Developmental Science

Developmental Science 14:2 (2011), pp 372–384 DOI: 10.1111/j.1467-7687.2010.00986.x

PAPER

Adaptation to novel accents by toddlers

Katherine S. White and Richard N. Aslin

Department of Brain & Cognitive Sciences, University of Rochester, USA

Abstract

Word recognition is a balancing act: listeners must be sensitive to phonetic detail to avoid confusing similar words, yet, at the same time, be flexible enough to adapt to phonetically variable pronunciations, such as those produced by speakers of different dialects or by non-native speakers. Recent work has demonstrated that young toddlers are sensitive to phonetic detail during word recognition; pronunciations that deviate from the typical phonological form lead to a disruption of processing. However, it is not known whether young word learners show the flexibility that is characteristic of adult word recognition. The present study explores whether toddlers can adapt to artificial accents in which there is a vowel category shift with respect to the native language. Nineteen-month-olds heard mispronunciations of familiar words (e.g. vowels were shifted from [a] to [æ]: 'dog' pronounced as 'dag'). In test, toddlers were tolerant of mispronunciations if they had recently been exposed to the same vowel shift, but not if they had been exposed to standard pronunciations or other vowel shifts. The effects extended beyond particular items heard in exposure to words sharing the same vowels. These results indicate that, like adults, toddlers show flexibility in their interpretation of phonological detail. Moreover, they suggest that effects of top-down knowledge on the reinterpretation of phonological detail generalize across the phono-lexical system.

Introduction

Word recognition in adults is remarkably efficient and resilient, given the variability of the speech signal. Adults easily recognize words that undergo large-scale acoustic-phonetic transformations due to differences in vocal tract length across speakers, speech rate, and even phonological context. One significant source of variability is that introduced by speakers of different dialects of the native language or by non-native speakers. In addition to sub-phonemic variability (sub-categorical changes in the realization of particular speech sounds), words pronounced by speakers of different dialects or languages may also involve phonemic changes, that is, full-scale category mismatches (e.g. as when the word 'pen' is pronounced as 'pin' in some dialects of American English). Despite the considerable challenges that such variability would seem to pose to listeners, in some cases adults require only a short period of exposure to adapt to novel accents. For example, after less than a minute of listening, English-speaking adults recognize words in Spanish- and Chinese-accented English as quickly as non-accented English words (Clarke & Garrett, 2004). The speed of this adaptation process varies, however, across accent types and contexts: for example, non-native accents may be difficult to adapt to under conditions of talker variability (Floccia, Goslin, Girard & Konopczynski, 2006).

Some explorations into the flexibility of lexical processing use natural productions from non-native speakers or speakers of non-native dialects (Clarke & Garrett, 2004). In such studies, deviations from the native sound system can exist at multiple levels (subphonemic, phonemic, higher-level prosodic). Another approach is to use artificial accents or pronunciations in order to control the particular changes introduced. In one such study, Maye, Aslin and Tanenhaus (2008) demonstrated that items initially classified as non-words in a lexical decision task (e.g. 'wetch') were reclassified as words after a 20-minute period of exposure to the vowel shift [I] → [E] in the context of a story. With this type of phonological variability – a mismatch with the prototypical vowel category – lexical knowledge (i.e. top-down knowledge that 'wetch' maps onto the lexical item 'witch') appears critical for recovering the intended form and for successful adaptation.

Although post-adaptation lexical decisions in this case could be strategic (i.e. participants consciously accepting recently heard pronunciations), there is evidence to indicate the automatic nature of adaptation effects. For example, implicit exposure to a non-standard speech sound in the context of a lexical decision task affects later judgments in an unrelated categorization task (Kraljic & Samuel, 2005; Norris, McQueen & Cutler, 2003). In some cases, this holds even when new phonemes or tokens from a new speaker are presented during the categorization task (Kraljic & Samuel, 2006). Moreover, hearing non-standard pronunciations can affect the recognition of phonologically related words not heard during the exposure period (Maye *et al.*, 2008; McQueen, Cutler & Norris,

Address for correspondence: Katherine White, Department of Psychology, University of Waterloo, Waterloo, ON, Canada N2L 3G1; e-mail: white@uwaterloo.ca

2006) and can alter the nature of lexical neighborhoods (Dahan, Drucker & Scarborough, 2008), demonstrating that adaptation effects do not simply involve memorization of items heard during the exposure phase, but rather more generalized tuning of phonological categories or re-mappings in phono-lexical space.

These findings with adults highlight two properties of a remarkably flexible word recognition system: first, top-down feedback helps drive the reinterpretation of deviant phonetic detail and, second, this detail is apparently reinterpreted, not on a lexically specific basis, but in a systematic way across the phono-lexical system. Whether such flexibility is present at earlier stages of development is unclear. On the one hand, children might be more willing to accept pronunciations that deviate from stored lexical representations because they are less confident about their phonetic analyses of the speech stream (Swingley & Aslin, 2007); that is, they might weight lexical knowledge more heavily. On the other hand, top-down knowledge is weaker in young learners (by virtue of knowing fewer words and knowing them less well). Therefore, lexical knowledge might exert less influence on the interpretation of phonetic detail, particularly if the system is initially less interactive or more dependent on bottom-up cues (Mattys, White & Melhorn, 2005; see also Trueswell, Sekerina, Hill & Logrip, 1999, for suggestions that bottom-up lexical biases dominate early syntactic processing). Additional support for the notion that accent variability might pose difficulties comes from the limited tolerance young learners show for acoustic variability during lexical processing.

At the earliest stages of word recognition, infants show a dependence on precise acoustic-phonetic form, failing to recognize that words are the same when they are spoken by different talkers, or when they differ in pitch or affect (Houston & Juszcyk, 2000; Singh, Morgan & White, 2004; Singh, White & Morgan, 2008). It is only later in the first year that infants are able to generalize across such non-linguistic acoustic changes and focus on phonological form (Houston & Jusczyk, 2000; Singh *et al.*, 2008). In the second year, toddlers continue to focus on phonological form, exhibiting sensitivity to phonological detail in referential word recognition and word learning tasks (i.e. those that involve mapping words onto visual or real-world referents). In such tasks, toddlers have difficulty recognizing words that are mispronounced by a single consonant or vowel (e.g. Mani & Plunkett, 2007; Nazzi, 2005; Swingley & Aslin, 2000; White & Morgan, 2008), showing that a rather precise match between heard token and stored representation is critical. This is true even when there is no alternative referent onto which to map the mispronounced label – that is, when the visual context should bias them towards accepting the mispronunciation as an instance of the target name. The problem this poses is that dialect or accent variation sometimes introduces phonological changes that are easily discriminable; if learners focus too rigidly on phonological detail, they run the risk of overlexicalizing – treating deviant tokens of familiar words as novel lexical items. Thus, the child's dilemma is one of interpretation: when a token deviates from a familiar pronunciation, should children allow their phonetic analyses or their lexical knowledge to exert more of a role on their interpretation?

