NIH Public Access
Author Manuscript
J Mem Lang. Author manuscript; available in PMC 2014 May 01.

Published in final edited form as:
J Mem Lang. 2013 May 1; 68(4): 362–378. doi:10.1016/j.jml.2013.01.003.

Adults show less sensitivity to phonetic detail in unfamiliar words, too

Katherine S. White[a,*], **Eiling Yee**[b], **Sheila E. Blumstein**[c], and **James L. Morgan**[c]
[a]Department of Psychology, University of Waterloo, Canada
[b]Basque Center on Cognition, Brain and Language, Spain
[c]Department of Cognitive, Linguistic & Psychological Sciences, Brown University, United States

Abstract

Young word learners fail to discriminate phonetic contrasts in certain situations, an observation that has been used to support arguments that the nature of lexical representation and lexical processing changes over development. An alternative possibility, however, is that these failures arise naturally as a result of how word familiarity affects lexical processing. In the present work, we explored the effects of word familiarity on adults' use of phonetic detail. Participants' eye movements were monitored as they heard single-segment onset mispronunciations of words drawn from a newly learned artificial lexicon. In Experiment 1, single-feature onset mispronunciations were presented; in Experiment 2, participants heard two-feature onset mispronunciations. Word familiarity was manipulated in both experiments by presenting words with various frequencies during training. Both word familiarity and degree of mismatch affected adults' use of phonetic detail: in their looking behavior, participants did not reliably differentiate single-feature mispronunciations and correct pronunciations of low frequency words. For higher frequency words, participants differentiated both 1- and 2-feature mispronunciations from correct pronunciations. However, responses were graded such that 2-feature mispronunciations had a greater effect on looking behavior. These experiments demonstrate that the use of phonetic detail in adults, as in young children, is affected by word familiarity. Parallels between the two populations suggest continuity in the architecture underlying lexical representation and processing throughout development.

Keywords

Artificial lexicon; Spoken word recognition; Phonetic sensitivity; Word familiarity; Developmental continuity; Visual world paradigm

Introduction

Do adults and infants share the same mechanisms for representing and processing words? To date, research on adult lexical processing has proceeded largely without reference to developmental issues; prominent models of spoken word recognition (McClelland & Elman, 1986) often omit any provision for word learning. The converse is also true: findings from adults rarely constrain models of acquisition (Werker & Curtin, 2005). This may be due in part to differences in methods and measures that make results difficult to compare. But the

© 2013 Elsevier Inc. All rights reserved.

* Corresponding author. Address: Dept. of Psychology, University of Waterloo, 200 University Ave., West Waterloo, ON, Canada N2L 3G1.white@uwaterloo.ca. .

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

absence of mutual constraint also reflects, at least implicitly, a theoretical commitment to a lack of developmental continuity. If indeed there are qualitative changes in mechanisms or processes between infancy and adulthood, then it is fitting that research with these two populations should proceed independently. However, if the same architecture underlies processing throughout development, then greater interaction between these two bodies of research is in order. In this article, we investigate to what extent fine phonetic detail is encoded in word representations and used during word processing across the lifespan.

The adult lexicon contains many highly similar words – words that differ from one another by as little as one phonetic feature (e.g., parrot/carrot; jello/cello; pear/bear). Fortunately, the mature lexical processing system is extremely sensitive to such differences. A large literature has documented that during processing, adults exhibit rapid and robust sensitivity to differences between the acoustic signal and stored lexical representations (e.g., Allopenna, Magnuson, & Tanenhaus, 1998; Andruski, Blumstein, & Burton, 1994; Connine, Titone, Deelman, & Blasko, 1997; Dahan, Magnuson, & Tanenhaus, 2001; Magnuson, Dixon, Tanenhaus, & Aslin, 2007; McMurray, Tanenhaus, Aslin, & Spivey, 2003; Milberg, Blumstein, & Dworetzky, 1988). In contrast to this robust sensitivity, young children and toddlers often demonstrate less sensitivity to phonetic detail in words, particularly when they need to simultaneously pay attention to meaning (e.g., in picture selection or word-object association tasks; Barton, 1976, 1980; Eilers & Oller, 1976; Garnica, 1973; Kay-Raining Bird & Chapman, 1998; Schvachkin, 1973; Stager & Werker, 1997; Werker, Fennell, Corcoran, & Stager, 2002). For example, two-year-old children have difficulty using phonological information to differentiate minimal pairs and resolve reference correctly (Eilers & Oller, 1976).

This lack of sensitivity is somewhat surprising, because early in life, infants are quite good at detecting phonetic detail in syllables and word forms (Eimas, 1974; Eimas, Siqueland, Jusczyk, & Vigorito, 1971; Jusczyk & Aslin, 1995; Miller & Eimas, 1983). Moreover, by 12 months, this sensitivity is tuned to phonetic contrasts that are relevant in the native language (Anderson, Morgan, & White, 2003; Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992; Werker & Tees, 1984). Thus, the apparent lack of sensitivity in toddlers and young children seems to suggest that the phonetic organization acquired in the first year is not initially applied in the representation and/or processing of meaningful words. Consistent with this, 14-month-olds sometimes fail to discriminate the very same contrast in the context of a referent that they can discriminate in a non-referential situation (Stager & Werker, 1997).

The difficulty that young word learners display in detecting phonetic contrasts in certain contexts supports arguments that word representations are restructured over development. One such claim is that, despite sensitivity to phonetic dimensions in early infancy, these dimensions are not used in the initial stages of representing meaningful words. The Lexical Restructuring Model (Metsala & Walley, 1998) posits that vocabulary growth produces significant structural change in phonological representations well into childhood. Because young learners do not use a mature set of dimensions to represent words, the amount of phonetic detail represented is idiosyncratic (item-specific), and depends on the amount of experience (familiarity) the learner has had with a particular word, as well as the number of phonologically similar words in the lexicon (Storkel, 2002). Thus, for young learners (unlike for adults), the representations of newly learned words are less specified/more holistic, whereas more familiar words include more phonetic detail (Fowler, 1991; Garlock, Walley, & Metsala, 2001; Metsala & Walley, 1998 but cf. Kay-Raining Bird & Chapman, 1998).

