and infants (Best, Tyler, Gooding, Orlando & Quann, 2009; Schmale *et al*., 2010; Schmale & Seidl, 2009). Accent-unfamiliarity may interfere with processing at multiple levels. It may impede segmentation – locating word boundaries in the speech stream. In addition, even when listeners locate word boundaries, accent-variants might register as *unfamiliar* words, which might trigger *mutual exclusivity*, a process thought to support vocabulary acquisition. Specifically, when a child sees an object with a known name and an object without a known name, and hears a novel word, the child tends to map the novel word to the novel object (Golinkoff, Mervis & Hirsh-Pasek, 1994; Halberda, 2003; Markman & Wachtel, 1988). If a child assumes that accent-variants are novel words, they may not only fail to recognize words, but also may incorrectly map words to novel referents. Interestingly, children growing up in multilingual environments appear not to use mutual exclusivity in hearing novel words (Byers-Heinlein & Werker, 2009), suggesting that mutual exclusivity itself may not be used frequently in word learning when a category can have multiple labels across languages.

Mispronunciation sensitivity in childhood

Recent work explores the phonological specificity of young children's word representations by presenting correctly pronounced vs. mispronounced words. Mispronunciations are analogous to some accented pronunciations – specifically, those accented pronunciations that result in misidentification of the intended phoneme. This literature suggests that children's representations of familiar word forms include phonetic detail shortly after the first year of life (e.g. Mani & Plunkett, 2007; Swingley & Aslin, 2000, 2002; White & Morgan, 2008). For instance, Swingley and Aslin (2002) showed 15-month-olds pairs of pictures (e.g. a dog and a sock). Children visually fixated the dog more when they heard *Look at the* <u>*doggy*</u> (a canonical pronunciation) than when they heard *Look at the* <u>*toggy*</u> (an atypical pronunciation). However, even *toggy* generated more dog-looks than sock-looks, suggesting sensitivity to partial similarity between *toggy* and *doggy*. White and Morgan found that 19-month-olds in a similar task were sensitive to how many phonological features differed from the original word. Duta, Styles and Plunkett (in press) presented 14-month-olds with pictures accompanied by spoken words or nonwords. Infants showed a phonological-mismatch ERP, but *not* a semantic-mismatch ERP, to vowel mispronunciations of words, suggesting that though they registered a mismatch at the perceptual level, they may not have registered mispronunciations as incorrect picture-labels.

Older children also seem sensitive to gradient word similarity. Merriman and colleagues (Jarvis, Merriman, Barnett, Hanba & Van Haitsma, 2004; Merriman & Schuster, 1991; see also Merriman & Marazita, 1995, and Swingley, 2009, on 2-year-olds) explored novelty responding in preschoolers. Merriman and Schuster found that 4-year-olds given one novel and one familiar object were more likely to choose the familiar word when the novel word (e.g. 'span') sounded similar to the familiar object's name (*spoon*; 34% novel) than when it did not (*wagon*; 75% novel). Other researchers have examined preschoolers' explicit judgments of similarity or detection of mispronunciation (Cole, 1981; Gerken, Murphy & Aslin, 1995; Storkel, 2002). Children's likeness judgments for a word (e.g. *nick*) compared to a standard (*lick*) are lower than identity judgments (*lick*-*lick*), but are higher when there are more overlapping phonemes (Gerken *et al*., 1995; Storkel, 2002). Cole additionally found a location-of-change effect: 4–5-year-olds, like college students, detected mispronunciations better in word-initial position (see also Creel, Aslin & Tanenhaus, 2006; Creel & Dahan, 2010). These studies together suggest that infants and preschoolers are sensitive to similarity and dissimilarity of word forms to existing word representations.

An additional consideration in specificity of children's word representations is segment type (consonants vs. vowels). Work by Nazzi and others using a naming-similarity task (Havy, Bertoncini & Nazzi, 2011; Havy & Nazzi, 2009; Nazzi, 2005; Nazzi & Bertoncini, 2009; Nazzi, Floccia, Moquet & Butler, 2009) suggests that children from 16 to 36 months are more sensitive to consonant changes than vowel changes (see also Jusczyk, Goodman & Baumann, 1999, on 9-month-olds). However, using a different task, Mani and Plunkett (2007, 2008) find sensitivity to vowel changes from 14 to 21 months, and Havy *et al*. find roughly equivalent sensitivity at 4–5 years using the original Nazzi task. Greater flexibility to vowel changes is consistent with several adult studies (Creel *et al*., 2006; Cutler, Sebastián-Gallés, Soler-Vilageliu & Van Ooijen, 2000; Van Ooijen, 1996). While these data are not harmonious, they raise the possibility that children may tolerate some pronunciation changes more than others.

Accent processing

Like the study of phonetic specificity, accent processing is also concerned with how listeners register altered pronunciations of familiar words. However, rather than sensitivity to altered pronunciations, accent processing requires *flexibility*: can listeners tell they are hearing a variant of a familiar word, despite not having heard that exact variant before? Must they learn an accented variant as a separate word form (Sebastián-Gallés *et al*., 2005), or can they recognize based on partial similarity?

Many researchers have studied accent intelligibility before and after exposing adult listeners to accented speech (Bradlow & Bent, 2008; Clarke & Garrett, 2004; Maye, Aslin & Tanenhaus, 2008). Adults show gains in transcription accuracy and lexical decision to accented materials following exposure. For instance, Maye *et al*.

(2008) used a shifted-front-vowel 'accent' – for example, *witch* sounded like 'wetch'. Adults heard a passage in either a US-English accent or the shifted accent. Lexical decisions to vowel-shifted words (e.g. 'melk' [milk]) were higher after shifted exposure than after unshifted exposure. This can be interpreted as greater flexibility (in the direction of the accent shift) in recognizing accented words after exposure.

In many studies, children show less flexibility than adults to changes in word form representations. Nine-month-olds' recognition of a familiarized word is disrupted by a change in regional or foreign accent (Schmale *et al.*, 2010; Schmale & Seidl, 2009), with slightly older infants (12–13 months) recognizing word forms across accents. Research on preschool and young school-aged children's processing of unfamiliar regional accents (Nathan *et al.*, 1998) or foreign-language phonemes (Baker, Trofimovich, Flege, Mack & Halter, 2008; Flege, 1991; Flege, Yeni-Komshian & Liu, 1999; Oh, Guion-Anderson, Aoyama, Flege, Akahane-Yamada & Yamada, 2011) suggests that children well past infancy are more sensitive, or less flexible, than adults when interpreting accented speech sounds. This may occur because their native phonology is less entrenched, leading to difficulty mapping accented words and sounds to native categories (e.g. Flege *et al.*). Nonetheless, White and Aslin (2011) recently found that 19-month-olds who were briefly exposed to a shifted vowel 'accent' (dog became 'dag') were induced to fixate the familiar picture (a dog) rather than a novel picture upon hearing the mispronounced token. This suggests that very young children (1.5 years) can learn to accept a small change in pronunciation.

Though children seem to have difficulty with accented input, few studies have explored directly how unfamiliar accents might impede children's word recognition. A primary question is whether children recognize the similarity between accented productions and familiar-word representations: what sound changes make a word *not that word* anymore? Existing literature answers these questions only indirectly. Infant looking time studies suggest that children react differently to familiar pronunciations from atypical pronunciations, but it is not clear if weaker recognition of a changed word form indicates a novel-word interpretation. Explicit-similarity-judgment studies in preschoolers (Gerken *et al.*, 1995; Storkel, 2002) suggest that children recognize both similarity and discrepancy between phonologically related word forms, but do not show whether children interpret these forms as familiar words. Merriman and Schuster's (1991) work suggests that sound similarity affects novelty decisions, but does not systematically explore particular types of sound changes.

The current study

This study presents 'accented' (mis)pronunciations as a first pass at understanding preschool children's processing of accented speech. Do preschoolers recognize a form similar to a familiar word as being the word itself? This relates most closely to Merriman and Schuster's (1991) work, but expands greatly upon it in three ways. First, mispronunciations here were more subtle (1–2 phonological features), putting them closer to the range of accent variability than Merriman and Schuster, who used starker sound changes (e.g. additional syllables and distant phoneme changes, e.g. /w/→/s/, /ei/→/u/). Second, this study presents another response alternative (phonologically unrelated familiar pictures) to measure what happens when children do not select the target – how much a novel-object response reflects *choosing* the novel picture vs. *rejecting* the familiar one. Third, the current study obtains a fine-grained implicit measure of recognition – eye tracking – that allows linkage to an earlier literature in younger children (e.g. Swingley & Aslin, 2000, 2002; Ramon-Casas *et al.*, 2009).

Preschoolers' responses to familiar-word mispronunciations were explored in three eye-tracked spoken-language comprehension experiments. In the first, an artificial accent shifted vowel pronunciations to other English vowels (Houde & Jordan, 1998; Maye *et al.*, 2008). Experiment 2 presented the same accent, but added novel pictures to allow mutual-exclusivity interpretations. Experiment 3 explored effects of segment type and degree of phonological feature mismatch on recognition difficulty and novelty responding. All experiments measured visual fixations (looks to depicted objects) and accuracy (selecting referents of atypically pronounced words) to gauge moment-by-moment processing and asymptotic recognition. This study provides insight into three related issues. First, how does children's sensitivity to mispronunciations shape their recognition of a word? Second, how does sound similarity interact with mutual exclusivity? Third, do certain types of atypical pronunciations (consonant vs. vowel changes; single vs. multiple feature changes) affect recognition more strongly?

Experiment 1

The first experiment explored preschoolers' identification of canonically pronounced (CP) and atypically pronounced (AP) words in a highly-constraining context: a set of four familiar pictures, one of which had a name close to the AP. This allowed assessment of whether, in the easiest case possible, preschoolers were affected by APs. The accent was a shift in vowel space for front vowels, as in Maye *et al.* (2008). Those authors shifted 'higher' vowels – those produced with the tongue higher in the mouth – down to the next lower vowel (/ɪ/→/ɪ /→/ɛ/→/æ/→/ɑ/). For instance, 'witch' sounded like 'wetch', and 'fetch' like 'fatch'. The current experiment used downward (pit→pet→pat→pot; 'fish'→ 'fesh') *and* upward (Pete ← pit ← pet ← pat; 'fish'→'feesh') shifts of /ɪ /, /ɛ/, and /æ/ to allow for future studies that employ multiple accents.

If preschoolers are sensitive to vowel changes, then children hearing APs should look more slowly to the target, and may be less accurate, than children hearing CPs. Further, children hearing APs may become accustomed to sound changes over the course of the experiment – adapting to the accent – which would be reflected by faster looking times and higher accuracy in the second half of the experiment.

Method

Participants

$N = 48$ monolingual English-speaking children (22 female; ages 3.6–6.1 years, $M = 4.6$; one age not reported) from local preschools and day cares took part. An additional ten children were run but excluded for: missing multiple control trials (four); not finishing (two); reported exposure to a language besides English (four). Three children were included in accuracy analyses but not eye tracking analyses because looks to pictures was less than 75% during the tested time window (either the eye tracker could not detect a gaze position at all, or children were looking at something other than pictures.).[1]

Stimuli

Pictures. Pictures were pretested for identifiability on 16 or more children. Average accuracy for exact labeling – for instance, calling the lizard picture 'lizard' but not 'gecko' – was 83% (73% for control trials). Twelve experimental pictures had labels containing front vowels. Twelve control pictures had labels without front vowels.