Little work has explored the degree to which young word learners can process and adapt to novel accents. At very young ages, non-native accents appear to disrupt processing in non-referential word recognition tasks (Best, Tyler, Kitamura, Notley & Bundgaard-Nielsen, 2008; Best, Tyler, Gooding, Orlando & Quann, 2009; Schmale & Seidl, 2009). Using a listening preference procedure, Best *et al.* (2008) found that English-learning Australian 15-month-olds showed no preference for familiar over unfamiliar words spoken in a non-native (Jamaican) English dialect; Best *et al.* (2009) found similar results for English-learning American 15-month-olds listening to Jamaican English. In contrast, Schmale and Seidl (2009) used a familiarization-recognition task and showed that even 9-month-old English-learning infants can recognize newly familiarized Spanish-accented words under some conditions. The discrepancy between these findings could reflect the difference between accessing long-term representations (Best *et al.*, 2008, 2009) versus short-term acoustic matching (Schmale & Seidl, 2009), or differences in the acoustic-phonetic distance between the various dialect/accent pairs. For example, vowels and consonants in the Mesolect variety of Jamaican English used in Best *et al.* (2008, 2009) differ significantly from those in the toddlers' native American and Australian dialects (vowel changes include shifts, lengthening, and monophthongization; Wassink, 2006; Wells, 1982). In a study testing word comprehension, Mulak, Best, Irwin and Tyler (2008) used the intermodal preferential looking procedure to ask whether American toddlers could identify the referents of words spoken in the same Jamaican English Mesolect as above. Toddlers aged 19–20 months saw two familiar pictures and heard the name of the target picture in either American or Jamaican English (in blocks). They were only able to match the words to the appropriate pictures when the American speaker produced the words. Similarly, Nathan, Wells and Donlan (1998) found word recognition failures in 4- and 7-year-old children from London listening to words produced with a Scottish accent. These studies suggest a potentially catastrophic inability of toddlers and young children to compensate for some types of natural accent variation during word comprehension.

An inability to adapt to the current phonological environment could be very disruptive to young language learners in the process of building a lexicon. Given word learning biases that encourage unique mappings between words and referents (e.g. principle of contrast (Clark, 1987); mutual exclusivity (Markman, 1990); novel-name-nameless-category (Golinkoff, Mervis & Hirsh-Pasek, 1994)), a failure to recognize novel pronunciations of familiar words might lead to searches for other possible referents (or parts/properties of referents). In fact, in a referential context containing one name-known and one name-unknown object, toddlers may map phonologically deviant labels (e.g. 'var' for 'car') onto the name-unknown object (White & Morgan, 2008). Such mapping errors could lead to delays in vocabulary development.

Given that tolerating accent differences is important for successful word recognition and word learning, we ask whether there are conditions under which toddlers can adapt to novel pronunciations of familiar words. We tested toddlers' adaptation to a simplified novel 'accent' in which all [a] vowels (i.e. the vowel in 'dog'[1]) were shifted to [æ] (i.e. the vowel in 'bag'). As in Maye *et al.* (2008), the 'accent' introduced a simple vowel shift, in contrast to the natural variation tested by Mulak *et al.* (2008) or Clarke and Garrett (2004). Because even adults require some (albeit brief) period of adjustment to novel accents, we pre-exposed a group of toddlers to this accent in a clear referential context prior to testing, rather than asking whether the early lexical system is robust enough to compensate for accent variability 'on the fly' (Mulak *et al.*, 2008). During training, all toddlers saw pictures and heard them labeled. They were then tested on their recognition of both standard ([a]) and shifted ([æ]) pronunciations of the labels in sentences. In Experiment 1, two groups of toddlers were tested: the Control group was exposed to standard pronunciations of the labels during training; the Accent group heard labels with shifted ([æ]) vowels during training. The Control group was expected to show a penalty for the shifted vowels at test because they had no prior exposure to the accented words (Mani & Plunkett, 2007; Swingley & Aslin, 2000). If toddlers can adapt to new pronunciations, we expect that the Accent group should show a reduced penalty for these shifted pronunciations; if they cannot adapt, both groups should show a mispronunciation penalty of similar magnitude. An additional question was whether any adaptation effects would generalize beyond the particular items heard during exposure. To test generalization, we included a set of referents in the exposure phase that were not labeled. If toddlers generalize novel accents to phonologically similar items, mispronunciation penalties should be reduced for these items during test as well. This type of generalization would constitute important evidence that adaptation involves changes at the phonological level, rather than simply memorization of the acoustic-phonetics of particular lexical items.

Experiment 1

Subjects

Twenty-four 18–20-month-olds (11 females and 13 males, mean age = 574 days) were assigned to two experimental conditions. Twelve toddlers were assigned to the Control group; 12 were assigned to the Accent group. Within each group, six subjects were assigned to each of two counterbalancing conditions. An additional 19 subjects were tested, but their data were not analyzed for the following reasons: did not complete enough trials for analysis (explained below) due to fussiness or disinterest in the stimuli, or failure to look at both objects during the baseline phase (16);[2] sibling interference (one); language or hearing delay (two). Subsequent contacts with the families revealed that only a handful of toddlers in the final sample (of both Experiments 1 and 2) had exposure to non-local dialects or languages in the home; the number did not differ across conditions.

Stimuli

The familiar stimuli comprised six highly familiar words that are comprehended by the majority (> 50%) of infants by 14 months, according to parental report norms (Dale & Fenson, 1996). All familiar words contained the vowel [a] in the local dialect; based on parental report norms, the vowel [a] permitted the construction of an early acquired and highly picturable stimulus set. The six unfamiliar items selected for the study were real objects, similar in visual complexity to the known-label objects. The names for the unfamiliar objects are not included on lists of familiar words on the infant or toddler versions of the MacArthur CDI (Dale & Fenson, 1996).

Audio stimuli

Natural tokens of child-directed speech were recorded in a soundproof booth by a female, native speaker of English from the same geographic region as the participants. Stimuli were recorded at a sampling rate of 44,100 Hz and later measured and edited using Praat (Boersma & Weenink, 2009). For the exposure stimuli, words were produced in isolation: standard tokens of the familiar words contained the vowel [a] and shifted tokens contained the vowel [æ]. Shifted versions were non-words or words judged to be unknown or relatively unfamiliar to children of this age. The MacArthur CDI lists only one shifted version of the exposure words we used in our experiments ('sick', used in Experiment 2); it is reported to be familiar to a small percentage of children (12.5% by 16 months). Our parental questionnaire (below) served as an additional check on the selection of stimuli. Phonologically, [a] is characterized as a low (height), mid-front vowel, whereas [æ] is characterized as a low-mid (height), front vowel. Average F1 and F2 values for the auditory stimuli are provided in Table 1. Formant values were determined at approximately the midpoint of the vowel.

[1] In the regional dialect of participants in the present study, words like 'dog' are pronounced with the vowel [a], rather than [open o].

[2] Baby noises were added to the silent attention getter after approximately half of the participants had been run. This change led to a considerable reduction in subject attrition.

Three tokens of each version were chosen for use during the exposure phase (mean length standard tokens = 617 msec; mean length shifted tokens = 644 msec). Tokens of 'look' and 'wow' were also recorded. For the test phase, two instruction frames ('*Find the X!*' and '*Do you see the X*?') were recorded, each with standard and shifted versions of each word. Then, for half of the familiar items, sentence frames with the standard pronunciation of the target word were chosen; for the other half of the items, sentence frames with the shifted version were chosen. For each, a matched sentence with the alternative form of the word was created by replacing the target word (i.e. for the base sentence '*Find the dog*!', the shifted sentence was created by replacing the word 'dog' with its shifted counterpart, 'dag'). Thus, the standard and shifted versions of each familiar word occurred with the identical preceding context. The average duration of the context prior to the target words was 1105 msec. The average duration of the target words in these frames was 645 msec (standard pronunciations: 628 msec; shifted pronunciations: 662 msec).