One piece of evidence in support of this model is the fact that young word learners show better discrimination of phonetic contrasts embedded in familiar words than the same contrasts embedded in novel words (Bailey & Plunkett, 2002; Barton, 1980; Fennell &

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Werker, 2003; Mani & Plunkett, 2007; Stager & Werker, 1997; Swingley & Aslin, 2000, 2002; Walley & Metsala, 1990; White & Morgan, 2008; but see Ballem & Plunkett, 2005; Yoshida, Fennell, Swingley, & Werker, 2009). For example, after training on a novel-object/novel-label pairing, 14-month-olds fail to notice a minimal phonetic change in the label (e.g., from *bih* to *dih*; Stager & Werker, 1997). However, in the same task, when habituated to an already familiar pairing (e.g., the label *dog* and a picture of a dog), 14-month-olds successfully detect the same phonetic change (e.g., from *dog* to *bog*; Fennell & Werker, 2003).

An alternative hypothesis for why item familiarity affects children's ability to detect phonetic contrasts is that children are inexperienced word learners, generally unfamiliar with the process of mapping labels to referents. As a result, they may have difficulty juggling the demands on their attention (from the referent, the label, the context, etc.), particularly when learning new words. Moreover, toddlers might not know which aspects of the phonetic form should be attended to during word processing because they lack knowledge of the phonemic distinctions in their language (Werker & Curtin, 2005). Familiarity with the label-referent mapping, the phonological form alone, or the visual form alone may reduce the difficulty of the task enough to enable young learners to detect phonetic differences (Fennell, 2012; Fennell & Werker, 2003; Stager & Werker, 1997; Werker & Curtin, 2005).

The discussion above highlights that current explanations for the lack of sensitivity to phonetic detail in some word processing tasks appeal to factors that are specific to young learners: immature phonological knowledge and inexperience with the process of mapping phonological forms to referents. However, what we know of adults' sensitivity to phonetic variation comes from their processing of familiar words. At present, it is simply assumed that adults would not show similar effects of word familiarity, as their knowledge of native language phonology would allow them to detect phonetic changes, even in relatively recently learned words, and to interpret single feature changes as novel words. If, instead, adults and young word learners show similar patterns of performance, this raises the possibility that these effects are driven by common mechanisms across development.

Lexical familiarity in adults

In adults, word familiarity (as indexed by frequency) affects the speed and accuracy of word processing (Grosjean, 1980; Marslen-Wilson, 1987, 1990). Frequency effects are therefore accommodated by most models of spoken word recognition (Gaskell & Marslen-Wilson, 1997; Luce & Pisoni, 1998; Marslen-Wilson, 1990; McClelland & Elman, 1986; Morgan, 2002). In addition, frequency affects the amount of input needed for recognition (Grosjean, 1980) and high frequency words experience less phonological competition than lower frequency words (Dahan et al., 2001; Goldinger, Luce, & Pisoni, 1989; Marslen-Wilson, 1987). While these studies demonstrate that word frequency influences the amount of phonological input needed to recognize a word and the degree to which other words become active (see also Magnuson et al., 2007), the effect of word frequency on adults' use of phonetic detail has received little attention. Furthermore, even the low frequency words in these studies are likely to be more familiar to adults than recently learned words are to young children. A legitimate comparison of familiarity effects in adults with those observed in children would require testing adults using extremely low frequency words. However, the difficulties of controlling for other lexical factors, such as lexical neighborhood density and phonotactic probabilities, make the use of real words problematic. Moreover, many of the studies showing a lack of phonetic sensitivity in infants and children involve words learned during the experimental session. Training adults on novel, artificial lexicons thus allows both for control over complicating lexical factors and for more direct comparison with developmental work.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Magnuson, Tanenhaus, Aslin, and Dahan (2003) demonstrated the utility of the artificial lexicon approach for studying lexical processing. Magnuson et al. used an artificial lexicon of word-object pairs, in order to strictly control word frequency (not possible in more naturalistic studies, which assume frequencies based on written or spoken corpora). Magnuson et al. trained subjects on a novel lexicon and monitored their eye movements in two sessions as they looked at visual displays containing these trained objects. They found that in the test session following the first half of training, subjects showed a large rhyme competitor effect. That is, subjects were more likely to fixate on an object whose name rhymed with the target than on objects with phonologically unrelated names. Intriguingly, the rhyme effect was larger than is typically observed in eye-tracking studies using familiar words (Allopenna et al., 1998; McMurray et al., 2003). Magnuson et al. suggested that weak lexical representations early in training (when words were relatively unfamiliar) led to failures to commit to lexical hypotheses and, consequently large rhyme effects. Another consequence of weak lexical representations might be an apparent reduction in sensitivity to phonetic detail. If word familiarity indeed affects phonetic sensitivity in adults, it is possible that young word learners' relative lack of sensitivity to phonetic mismatch results from the fact that they are, in general, less familiar with words. As Magnuson et al. (2003) argue, "This suggests the possibility that children's early lexical representations may not be fundamentally different from adults' … Evidence suggesting holistic processing may instead reflect weak lexical representations. This hypothesis makes strong predictions about the conditions under which apparent holistic processing should be observed, for example, when a word has recently been introduced to the lexicon or is infrequent" (page 224).

In the current studies, we test this hypothesis directly, using an artificial lexicon paradigm to ask whether adults show less use of phonetic detail in less familiar words. The use of an artificial lexicon has several advantages, allowing for strict control over such factors as word and visual familiarity. Furthermore, Magnuson et al. demonstrated that the processing of words from artificial lexicons is similar to the processing of real words, showing incremental processing, the importance of lexical factors like frequency, and competition effects. We first taught adults artificial lexicons in which the number of training presentations was manipulated. Then, after training, adults were presented with test displays containing one trained and one untrained object (analogous to displays used in some infant studies) and they heard either a label for the trained object, an onset mispronunciation of the trained object's label, or a label for the untrained object. In Experiment 1, single-feature (place, voicing) mispronunciations were presented; in Experiment 2, two-feature (place + voicing) mispronunciations were presented. Participants' eye movements were monitored as they selected an object. (Note that although we use the term "mispronunciations", because these labels differed from the training labels by a phonetic feature, we could also have referred to them as similar words, or phonological competitors.) The critical question was whether adults would correctly map mispronounced labels to the untrained object at all levels of familiarity. If, instead, adults' behavior parallels that of younger word learners and they fail to detect mispronunciations of less familiar words, this would suggest that such behavior is a signature of the lexical processing system throughout the lifespan.