Recordings. Recordings were made by the author, a native speaker of US English, in a sound-treated chamber. Altered words (Table 1) were produced naturally in a neutral sentence context which contained none of the shifting vowels ('Point to the ___'). Average duration of down-shifted words (670 ms) was significantly longer than unshifted (628 ms; $p < .05$), but neither differed from up-shifted words (645 ms). Acoustic analyses of vowels reflected the intended shifts in the first two formants (F1, F2) and vowel durations (see the Appendix for details).

Procedure

Testing took place in a quiet room in the child's preschool/day care facility. The eye tracking computer monitor (attached to a Dell tower running DOS) and the experimental display monitor (attached to a Mac mini) were positioned facing directly away from each other. The Mac ran Matlab, which presented picture and sound stimuli via PsychToolbox3 (Brainard, 1997). Children sat in an unbuckled car seat to maintain a consistent distance from the display computer monitor and the remote eye tracker. The eye tracker, an Eyelink 1000 Remote (sr-research.com), was positioned just beneath the monitor. This eye tracker has a 4-ms sampling interval, and gaze position accuracy of .5° visual angle. Calibration used adult routines in the Eyelink Toolbox (Cornelissen, Peters & Palmer, 2002); children were told that this was a 'follow-the-dot game'. One experimenter sat next to the child and controlled the mouse; a second experimenter sat in front of the eye tracking computer monitor to recalibrate if necessary.

Children wore child-sized KidzGear headphones (http://www.gearforkidz.com). They were told that a lady would ask for some things on the screen, and they were to point to the thing asked for. Four pictures appeared at a time, and 500 milliseconds (ms) later, an instruction

Table 1 *Words used in Experiment 1*

Type	Original word	Klattese	Summed phone probabilities[a]	Summed biphone probabilities[a]	Word Length (ms)
standard	apple	@pL	0.0531	0.0017	531
	bed	bEd	0.1621	0.0069	606
	bell	bEl	0.1978	0.0119	597
	brick	brIk	0.2198	0.0242	649
	candle	k@ndL	0.3351	0.0408	771
	candy	k@ndi	0.3489	0.0420	730
	elephant	Elxfxnt	0.3765	0.0344	697
	fish	fIS	0.1505	0.0060	636
	hat	h@t	0.1848	0.0111	539
	lizard	lIzXd	0.2341	0.0136	557
	pig	pIg	0.1985	0.0083	627
	teddy	tEdi	0.1985	0.0137	590
down	opple	apL	0.0351	0.0015	528
	bad	b@d	0.1686	0.0082	652
	bal	b@l	0.2043	0.0144	656
	breck	brEk	0.2103	0.0182	650
	condle	kandL	0.3161	0.0429	915
	condy	kandi	0.3300	0.0441	732
	alaphant	@lxfxnt	0.3891	0.0358	722
	fesh	fES	0.1272	0.0029	703
	hot	hat	0.1659	0.0063	553
	lezzard	lEzXd	0.2108	0.0087	537
	peg	pEg	0.1752	0.0061	727
	taddy	t@di	0.2051	0.0100	663
up	epple	EpL	0.0405	0.0010	702
	bid	bId	0.1854	0.0073	531
	bill	bIl	0.2211	0.0131	592
	breek	brik	0.2036	0.0167	661
	kendle	kEndL	0.3285	0.0303	718
	kendy	kEndi	0.3423	0.0315	720
	illaphant	Ilxfxnt	0.4213	0.0341	732
	feesh	fiS	0.0861	0.0021	722
	het	hEt	0.1782	0.0090	513
	leezard	lizXd	0.1697	0.0072	573
	peeg	pig	0.1341	0.0029	663
	tiddy	tIdi	0.2219	0.0095	607

[a] Phonotactic probabilities were obtained from the Phonotactic Probability Calculator (Vitevitch & Luce, 2004).

[1] Choosing a cutoff value is somewhat arbitrary, so a round number value of 75% was chosen across experiments (this criterion is typically higher for adults, who are better at sitting still). Poorly tracked data are usually discarded because they may introduce unnecessary noise. In the current study, though, including these participants did not change overall data patterns.

was spoken (e.g. 'Point to the fish'). Pictures stayed on-screen until the child pointed and the experimenter clicked on that selection. Four-picture groupings were the same across participants, and were carefully selected so that no pictures with phonologically similar or semantically similar names appeared together. Each child saw each picture set four times, but each picture was only a target once. Four unique trial orders were pre-randomized with the following constraints: the same picture set did not occur on successive trials; each picture was equally likely to be a target in the first half of the experiment (half of participants) or the second half of the experiment (other half of participants). Two trial orders were mirror-reverses of the other two. Pronunciation varied between participants, with 16 children each hearing up-shifted, down-shifted, or unshifted front vowels. All children heard control words pronounced canonically.

Results

Accuracy

Accuracy was extremely high (Figure 1a), well above chance (1 in 4 = .25) in unshifted (99.5% ± 2.1%; $t1(15) = 143$, $p < .0001$; $t2(11) = 143$, $p < .0001$; $d = 35.75$), up-shifted (96.9% ± 4.1%; $t1(15) = 69$, $p < .0001$; $t2(11) = 59.09$, $p < .0001$; $d = 17.25$), and down-shifted (95.3 ± 6.8; $t1(15) = 41.46$, $p < .0001$; $t2(11) = 36.93$, $p < .0001$; $d = 10.37$) conditions. ANOVAs were computed with accuracy as the dependent variable, with independent variables Pronunciation (canonical, shifted-up, shifted-down; between-participants), and Stimulus Type (shifted words, controls; within-participants). Effect sizes for ANOVAs are reported as generalized eta-squared (η^2_G), which equates variability better than η^2_P across within- and between-participants designs (Bakeman, 2005; Olejnik & Algina, 2003). Pronunciation ($F1(2, 45) = 3.28$, $p = .047$; $F2(1, 22) = 3.09$, $p = .09$; $\eta^2_G = .07$) and Stimulus Type ($F1(1, 45) = 4.97$, $p = .03$; $F2(2, 44) = 2.26$, $p = .12$; $\eta^2_G = .05$) were significant by participants, as was the Pronunciation × Stimulus Type interaction ($F1(2, 45) = 3.63$, $p = .03$; $F2(2, 44) = 2.46$, $p = .10$; $\eta^2_G = .07$). The down-shifted condition showed lower accuracy on experimental than control trials (D = 4.7% ± 6.8%; $t1(15) = 2.76$, $p = .01$; $t2(22) = 2.46$, $p = .02$; $d = 1.38$), but up-shifted (D = 0.5% ± 5.7%; $t1(15) = 0.37$, $p = .72$; $t2(22) = .26$, $p = .80$; $d = .13$) and CP conditions (D = 0.0% ± 3.0%; $t1(15) = 0$, $p = 1$; $t2(22) = 0$, $p = 1$; $d = .00$) did not. Accuracy in AP conditions did not correlate significantly with age ($r = -.15$, $t(29) = .8$, $p = .43$).

Visual fixations to target

To parallel studies of mispronunciations on younger children (e.g. Swingley & Aslin, 2000, 2002; Swingley, 2009) who did not make overt responses, fixations in the

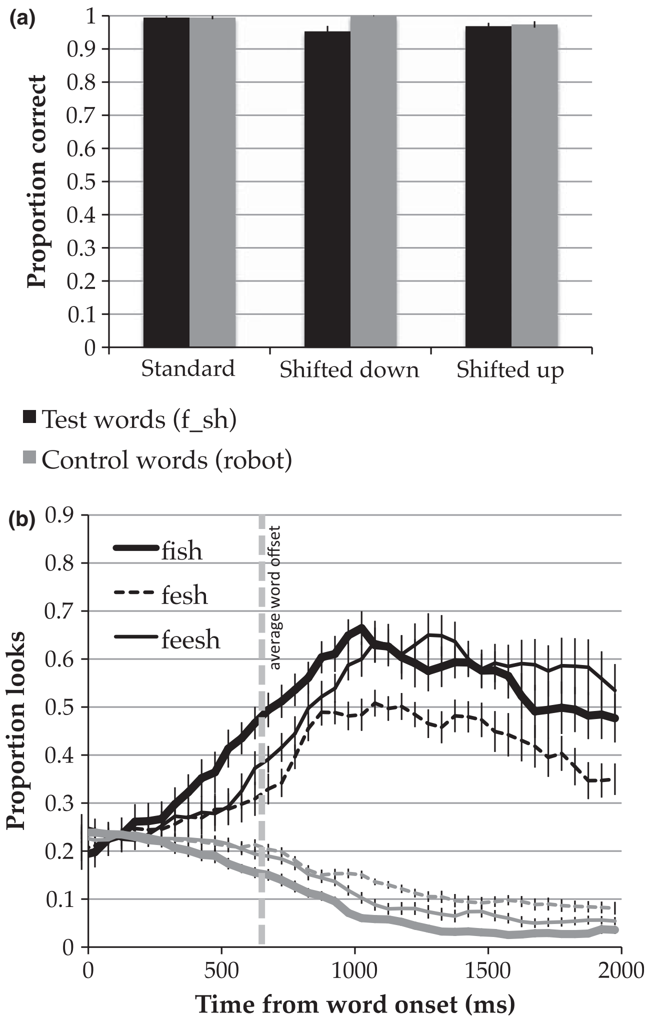


Figure 1 *Experiment 1, (a) proportion correct ± standard errors, and (b) looks to target pictures ± standard errors (experimental trials only); black: looks to targets; gray: averaged looks to other pictures.*

current study were assessed *without* removing trials with non-target responses.

Children tended to look at the correct picture more than the others as time from word onset increased (Figure 1b). However, target looks increased more slowly for APs than for CPs. Looking proportions during the first 1000 milliseconds (ms) were compared across conditions. After this time window, children began to point to pictures, which obscured the camera, causing an increase in eye track loss.[2] As is customary in visual world eye tracking paradigms, the window was shifted forward in time 200 ms, from 0–1000 to 200–1200 ms, accounting for time to program and launch an eye movement based on heard material (estimated at 200 ms; Hallett, 1986). Looks were corrected for non-normal distribution using the empirical logit transformation prior to analysis (Barr, 2008).

[2] A few trials ended prior to the end of the analysis window (< 0.1%) or the end of the graph display (2000 ms; < 10%). In these cases, the child was counted as having looked at their final look location until 2000 ms.

An ANOVA was conducted on transformed looking proportions in the experimental condition only, with Pronunciation (correct, shifted upward, shifted downward) as a between-participants variable and Experiment Half as a within-participants variable. The effect of Shift was significant (*F1*(2, 42) = 6.32, *p* = .004; *F2*(2, 22) = 4.89, *p* = .017; η^2_G = .23), indicating fewer looks to target when pronunciations were shifted. Planned comparisons examined the CP condition vs. upward and downward shifts separately. The upward shift (42.3% ± 8.4%) generated only a marginal decrease in target looks relative to the CP condition (48.3% ± 9.0%; *t1*(28) = 1.90, *p* = .07; *t2*(11) = 1.88, *p* = .09; *d* = .69), while the downward shift (37.2% ± 7.8%) generated significant target-look decreases (*t1*(28) = 3.45, *p* = .002; *t2*(11) = 3.29, *p* = .007; *d* = 1.26). Target looks did not correlate significantly with age (*r* = −.04, *t*(28) = .18, *p* = .86).