Exposure phase

The exposure phase contained three animated displays, each with four familiar objects (see Figure 1). Displays were created using Adobe Macromedia Flash. Across the three displays, six unique familiar objects were presented. Each familiar object appeared in two displays (in different positions in the array). Two seconds after the display initially appeared, one of the objects was labeled (e.g. '*Dog!*') and labeling of objects in the display continued every 2 seconds, to a total of 10 labeling events. During naming, the labeled object loomed (i.e. expanded and contracted in size) to highlight the relationship between the label and the object. In each display, two objects ('labeled' items) were labeled four times. The other two objects ('unlabeled' items) each loomed a single time, accompanied by '*Wow!*' or '*Look!*' These objects were included in the displays so that they would be visually familiar to subjects during test; they were unlabeled so that subjects would have no evidence about how the talker produced these words.

The two groups of subjects (Control and Accent) differed with respect to the labels heard during exposure (see Table 1). The Control group heard the objects labeled with the standard vowel (e.g. 'dog'); the Accent group heard them labeled with the shifted vowel (e.g. 'dag'). Within each group, there were two counterbalancing groups. For counterbalancing group 1, the items *ball*, *dog*, and *block* were labeled and the items *car*, *sock*, and *bottle* were unlabeled. The reverse was true for counterbalancing group 2. As described above, only the labeled objects were named during exposure (eight times total); unlabeled objects were seen in the same number of displays as the labeled objects, but loomed only twice, accompanied by '*Look!*' or '*Wow!*' Each display was followed by a darkened screen and an attention-getter (a laughing baby). Once the child was judged to be attending, the experimenter pressed a button to advance to the next display.

Test phase

Test displays were static images containing one familiar (labeled or unlabeled) item paired with one unfamiliar item (which was not seen during training). All participants received the same six familiar–unfamiliar pairings (see Figure 1). Each test trial consisted of two phases: during the 3-second baseline phase, the two objects were presented side-by-side in silence to establish baseline looking preferences. After 3 seconds, a sentence instructed the child to look at the familiar object using one of the two carrier phrases ('*Find the X!*', '*Do you see the X?*'). The display remained on the screen for 5 seconds following the onset of the carrier phrase.

There were 24 total test trials (four involving each of the object pairs). Twelve trials contained the standard pronunciation of a familiar object's name; 12 contained a shifted pronunciation. Thus, each familiar item was heard twice with the standard pronunciation (once per sentence frame) and twice with the shifted pronunciation (once per sentence frame). Within each test block (12 trials), each familiar item was named once with the standard and once with the shifted pronunciation; these trials involved the same sentence frame. Half of the familiar items occurred in each sentence frame during each test block. Additionally, within each block, the side of presentation was the same for the two trials involving each familiar item (and reversed for the second block). Across items, familiar objects appeared on the left and right sides of the screen equally often. The order of presentation was random, with two constraints: the familiar item could not appear on the same side more than three trials in succession; nor could the

Table 1 *Exposure and test vowels, by experiment. Values in parentheses give average F1 and F2 values (in Hz) for vowels in exposure words and vowels in sentence target words*

	Training vowels	Test vowels
Experiment 1		
Control Group:	[a] (F1: 873Hz, F2: 1388Hz)	[a] (F1: 884Hz, F2: 1421Hz), [æ] (F1: 991Hz, F2: 1866Hz)
Accent Group:	[æ] (F1: 953Hz, F2: 1742Hz)	[a] (F1: 884Hz, F2: 1421Hz), [æ] (F1: 991Hz, F2: 1866Hz)
Experiment 2		
Near Accent Group:	[æ] (F1: 953Hz, F2: 1742Hz)	[a] (F1: 868Hz, F2: 1284Hz), [ɛ] (F1: 800Hz, F2: 1963Hz)
Far Accent Group:	[æ] (F1: 953Hz, F2: 1742Hz)	[a] (F1: 951Hz, F2: 1323Hz), [I] (F1: 596Hz, F2: 2134Hz)

same sentence frame occur in more than three trials in a row.

Procedure

The parent sat with the child on his/her lap, listening to instrumental music over noise-attenuating headphones to mask the audio stimuli. Approximately 100 cm in front of the child was a large Panasonic plasma screen monitor (90 cm × 50 cm). Two speakers were located adjacent to the monitor, one to each side. On a table approximately 60 cm in front of the child was a pan/tilt model 504 camera from Applied Sciences Laboratories. Video output was routed to a monitor located behind a heavy, black curtain for viewing by the experimenter, as well as to a Sony digital video recorder for later off-line coding. Speech stimuli were played at conversation level (70 dB). Stimulus presentation was controlled by Psyscope X (Cohen, MacWhinney, Flatt & Provost, 1993; http://psy.ck.sissa.it).

The exposure phase began when the experimenter judged the child to be fixating the center of the screen. At this point, the experimenter pressed a key to bring up one of the exposure displays (in a randomly determined order for each participant) and the trial proceeded automatically. An attention-getter followed each trial. The next trial was initiated when the experimenter judged that the child was again attending centrally. Each exposure movie was 22 seconds long; the exposure phase lasted approximately 1 minute, 9 seconds (assuming 1-second delays between trials).

The test phase immediately followed exposure. Each trial was again initiated by a button press from the experimenter and continued automatically. Each trial began with presentation of the two objects without audio. After 3 seconds, one of the sentences played. One of two attention-getters followed each trial (the baby or a looming, checkered circle); each appeared with every item pair. Test trials were 8 seconds; the test phase lasted approximately 3 minutes, 36 seconds (assuming 1 second between trials).

Following the session, the parent completed a questionnaire on his/her toddler's comprehension and production of the stimulus words (familiar, shifted, and unfamiliar).

Validation of stimulus items

Results from the parental questionnaire indicated that the items we selected were appropriate for this sample of toddlers. On a scale of 1 (not visually familiar, label unknown), 2 (visually familiar, label unknown), 3 (visually familiar, label familiar), and 4 (visually familiar, label highly familiar), items used as familiar words received an average score of 3.73 (SD = .22) for toddlers assigned to the Control group and 3.76 (SD = .3) for toddlers assigned to the Accent group, indicating that they were very familiar to both groups of toddlers. Parents reported that children were also producing a number of the words themselves.

Figure 1 *(a) Sample exposure stimuli, (b) test stimulus pairs. Actual stimuli were in color.*

Due to constraints on stimulus selection (i.e. all stimuli were familiar, picturable objects whose name shared a vowel), some of the shifted productions resulted in real words ('battle', 'black', 'sack'). The word *battle* was not rated by parents as being familiar to any of the participants. 'Black' was rated as somewhat familiar (score of 3) to two participants; 'sack' was rated as somewhat familiar to three different participants. Removing trials in which parents judged the shifted versions to be familiar did not change the pattern of results. For the unfamiliar objects, familiarity ratings were 1.34 (SD = .27) and 1.26 (SD = .31) for the Control and Accent groups, respectively. These ratings indicate that most of the objects were visually novel and their names were unfamiliar. Parents reported that their children were not producing the names of any of the unfamiliar objects. Importantly, the two training groups had similar ratings for both familiar and unfamiliar items.