Experiment 1

Participants were trained on an artificial lexicon of nonsense object-label pairings (as in Magnuson et al. (2003), both the phonological forms and the visual objects were novel). During training, the number of presentations of each item (object-label pairing) was manipulated. During testing, participants were shown visual displays containing one training (familiar) object and one unfamiliar object (not from the training set).

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

We used this test procedure because it is analogous to the presentation of one familiar and one unfamiliar object in recent word recognition studies with toddlers (Mani & Plunkett, 2011; White & Morgan, 2008). For example, in White and Morgan, 19-month-olds were presented with displays consisting of two objects: the familiar object was one with which the child had familiarity outside of the lab setting (both the object and label were familiar); the unfamiliar object was one with which the child was unfamiliar (both the object and label were unfamiliar). Using an analogous procedure will allow us to compare the findings from adults with those observed in younger learners.

Participants heard the label for the familiar object pronounced either correctly or with an onset mispronunciation, or heard a novel label (with no phonological relationship to the familiar label). They were instructed to point to the object named by the accompanying auditory stimulus; their eye-movements were monitored as they performed the task. Looking behavior to the two objects was measured to assess listeners' online interpretations of the labels. The use of familiar–unfamiliar object displays provides a means of testing listeners' tolerance for deviation. Even small (1-feature) phonetic differences should be interpreted as novel words; the present design, with an object whose label is unknown, allows us to test whether this is indeed the case.

Method

Participants—Twenty-four participants were included in the final analyses. All were monolingual, native speakers of English with no history of hearing or language deficits and normal or corrected-to-normal vision. The data from four additional participants were discarded because of equipment failure (2) or failure to complete two experimental sessions (2).

Apparatus—During the session, participants were fitted with an SMI Eyelink I head-mounted eye-tracker. A camera imaged the left eye at a rate of 250 Hz as the participant viewed the stimuli on a 15-in. ELO touch-sensitive monitor and responded to the spoken instructions. Stimuli were presented with Psyscript (Bates & Oliveiro, 2003).

Design—Each participant was trained on 48 object-label pairings. To facilitate learning, given the size of the lexicon, a unique subset of 12 items was trained and then tested in each of four distinct blocks that were completed over 2 days. In this way participants learned 24 items per day (12 per block, see procedure). Lexical familiarity was manipulated during training *within* each block. As a consequence of this design, participants only heard mispronunciations of trained labels after training was completed for that set. This was an important feature of the design; periodic testing on mispronunciations throughout training could have disrupted the learning process, causing interference and uncertainty about the phonological form of the labels.

Within each block of 12 items, sets of four items were presented either once, five times, or eight times in training. Thus, across the four training blocks, 16 items were presented at each of the three levels of exposure. These exposure levels were chosen based on pilot work and the results of Magnuson et al. (2003).[1]

Each test block contained 18 trials. In six trials, trained objects were labeled correctly (two items from each exposure level), in six trials there was a mispronunciation of trained object

[1]In Magnuson et al., participants reached accuracy levels of ~55% and ~75% after one and seven exposures with training items, respectively (accuracy for intermediate training values was not reported). The levels used for the present experiment were chosen to capture the period of greatest learning; it was essential that participants learn each item to some extent, but that low exposure items not be learned to ceiling levels of performance.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

labels (two items (one place mispronunciation and one voicing mispronunciation) from each exposure level), and in six trials novel, phonologically unrelated, labels were presented. Thus, across the four test blocks, there were eight items in each of the nine exposure by pronunciation conditions. The assignment of training items to pronunciation (correct, mispronounced) and exposure conditions (one, five, eight) was counterbalanced across participants.

Visual stimuli—Visual stimuli were geometric shapes generated in MATLAB by randomly filling in the squares of a 9 × 9 grid. Eighteen squares of the grid were filled in, all either horizontally or vertically contiguous (see Magnuson et al. (2003) for more details). Ninety-six shapes were selected for use as stimuli. Forty-eight shapes were used as training stimuli; the remaining 48 appeared as unfamiliar shapes during test trials. All shapes were presented on the computer monitor within a 3 × 3 grid. Each cell in the grid was approximately 2 × 2 in. Because the participants were seated approximately 18 in. from the monitor, each cell in the grid subtended about 6.4° of visual angle. Training stimuli were presented alone in the center of the grid.

Test displays contained one trained object and one unfamiliar object. These stimuli were presented to the left and right of the grid center (see Fig. 1). Within each test block, half of the time the familiar object was located left of center; the other half of the time it was located right of center.

Auditory stimuli—Auditory stimuli were monosyllabic non-words composed of English phonemes. Stimuli were constructed such that no stimulus was an onset or rhyme competitor of any other stimulus. Across training labels, novel labels, and mispronunciations (see below) this resulted in 110 CVC non-words and 10 CCVC non-words; non-words began with stops and fricatives (114), affricates (5) and liquids (1). The full list is given in Appendix A. The auditory stimuli were recorded by a female native speaker of English in a sound-treated room. Each non-word was read in isolation with sentence-final intonation. Forty-eight non-words were assigned to the training objects as correct labels and 24 non-words were assigned to the unfamiliar objects.

Two mispronounced versions of each training label were produced. None of these mispronunciations were onset or rhyme competitors of any original stimulus or any of the other mispronunciations. One version involved a change in the place of articulation in onset position; a second version involved a change in voicing in onset position. The intonation, pitch, and duration were matched as closely as possible across correct and mispronounced versions of the same label. For Experiment 1, voicing mispronunciations for half of the training labels were used and place mispronunciations for the other half of the training labels were used. Measures of position-specific phonotactic probability confirmed that correct and mispronounced versions of the labels did not differ in terms of segment and biphone probabilities (for 1st segment, $t(47) = .7$, *ns*; for 1st biphone, $t(47) = 1.2$, *ns*). The average durations of training, mispronounced, and novel labels were 686 ms (sd = 91, range 520–891), 698 ms (sd = 93, range 499–908), and 699 ms (sd = 87, range 547–909), respectively.