Additional analyses included Experiment Half (first, second) as a factor to assess whether exposure to APs in the first half facilitated AP recognition in the second half. For accuracy, there was no effect of, and no interactions with, Experiment Half (Experiment Half: *F1*(1, 45) = 0, *p* = 1.00; *F2*(1, 22) = 0, *p* = 1.00; η^2 = .00; Experiment Half × Pronunciation, *F1*(2, 45) = 0.14, *p* = .87; *F2*(2, 44) = 0.28, *p* = .76; η^2_G = .00; Experiment Half × Stimulus Type, *F1*(1, 45) = 0, *p* = 1.00; *F2*(1, 22) = 0, *p* = 1.00; η^2_G = .00; Experiment Half × Pronunciation × Stimulus Type, *F1*(2, 45) = 1.03, *p* = .37; *F2*(2, 44) = 1.93, *p* = .16; η^2_G = .01) – that is, there was no evidence of adaptation. There was some evidence of adaptation in target looks (interaction of Experiment Half × Pronunciation, *F1*(2, 42) = 3.89, *p* = .03; *F2*(2, 22) = 2.80, *p* = .08; η^2_G = .08), with an increase in looks from the first to the second half for the downward-shifted words (*D* = 9.0% ± 13.5%; *F1*(1, 14) = 5.65, *p* = .03; *F2*(1, 11) = 7.22, *p* = .02; η^2 = .29), but not upward-shifted (*D* = −1.1% ± 14.0%; *F1*(1, 14) = 0.48, *p* = .50; *F2*(1, 11) = 0.18, *p* = .68; η^2 = .03) or unshifted (*D* = −5.2% ± 16.7%; *F1*(1, 14) = 1.48, *p* = .24; *F2*(1, 11) = 2.16, *p* = .17; η^2 = .10) words.

Discussion

Children were highly accurate, suggesting easy detection of similarity between APs and familiar word forms. However, visual fixation patterns suggested that children were sensitive to APs, looking to the target more slowly. One interpretation of this result is that children recognized the APs eventually because the context was so constraining: only four alternatives, and the target picture's name (but not the other pictures) was a close phonological match. The children may have initially parsed the APs as novel, but since no novel picture was present, they could not have selected it. A different possibility is that children actually recognized the words as familiar. It is impossible to know from Experiment 1 which of these was the case. Accordingly, the next experiment added a novel alternative.

Experiment 2

This experiment was similar to Experiment 1, but each trial contained a novel-picture alternative. Half of children heard CPs, and half APs, on experimental trials, and all children heard CPs on control trials. In addition, half of children (divided between CP and AP conditions) heard experimental words in supportive sentence contexts (e.g., 'I want to feed the f_sh') and the other half heard neutral sentence contexts ('I want to find the f_sh'). This introduced another source of contextual constraint, to assess whether children would incorporate additional contextual information into their novelty decisions.

If children in Experiment 1 initially parsed APs as unfamiliar but were foiled by the absence of an unfamiliar referent, adding an unfamiliar referent should cause a large decrease in target selections. If, instead, children recognized APs as acceptable variants of familiar words, then they should show high accuracy despite the novel alternative. An additional prediction is that if supportive sentence context biases children toward a familiar-object interpretation, then looking proportions and accuracy should be greater on supportive-sentence trials.

Method

Participants

N = 64 monolingual English-speaking preschoolers (26 female; *M* = 4.4 ± .5 years; range: 3.2–5.7) from the same pool as before took part. An additional 16 children took part but were excluded due to: multiple control-condition errors (seven); reported exposure to a language besides English (seven); reported learning disability (one); having been run in Experiment 1 (one). Seven children were included in response analyses but not eye tracking analyses by the same criterion as in Experiment 1.

Stimuli

Pictures. Familiar pictures were drawn from the same name-normed sources as before (90% recognition for experimental words, 83% for control words). APs that were also real nouns were eliminated (*hat*→hot, *pig*→peg), as well as one close to a highly-familiar noun (*bell*→bal≈*ball*), and *brick* because it was discovered to generate low accuracy in overt naming. *Lizard* was replaced with *milk* to decrease the high proportion of animate pictures, which made counterbalancing easier (preventing multiple semantically similar animates from occurring on-screen at once). Novel pictures were selected from a larger set of photographs of artwork, scientific devices, and non-child-oriented household devices found on-line. Candidate novel pictures were presented to 18 children for naming. The eight pictures selected were never identified correctly, and no more than two children agreed in calling one the same (erroneous) name. This ensured that novel pictures were unfamiliar and unidentifiable.

Recordings. The author recorded new phrases as before. Given that the AP effect in Experiment 1 was more robust for the downward shift, only the downward shift was used here. Phrases contained none of the three vowels that shifted (/ɪ/, /ɛ/, /æ/) other than in target words. APs and CPs were cross-spliced into a carrier recorded preceding a different CP token. Thus the APs or CPs in both consistent-verb and neutral-verb sentences were acoustically identical, and carrier phrases were acoustically identical in both CP and AP conditions. Durations of AP and CP words did not differ (Table 2). Acoustic analyses of the vowels reflected intended shifts in F1, F2, and vowel duration (Appendix).

Procedure

In a pre-exposure phase, eight novel pictures were presented twice each, along with the word ‘Look!’ This was done because Mather and Plunkett (2009) showed that 2-year-olds did not preferentially fixate novel pictures on the first appearance, but *did* on the second appearance. Pre-exposure thus aimed to maximize the potential for children to identify the novel object as a referent. (Note, though, that Evey & Merriman, 1998, found that repeated exposure to the presence of novel objects seemed to *decrease* overt novelty responding in 2-year-olds, and Merriman & Schuster, 1991, found fewer novelty responses after novel preview among 2- and 4-year-olds.) Pre-exposure trial length was controlled by the experimenter, who clicked on each picture at roughly 1-second intervals to advance to the next trial.

Each set of pictures contained the familiar (target) picture, one novel picture, and two other non-target (familiar) pictures. Each child saw each picture set twice, once in each of two blocks, but no picture was target more than once. As in Experiment 1, four-picture sets were the same for all participants, and were carefully selected so that no phonologically similar or semantically similar names cooccurred. Four picture-to-location assignments, two block orders, two verb types, and two pronunciation types yielded 32 lists, each of which was run twice. For each participant, random trial ordering within a block was determined at runtime.

Table 2 *Words used in Experiment 2*

Type	Original Word	Klattese	Summed phone probabilities	Summed biphone probabilities	Word Length (ms)
standard	apple	@pL	0.0531	0.0017	649
	bed	bEd	0.1621	0.0069	704
	candle	k@ndL	0.3351	0.0408	647
	candy	k@ndi	0.3489	0.0420	723
	elephant	Elxfxnt	0.3765	0.0344	724
	fish	fIS	0.1505	0.0060	746
	milk	mIlk	0.2693	0.0188	587
	teddy	tEdi	0.1985	0.0137	631
down	opple	apL	0.0351	0.0015	667
	bad	b@d	0.1686	0.0082	715
	condle	kandL	0.3161	0.0429	599
	condy	Kandi	0.3300	0.0441	796
	alaphant	@lxfxnt	0.3891	0.0358	851
	fesh	fES	0.1272	0.0029	725
	melk	mElk	0.2459	0.0162	649
	taddy	t@di	0.2051	0.0100	677

Results

Responses in Experiments 2–3 are classified as *target* responses (choosing the fish when hearing ‘fesh’ or ‘fish’); *novel*-picture responses; and *other* responses (choosing the guitar when hearing ‘fesh’ or ‘fish’). After running the experiment, it was discovered that the verbs for ‘elephant’ were switched (the consistent verb was presented to children hearing neutral verbs, and vice versa). As a precaution, it was eliminated from analyses. Preliminary analyses suggested that neither accuracy nor looking proportions changed from the first to the second half of trials, so further analyses were collapsed across Experiment Half.

Responses

Overall, children in AP conditions showed fewer target responses than children in CP conditions (Figure 2a). Nonetheless, children exceeded chance (at a conservative level of chance = .5) in all conditions: consistent-verb CP (98.2% ± 4.9%; $t1(15) = 39.52$, $p < .0001$; $t2(6) = 41.82$, $p < .0001$; $d = 9.88$), neutral-verb CP (98.2% ± 4.9%; $t1(15) = 39.52$, $p < .0001$; $t2(6) = 27$, $p < .0001$; $d = 9.88$), consistent-verb AP (92.0% ± 12.7%; $t1(15) = 13.17$, $p < .0001$; $t2(6) = 11.08$, $p < .0001$; $d = 3.29$), and neutral-verb AP (81.3% ± 19.3%; $t1(15) = 6.47$, $p < .0001$; $t2(6) = 5.12$, $p = .002$; $d = 1.62$). An ANOVA was conducted on target response proportions with Pronunciation (CP, AP) and Verb Bias (consistent, neutral) as between-participants factors and Stimulus Type (experimental, control) within-participants. Pronunciation was significant ($F1(1, 60) = 14.85$, $p = .0003$; $F2(1, 13) = 13.65$, $p = .003$; $\eta^2_G = .11$), as was Stimulus Type ($F1(1, 60) = 17.46$, $p < .0001$; $F2(1, 13) = 11.04$, $p = .006$; $\eta^2_G = .13$), and their interaction ($F1(1, 60) = 12.56$, $p = .0008$; $F2(1, 13) = 13.50$, $p = .003$; $\eta^2_G = .10$). This is consistent with less target-word recognition for AP words (86.6% ± 17.0%) than CP words (98.2 ± 4.8%; $F1(1, 62) = 13.82$, $p = .0004$; $F2(1, 6) = 13.34$, $p = .01$; $\eta^2_G = .18$), which was not present in the control-word (canonically pronounced) condition (98.8% ± 3.7% vs. 99.2% ± 3.1%; $F1(1, 62) = 0.21$, $p = .65$; $F2(1, 6) = 0.30$, $p = .60$; $\eta^2_G = .00$). Even limiting analysis to the experimental words, neither Verb Bias ($F1(1, 60) = 3.15$, $p = .08$; $F2(1, 6) = 1.55$, $p = .26$; $\eta^2 = .05$) nor the Verb Bias × Pronunciation interaction ($F1(1, 60) = 3.15$, $p = .08$; $F2(1, 13) = 2.84$, $p = .14$; $\eta^2 = .05$) quite reached significance, suggesting that the visible difference (Figure 2a, right half) in target word responses in the