Coding of looking times

Looking behavior in test trials was coded off-line frame-by-frame (1 frame = 33 msec) using the SuperCoder program (Hollich, 2005). No audio was available during coding. For each phase (baseline and naming), the proportion of looking towards each of the objects was computed over the total time the child spent looking at both objects. For the naming portion of the trial, only looks from 300 to 3000 msec post target word onset were included in the analyses. The 300 msec delay was included to allow for the time necessary for eye movement programming and execution based on the audio input (similar to procedures used in work with adults). Lags ranging from 0 to 400 msec have been used with toddlers in word recognition studies (Bailey & Plunkett, 2002; Ballem & Plunkett, 2005; Swingley & Aslin, 2000; White & Morgan, 2008). Given that the minimum latency to launch a saccade in younger infants viewing less complex displays is estimated to be 133 msec (with mean latencies of 300–400 msec; Canfield, Smith, Brezsnyak & Snow, 1997) and given target–distractor divergence points of 400 msec in previous studies (Swingley & Aslin, 2000), 300 msec seems a reasonable time delay.[3]

Results and discussion

For each subject, difference scores were computed for four word type conditions: labeled words/standard pronunciation, unlabeled words/standard pronunciation, labeled words/shifted pronunciation, unlabeled words/shifted pronunciation. The difference score measured the change in looking toward the familiar object after naming (proportion looking to familiar object during naming − proportion looking to familiar object during baseline). Comparison across baseline and naming phases allows us to use each stimulus pair as its own control, thereby controlling for any inherent preference for a particular stimulus in each pairing. In addition, preferences within a pair might change over the course of the experiment; these scores allow us to examine the effect of naming on a trial-by-trial basis.

Because of the importance of establishing baseline preferences, trials in which subjects did not look at both objects during the baseline phase were not included in the analysis. Across all included subjects, approximately 6% of trials were discarded for this reason. To be included in the analyses, subjects had to complete at least 16 useable trials (of 24), with at least four usable trials in each of the four word type conditions. In addition to providing enough trials for reasonable estimates of performance, this criterion ensured that responses in each condition were based on at least two of the three labeled items and two of the three unlabeled items.

To ensure that the two participant groups did not differ in their baseline preferences for the familiar vs. unfamiliar objects, we first examined looking to the familiar object during the silent baseline period of the test trials. Proportion of time spent looking at the familiar object did not differ significantly across the training groups (Control group mean = .52; Accent group mean = .49).

Naming-minus-baseline difference scores are depicted in Figure 2. A $2 \times 2 \times 2$ ANOVA was performed, including training group (Control, Accent) as a between-subjects factor and word type (labeled, unlabeled) and test pronunciation (standard, shifted) as within-subject factors. The effects of test pronunciation and training group were significant ($F(1, 22) = 10.57$, $p < .004$; $F(1, 22) = 11.68$, $p < .002$, respectively). Overall, looking to the familiar object was higher for standard pronunciations than shifted pronunciations, and was higher across pronunciations for the Accent group than the Control group. There was a trend towards an interaction of training group and test pronunciation ($F(1, 22) = 3.12$, $p < .09$).

Given the predictions of the study, we then examined the effects of test pronunciation and word type for each training group individually. In the Control group, there was a significant effect of test pronunciation ($F(1, 11) = 11.72$, $p < .006$), but no effect of word type (F(1, 11 = 1.82, $p < .2$) and no interaction ($F(1, 11) = .07$, *ns*). Word recognition is demonstrated by a significant increase in looking from the baseline to naming phase. Planned two-tailed *t*-tests comparing difference scores in each condition to zero (no change from baseline to naming) revealed significant increases for the standard conditions (labeled and unlabeled: $t(11) = 4.62$, $p < .001$ and $t(11) = 2.27$, $p < .04$, respectively), but not for the shifted conditions (labeled and unlabeled: $t(11) = .3$, *ns* and $t(11) = -1.5$, $p < .16$). Thus, only correctly pronounced words were recognized.

In the Accent group, in contrast, there was no significant effect of either test pronunciation or word type ($F(1, 11) = 1.19$, $p < .3$ and $F(1, 11) = 1.25$, $p < .29$) . Planned two-tailed *t*-tests comparing difference scores to zero revealed significant increases for all conditions, both the standard conditions (labeled and unlabeled: $t(11) = 5.12$, $p < .001$ and $t(11) = 3.61$, $p < .004$) and the shifted conditions (labeled and unlabeled: $t(11) = 2.88$, $p < .01$ and $t(11) = 3.02$, $p < .01$). Therefore, unlike the Control group, the Accent group recognized both correct and shifted pronunciations of the familiar words. This was true not only for words that they had heard with those pronunciations in training, but it also generalized to phonologically related words.

Because each item pair was repeated during testing (twice with each pronunciation type), there was an opportunity for adaptation to occur during the test phase alone. To address this possibility, we compared recognition scores for shifted words in the two test blocks. In the Control group, recognition scores were significantly negative in the first block, but became (non-significantly) positive by the second block (block 1 mean = −.05; block 2 mean = .03). A paired *t*-test showed a significant increase in recognition scores between blocks ($t(11) = 3.11$, $p < .01$). As described above, the Accent group had significant recognition scores for shifted words even in the

[3] As a check, analyses were also conducted using the full naming window (starting at target onset, with no time lag). This did not alter the pattern of results.

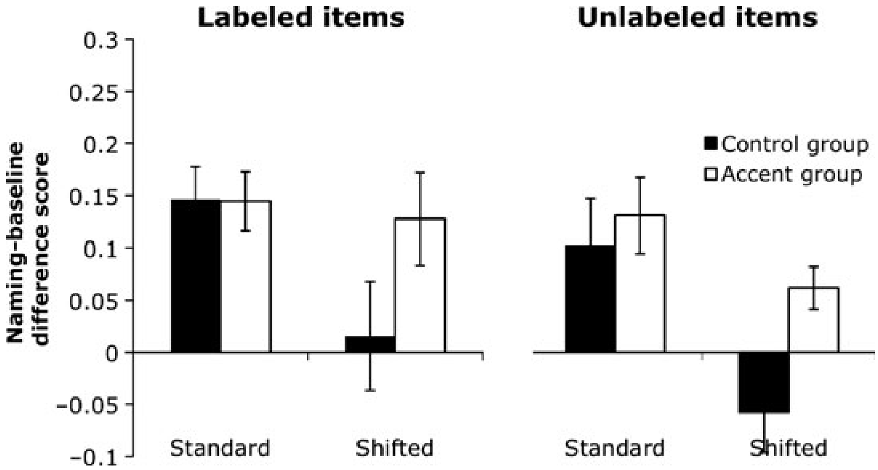

Figure 2 *Difference scores and standard errors, Experiment 1. Test condition is on the x-axis. The y-axis gives the difference between proportion looking at the familiar object in the naming phase and proportion looking at the familiar object in the baseline phase. Black bars correspond to the Control Group, white bars to the Accent Group. The left panel shows results for the labeled items; the right panel shows results for the unlabeled items.*

first block; these became (non-significantly) higher by the second block (block 1 mean = .06; block 2 mean = .13, $t(11) = 1.01$, $p < .33$).