Auditory-visual pairings—Each participant received the same pairings, but there were six different stimulus lists (four participants were assigned to each list). The assignment of the training items to training exposure (one, five, eight) and test pronunciation conditions (correct, mispronounced) was counterbalanced across lists. Half of the training items had voicing changes when mispronounced; the other half of the training items had place changes when mispronounced. Pronunciation type was therefore not completely counterbalanced.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Procedure—Participants completed two 30-min sessions separated by at least one day. This separation was included to minimize potential interference across sessions. Each session contained two blocks of training and test. For each participant, the blocks were presented in a randomly determined order. The eye-tracker was calibrated prior to the first training block. During training, a single object appeared in the center of the grid and was accompanied by the auditory presentation of the object's label. Training was selfpaced; the participant pressed the touch-sensitive monitor when ready to proceed to the next trial. There were 56 training trials for each block, presented in random order.

After training, the subject completed two practice test trials with real objects. The sequence during practice and test trials was as follows: first, the two objects were displayed for one second. The objects then disappeared and a red square appeared in the center. The participant touched the red square, which triggered the disappearance of the square, the reappearance of the two objects, and the synchronized presentation of an audio file. The participant then selected the object he believed was being named by touching it on the screen. Once the participant selected one of the objects, the screen went blank and the next trial began. After the practice trials, the test trials began. In test trials, the audio file either labeled the familiar object in the display, was a mispronunciation of the familiar object's label, or was a completely novel label. There were 18 test trials for each block, presented in random order. Following a short break, the participant proceeded to the second training block and test.

Dependent measures—Two types of data were collected during the test trials. The first was the participants' behavioral response – the object selected. The second was looking behavior. For each trial, eye movements were recorded beginning when the objects were first displayed on the screen and ending when the participant selected an object. Eye movements were analyzed starting 200 ms after the onset of the auditory stimulus. Two regions of interest were defined: each contained the cell with the object as well as the surrounding 2° (included to avoid data loss when tracking accuracy was imperfect). Fixations of at least 100 ms were determined within these regions using a customized analysis program. The length of a fixation was defined as starting with the saccade that moved to that region and ending with the saccade that exited the region.

Results

Object choice—Fig. 2a displays the proportion of trials in which participants selected the familiar object minus the proportion of trials in which participants selected the novel object. In other words, this difference score reflects the bias to select the familiar object, i.e., how much **more** participants were considering the familiar object (we use this measure to maintain consistency with the dependent measure in the eye-tracking data described below). In the correct conditions, the familiar object is the appropriate response; in both the novel and mispronunciation conditions, the unfamiliar item is the appropriate response.

Participants' object selections were analyzed using a repeated measures ANOVA with two within-subjects factors (pronunciation type: three levels; exposure: three levels). The dependent measure for each pronunciation × exposure condition was the selection bias for the familiar object, described above. This analysis revealed significant main effects of pronunciation type, $F_1(2,46) = 180.8$, $p < .001$, $F_2(2,94) = 390.6$, $p < .001$ and of exposure $F_1(2,46) = 8.0$, $p < .001$, $F_2(2,94) = 6.0$, $p < .004$. More importantly, there was also a significant interaction between the two factors, $F_1(4,92) = 20.9$, $p < .001$, $F_2(4,188) = 23.9$, $p < .001$, indicating that the effect of exposure differed for the different pronunciation types. When testing session (blocks 1 and 2 vs. blocks 3 and 4) was added as a factor in the

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

analyses, there was no interaction involving session, demonstrating that participants exhibited the same patterns of performance on both days.

To test our primary question, whether familiarity affects the *differentiation* of close phonological variants, as it does in young word learners (Fennell & Werker, 2003; Stager & Werker, 1997), we examined whether familiarity affected our participants' ability to differentiate correct and mispronounced labels. In all three exposure conditions there were significant differences in the way participants responded to correct vs. mispronounced words (all *p's* < .001). However, this difference was larger at five exposures than it was at one exposure $t_1(23) = 3.5$, $p < .002$, $t_2(47) = 3.7$, $p < .001$ (there was no change in the difference between five and eight exposures, $t_1(23) = .1$, $p < .92$, $t_2(47) = .4$, $p < .68$). Thus, participants showed greater sensitivity to single feature differences for more familiar words.

Examining the effect of exposure for the correct and mispronunciation conditions separately revealed that the increasing differentiation of these labels was due to changes in the correct condition: the percentage bias for the familiar object increased significantly between one and five exposures, $t_1(23) = 4.8$, $p < .001$, $t_2(47) = 6.2$, $p < .001$ (there was no significant change in performance between five and eight exposures). In the mispronunciation condition, there was no significant decrease in the bias for the familiar object between one and five exposures, $t_1(23) = -1.4$, $p < .17$, $t_2(47) = -1.3$, $p < .19$, nor was there a difference between one and eight exposures, $t_1(23) = -1.7$, $p < .11$, $t_2(47) = -1.5$, $p < .13$.[2]

Looking behavior: proportion data—Object selection responses reflect the end-state of higher-level decision processes. In contrast, eye-movements provide more continuous information about the on-line interpretation of the auditory stimulus as it is mapped to the visual display. Fixations were analyzed over a window that began 200 ms post target onset (because it takes an average of about 180 ms to initiate a saccade to a target in response to linguistic input when the specific target is not known ahead of time but the possible locations of the target are known; Altmann & Kamide, 2004), until 1700 ms post target onset, the end of the time bin closest to the median response time of 1667 ms. We used the median response time, as mean response times varied considerably across conditions and exposures. Median reaction times for one, five, and eight exposures, respectively were: Correct: 1970 ms, 1178 ms, 931 ms; Mispronounced: 2307 ms, 1956 ms, 1818 ms; Novel: 2146 ms, 1565 ms, 1347 ms.