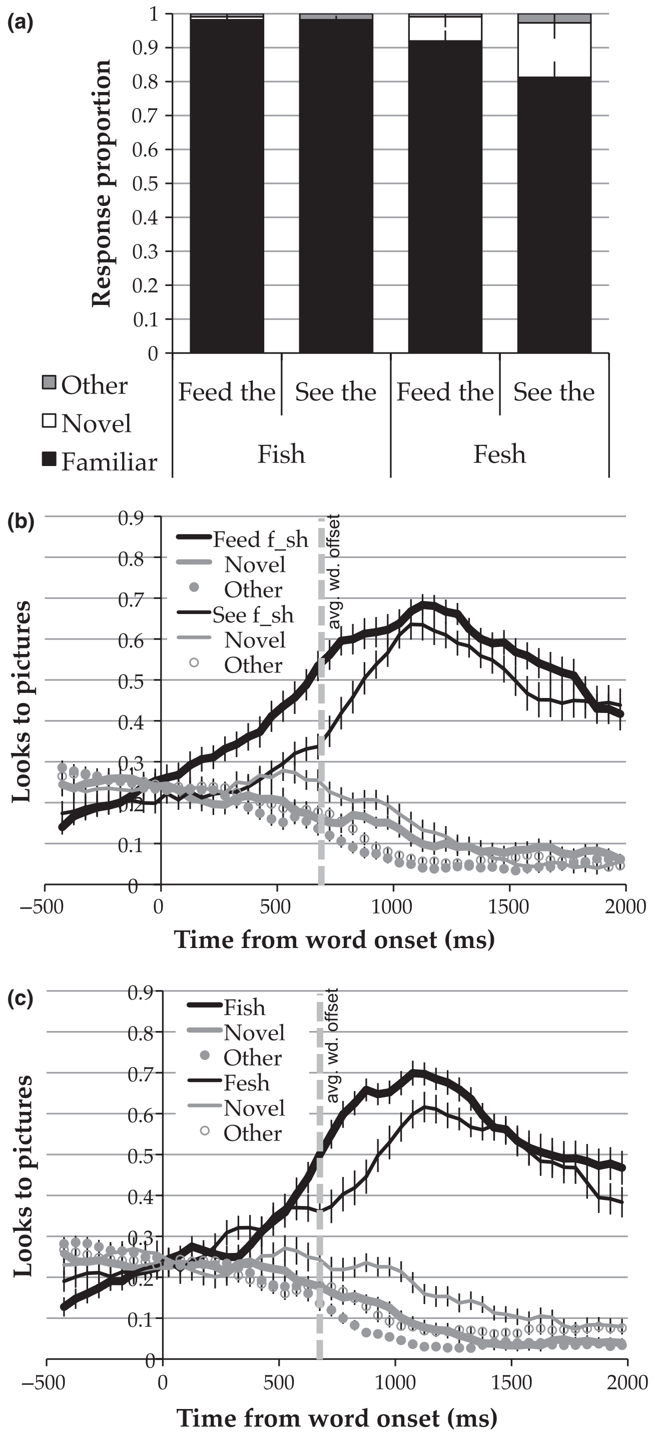


Figure 2 *Experiment 2, (a) response proportions ± standard errors, and visual fixations plus standard errors for (b) verb effect, (c) pronunciation effect.*

consistent and neutral AP conditions was not strong. Accuracy in AP conditions did not correlate significantly with age (r = .00).

Visual fixations

There were faster visual fixations to target pictures for target-consistent verbs (Figure 2b), and more decisive visual fixations to target pictures for CPs than APs (Figure 2c). Fixation analyses were confined to experimental trials for simplicity. An analysis on empirical-logit-transformed target-picture fixations (200–1200 ms after word onset) with Pronunciation and Verb Bias as between-participants and within-items factors showed an effect of Verb Bias ($F1(1, 53) = 13.88$, $p = .0005$; $F2(1, 6) = 13.99$, $p = .01$; $\eta^2_G = .21$), with more target looks for consistent verbs (51.6% ± 11.5%) than neutral verbs (40.4% ± 11.5%). There was also an effect of Pronunciation by participants ($F1(1, 53) = 6.13$, $p = .02$; $F2(1, 6) = 4.04$, $p = .09$; $\eta^2_G = .10$) such that there were more fixations to the target picture on CP trials (49.4% ± 12.0% vs. 42.1% ± 12.6% on AP trials). There was no interaction ($F1(1, 53) = 0.12$, $p = .72$; $F2(1, 6) = 0.14$, $p = .72$; $\eta^2_G = .00$). Target looks in AP conditions did not correlate significantly with age ($r = .02$, $t(25) = .25$, $p = .90$).

Were children more likely to fixate the novel picture when they heard an AP? An ANOVA was conducted on looks to novel vs. other (non-target) pictures, with Pronunciation, Verb Bias, and Picture (novel, other) as factors. There was an effect of Pronunciation ($F1(1, 53) = 15.29$, $p = .0003$; $F2(1, 6) = 6.84$, $p = .04$; $\eta^2_G = .10$), reflecting more looks to non-target pictures (both novels and 'others') for children hearing APs (17.5% ± 2.8%) than children hearing CPs (14.1% ± 3.9%). An effect of Verb Bias ($F1(1, 53) = 18.25$, $p < .0001$; $F2(1, 6) = 17.02$, $p = .006$; $\eta^2_G = .12$) reflected more non-target looks for neutral verbs (17.4% ± 4.0%) than for consistent verbs (13.9% ± 3.6%). An effect of Picture ($F1(1, 53) = 19.77$, $p < .0001$; $F2(1, 6) = 30.11$, $p = .002$; $\eta^2_G = .19$) reflected more looks overall to the novel picture (19.3% ± 7.3%) than the (averaged) other pictures (13.9% ± 5.3%). Picture did not interact with Pronunciation ($F1(1, 53) = 0.65$, $p = .42$; $F2(1, 6) = .84$, $p = .39$; $\eta^2_G = .01$), suggesting that the likelihood of fixating the novel picture did not increase when alternative pronunciations occurred. The three-way interaction of Pronunciation × Verb Bias × Picture was significant by participants ($F1(1, 53) = 5.65$, $p = .02$; $F2(1, 6) = 1.6$, $p = .18$; $\eta^2_G = .06$). This reflected larger novel vs. other differences for CPs when the verb was consistent (5.4% ± 8.5% vs. 2.5% ± 6.6%), but larger novel–other differences for APs when the verb was neutral (9.1% ± 11.4%) rather than consistent (5.2% ± 9.4%). However, given the smaller amount of variance accounted for relative to main effects, this result should be interpreted with caution. It may reflect large individual variability in tendency to fixate novel pictures.

Discussion

Answering the question left open by Experiment 1, non-target selections (mainly novel-object selections) were more frequent for APs (11.6%) than for CPs (0.4%), suggesting that children occasionally regarded APs as novel words. Children hearing APs also looked less to targets, and more to novel and other pictures, than children hearing CPs. Children hearing neutral verbs also looked less to targets, and more to novel and other pictures, than children hearing consistent verbs. This pattern of results suggests that as uncertainty increases, children are more likely to visually explore the array of pictures, including but not limited to the novel picture. Interestingly, looks to the novel object were elevated over looks to 'other' objects. That is, children visually explored the novel pictures more than the other non-target pictures even in CP conditions. One explanation for this pattern is that all children had some level of interest in the novel object – possibly due to novel-picture pre-exposure. If children did find the novel pictures especially interesting because of pre-exposure, it is particularly impressive that they so often chose the familiar pictures instead. Another explanation is that the novel pictures were simply more visually interesting than the 'other' pictures, though if this were the case one would expect that looks to novel pictures should be elevated prior to target word onset, yet they are not.

Perhaps the most striking result is the high accuracy – 87% of APs resulted in selection of the target picture, even though recognition was slowed. This means that, for apparent single-feature alterations in accented speech, children may not automatically assume that they are hearing a novel name. This is consistent with Merriman and Schuster's (1991) findings that words similar to the familiar object label led to lower mutual-exclusivity responding than dissimilar labels. By tracking eye movements to unrelated pictures, the current study additionally suggests that deviations from familiarity not only drive children toward a novel object, but may also increase uncertainty generally.

However, it could be that children are simply very permissive of vowel changes, consistent with Nazzi's data on younger children (though, interestingly, not the current age group; see Havy *et al.*, 2011), and with adult data (Creel *et al.*, 2006; Cutler *et al.*, 2000; Van Ooijen, 1996). A second, not exclusive, possibility is that children are accepting these vowel APs because they are very *subtle* APs – the shifted-to vowels are close to each other in formant-frequency space (phonetically, vowel height).

The final experiment examines both questions by parametrically varying the segment type, segment location, and acoustic–phonetic distance of the APs presented. Instead of hearing only vowel APs, children heard words with changed onset consonants, vowels, or coda consonants. Each AP was either *close* (varying by a single feature – consonant voicing or vowel height) or *distant* (varying by voicing and place for consonants, and by a large shift across F1–F2 space for vowels). If children are simply more permissive of vowel changes, there should be higher target-word responding to vowel APs than consonant APs. If children are affected by degree of acoustic–phonetic mismatch, there should be more novel-picture responses for distant APs than close APs. Finally, inclusion of both onset and coda consonant changes asks whether children are also sensitive to the *location* of consonant mismatch, as found in some studies (children: Cole, 1981; adults: Creel *et al.*, 2006; Creel & Dahan, 2010) but not others (Nazzi & Bertoncini, 2009; Swingley, 2009).

Experiment 3

This experiment presented multiple AP types within-participants. This allowed each child to serve as their own canonical-pronunciation control. Each participant heard six CPs, six vowel APs, six onset, and six coda. Onset, vowel, and coda APs were evenly split between close and distant, both for each participant and across the entire set of words.

Method

Participants

$N = 32$ children (14 female; $m = 4.2 \pm .5$ years; range: 3.1–5.4, two ages not reported) from the same pool as the first two experiments took part. Twelve additional children were replaced due to: not finishing (one); multiple control-trial errors (two); technical problems (two); refusal to do post-experiment picture naming (one); teacher indicated comprehension problems (one); exposure to language besides English (four); speaking an English dialect not found in the US (one). Three children were dropped from eye movement analyses by the same criterion as in previous experiments.

Stimuli

Pictures were selected from the same sources as before (mean identification accuracy: 90%). New words (Table 3) were recorded in the carrier 'Point to the X'. To avoid hesitation cues to mispronunciation in the introductory phrase (see Kidd, White & Aslin, 2011), original words, close APs, and distant APs were all spliced onto a carrier from a different token of the original word. For onset-change words, words beginning with vowels (elephant, apple) and the words *rocks* and *milk*, the splice occurred just before the word 'the' to ensure natural-sounding coarticulation. After splicing, each sentence was normalized to 70 dB SPL in Praat and exported to a .wav file. Duration did not differ significantly across conditions ($M = 736$ ms). Acoustic analyses were conducted on vowel properties (F1, F2, duration) for vowel-changed words, for voice onset time (VOT) for onset-changed words, and vowel duration (a strong correlate of coda voicing in English) of coda-changed words (Appendix). Results were consistent with the intended mispronunciations.