These results suggest that brief exposure helps toddlers learn about a new accent: toddlers exposed to novel pronunciations of familiar words recognized these pronunciations later; the same pronunciations were interpreted by non-exposed toddlers as mispronunciations (although this group showed some learning over the test phase itself). Note that these results do not reveal *what* exactly toddlers in the Accent group learned about the vowel shift heard in exposure. For example, they might have learned that [a] → [æ] in restricted syllabic contexts, that [a] → [æ] in all environments, or even a more general vowel raising process. Before tackling the issue of what particular vowel shift these toddlers might have learned, however, it is important to acknowledge still another possibility: perhaps the Accent group learned nothing about the vowel shift at all, but simply became more tolerant of this talker's mispronunciations because they found her pronunciations odd during exposure. That is, upon observing that this talker pronounced familiar words in a deviant manner, perhaps they subsequently dismissed her vowels as being irrelevant to lexical identity. In order to determine whether toddlers were learning something specific about the exposure accent, and not simply to ignore the vowels, in Experiment 2 we exposed toddlers to the same shifted [æ] pronunciations as in Experiment 1, but changed the nature of the test mispronunciations. Across two groups of participants, we used two new types of mispronunciations to gauge the specificity of the learning. One group (the Near Accent group) was exposed to [æ] pronunciations in training, but heard a closely related vowel ([E]) during test; a second group (the Far Accent group) was also exposed to [æ] pronunciations in training, but heard a more distant vowel ([I]) during test. If hearing non-standard pronunciations simply causes toddlers to be more tolerant of any deviant pronunciations from this talker, then both groups should show a reduced penalty for the shifted test items, just as the Accent group did in Experiment 1. However, if learning is restricted to the particular vowel shift heard during exposure, then neither group should show a reduced mispronunciation penalty, and their behavior should resemble that of the Control group in Experiment 1. Finally, if there is learning of the exposure shift to [æ], but tolerance for similar pronunciations, the Near Accent group might recognize shifted versions during test.

Experiment 2

Subjects

Twenty-four 18–20-month-olds (12 females and 12 males, mean age = 574 days) were assigned to two experimental conditions. Twelve toddlers were assigned to the Near Accent group; 12 were assigned to the Far Accent group. Within each group, six subjects were assigned to each of two counterbalancing conditions. An additional nine subjects were tested, but their data were not analyzed for the following reasons: did not complete enough trials for analysis due to fussiness or disinterest in the stimuli or failure to look at both objects during the baseline phase (four); not codable due to laughing (one); prematurity or language delay (two); mothers had strong non-native accents (two).

Visual stimuli

The same stimuli were used as in Experiment 1.

Audio stimuli

The exposure stimuli were identical to those played to the Accent group in Experiment 1. Test stimuli were produced for Experiment 2 by the same speaker and were of the same form as in Experiment 1. Standard and shifted versions of the target words in the same two sentence frames ('*Find the X!*' and '*Do you see the X?*') were recorded. For the Near Accent group, the target words in the shifted test sentences were pronounced with the vowel [E]; for the Far Accent group, the target words in the shifted test sentences were pronounced with [I]. Phonologically, [E] is characterized as a mid (height), front vowel, whereas [I] is characterized as a mid-high (height), front vowel. Average formant frequencies are provided in Table 1. The procedure described above was used to record and create the final matched test sentences. The average duration of the context prior to the target words was 1318 msec for the Near Accent group and 1267 msec for the Far Accent group. The average duration of the target words in these frames was 639 msec for the Near Accent group (standard pronunciations: 643 msec; shifted pronunciations: 636 msec) and 642 msec for the Far Accent group (standard pronunciations: 660 msec; shifted pronunciations: 623 msec).

Exposure phase, test phase, procedure, and scoring

These were identical to Experiment 1.

Validation of stimulus items

On the parental questionnaire, familiar words received an average score of 3.76 (*SD* = .28) for toddlers in the Near Accent group and 3.91 (*SD* = .15) for toddlers in the Far Accent group. Parents also indicated that children were producing a number of the words themselves. For the Near Accent group, five toddlers were rated as being familiar with 'bell', a partially overlapping set of four toddlers were rated to be familiar with 'black', and one of these was also rated as being familiar with 'sack'. Removing trials in which parents judged the shifted versions to be familiar did produce one change in the pattern of results (see below). For the Far Accent group, four toddlers were rated as being familiar with 'sick', and partially overlapping sets of four and three toddlers were rated to be familiar with 'dig', and 'black', respectively. Removing trials in which parents judged the shifted versions to be familiar did not alter the pattern of results. Unfamiliar items received familiarity ratings of 1.34 (*SD* = .43) and 1.32 (*SD* = .29) for toddlers in the Near Accent and Far Accent groups, respectively. Parents reported that their children were not producing the names of any of the unfamiliar objects.

Results and discussion

As in Experiment 1, trials in which subjects did not look at both objects during the baseline phase were not analyzed. Across subjects, approximately 8% of trials were discarded for this reason.

Preference for the familiar object during the baseline period did not differ across training groups or test trial type (Near Accent group mean proportion = .51; Far Accent group mean proportion = .49). Naming-minus-baseline difference scores are depicted in Figure 3.

A 2 × 2 × 2 ANOVA was performed, including group (Near, Far) as a between-subjects factor and word type (labeled, unlabeled) and test pronunciation (standard, shifted) as within-subject factors. The only significant effect was that of test pronunciation ($F(1, 22) = 24.67$, $p < .001$). Overall, looking to the familiar object was higher for standard pronunciations than shifted pronunciations. There was no effect of word type ($F(1, 22) = 1.15$, $p < .3$) or group ($F(1, 22) = .02$, *ns*), nor were there any interactions.

The effects of test pronunciation and word type were then assessed for each training group. In the Near Accent group, there was a significant effect of test pronunciation ($F(1, 11) = 5.4$, $p < .04$) and no effect of word type ($F(1, 11 = .15$, *ns*) or interaction of test pronunciation and word type ($F(1, 11) = 3.2$, $p < .1$). Planned two-tailed *t*-tests comparing difference scores to zero revealed significant increases for the standard conditions (labeled and unlabeled: $t(11) = 3.71$, $p < .003$ and $t(11) = 2.94$, $p < .01$, respectively), but not for the shifted conditions

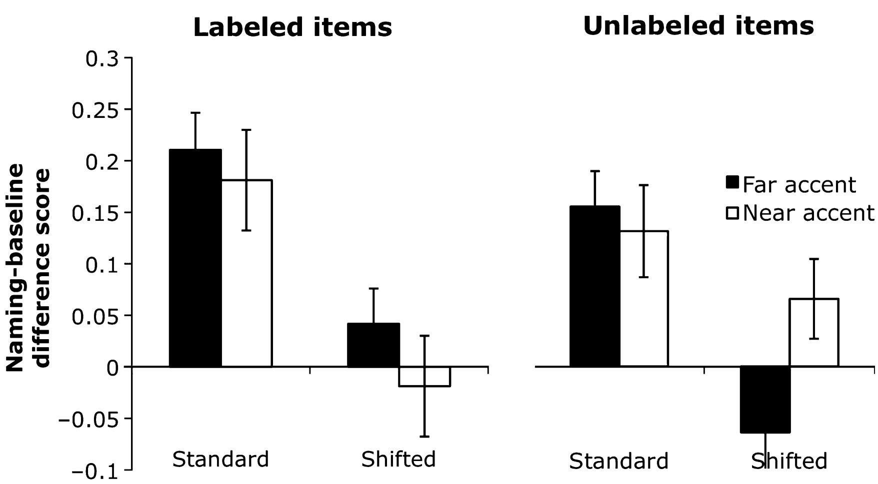

Figure 3 *Difference scores and standard errors, Experiment 2. Test condition is on the x-axis. The y-axis gives the difference between proportion looking at the familiar object in the naming phase and proportion looking at the familiar object in the baseline phase. Black bars correspond to the Far Accent Group, white bars to the Near Accent Group. The left panel shows results for the labeled items; the right panel shows results for the unlabeled items.*