As described earlier, we had two primary regions of interest, each encompassing one of the objects in the display. Looks outside of these regions were uniformly low for all conditions (ranging from 5% to 7%), as confirmed by a repeated-measures ANOVA that found no differences across conditions and no interactions. Given that looks almost exclusively occurred in the regions of interest, we used a difference score as our measure of interest, as it simplifies the analysis and reflects the bias for looking at one object or the other. Fig. 3a plots the average difference in looking to the familiar object minus looking to the unfamiliar object (i.e., the familiar object bias) for the entire duration of this time window. (Appendix C, top half, provides raw looking proportions for Experiment 1.)

Looking behavior was analyzed using a repeated measures ANOVA with two within-subjects factors (pronunciation type: three levels; exposure: three levels). This analysis revealed significant main effects of pronunciation type, $F_1(2,46) = 79.8$, $p < .001$, $F_2(2,94) =$

[2]Considering the two types of mispronunciations separately, we found that the bias for the familiar object did not change significantly between one and five exposures for either place or voice mispronunciations ($t(23) = -1.52$, $p < .14$ and $t(23) = -.5$, $p < .62$, respectively). There was a significant decrease in the bias between one and eight exposures for place mispronunciations only ($t(23) = -2.24$, $p < .04$), indicating somewhat greater sensitivity to the specific place mispronunciations used.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

138.1, $p < .001$ and of exposure, $F_1(2,46) = 12.4$, $p < .001$, $F_2(2,94) = 7.4$, $p < .001$. There was also a significant interaction between the two factors, $F_1(4,92) = 10.97$, $p < .001$, $F_2(4,188) = 11.7$, $p < .001$.

As before, our primary question is whether familiarity affected our participants' ability to differentiate correct and mispronounced labels. Therefore, we examined this for each exposure level. At the one-exposure level, the correct and mispronunciation conditions were not significantly different from one another, $t_1(23) = .87$, $p < .39$, $t_2(47) = 1.1$, $p < .3$. At the five- and eight-exposure levels, they were (p's $< .001$). As with object choice above, an effect of familiarity on the use of phonetic detail was reflected by a significant increase in the difference between correct and mispronunciation trials from one to five exposures, $t_1(23) = 3.67$, $p < .001$, $t_2(47) = 4.0$, $p < .001$. There was no significant change between five and eight exposures, $t_1(23) = .85$, $p < .4$, $t_2(47) = .58$, $p < .57$.[3] The same pattern of exposure-based changes in sensitivity to the difference between correct and mispronounced labels held when we considered only trials in which participants selected the familiar object, demonstrating that the changes across exposure conditions did not simply reflect an averaging of trials with different selection profiles. Even for this subset of trials, participants' looking behavior showed differentiation of correct and mispronounced labels at higher levels of exposure, but not at the lowest level.

Thus, with additional exposure, participants were better able to differentiate correct and mispronounced labels. Examining the effect of exposure for the correct and mispronunciation conditions separately revealed that this was entirely due to changes in the correct condition: in the correct condition, the bias for the familiar object increased significantly with exposure, $F_1(2,46) = 48.4$, $p < .001$, $F_2(2,94) = 27.1$, $p < .001$, with a significant linear trend, $F_1(1,23) = 95.1$, $p < .001$, $F_2(1,47) = 45.65$, $p < .001$. There was no significant effect of exposure in the mispronunciation condition, $F_1(2,46) = .25$, $p < .78$, $F_2(2,94) = .22$, $p < .81$.

It has been argued that a signature of "immature" phonological sensitivity in toddlers is that they often continue to look at familiar target objects at above-chance levels when hearing single feature mispronunciations. To see whether our adult participants behaved the same way, we compared looking in the mispronunciation conditions to chance. Recall that our primary measure is a difference score (looks to familiar–looks to unfamiliar). Therefore, scores higher than zero indicate greater looking to the familiar object; looks less than zero indicate greater looking to the unfamiliar object. We found that for all three exposure conditions, adults spent significantly more time looking at the familiar object (1 exposure: $t_1(23) = 3.38$, $p < .003$, $t_2(47) = 4.03$, $p < .001$; 5 exposures: $t_1(23) = 4.01$, $p < .001$, $t_2(47) = 4.93$, $p < .001$; 8 exposures: $t_1(23) = 3.72$, $p < .001$, $t_2(47) = 4.05$, $p < .001$). Thus, despite knowing the names of the familiar objects (at least in the 5 and 8 exposure conditions), and the presence of a viable alternative referent onto which the mismatching label could be mapped, adults continued to look at the familiar object when they heard a mispronunciation. This was true for both types of mispronunciations individually (1 exposure place $t_1(23) = 2.72$, $p < .012$, $t_2(23) = 2.86$, $p < .009$; voice $t_1(23) = 2.29$, $p < .031$, $t_2(23) = 2.84$, $p < .009$; 5 exposure place $t_1(23) = 2.87$, $p < .009$, $t_2(23) = 3.67$, $p < .001$; voice $t_1(23) = 3.42$ $p <$.

[3]As with object choice, we considered whether the patterns described above held for both types of mispronunciations. Each type of mispronunciation individually became significantly more distinct from the correct condition between one and five exposures (place: ($t(23) = 2.9$, $p < .008$; voice: $t(23) = 3.13$, $p < .005$), but not between five and eight exposures (place: $t(23) = 1.79$, $p < .09$; voice: $t(23) < 1$). Comparing this (correct-mispronunciation) difference for the place and voice mispronunciations, the two mispronunciations had equivalent effects at one and five exposures ($t(23) = .4$, $p < .69$ and $t(23) = .11$, $p < .91$, respectively). However, the difference between correct and place mispronunciations was larger at eight exposures than the difference between correct and voice mispronunciations ($t(23) = 2.58$, $p < .02$). Thus, over time, participants became better able to detect place deviations than voicing deviations.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

002, $t_2(23) = 3.28$, $p < .003$; 8 exposure place $t_1(23) = 2.15$, $p < .042$, $t_2(23) = 1.39$, $p < .178$; voice $t_1(23) = 3.66$, $p < .001$, $t_2(23) = 4.79$, $p < .001$).