Table 3 *Words used in Experiment 3*

Type	Change Location	Original Word	Klattese	Summed phone probabilities	Summed biphone probabilities	Word Length (ms)
Standard	Onset	Broom	brum	0.1868	0.0135	735
		Button	b^tN	0.1665	0.0082	642
		Cows	kWz	0.1225	0.0023	956
		Fork	fcrk	0.1837	0.0072	649
		Grapes	greps	0.2227	0.0145	758
		Puzzle	p^zL	0.1665	0.0043	713
		Truck	tr^k	0.1895	0.0167	688
		Zebra	zibrx	0.1848	0.0124	753
	Vowel	Apple	@pL	0.0531	0.0017	682
		Bell	bEl	0.1978	0.0119	649
		Candle	k@ndL	0.3351	0.0408	815
		Candy	k@ndi	0.3489	0.0420	819
		elephant	Elxfxnt	0.3765	0.0344	729
		Fish	fIS	0.1505	0.0060	778
		Milk	mIlk	0.2693	0.0188	629
		Teddy	tEdi	0.1985	0.0137	635
	Coda	Clock	klak	0.2000	0.0106	713
		Couch	kWC	0.1103	0.0020	752
		Fries	frYz	0.1619	0.0114	889
		Frog	frag	0.1719	0.0133	790
		Keys	kiz	0.1446	0.0021	817
		Nose	noz	0.0933	0.0047	790
		Rocks	raks	0.2142	0.0094	753
		Shirt	SRt	0.1004	0.0026	563
Close	Onset	Proom	prum	0.2199	0.0300	701
		Putton	p^tN	0.1997	0.0072	642
		Gows	gWz	0.0558	0.0008	913
		Vork	vcrk	0.1595	0.0059	643
		Crapes	kreps	0.2894	0.0158	830
		Buzzle	b^zL	0.1333	0.0053	704
		Pruck	pr^k	0.2294	0.0282	661
		Sebra	sibrx	0.2846	0.0148	785
	Vowel	Opple	apL	0.0351	0.0015	749
		Bal	b@l	0.2043	0.0144	683
		Condle	kandL	0.3161	0.0429	822
		Condy	kandi	0.3300	0.0441	796
		alaphant	@lxfxnt	0.3891	0.0358	780
		Fesh	fES	0.1272	0.0029	814
		Melk	mElk	0.2459	0.0162	584
		Taddy	t@di	0.2051	0.0100	700
	Coda	Clog	klag	0.1714	0.0092	784
		Couge	kWJ	0.1131	0.0018	840
		Frice	frYs	0.1999	0.0120	824
		Frock	frak	0.2005	0.0148	651
		Keece	kis	0.2033	0.0024	722
		Noce	nos	0.1520	0.0058	690
		Roggs	ragz	0.1406	0.0018	802
		Shird	SRd	0.0723	0.0015	683
Far	Onset	Croom	krum	0.2282	0.0155	699
		cotton	k^tN	0.2080	0.0091	645
		Dows	dWz	0.0816	0.0014	871
		Zork	zcrk	0.1397	0.0057	647
		Trapes	treps	0.2412	0.0188	823
		Guzzle	g^zL	0.1080	0.0035	680
		Bruck	br^k	0.1962	0.0117	633
		Shebra	Sibrx	0.1919	0.0128	749
	Vowel	Eeple	ipL	0.0264	0.0001	672
		Bool	bul	0.1470	0.0027	690
		Keendle	kindL	0.2874	0.0178	771
		Keendy	kindi	0.3013	0.0190	829
		oolaphant	ulxfxnt	0.3592	0.0329	731
		Fosh	faS	0.1148	0.0021	847
		Malk	malk	0.2336	0.0139	616
		Toody	tudi	0.1478	0.0075	656
	Coda	Clod	klad	0.1980	0.0099	836
		Cowve	kWv	0.1260	0.0018	838
		Fryshe	frYS	0.1590	0.0111	806
		Frot	frat	0.2476	0.0144	640
		Keesh	kiS	0.1322	0.0008	667
		Nofe	nof	0.0928	0.0036	664
		Rodge	raJ	0.1214	0.0020	763
		Shirb	SRb	0.0603	0.0015	732

Procedure

As in Experiment 2, children were pre-exposed to novel pictures. They then completed 24 test trials. Each child saw a particular picture set three times, with a different familiar picture as the target each time. Pictures presented together were the same across participants, and were selected to prevent co-occurrence of phonologically similar or semantically similar pictures. Trials were presented in three blocks without breaks. Each block contained eight trials: two intact words, and two each of onset, vowel, and coda APs (one each of close and distant APs), and no picture set was repeated in a block. For a given child, the same picture appeared in the same location, but across children, that picture appeared equally often in each of the four screen positions. No child heard multiple versions of the same base word, necessitating eight different stimulus lists. Crossed with these lists were four picture-to-location assignments, and three orderings of stimulus blocks (approximately counterbalanced), yielding 32 unique lists. Test trials were randomly ordered with the constraint that the same picture set could not appear in the following two trials. After test trials, each child saw all target pictures in a fixed random order, and was asked to name them. Verbal responses were recorded by an experimenter.

Results

To provide the strongest test of novelty responding based on recognition failure, analysis was restricted to trials containing targets that the child named correctly at posttest, eliminating 10.3% of trials (results were qualitatively similar with these trials included). This means that failure to select the target picture is not the result of not recognizing the picture – a possible alternative explanation in Experiments 1-2. Preliminary analyses indicated no effects of Block (first, second, third), so this factor was dropped.

Responses

Figure 3 suggests that increasing phonological distance led to a decrease in target responses, and a corresponding increase in novel-picture responses. It also increased other-picture responses. Children were equally likely to accept vowel APs as consonant APs, and to accept consonant APs in onset or coda position. ANOVAs were computed on Target responses and Novel responses separately, with Feature (CP, close AP, distant AP) and Position (onset, vowel, coda) as factors. Three children were dropped from participant analyses because they did not have data in all cells.

Target response rates differed by Feature ($F1(2, 56) = 36.78$, $p < .0001$; $F2(2, 42) = 35.79$, $p < .0001$; $\eta^2_G = .29$), with higher target response rates in CPs (*broom*; 97.6% ± 6.2%) than close APs (*proom*; 90.6% ± 11.2%; $t1(28) = 3.04$, $p = .005$; $t2(23) = 2.65$, $p = .01$; $d = .84$), and higher rates in close APs than distant APs (*croom*; 64.9% ± 22.1%; $t1(28) = 5.29$, $p < .0001$; $t2(23) = 6.19$, $p < .0001$; $d = .32$). Nonetheless, target responding exceeded chance (calculated with the conservative criterion of $\mu = .5$) in all three feature conditions: CP ($t1(28) = 43.91$, $p < .0001$; $t2(23) = 47.45$, $p < .0001$;

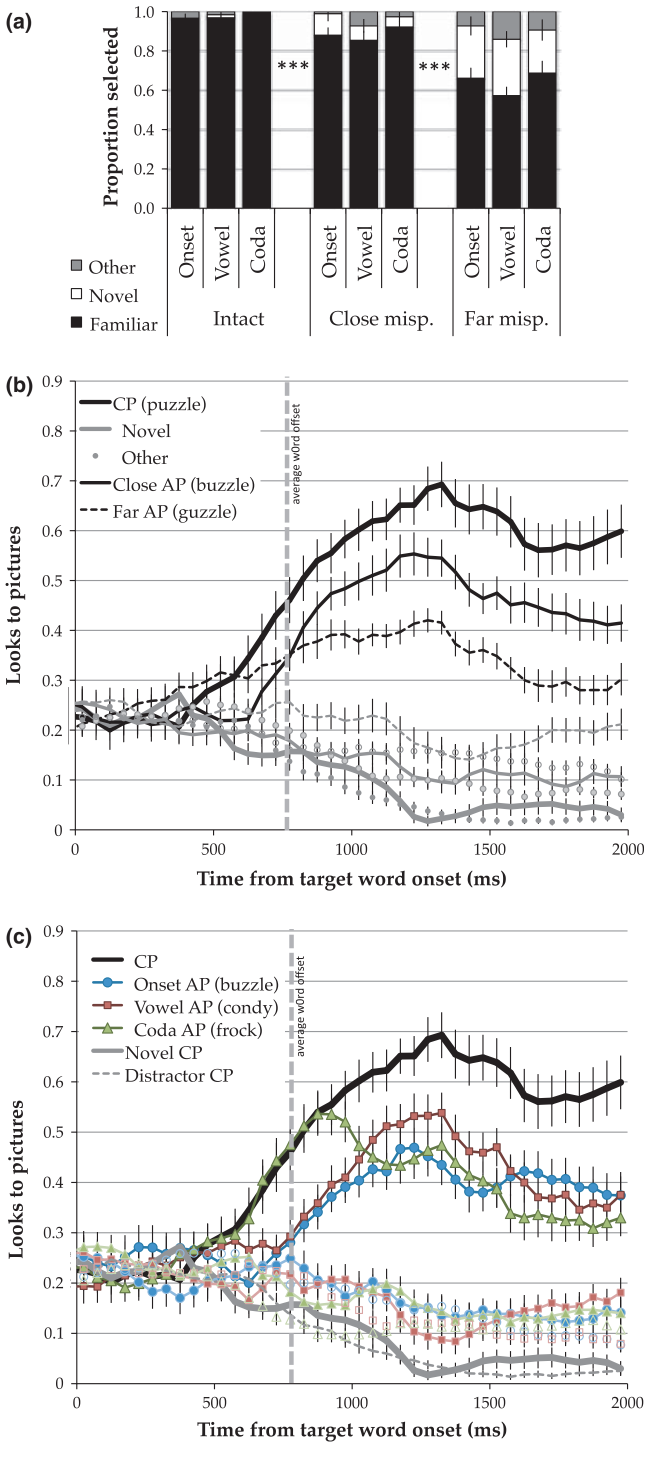


Figure 3 *Experiment 3 (n = 32), (a) picture selection ± standard errors for trials where children named the target later exactly; (b) visual fixations ± standard errors for CPs, and for APs of different feature distances; (c) visual fixations ± standard errors for CPs, and for APs of different syllable positions.*

d = 8.16), close AP ($t1(28)$ = 18.29, $p < .0001$; $t2(23)$ = 12.17, $p < .0001$; d = 3.40), and far AP ($t1(28)$ = 3.64, p = .001; $t2(23)$ = 2.71, p = .01; d = .68).

The reverse pattern was found for novel-picture responding ($F1(2, 56)$ = 29.21, $p < .0001$; $F2(2, 42)$ = 29.49, $p < .0001$; η^2_G = .25): CPs showed *fewer* novel responses (0.6% ± 3.1%) than close APs (6.8% ± 9.5%; $t1(28)$ = 3.50, p = .002; $t2(23)$ = 3.69, p = .001; d = 1.00), which in turn showed fewer novel responses than distant APs (25.3% ± 19.9%; $t1(28)$ = 4.55, $p < .0001$; $t2(23)$ = 5.11, $p < .0001$; d = 1.16). Interestingly, rate of selecting *other* pictures also rose as feature distance increased ($F1(2, 56)$ = 8.45, $p < .0001$; $F2(2, 42)$ = 7.51, p = .002; η^2_G = .06), carried by the difference between close and distant AP conditions (0.9% ± 2.8% vs. 4.9% ± 6.3%; $t1(28)$ = 3.39, p = .002; $t2(23)$ = 2.85, p = .009; d = 1.04; CP vs. close AP conditions [1.3% ± 3.1%] did not differ; $t1(28)$ = .79, p = .43; $t2(23)$ = .65, p = .52; d = 0.21). Of course, these selections could have reflected that children did not know the names of those phonologically unrelated pictures (meaning they were actually novelty responses), so the last t-test was repeated with trials where the child named all three familiar pictures exactly (62.9% of trials). The effect held (1.1% ± 4.0% vs. 4.2% ± 6.0%; $t1(25)$ = 3.21, p = .004; $t2(23)$ = 4.32, p = .0003; d = 1.10), suggesting that these were genuine confusion responses rather than novelty responses. Nonetheless, novel responses (12.8% ± 9.0%) outnumbered other responses overall (2.8% ± 3.1%; $t1(31)$ = 5.64, $p < .0001$; $t2(23)$ = 4.99, $p < .0001$; d = 1.67). There was a mild positive correlation between target selections on AP trials and age (r = .38, $t(28)$ = 3.17, p = .04). Though numerically, vowel changes decreased accuracy the most, Position did not reach significance for target ($F1(2, 56)$ = 2.82, p = .07; $F2(2, 21)$ = 0.47, p = .63; η^2_G = .01), novel ($F1(2, 56)$ = 1.30, p = .28; $F2(2, 21)$ = 0.34, p = .71; η^2_G = .01), or other ($F1(2, 56)$ = 2.43, p = .10; $F2(2, 21)$ = 0.55, p = .59; η^2_G = .02) responses.