(labeled and unlabeled: $t(11) = -.38$, *ns* and $t(11) = 1.69$, $p < .12$). Thus, only correctly pronounced words were recognized.[4]

In the Far Accent group, there were significant effects of test pronunciation ($F(1, 11) = 35.07$, $p < .001$) and word type ($F(1, 11) = 4.91$, $p < .05$), but no interaction ($F(1, 11) = .54$, *ns*). Planned two-tailed *t*-tests comparing difference scores to zero again revealed significant increases for the standard conditions only (labeled and unlabeled: $t(11) = 5.84$, $p < .001$ and $t(11) = 4.51$, $p < .001$), and not for the shifted conditions (labeled and unlabeled: $t(11) = 1.22$, $p < .25$ and $t(11) = -1.48$, $p < .17$). The pattern of performance was similar for the Near and Far Accent groups: like the Control group of Experiment 1 (and unlike the Accent group), they recognized only the correct test pronunciations.

We again assessed whether there were any learning effects in the test phase by comparing recognition scores for shifted words in the two test blocks. In the Near Accent group, recognition scores for the shifted words did not differ from zero in either block (block 1 mean = .01; block 2 mean = .02). A paired *t*-test showed no change in recognition scores between blocks ($t(11) = .14$, *ns*). In the Far Accent group, recognition scores for shifted words were 0 in the first block and −.02 in the second block, again non-significantly different ($t(11) = -.25$, *ns*). Thus, unlike the toddlers in Experiment 1, toddlers in Experiment 2 showed no evidence of adapting to the shifted pronunciations over the course of the test phase alone.

General discussion

In two experiments, we assessed the role of immediately preceding exposure on 18–20-month-olds' recognition of standard and non-standard pronunciations of familiar words. All groups recognized the standard pronunciations, even when faced with evidence that the talker had an accent. Similarly, in Maye *et al.* (2008), adults judged standard versions to be lexical items, even after exposure to the items in shifted form from the same talker. We used displays containing one familiar target object and one unfamiliar object in order to ask about toddlers' tolerance for the deviant pronunciations (White & Morgan, 2008): when an unfamiliar competitor is present, the lexical processing system must determine whether the deviation is substantial enough to constitute a new lexical entry. We were particularly interested in how toddlers with different types of prior exposure would treat shifted pronunciations. In Experiment 1, toddlers who heard standard pronunciations during an immediately preceding exposure phase later interpreted shifted pronunciations involving vowel changes as mispronunciations. In contrast, toddlers who had brief exposure to these new pronunciations (and their referent mappings) did not interpret them later as mispronunciations. In Experiment 2, toddlers who heard the talker produce words with one atypical vowel in exposure and a different atypical vowel in test interpreted the shifted test pronunciations as mispronunciations. Below we discuss three aspects of these results, and implications for the nature of the early lexical processing system.

First, similar to adults, toddlers require some (albeit brief) period of exposure in order to adapt to novel pronunciations. This is demonstrated by the contrast between the results of the Accent group in Experiment 1 and previous work. Recall that American toddlers encountering a non-native dialect for the first time show a complete failure to recognize familiar words in a similar task (Mulak *et al.*, 2008). Further evidence for the role of exposure comes from studies of mispronunciation detection in infants and toddlers. Across multiple studies, mispronunciations of similar magnitude to the vowel shifts used here have caused recognition difficulties in children aged 14 to 24 months (Mani & Plunkett 2007, 2008). Thus, hearing similarly deviant pronunciations in the absence of prior exposure disrupts word recognition (see Nazzi, 2005, and subsequent work, however, for evidence that toddlers are less sensitive to vowel contrasts during word learning).

Our findings raise the possibility that unambiguous or salient evidence for the mapping between novel pronunciations and referents aids adaptation. If exposure to the deviant forms alone were enough, then toddlers in the Mulak *et al.* study should have recognized the words in later trials. We might also expect performance to improve in mispronunciation studies that have included item repetition, even when attention has not been explicitly directed to the new mappings. Most of these studies have not assessed repetition effects. However, Ballem and Plunkett (2005) reported that mispronunciation penalties found for familiar items in the first block of testing (where only correct pronunciations were recognized) had disappeared by the second block (where both correct and mispronounced versions were recognized). A hint that repetition alone might be sufficient in some contexts is also provided by the block effects we found in Experiment 1. We observed two types of effect: an effect of explicit training (the advantage for the Accent group starting in block 1), as well as an effect of repetition (the increase in recognition scores for the Control group between blocks 1 and 2). Our explicit training effects are similar to lexical effects found in adults' adaptation to novel accents. Adults use both linguistic and non-linguistic context to resolve phonemic category mismatches.

[4] After removing trials for which parents judged the shifted versions to be familiar, the overall pattern of results remained the same (correctly pronounced words were recognized, but shifted versions were not). However, difference scores for labeled words heard in shifted pronunciations became more negative, whereas difference scores for unlabeled words heard with shifted pronunciations became more positive. In fact, the scores for shifted versions of unlabeled words differed significantly from zero, suggesting that these words were, in fact, recognized (although the removal of these trials meant that some participants no longer met the 4-trial per condition criterion).

For example, in Maye *et al*. (2008), adults were able to use both the linguistic context and their knowledge that certain items (e.g. 'wetch') were not words in use in English. Here, we provided very salient and pictorial evidence of the new mappings; it is unclear how well young learners would be able to use other cues, such as linguistic context, to adapt to novel pronunciations of familiar words.

Second, the results of Experiment 1 demonstrate that participants did not simply memorize how particular items were pronounced during the exposure phase. Toddlers in the Accent group accepted not only shifted pronunciations of words heard in the exposure phase, but also shifted versions of words not heard during exposure. Thus, learning was not specific to particular lexical items, but rather generalized to words with similar phonological properties. This pattern is similar to the performance observed when adults encounter novel accents (Clarke & Garrett, 2004). Recall that Maye *et al*. (2008) found that adults exposed to a novel accent in which [i] was pronounced as [E] subsequently judged shifted versions of both heard and not-heard items as acceptable words. Similarly, McQueen *et al*. (2006) found that hearing an ambiguous [f-s] fricative during a lexical decision task affected subsequent recognition of words not heard during the lexical decision phase.

The ability to generalize beyond specific lexical items is critical for efficient accent adaptation, for toddlers and adults alike. But how top-down feedback alters representations and/or processing in such a generalized manner is unclear. For example, it is possible that recognizing [dæg] and [bæl] as instances of [dag] and [bal] causes an expansion of the [a] vowel category such that [æ] comes to be perceived as more similar to [a]. This would account quite easily for the generalization to phonologically similar items that were not heard in exposure. In work on perceptual adaptation in adults, it appears that categories are broadened following exposure to ambiguous segments in lexical contexts (e.g. expansion of [s] to include [ʃ] tokens: Kraljic & Samuel, 2005; Norris *et al*., 2003). A broadening account would also explain why toddlers continue to recognize the original pronunciations following exposure to the accented productions in the present study.