A striking aspect of the looking behavior is that in the 1-exposure condition, participants looked at the familiar object just as much when its label was mispronounced as when it was correctly pronounced. Although this suggests that participants had difficulty discriminating between the correct and mispronounced labels, it is possible that participants did initially preferred the unfamiliar object in this condition, but ultimately settled on the familiar object. This type of behavior could produce the similarity observed in overall looking for correct and mispronounced trials.

Fig. 4, which plots the familiar object preference over time, reveals that in one-exposure trials (top left panel) participants did *not*, in fact, prefer the unfamiliar object early in mispronunciation trials. Rather, from the earliest point at which the acoustic input can affect eye movements (approximately 200 ms after word onset), participants' looking behavior was very similar in correct and mispronunciation trials, suggesting that participants initially processed these labels similarly. In contrast, looking on novel trials differed from the other two conditions almost immediately in response to hearing the mismatching acoustic input. Five and eight exposure trials (middle and bottom left panels, respectively) show a different pattern: participants differentiated mispronounced and correct labels early in the trial.

Summary of Experiment 1—Results of Experiment 1 suggest that lexical familiarity, as instantiated by frequency, is an important factor affecting adults' use of phonetic detail. This was evidenced by the fact that, in both object selection and in looking proportions, the difference between responses to correct vs. mispronounced labels increased between one and five exposures. Most importantly, eye movements revealed that adults were relatively insensitive to mispronunciations of the least familiar words. These findings parallel the results of a number of developmental studies (e.g., Werker et al., 2002) in which toddlers show insensitivity to phonetic contrasts in word learning contexts. Previously, such developmental findings have been explained by positing that young learners have immature (e.g., holistic) lexical representations or are overwhelmed by the demands of the mapping problem because they are inexperienced word learners. However, in the current work, experienced (adult) word learners displayed a pattern similar to that found in young children. An additional finding was that, even at higher levels of exposure, adults looked more to familiar objects in response to mispronunciations, even though mispronunciations should have been mapped to the unfamiliar referents.

In Experiment 2, we asked whether adults would similarly map labels with larger changes in onset position to the familiar object. If so, this would suggest that at these levels of familiarity, the onset's role in lexical discrimination is quite limited. If, instead, adults show graded sensitivity – by looking more at the unfamiliar object than they did in Experiment 1 – this would demonstrate that the onset consonant does play a role in processing. Thus, the goal of Experiment 2 was to determine adults' sensitivity to and interpretation of larger changes in onset position.

Experiment 2

The results of Experiment 1 reveal that when adults are relatively unfamiliar with mappings between words and referents, they often map single-feature mispronunciations of these words onto the same referents, despite the existence of alternative, novel referents onto which these mispronunciations should be mapped. This was demonstrated both by the looking data, in which looks to the familiar and novel objects were similar for correct and mispronounced words, and by the object selection data, in which adults were at chance in

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

choosing between the familiar and novel objects when hearing mispronunciations. These results suggest that adults have some difficulty using phonetic detail in newly learned words. It is less clear what is encoded about the phonological forms of these words. In Experiment 2, participants were presented with two-feature mispronunciations in initial position. If adults exhibit more sensitivity to these larger mispronunciations, this would suggest that they encode fine detail about newly learned words. In this case, participants should be less likely to select and look at trained objects when the labels are mispronounced by two features (Experiment 2) than by one feature (Experiment 1). If, instead, the encoding is coarser, or the onset plays little role in processing at this stage, responses should be similar across the two experiments.

Method

Participants—Twenty-four participants were included in the data analyses. All were monolingual, native speakers of English with no history of hearing or language deficits and normal or corrected-to-normal vision. The data from three additional participants were discarded because of failure to complete two experimental sessions (1) or failure to successfully complete calibration (2).

Design—The design was the same as in Experiment 1, except that all mispronunciations were two-feature (place + voicing) mispronunciations in onset position.

Visual stimuli—The same 96 objects were used as stimuli, with the same assignment of objects to the training and unfamiliar stimuli sets. In addition, the same test displays (pairings of familiar and unfamiliar objects) were used.

Auditory stimuli—Training labels and mispronunciations were drawn from the set of place and voicing mispronunciations in Experiment 1 in such a way that the training and mispronounced versions in Experiment 2 differed by two phonetic features. For example, for the Experiment 1 training label /tib/, the voicing mispronunciation was /dib/ and the place mispronunciation (not used) was /kib/. In Experiment 2, /dib/ was chosen as the training label and the mispronounced version was /kib/. The average lengths of the training and mispronounced items for this experiment were 696 ms (sd = 101, range 416–921) and 694 ms (sd = 97, range 499–908), respectively. The same novel labels were used as in Experiment 1. Appendix B includes the list of auditory stimuli.

Auditory-visual pairings—The pairings were the same as in Experiment 1. For example, the object associated with the test stimulus /tib/ in Experiment 1 was associated with the test stimulus /dib/ in Experiment 2.

Results

Object choice—Fig. 2b displays the proportion of trials in which participants selected the familiar object minus the proportion of trials in which participants selected the novel object. The pattern of results for the correct and novel trials is very similar to that observed in the object choice data in Experiment 1. In contrast, results for the mispronunciation trials were markedly different across the two experiments.

Object selections were analyzed using a repeated measures ANOVA with two within-subjects factors (pronunciation type: three levels; exposure: three levels). The dependent measure was the bias for the familiar object (familiar object–unfamiliar object). As in Experiment 1, this analysis revealed significant main effects of pronunciation type, $F_1(2,46) = 205.7$, $p < .001$, $F_2(2,94) = 471.3$, $p < .001$, and of exposure, $F_1(2,46) = 7.7$, $p < .005$, $F_2(2,94) = 7.8$, $p < .001$. There was also a significant interaction between the two factors,

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

$F_1(4,92) = 30.7$, $p < .001$, $F_2(4,188) = 24.2$, $p < .001$, indicating that the effect of exposure differed for the different pronunciation types.