Visual fixations

Overall, children looked more to the target for CPs than for APs. An initial evaluation suggested that target looks varied inversely with novel- and other-picture looks and, unlike Experiment 2, there were not above-baseline looks to novel pictures in the test time window, a point returned to in the discussion. Thus, for simplicity, looking time analyses were confined to target looks. An ANOVA on transformed target looks showed an effect of Feature ($F1(2, 50)$ = 4.66, p = .01; $F2(2, 42)$ = 5.48, p = .008; η^2_G = .04) and a Feature × Position interaction ($F1(4, 100)$ = 5.66, p = .0004; $F2(4, 42)$ = 5.38, p = .001; η^2_G = .07). This resulted from a strong effect of Feature for the onset position ($F1(2, 50)$ = 21.27, p = .0001; $F2(2, 14)$ = 19.96, $p < .0001$; η^2_G = .28) but not the other two positions (vowel: $F1(2, 50)$ = 0.67, p = .52; $F2(2, 14)$ = 0.43, p = .66; η^2_G = .01; coda: $F1(2, 50)$ = 0.12, p = .89; $F2(2, 14)$ = 0.30, p = .75; η^2_G = .00). This is expected, in that the onset APs will be the earliest APs evident. To explore explicit hypotheses about degree and location of change systematically, two different ANOVAs were conducted: one with Feature as a within-participants and within-items variable, and the other with Position as a within-participants and between-items variable. In order to remove item-related variance, the Position ANOVA for items was calculated as a one-way ANOVA on the *difference score* between the CP version of a word and the AP versions of that word (for instance, looks to *broom* when children heard 'broom' vs. looks to *broom* when children heard 'proom' or 'kroom').

There was an overall effect of Feature ($F1(2, 54)$ = 4.16, p = .02; $F2(2, 46)$ = 3.97, p = .026; η^2_G = .07). This stemmed mainly from significantly greater target looks in the CP trials (40.9% ± 13.5%) vs. far AP trials (32.1% ± 10.6%; $t1(27)$ = 2.89, p = .007; $t2(23)$ = 2.45, p = .02; d = .62), with a marginal difference between CP and close AP trials (34.2% ± 11.0%; $t1(27)$ = 2.24, p = .03; $t2(23)$ = 1.78, p = .09; d = .53) and no difference between close and far AP trials ($t1(27)$ = 0.29, p = .77; $t2(23)$ = 0.66, p = .52; d = .08). There was also an overall effect of Position ($F1(3, 81)$ = 5.39, p = .002; $F2(2, 21)$ = 10.38, p = .0007; η^2_G = .09). To measure the effect of altering pronunciation at different word locations, planned t-tests compared looks in each AP position to looks in CP trials. There were significantly fewer looks in onset AP trials (30.3% ± 10.9%) than on CP trials ($t1(27)$ = 3.50, p = .002; $t2(7)$ = 7.21, p = .0002; d = .76), fewer looks by participants only on vowel AP trials (32.1% ± 10.1%; $t1(27)$ = 2.74, p = .01; $t2(7)$ = 0.68, p = .52; d = .60), and no difference between coda AP trials (37.1% ± 12.6%) and CP trials ($t1(27)$ = 0.91, p = .37; $t2(7)$ = .5, p = .63; d = .22). Note that there is a drop in looks to coda AP trials relative to CP trials in the second 1000-ms time window (1200–2200 ms, 35.8% ± 14.9% vs. 58.0% ± 21.6%; post-hoc test significant by participants: $t1(27)$ = 4.68, $p < .0001$; $t2(7)$ = 1.73, p = .12; d = 1.15), consistent with a later divergence from the target word (see Swingley, 2009, for similar temporal effects in toddlers). The correlation of age with target looks on AP trials was not significant (r = .25, $t(24)$ = 1.28, p = .21).

Discussion

Children recognized words with both vowel and consonant changes. Novelty responses increased as phonological feature distance increased. Interestingly, other-object responses also increased slightly with phonological distance. This may reflect increased uncertainty in the presence of an alternative pronunciation. As in the first two experiments, visual fixation data reflected less rapid looks to the target picture when altered pronunciations were heard. The timing of the feature alteration (early vs. late in the word) shaped the time course of looking patterns, but did not significantly affect novelty responding.

This experiment replicates and extends Experiment 2, showing that larger pronunciation alterations increase novelty responding. It also verifies that high rates of novelty responding are likely *not* the result of target unfamiliarity – all novelty responses reported are in cases where children named the picture exactly in the post-test. Like Experiment 2, the target picture was the most likely selection even in conditions with the highest novelty responding. The one notable difference from Experiment 2 is the lack of a visual preference for the novel picture among the non-target pictures. This may have resulted from a higher tendency to explore the display, perhaps due to the presence of more numerous and more striking mispronunciations than in Experiment 2. On the other hand, picture *selections* indicated a pattern similar to that in Experiment 2: as target responses decreased, both novel-picture *and* other-picture responses increased. Taken together, the latter two experiments suggest that when children encounter an unfamiliar pronunciation of a familiar word, they are prone not only to mutual-exclusivity responses, but also to general uncertainty.

General discussion

Three experiments explored preschool-aged children's processing of atypical pronunciations (APs) versus canonical pronunciations (CPs) of familiar words. Children were highly accurate at selecting the intended familiar object from four familiar alternatives (Experiment 1) when they heard a vowel-shifted AP, though visual fixations suggested slower recognition for APs than CPs. Experiment 2 replaced one of the familiar pictures with a novel picture, allowing mutual-exclusivity responses to APs. Still, children mostly selected target referents, again with slower target looks for APs than CPs. In addition, heightened looks to novel objects were observed across the board. In Experiment 3, phonological feature distance predicted novelty responding, but segment type (consonant, vowel) and segment position did not. Even for multiple-feature-change APs, though, children still selected the target most often. Other-picture selection also increased as feature distance increased (though never outweighing novel-picture selections), suggesting that APs generated not just novelty responses but general uncertainty.

Phonological knowledge

Like younger children, preschoolers are sensitive to atypical pronunciations. However, preschoolers usually conclude that they are hearing a familiar word. Factors including the set of alternative referents and phonological distance, as well as the individual child, may all influence familiarity vs. novelty decisions. Relating this to looking-while-listening studies with infants, the current results imply that elevated looks to a non-target picture do not necessarily indicate that listeners (children) interpret a word as novel. That is, looks to something other than a dog when hearing 'tog', for example, may indicate heightened uncertainty more than they indicate mutual-exclusivity responding.

For the words tested, the type of segment altered (consonant or vowel) did not strongly influence novelty responding. Children are as affected by vowel changes as consonant changes. This is consistent with Havy *et al.*'s (2011) findings with older preschoolers in a different task, but it differs from Nazzi's (2005; Nazzi *et al.*, 2009) studies on younger children, and numerous adult studies (Creel *et al.*, 2006; Cutler *et al.*, 2000; Van Ooijen, 1996), which all suggest greater sensitivity to consonant changes than vowel changes. This may indicate a U-shaped function of consonant bias over development. On the other hand, it may simply indicate that tasks used at different ages are insufficiently similar, and thus different constructs are being probed at different ages. Further, equating differences in onsets, vowels, and codas is challenging, as the acoustic-phonetic realizations of each differ in many respects, and acoustics (or their perceptual consequences), rather than segment type itself (Macmillan, Goldberg & Braida, 1998), may drive differences. A broader age range should be tested on the same task before drawing strong conclusions about the relative importance of consonants and vowels to word identity across age.

Accent processing

What do these results imply for children's processing of accented speech? Primarily, they suggest that gradient similarity to the familiar form governs whether children accept it as a familiar word. This is consistent with looking-while-listening data (Swingley & Aslin, 2000, 2002; White & Morgan, 2008) and ERP data (Duta *et al.*, in press) in younger children: a one- or two-feature difference may be regarded as atypical, but not necessarily a new word. While this might seem like a negative in terms of maximizing discriminability of minimal pairs, it is a positive in cases where one routinely hears speakers using alternative pronunciations (e.g. bilingual or sociolectal situations; see Sebastián-Gallés *et al.*, 2005, 2009). Of course, the alternative pronunciations here consisted of another native-language (native-accent) speech sound. A true accented sound might be perceived as the native sound, as an accented but still 'correct' sound, a different native sound, or as completely ambiguous. Since all of these patterns occur in non-native phonological input (e.g. Best, McRoberts & Sithole, 1988; Flege *et al.*, 1999), effects should be replicated in a real-world accent. Further, because children may not assimilate accented input to their native phonology effectively (Flege *et al.*, 1999; Nathan *et al.*, 1998), they may be less likely to interpret ambiguous or atypical-but-same-category sounds as native ones, possibly resulting in higher novelty responding than seen here.

Second, these data suggest that there *is* some risk of novel-word interpretations when hearing an unfamiliar pronunciation of a familiar word, implying that accent-variants might need to be partially re-learned. This is consistent with Sebastián-Gallés *et al*.'s account (2005, 2009) that accent-variants are stored as separate lexical items. Further, accent-variants may generate uncertainty as well as novelty-responding. General uncertainty might increase with the rate of similar-sounding words heard in one's environment. For listeners in multi-accent or multilingual environments, long-term exposure to similar-sounding word variants might lead to increased uncertainty rather than mutual-exclusivity responding (as in Byers-Heinlein & Werker's, 2009, bilingual and trilingual toddlers).

Third, despite high familiarity responding for single-feature changes, altered pronunciations clearly slowed children's processing of words. While the current study limited altered pronunciations just to target words, leaving carrier phrases canonically pronounced, realistically, multiple words in a sentence will be accented. Thus, each word might generate an additional lag in comprehension. If so, child listeners might have difficulty keeping up with an accented speaker at normal speaking rates.

The current study leaves open two issues in child accent processing: the role of adaptation, and the role of sentence context. This study did not find strong evidence of adaptation after a small number of exposures, where White and Aslin (2011) did. However, there are several differences between studies. First, White and Aslin used younger children (19-month-olds), who may be more swayed by short-term input than preschoolers are. Second, White and Aslin used 24 exposures to a single vowel change before testing adaptation, while the current study only gave children 4–6 exposures (1–2 per vowel) before testing adaptation (Experiments 1–2). Additional exposure might be necessary for adaptation to a multiple-vowel change.

Regarding contextual cues, manipulation of sentence context (Experiment 2) only marginally affected AP accuracy. However, this may represent a ceiling effect – children often chose familiar targets even in the neutral-verb/AP condition (81%), and six of 16 children in that condition made *no* novel-picture responses. With a higher base rate of errors, sentence-context effects might be more facilitative. Alternatively, children may have difficulty integrating contextual information into word recognition, as found by Trueswell, Sekerina, Hill, and Logrip (1999) for children's use of visual-scene cues in resolving syntactic ambiguity. Yet adults use context to interpret accented speech (e.g. Bradlow & Bent, 2008), so children must eventually develop this ability.