Third, the pattern of results across our two experiments suggests that toddlers learned about the particular accent heard during exposure. While toddlers in the Accent group (Experiment 1) recognized shifted versions of the target words, toddlers in the Near and Far Accent groups (Experiment 2) did not. Hearing deviant pronunciations during the exposure period did not cause toddlers to become more tolerant of phonological deviation in general; they did not simply assume that because this person talks 'funny', they should accept any variant she produced (dismissing her vowels as being irrelevant to lexical identity). Rather, once a particular shift was learned in exposure ([a] → [æ]), toddlers expected subsequent pronunciations by this talker to match this shift.[5] This specificity is demonstrated by the Near Accent group's failure (after familiarization with [æ]) to recognize items pronounced with the highly similar vowel [E] during test. One caveat is that, after we removed trials for which parents indicated some familiarity with the shifted versions, this group showed recognition of unlabeled words heard in shifted form. This would seem to suggest either some lack of specificity, or a case of misperception. In adults, the [æ]-[E] pairing shows the highest degree of confusability among vowel pairs (Hillenbrand, Getty, Clark & Wheeler, 1995; Neel, 2008). Perhaps toddlers in the Near Accent group misinterpreted some of the [E] test pronunciations as containing the [æ] vowel to which they had been exposed in training (note that this would imply generalization of the [a] → [æ] vowel shift to unlabeled items). If it were a case of misperception alone, however, toddlers in this group should have accepted [E] test pronunciations of labeled items as well, which was not the case. It remains to be seen whether this pattern of results will replicate, as it was found in only a subset of trials.

Our results highlight specificity once a particular shift has been learned, but do not address what constraints there might be on the initial learning process. The most likely shift learned by toddlers in our experiments ([a]–[æ]) is fairly natural in that the distance between these vowel categories is small in acoustic-phonetic space and vowel raising is a frequently occurring phonological process. However, our results do not rule out the possibility that more extreme (less natural) shifts are learnable. In fact, learning might be quite unconstrained, in that any type of shift is learnable. If toddlers in Experiment 2 had heard [E] or [I] variants initially and had been tested with the same vowels, perhaps their performance would have resembled that of the Accent group in Experiment 1. Their failure may have been a result of having learned one vowel shift in exposure ([a] → [æ]) and encountering a novel shift in test. Perhaps toddlers can only cope with a single shift per talker (either per word, or per vowel). Note that, even after a block of testing, toddlers showed no evidence of adapting to the [E] and [I] forms. Whether learning is constrained by the naturalness of the shift therefore remains an open question.

The present work raises a number of additional questions for future research. First, our design exposed toddlers to an artificial accent in which the vowel in a small set of words was systematically shifted. A major difference between the current study and that of Mulak *et al.*

[5] As pointed out by an anonymous reviewer, it is not clear exactly what shift toddlers learned. Our results are compatible with toddlers learning the shift [a] → [æ] (either in particular syllable contexts or more generally). But they are also compatible with the learning of a more generalized raising process. In the latter case, toddlers exposed initially to an [a] → [æ] shift ('dog' → 'dag') might later also recognize words originally containing [æ] when they are pronounced with [E] ('blanket' → 'blenket'; a similar type of raising).

(2008), in addition to the absence of an exposure period, is that they used a natural dialect difference, that between American and Jamaican English. Of course, natural dialects contain much more variability than a single vowel shift. It is not clear that the type and extent of exposure here would suffice to familiarize toddlers with natural dialects. In addition, our artificial accent involved only vowels, the processing of which might be more flexible than that of consonants. Vowel categories, even within a single dialect, are quite variable; it has also been argued that they are less critical to lexical identity and word learning than consonants (Havy & Nazzi, 2009; Nazzi, 2005; Nazzi, Floccia, Moquet & Butler, 2009; Nespor, Pena & Mehler, 2003). Dialect variants often involve shifts, divisions, or mergers of vowel categories, as in the distinction between [a] and [open o] found in dialects of northeastern American English, or the shift from [E] to [I] found in other dialects (Labov, Ash & Boberg, 2006). Perhaps previous exposure to these types of vowel variations by the age of 18 months facilitated the adaptation to the vowel shift we observed in our Accent group. It is not clear whether toddlers would be so quick to adapt to novel accents involving cross-category consonant shifts, either because consonants are more critical to lexical identity, or because consonant categories are narrower, both within and across dialects. One possible explanation for why non-native accents can lead to greater processing difficulty than regional dialects (Floccia *et al.*, 2006; Floccia, Butler, Girard & Goslin, 2009) is that non-native accents can involve deviations of both consonant and vowel categories, whereas regional accents typically involve changes only to the latter (Floccia *et al.*, 2009).

A second question for future research is how previous experience may create individual differences in flexibly dealing with new accents. Many children are exposed to bi-dialectal, if not bilingual, input and therefore hear multiple mappings between phonological forms and referents. Recall that research with monolingual children demonstrates the presence of word learning biases that encourage unique mappings between words and referents (Clark, 1987; Golinkoff *et al.*, 1994; Markman, 1990). The presence of translation equivalents (words that have the same referent) in bilingual children (Pearson, Fernandez & Oller, 1995; Petitto, Katerelos, Levy, Gauna, Tétreault & Ferraro, 2001) argues that these lexical biases operate within, and not across, languages. Recent research has also demonstrated that bilingual children might abandon, or rely less heavily on, these biases than monolingual children (Byers-Heinlein & Werker, 2009). In addition to affecting the operation of word learning biases, exposure to multiple languages alters the nature of early phonological sensitivity. For example, bilingual Spanish-Catalan toddlers are less sensitive to mispronunciations involving /e/ and /E/ than Catalan monolinguals (Ramon-Casas, Swingley, Sebastian-Galles & Bosch, 2009). This is likely due to the pattern of vowel alignment across these languages: the high frequency /e/ category in Spanish overlaps with the lower frequency /e/ and /E/ categories in Catalan. The effects of bilingualism on word learning biases and phonological processing lead to the intriguing possibility that exposure to multiple dialects *of the same language* might also have effects on the interpretation of novel pronunciations. The effects of dialect exposure may be further modulated by the type of exposure (listening only vs. listening + self-productions), as has been demonstrated in work with adults (Sumner & Samuel, 2009).

Finally, although we have discussed the specificity of learning with respect to the phonological changes involved, there is another sense in which this type of learning might be specific: if children require evidence that a particular talker uses non-standard pronunciations, then switching the talker between exposure and test should eliminate the exposure effect. Previous work has demonstrated that adaptation effects with stops generalize across talkers (Kraljic & Samuel, 2006), but effects with fricatives do not (Kraljic & Samuel, 2005; Eisner & McQueen, 2005). One possible explanation for this discrepancy is that fricatives contain spectral information relevant to speaker identity, and so are more speaker-specific (Kraljic & Samuel, 2006). If this account is correct, then vowels would likewise be expected to show speaker-specificity, as they contain even more information about talker identity. We are currently exploring the role of talker identity in toddlers' adaptation to novel accents.

The present study demonstrates that many of the features that characterize the adult lexical processing system are operational in early childhood: despite adult-like sensitivity to the phonological properties of known words, toddlers are able to update their representations to reflect new experiences. Moreover, this updating occurs not only for heard items, but also for phonologically similar items, reflecting a generalization process across the phono-lexical system.