Our primary question, as in Experiment 1, was whether amount of exposure affects adults' ability to differentiate correct and mispronounced labels. In all three exposure conditions there were significant differences in the way participants responded to correct vs. mispronounced words (all *ps* < .001). However, this difference increased between one and five exposures $t_1(23) = 8.8$, $p < .001$, $t_2(47) = 5.9$, $p < .001$. Thus, as in Experiment 1, adults' ability to differentiate mispronunciations from correct pronunciations increased with greater familiarity. There was no difference between five and eight exposures, $t_1(23) = 1.3$, $p < .22$, $t_2(47) = 0.7$, $p < .48$.

Examining the effect of exposure for the correct and mispronunciations revealed that, in contrast to Experiment 1, the increasing difference in response to these labels was due to changes in both the correct and mispronunciation conditions: on correct trials, the bias for the familiar object increased significantly between one and five exposures, $t_1(23) = 9.4$, $p < .001$, $t_2(47) = 5.8$, $p < 001$. On mispronunciation trials, there was a significant *decrease* in the bias for choosing the familiar object between one and five exposures, $t_1(23) = -5.0$, $p < .001$, $t_2(47) = -4.1$, $p < .001$. There were no changes for any of the pronunciation types between five and eight exposures.

Looking behavior: proportion data—Fixations were analyzed over a window that began 200 ms post target onset and ended at 1700 ms, the end of the time bin closest to the median response time of 1670 ms. Median reaction times for one, five, and eight-exposures, respectively were: Correct: 2328 ms, 1031 ms, 1012 ms; Mispronounced: 2787 ms, 2068 ms, 1667 ms; Novel: 2419 ms, 1524 ms, 1340 ms. As in Experiment 1, there were very few fixations outside of the two regions of interest and these were uniformly distributed across conditions, as confirmed by a repeated-measures ANOVA that found no differences across conditions and no interactions. Fig. 3b plots the average difference in looking to the familiar object minus looking to the unfamiliar object for the duration of this time window. Again, this difference score reflects the tendency to fixate on the familiar object over the unfamiliar object given the acoustic input. (Appendix C, bottom half, provides raw looking proportions for Experiment 2.)

Looking behavior was analyzed using a repeated measures ANOVA with two within-subjects factors (pronunciation type: three levels; exposure: three levels). This analysis revealed significant main effects of pronunciation type, $F_1(2,46) = 38.4$, $p < .001$, $F_2(2,94) = 97.4$, $p < .001$, and of exposure, $F_1(2,46) = 4$, $p < .03$, $F_2(2,94) = 23.67$, $p < .001$. There was a significant interaction between the two factors, $F_1(4,92) = 12.4$, $p < .001$, $F_2(4,188) = 13.13$, $p < .001$.

We again explored our primary question of interest by comparing looks in the correct and mispronunciation conditions at each level of exposure. At the one-exposure level, in contrast to Experiment 1, the correct and mispronunciation conditions differed from one another, $t_1(23) = 2.6$, $p < .016$, $t_2(47) = 2.0$, $p < .051$. They also differed from one another at the five- and eight-exposure levels (*ps* < .001). There was a significant increase in the difference between correct and mispronunciation trials between one and five exposures, $t_1(23) = 3.7$, $p < .001$, $t_2(47) = 4.58$, $p < .001$. There was no significant change between five and eight exposures $t_1(23) = 1.0$, $p < .32$, $t_2(47) = 1.02$, $p < .32$. The same patterns held when we considered only trials in which participants selected the familiar object, demonstrating that the changes across exposure conditions did not simply reflect an averaging of trials with different selection profiles. The fact that looking behavior differed for correct and

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

mispronunciation trials in the one-exposure condition of Experiment 2, but not Experiment 1, suggests that adults were more sensitive to the larger deviations of Experiment 2.

Examining the effect of exposure for the correct and mispronunciation conditions separately revealed that, as in Experiment 1, the increasing differentiation across exposure levels was due to changes in the correct condition only: there was a significant effect of exposure for the correct condition $F_1(2,46) = 23.2$, $p < .001$, $F_2(2,94) = 24.17$, $p < .001$, with a significant linear trend, $F_1(1,23) = 48.8$, $p < .001$, $F_2(1,47) = 40.37$, $p < .001$. There was no significant effect of exposure in the mispronunciation condition, $F_1(2,46) = .6$, $p < .56$, $F_2(2,94) = .6$, $p < .55$.

We also compared looking behavior in the mispronunciation conditions to chance (equal looking to the familiar and unfamiliar objects). We found that, as in Experiment 1, adults looked at the familiar object at above chance levels in the one-exposure condition ($t_1(23) = 3.36$, $p < .003$, $t_2(47) = 2.86$, $p < .006$). But in contrast to the single-feature mispronunciations, looking behavior at higher exposure levels did not differ from chance (5 exposures t_1 (23) = 1.03, $p < .31$, $t_2(47) = .97$, $p < .34$; 8 exposures $t_1(23) = 1.62$, $p < .12$, $t_2(47) = 1.91$, $p < .06$).

As discussed with respect to the 1-feature mispronunciations of Experiment 1, it is possible that participants' preference changed from the unfamiliar object to the familiar object over the course of the trial, and that this produced chance looking behavior for mispronunciations when averaged across the entire trial. However, Fig. 4 (top right panel) reveals that participants' preference remained relatively constant (and near zero) across the trial.

Summary of Experiment 2—These results support those of Experiment 1 in showing that lexical familiarity, as instantiated by frequency of occurrence, influences adults' use of phonetic detail. As in Experiment 1, this was demonstrated by the fact that, in both object selection and in looking proportions, there was greater differentiation of correct vs. mispronounced labels at five exposures than at one exposure. However, there were also some indications that participants were more sensitive to these two-feature changes than the one-feature changes of Experiment 1: object choice differed more in the correct and mispronunciation conditions of Experiment 2 than in Experiment 1 ($t_1(46) = 1.91$, $p < .062$; $t_2(94) = 5.23$, $p < .001$). Further, looking proportions differed more in the correct and mispronunciation conditions of Experiment 2 than in Experiment 1, although this comparison was not statistically significant ($t_1(46) = 1.2$, $p < .23$, $t_2(94) = 1.68$, $p < .097$). This cross-experiment difference was a result of differences in the mispronunciation conditions: overall, there were fewer looks to the familiar object for the two-feature changes of Experiment 2 than for the one-feature changes of Experiment 1 ($t_1(46) = 2.23$, $p < .03$, $t_2(94) = 2.73$, $p < .008$). Similarly, there were fewer choices of the familiar object for mispronunciations in Experiment 2 than in Experiment 1.