Mutual exclusivity

What do these data say about mutual exclusivity? As noted above, it is a real risk in accented-speech perception, dependent on the degree of phonological mismatch. This study also assesses competing explanations for novel-object selections: target rejection vs. true novel-word responses. Children who heard a mispronunciation and eliminated the target item could have chosen the novel object exclusively, but, particularly in Experiment 3, they sometimes did not. This is somewhat consistent with process-of-elimination accounts of novelty responding (Halberda, 2006) – children reject the target before selecting the novel item. The additional wrinkle here is that novelty responding is probabilistic – occasionally, children erroneously chose a familiar picture (even though they knew its name) rather than the novel one. This implies that uncertainty does not uniformly lead to novelty responding. Of course, children here may have been stymied by the phonological similarity of the target's label to the mispronunciation (atypical of most mutual-exclusivity experiments), leading to response patterns reflecting greater uncertainty than seen in most mutual-exclusivity experiments.

Conclusion

This study explored various influences on preschool children's comprehension of altered pronunciations of familiar words. Seen from one perspective, children are good at recognizing familiar words despite accent-like alterations, implying flexible word recognition. However, from another perspective, results suggest that children perceive accented speech with more difficulty – they are slower to recognize words, and are increasingly prone to novel-word interpretations, as distance from the familiar form increases. This has implications for children growing up in multi-accent and multilingual environments. Future work will extend to more naturalistic accents, which provide greater variability in phonological and syntactic features.

Acknowledgements

Thanks to Heather Pelton for running participants, Robin Blumberg for work on testing materials, and to the children, parents, and schools who participated. SCC was supported by a Hellman Family Foundation award.

References

Bakeman, R. (2005). Recommended effect size statistics for repeated measures designs. *Behavior Research Methods*, **37** (3), 379–384.

Baker, W., Trofimovich, P., Flege, J.E., Mack, M., & Halter, R. (2008). Child–adult differences in second-language phonological learning: the role of cross-language similarity. *Language and Speech*, **51** (4), 317–342. doi:10.1177/0023830908099068

Barr, D.J. (2008). Analyzing 'visual world' eyetracking data using multilevel logistic regression. *Journal of Memory and Language*, **59** (4), 457–474. doi:10.1016/j.jml.2007.09.002

Best, C.T., McRoberts, G.W., & Sithole, N.M. (1988). Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants. *Journal of Experimental Psychology: Human Perception and Performance*, **14** (3), 345–360.

Best, C.T., Tyler, M.D., Gooding, T.N., Orlando, C.B., & Quann, C.A. (2009). Development of phonological constancy: toddlers' perception of native- and Jamaican-accented words. *Psychological Science*, **20** (5), 539–542. doi:10.1111/j.1467-9280.2009.02327.x

Bosch, L., & Sebastián-Gallés, N. (2003). Simultaneous bilingualism and the perception of a language-specific vowel contrast in the first year of life. *Language and Speech*, **46** (2–3), 217–243. doi:10.1177/00238309030460020801

Bradlow, A.R., & Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition*, **106**, 707–729. doi:10.1016/j.cognition.2007.04.005

Brainard, D.H. (1997). The Psychophysics Toolbox. *Spatial Vision*, **10**, 433–436.

Byers-Heinlein, K., & Werker, J.F. (2009). Monolingual, bilingual, trilingual: infants' language experience influences the development of a word-learning heuristic. *Developmental Science*, **12** (5), 815–823. doi:10.1111/j.1467-7687.2009.00902.x

Clarke, C.M., & Garrett, M.F. (2004). Rapid adaptation to foreign-accented English. *Journal of the Acoustical Society of America*, **116** (6), 3647–3658. doi:10.1121/1.1815131

Cole, R.A. (1981). Perception of fluent speech by children and adults. *Annals of the New York Academy of Sciences*, **379**, 92–109. doi:10.1111/j.1749-6632.1981.tb42001.x

Cornelissen, F.W., Peters, E.M., & Palmer, J. (2002). The Eyelink Toolbox: eye tracking with MATLAB and the Psychophysics Toolbox. *Behavior Research Methods, Instruments, & Computers*, **34** (4), 613–617. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12564564

Creel, S.C., Aslin, R.N., & Tanenhaus, M.K. (2006). Acquiring an artificial lexicon: segment type and order information in early lexical entries. *Journal of Memory and Language*, **54**, 1–19.

Creel, S.C., & Dahan, D. (2010). The effect of the temporal structure of spoken words on paired-associate learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **36** (1), 110–122. doi:10.1037/a0017527

Cutler, A., Sebastián-Gallés, N., Soler-Vilageliu, O., & Van Ooijen, B. (2000). Constraints of vowels and consonants on lexical selection: cross-linguistic comparisons. *Memory & Cognition*, **28** (5), 746–755.

Duta, M.D., Styles, S.J., & Plunkett, K. (in press). ERP correlates of unexpected word forms in a picture–word study of infants and adults. *Developmental Cognitive Neuroscience*. doi:10.1016/j.dcn.2012.01.003

Eckert, P. (2008). Where do ethnolects stop? *International Journal of Bilingualism*, **12** (1–2), 25–42. doi:10.1177/13670069080120010301

Evey, J.A., & Merriman, W.E. (1998). The prevalence and the weakness of an early name mapping preference. *Journal of Child Language*, **25** (1), 121–147. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9604571

Flege, J.E. (1991). Age of learning affects the authenticity of voice-onset time (VOT) in stop consonants produced in a second language. *Journal of the Acoustical Society of America*, **89** (1), 395–411.

Flege, J.E., Yeni-Komshian, G.H., & Liu, S. (1999). Age constraints on second-language acquisition. *Journal of Memory and Language*, **41** (1), 78–104. doi:10.1006/jmla.1999.2638

Fought, C. (1999). A majority sound change in a minority community: /u/-fronting in Chicano English. *Journal of Sociolinguistics*, **3** (1), 5–23. doi:10.1111/1467-9481.t01-1-00060

Gerken, L.A., Murphy, W.D., & Aslin, R.N. (1995). Three- and four-year-olds' perceptual confusions for spoken words. *Perception & Psychophysics*, **57** (4), 475–486.

Golinkoff, R.M., Mervis, C.B., & Hirsh-Pasek, K. (1994). Early object labels: the case for a developmental lexical principles framework. *Journal of Child Language*, **21**, 125–155.

Green, L. (2002). A descriptive study of African American English: research in linguistics and education. *International Journal of Qualitative Studies in Education*, **15** (6), 673–690. doi:10.1080/0951839022000014376

Halberda, J. (2003). The development of a word-learning strategy. *Cognition*, **87**, B23–B34. doi:10.1016/S0

Halberda, J. (2006). Is this a dax which I see before me? Use of the logical argument disjunctive syllogism supports word-learning in children and adults. *Cognitive Psychology*, **53** (4), 310–344. doi:10.1016/j.cogpsych.2006.04.003

Hallett, P.E. (1986). Eye movements. In K.R. Boff, L. Kaufman & J.P. Thomas (Eds.), *Handbook of perception and human performance* (pp. 10.1–10.112). New York: Wiley.

Havy, M., & Nazzi, T. (2009). Better processing of consonantal over vocalic information in word learning at 16 months of age. *Infancy*, **14** (4), 439–456. doi:10.1080/15250000902996532

Havy, M., Bertoncini, J., & Nazzi, T. (2011). Word learning and phonetic processing in preschool-age children. *Journal of Experimental Child Psychology*, **108** (1), 25–43. doi:10.1016/j.jecp.2010.08.002

Hirschfeld, L.A., & Gelman, S.A. (1997). What young children think about the relationship between language variation and social difference. *Cognitive Development*, **12**, 213–238.

Houde, J.F., & Jordan, M.I. (1998). Sensorimotor adaptation in speech production. *Science*, **279**, 1213–1216.

Houston, D.M., & Jusczyk, P.W. (2000). The role of talker-specific information in word segmentation by infants. *Journal of Experimental Psychology: Human Perception and Performance*, **26** (5), 1570–1582. doi:10.10371/0096-1523.26.5

Jarvis, L.H., Merriman, W.E., Barnett, M., Hanba, J., & Van Haitsma, K.S. (2004). Input that contradicts young phonological and semantic interpretation of other novel words. *Journal of Speech, Language, and Hearing Research*, **47**, 392–406.

Jusczyk, P.W., Goodman, M.B., & Baumann, A. (1999). Nine-month-olds' attention to sound similarities in syllables. *Journal of Memory and Language*, **40**, 62–82.

Kidd, C., White, K.S., & Aslin, R.N. (2011). Toddlers use speech disfluencies to predict speakers' referential intentions. *Developmental Science*, **14** (4), 925–934. doi:10.1111/j.1467-7687.2011.01049.x

Kinzler, K.D., Dupoux, E., & Spelke, E.S. (2007). The native language of social cognition. *Proceedings of the National Academy of Sciences of the United States of America*, **104** (30), 12577–12580. doi:10.1073/pnas.0705345104

Kohler, C.T., Bahr, R.H., Silliman, E.R., Bryant, J.B., Apel, K., & Wilkinson, L.C. (2007). African American English dialect and performance on nonword spelling and phonemic awareness tasks. *American Journal of Speech-Language Pathology*, **16**, 157–168.

Labov, W., Ash, S., & Boberg, C. (2006). *Atlas of North American English*. New York: Mouton de Gruyter.

Macmillan, N.A., Goldberg, R.F., & Braida, L.D. (1988). Resolution for speech sounds: basic sensitivity and context memory on vowel and consonant continua. *Journal of the Acoustical Society of America*, **84** (4), 1262–1280. doi:10.1121/1.396626

Mani, N., & Plunkett, K. (2007). Phonological specificity of vowels and consonants in early lexical representations. *Journal of Memory and Language*, **57** (2), 252–272. doi:10.1016/j.jml.2007.03.005

Mani, N., & Plunkett, K. (2008). Fourteen-month-olds pay attention to vowels in novel words. *Developmental Science*, **11** (1), 53–59. doi:10.1111/j.1467-7687.2007.00645.x

Markman, E.M., & Wachtel, G.F. (1988). Children's use of mutual exclusivity to constrain the meanings of words. *Cognitive Psychology*, **20**, 121–157.

Mather, E., & Plunkett, K. (2009). Learning words over time: the role of stimulus repetition in mutual exclusivity. *Infancy*, **14** (1), 60–76. doi:10.1080/15250000802569702

Maye, J., Aslin, R.N., & Tanenhaus, M.K. (2008). The weckud wetch of the wast: lexical adaptation to a novel accent. *Cognitive Science*, **32**, 543–562.