Acknowledgements

This research was supported by NIH grants R01 HD037082 to RNA and R01 DC05071 to Michael Tanenhaus and RNA. The first author was supported on NIH grant T32 MH19942 to the Department of Brain & Cognitive Sciences at the University of Rochester. We thank the members of the Rochester Baby Lab for help with subject recruitment and all of the families who took time to participate. We also thank Eiling Yee, Mohinish Shukla and two anonymous reviewers for comments on a previous draft.

References

Bailey, T.M., & Plunkett, K. (2002). Phonological specificity in early words. *Cognitive Development*, **17**, 1265–1282.

Ballem, K.D., & Plunkett, K. (2005). Phonological specificity in children at 1;2. *Journal of Child Language*, **32**, 159–173.

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**, 539–542.

Best, C.T., Tyler, M.D., Kitamura, C., Notley, A., & Bundgaard-Nielsen, R. (2008). Phonetic specificity of early words? Australian toddlers' perception of Australian versus Jamaican English pronunciations. International Conference on Infant Studies, Vancouver BC, Canada.

Boersma, P., & Weenink, D. (2009). Praat: doing phonetics by computer (Version 5.1.05) [Computer program]. Retrieved 1 May 2009 from http://www.praat.org/

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

Canfield, R.L., Smith, E.G., Brezsnyak, M.P., & Snow, K.L. (1997). Information processing through the first year of life: a longitudinal study using the visual expectation paradigm. *Monographs of the Society for Research in Child Development*, **62** (2, Serial No. 250).

Clark, E.V. (1987). The principle of contrast: a constraint on language acquisition. In B. MacWhinney (Ed.), *Mechanisms of language acquisition* (pp. 1–33). Hillsdale, NJ: Lawrence Erlbaum Associates.

Clarke, C.M., & Garrett, M. (2004). Rapid adaptation to foreign accented speech. *Journal of the Acoustical Society of America*, **116**, 3647–3658.

Cohen, J.D., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: an interactive graphical system for designing and controlling experiments in the psychology laboratory using Macintosh computers. *Behavior Methods, Instruments, & Computers*, **25**, 257–271.

Dahan, D., Drucker, S.J., & Scarborough, R.A. (2008). Talker adaptation in speech perception: adjusting the signal or the representations? *Cognition*, **108**, 710–718.

Dale, P.S., & Fenson, L. (1996). Lexical development norms for young children. *Behavior Research Methods, Instruments, & Computers*, **28**, 125–127.

Eisner, F., & McQueen, J.M. (2005). The specificity of perceptual learning in speech processing. *Perception & Psychophysics*, **67**, 224–238.

Floccia, C., Butler, J., Girard, F., & Goslin, J. (2009). Categorization of regional and foreign accent in 5- to 7-year-old British children. *International Journal of Behavioral Development*, **33**, 366–375.

Floccia, C., Goslin, J., Girard, F., & Konopczynski, G. (2006). Does a regional accent perturb speech processing? *Journal of Experimental Psychology: Human Perception and Performance*, **32**, 1276–1293.

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.

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

Hillenbrand, J., Getty, L.A., Clark, M.J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *Journal of the Acoustical Society of America*, **97**, 3099–3111.

Hollich, G. (2005). *Supercoder: A program for coding preferential looking (Version 1.5). [Computer Software]*. West Lafayette, IN: Purdue University.

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

Kraljic, T., & Samuel, A.G. (2005). Perceptual learning for speech: is there a return to normal? *Cognitive Psychology*, **51**, 141–178.

Kraljic, T., & Samuel, A.G. (2006). Generalization in perceptual learning for speech. *Psychonomic Bulletin & Review*, **13**, 262–268.

Labov, W., Ash, S., & Boberg, C. (2006). *The atlas of North American English: Phonetics, phonology and sound change*. Berlin: Mouton/de Gruyter.

McQueen, J.M., Cutler, A., & Norris, D. (2006). Phonological abstraction in the mental lexicon. *Cognitive Science*, **30**, 1113–1126.

Mani, N., & Plunkett, K. (2007). Phonological specificity of vowels and consonants in early lexical representations. *Journal of Memory and Language*, **57**, 252–272.

Mani, N., & Plunkett, K. (2008). Fourteen-month-olds pay attention to vowels in novel words. *Developmental Science*, **11**, 53–59.

Markman, E.M. (1990). Constraints children place on word meanings. *Cognitive Science*, **14**, 57–77.

Mattys, S.L., White, L., & Melhorn, J.F. (2005). Integration of multiple speech segmentation cues: a hierarchical framework. *Journal of Experimental Psychology: General*, **134**, 477–500.

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.

Mulak, K.E., Best, C.T., Irwin, J.R., & Tyler, M.D. (2008). The effect of dialect on toddler identification of depicted words. International Conference on Infant Studies, Vancouver, BC, Canada.

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

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

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**, 522–537.

Neel, A.T. (2008). Vowel space characteristics and vowel identification accuracy. *Journal of Speech, Language, and Hearing Research*, **51**, 574–585.

Nespor, M., Pena, M., & Mehler, J. (2003). On the different roles of vowels and consonants in speech processing and language acquisition. *Lingue e Linguaggio*, **2**, 221–247.

Norris, D., McQueen, J.M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology*, **47**, 204–238.

Pearson, B.Z., Fernandez, S., & Oller, D.K. (1995). Cross-language synonyms in the lexicons of bilingual infants: one language or two? *Journal of Child Language*, **22**, 345–368.

Petitto, L.A., Katerelos, M., Levy, B.G., Gauna, K., Tétreault, K., & Ferraro, V. (2001). Bilingual signed and spoken language acquisition from birth: implications for the mechanisms underlying bilingual language acquisition. *Journal of Child Language*, **28**, 1–44.

Ramon-Casas, M., Swingley, D., Sebastian-Galles, N., & Bosch, L. (2009). Vowel categorization during word recognition in bilingual toddlers. *Cognitive Psychology*, **59**, 96–121.
Schmale, R., & Seidl, A. (2009). Accommodating variability in voice and foreign accent: flexibility of early word representations. *Developmental Science*, **12**, 583–601.
Singh, L., Morgan, J.L., & White, K.S. (2004). Preference and processing: the role of speech affect in early spoken word recognition. *Journal of Memory and Language*, **51**, 173–189.
Singh, L., White, K.S., & Morgan, J.L. (2008). Building a word-form lexicon in the face of variable input: influences of pitch and amplitude on early spoken word recognition. *Language Learning and Development*, **4**, 157–178.
Sumner, M., & Samuel, A.G. (2009). The effect of experience on the perception and representation of dialect variants. *Journal of Memory and Language*, **60**, 487–501.
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. (2007). Lexical competition in young children's word learning. *Cognitive Psychology*, **54**, 99–132.
Trueswell, J., Sekerina, I., Hill, N., & Logrip, M. (1999). The kindergarten-path effect: studying online processing in young children. *Cognition*, **73**, 89–134.
Wassink, A.B. (2006). A geometric representation of spectral and temporal vowel features: quantification of vowel overlap in three varieties. *Journal of the Acoustical Society of America*, **119**, 2334–2350.
Wells, J.C. (1982). *Accents of English (3): Beyond the British Isles.* Cambridge: Cambridge University Press.
White, K.S., & Morgan, J.L. (2008). Sub-segmental detail in early lexical representations. *Journal of Memory and Language*, **59**, 114–132.

Received: 20 September 2009
Accepted: 8 April 2010