Merriman, W.E., & Marazita, J.M. (1995). The effect of hearing similar-sounding words on young 2-year-olds' disambiguation of novel reference. *Developmental Psychology*, **31** (6), 973–984. doi:10.1037//0012-1649.31.6.973

Merriman, W.E., & Schuster, J.M. (1991). Young children's disambiguation of object name reference. *Child Development*, **62** (6), 1288–1301. doi:10.2307/1130807

Nathan, L., Wells, B., & Donlan, C. (1998). Children's comprehension of unfamiliar regional accents: a preliminary investigation. *Journal of Child Language*, **25** (2), 343–365. doi:10.1017/S0305000998003444

Nazzi, T. (2005). Use of phonetic specificity during the acquisition of new words: differences between consonants and vowels. *Cognition*, **98** (1), 13–30. doi:10.1016/j.cognition.2004.10.005

Nazzi, T., & Bertoncini, J. (2009). Phonetic specificity in early lexical acquisition: new evidence from consonants in coda positions. *Language and Speech*, **52** (4), 463–480. doi:10.1177/0023830909336584

Nazzi, T., Floccia, C., Moquet, B., & Butler, J. (2009). Bias for consonantal information over vocalic information in 30-month-olds: cross-linguistic evidence from French and English. *Journal of Experimental Child Psychology*, **102** (4), 522–537. doi:10.1016/j.jecp.2008.05.003

Oh, G.E., Guion-Anderson, S., Aoyama, K., Flege, J.E., Akahane-Yamada, R., & Yamada, T. (2011). A one-year longitudinal study of English and Japanese vowel production by Japanese adults and children in an English-speaking setting. *Journal of Phonetics*, **39** (2), 156–157. doi:10.1016/j.wocn.2011.01.002

Olejnik, S., & Algina, J. (2003). Generalized eta and omega squared statistics: measures of effect size for some common research designs. *Psychological Methods*, **8** (4), 434–447. doi:10.1037/1082-989X.8.4.434

Ramon-Casas, M., Swingley, D., Sebastián-Gallés, N., & Bosch, L. (2009). Vowel categorization during word recognition in bilingual toddlers. *Cognitive Psychology*, **59** (1), 96–121. doi:10.1016/j.cogpsych.2009.02.002

Schmale, R., Cristià, A., Seidl, A., & Johnson, E.K. (2010). Developmental changes in infants' ability to cope with dialect variation in word recognition. *Infancy*, **15** (6), 650–662. doi:10.1111/j.1532-7078.2010.00032.x

Schmale, R., & Seidl, A. (2009). Accommodating variability in voice and foreign accent: flexibility of early word representations. *Developmental Science*, **12** (4), 583–601. doi:10.1111/j.1467-7687.2009.00809.x

Sebastián-Gallés, N., & Bosch, L. (2009). Developmental shift in the discrimination of vowel contrast in bilingual infants: is the distributional account all there is to it? *Developmental Science*, **12** (6), 874–887.

Sebastián-Gallés, N., Echeverria, S., & Bosch, L. (2005). The influence of initial exposure on lexical representation: comparing early and simultaneous bilinguals. *Journal of Memory and Language*, **52** (2), 240–255. doi:10.1016/j.jml.2004.11.001

Sebastián-Gallés, N., Vera-Constán, F., Larsson, J.P., Costa, A., & Deco, G. (2009). Lexical plasticity in early bilinguals does not alter phoneme categories: II. Experimental evidence. *Journal of Cognitive Neuroscience*, **21** (12), 2343–2357. doi:10.1162/jocn.2008.21152

Singh, L. (2008). Influences of high and low variability on infant word recognition. *Cognition*, **106**, 833–870.

Storkel, H.L. (2002). Restructuring of similarity neighbourhoods in the developing mental lexicon. *Journal of Child Language*, **29** (2), 251–274.

Swingley, D. (2009). Onsets and codas in 1.5-year-olds' word recognition. *Journal of Memory and Language*, **60** (2), 252–269. doi:10.1016/j.jml.2008.11.003

Swingley, D., & Aslin, R.N. (2000). Spoken word recognition and lexical representation in very young children. *Cognition*, **76**, 147–166.

Swingley, D., & Aslin, R.N. (2002). Lexical neighborhoods and the word-form representations of 14-month-olds. *Psychological Science*, **13**, 480–484.

Trueswell, J.C., Sekerina, I., Hill, N.M., & Logrip, M.L. (1999). The kindergarten-path effect: studying on-line sentence processing in young children. *Cognition*, **73** (2), 89–134.

Van Ooijen, B. (1996). Vowel mutability and lexical selection in English: evidence from a word reconstruction task. *Memory & Cognition*, **24** (5), 573–583.

Vitevitch, M.S., & Luce, P.A. (2004). A web-based interface to calculate phonotactic probability for words and nonwords in English. *Behavior Research Methods, Instruments, & Computers*, **36** (3), 481–487.

Werker, J.F., & Tees, R. (1984). Cross-language speech perception: evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, **7** (1), 49–63. doi:10.1016/S0163-6383(84)80022-3

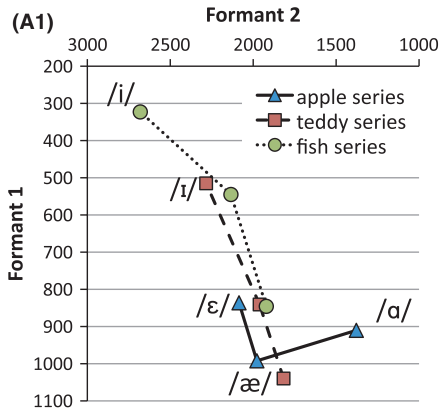
White, K.S., & Aslin, R.N. (2011). Adaptation to novel accents by toddlers. *Developmental Science*, **14** (2), 372–384. doi:10.1111/j.1467-7687.2010.00986.x

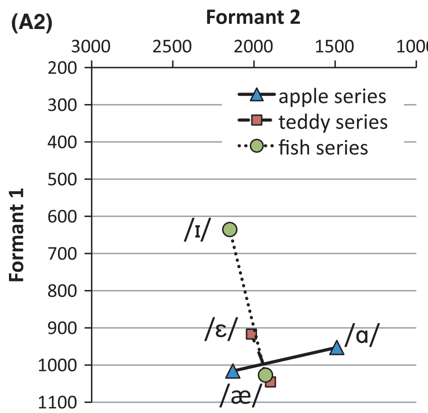
White, K.S., & Morgan, J.L. (2008). Sub-segmental detail in early lexical representations. *Journal of Memory and Language*, **59**(1), 114–132. doi:10.1016/j.jml.2008.03.001

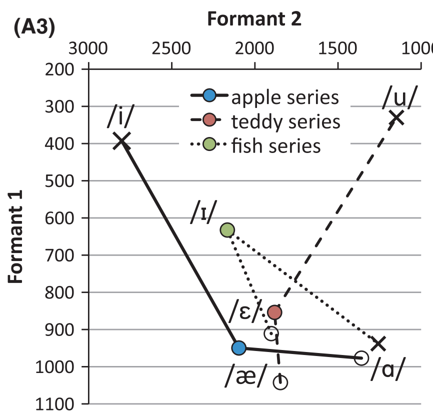
Received: 21 December 2011
Accepted: 11 May 2012

Appendix

Acoustic analyses of vowels and consonants







Note: Formants were measured in steady-state portions in the middle of each vowel.
Figure A1 *Vowels, Experiment 1.*
Figure A2 *Vowels, Experiment 2.*
Figure A3 *Vowels, Experiment 3. O = close mispronunciation, X = far mispronunciation.*

Table A1 *Vowel durations, Experiment 1. Note that /ɛ/ and /æ/, which are similar in F1 and F2, differ greatly in duration (/æ/ is longer; bolded)*

	Duration (ms)		
Series	original	raised	lowered
fish	156	203	193
bed	**201**	172	**248**
apple	**177**	**108**	175

Table A2 *Vowel durations, Experiment 2. Again, /ɛ/ and /æ/ differ in duration (bolded)*

	Duration (ms)	
Series	Original	Lowered
apple	176	174
bed	**199**	**262**
fish	165	154

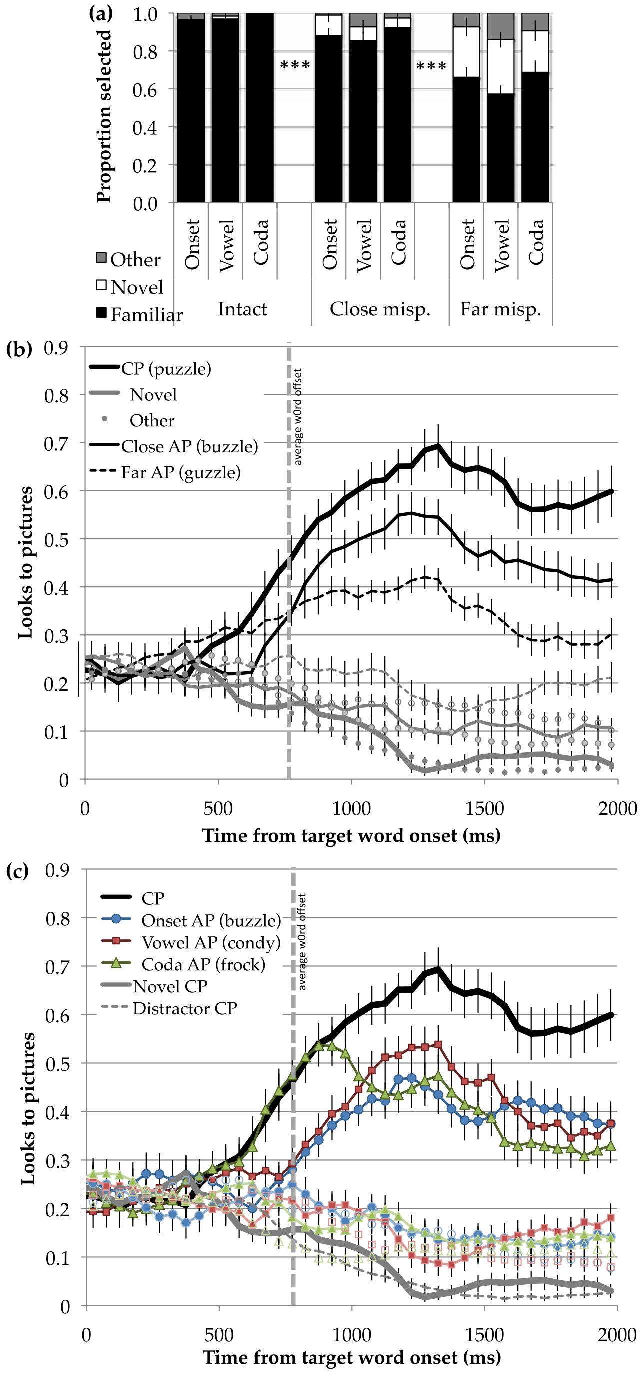
Table A3 *Durations, Experiment 3. Again, /ɛ/ and /æ/ differ in duration (bolded)*

	Original	Close	Far
Series	Vowels: Vowel duration (ms)		
fish	175	216	248
teddy	**149**	**195**	166
apple	179	179	141
	Onset voicing: VOT (ms)*		
cows (−voice → +voice)	109	17	15
button (+voice → −voice)	19	91	111
	Coda voicing: Vowel duration (ms)		
clock (−voice → +voice)	258	411	447
frog (+voice → −voice)	472	287	272

* Excludes zebra, fork, and their mispronunciations (inspection confirmed that the presence and absence of voicing were as expected).

Graphical Abstract

The contents of this page will be used as part of the graphical abstract of html only. It will not be published as part of main article.



Recent research has considered the phonological specificity of children's word representations, but few studies have examined the *flexibility* of those representations. Tolerating acoustic–phonetic deviations has been viewed as a negative in terms of discriminating minimally different word forms, but may be a positive in an increasingly multicultural society where children encounter speakers with variable accents. Here, 3-5-year-olds showed fairly accurate but slowed recognition of accent-like pronunciations.