General discussion

The goal of the present study was to investigate the effects of lexical familiarity on adults' sensitivity to and use of phonetic detail. In two experiments, adults were trained on a novel lexicon of word-object associations in which items were presented once, five, or eight times during training. At test, participants saw pairs of trained and unfamiliar objects and heard a label which was either a correct pronunciation of the trained object's label, a mispronunciation of the trained object's label, or a completely novel label. In Experiment 1, adults heard single-feature (voicing or place) onset mispronunciations. In Experiment 2, adults heard two-feature (voicing + place) onset mispronunciations. Our primary finding was that when items had been experienced only once in training, adults' looking behavior

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

did not differentiate mispronunciations from correct pronunciations. Indeed, even two-feature mispronunciations were mapped to the familiar object more often than chance after a single exposure. This bias is notable because there was a viable alternative referent onto which the mispronunciation could (and should) have been mapped. These findings – the apparent lack of sensitivity to single-feature mispronunciations in the eye movement data, as well as the familiar object bias – are strikingly parallel to behavior observed in younger learners (Stager & Werker, 1997). Further, even with more exposure, adults continued to look at the familiar object at greater than chance levels when they heard single-feature mispronunciations. Thus the present results suggest that, even with a fully developed phonological system, mismatching phonetic detail does not always prevent adults from treating mispronunciations (i.e., phonological competitors) as known words.

At the same time, our results are compatible with participants having *encoded* acoustic–phonetic detail even about the onsets of labels they had heard only once. Although neither effect was significant, in both object choice and looking behavior, there was a larger preference for the familiar object with one-feature changes than two-feature changes.

One notable difference between the two experiments is the nature of the familiarity effect. In Experiment 1, the primary change with increasing familiarity was the response (both in eye movements and selection) to correct pronunciations, whereas the relative proportion of responses to the two objects remained constant for mispronunciation trials. (This pattern is consistent with that reported by Mayor and Plunkett (submitted for publication) in their reanalysis of toddler data from Bailey & Plunkett, 2002: older toddlers showed more of a differentiation between correct and mispronounced labels than younger toddlers – but this was due to an increase in the response to correct pronunciations in the older age group, not a decrease in the response to mispronunciations.) The lack of change in the mispronunciation trials of Experiment 1 might suggest that increasing familiarity does *not*, in fact, increase sensitivity to phonetic mismatch – for if it did, one might have expected the familiar object preference to decrease for mispronunciations at higher levels of exposure. However, in Experiment 2, selection of the familiar object in mispronunciation trials *did* decrease as familiarity increased, demonstrating that for larger degrees of phonetic difference, familiarity does increase sensitivity to mismatch. Thus, although we must remain somewhat cautious in concluding that familiarity causes a general increase in sensitivity to phonetic detail (regardless of degree of mismatch), it is nevertheless clear that (1) when the phonetic difference is sufficiently large, familiarity does increase sensitivity to mismatch, and (2) when the same metric that has been applied to infants is applied, adults, like infants, appear relatively insensitive to phonetic detail at low levels of familiarity.

These findings raise questions about the mechanisms by which word familiarity affects phonetic sensitivity and the nature of lexical representation over development. In the Introduction, we reviewed studies showing a lack of sensitivity to phonetic contrasts in the early stages of word learning in infants and children. These studies have been advanced to support arguments that either the format of lexical representations undergoes significant change into childhood, or that inexperienced word learners cannot attend to phonetic detail. However, the present results demonstrate that adults similarly show reduced use of phonetic detail when words are not familiar. These effects in adults cannot be attributed to immaturity in the representational system. Nor can they be attributed to a general lack of experience with the process of mapping phonological forms to referents.

The parallels between the effects of familiarity in infants and adults raise the possibility that a common mechanism underlies the effects in both populations. Some possible explanations for familiarity effects that are consistent with this claim are described below.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Information processing overload

One possibility is that learning new words is just a demanding task, for anybody. During a word-learning situation, learners must juggle multiple sources of information – different tokens of the word to be learned, the phonological properties of the word, and properties of the referent. It may be that, whenever the processing system is stressed, responses are based on more global matches (in either the visual referent or phonological label) because it is difficult to focus on detail. Consider the task faced by the adults in the present studies: the visual referentswere geometric shapes, all fairly similar to one another, the labelswere allmonosyllabic, and participantswere asked to learn 12 items in each block. Word familiarity, then, might alleviate processing load, allowing attention to be directed to phonetic detail. This explanation has been proposed for young word learners (Fennell, 2012; Fennell & Werker, 2003), as it is primarily in referential tasks in which there are mappings between labels and referents that infants and children fail to demonstrate sensitivity to phonetic detail in newly learned words.

Confidence

Another possibility is that the apparent reduced sensitivity to single-segment changes in less familiar words arises because of a lack of confidence in the fidelity of the encoding. Thus, although the mismatch is detected, altered pronunciations are accepted because of this uncertainty. This argument has been made in a study with toddlers (Yoshida et al., 2009) and would also be consistent with models where frequency effects are instantiated at the level of a decision bias (e.g., Luce & Pisoni, 1998). In its simplest form, however, a (lack of) confidence hypothesis would predict greater competition early in learning irrespective of the position (e.g., onset or rhyme) of the change. Yet, Magnuson et al.'s (2003) artificial lexicon study found that while rhyme competition was largest early in learning and decreased with training – a pattern compatible with increased confidence – cohort competition did not decrease with learning and, in fact, was larger after more training. Hence, the position of the change appears to interact with the time course of spoken word recognition, making rhyme competitors particularly strong competitors early in learning (and onset changes hard to detect). On a simple confidence account, it is not clear why such a positional asymmetry would exist. However, if confidence affects not only whether altered pronunciations are accepted, but also the degree to which a correct pronunciation activates the stored form itself, uncertainty could lead to weak activation of a stored form even if a correct pronunciation is heard. As we describe next, weakly activated lexical representations can account for the patterns that have been observed both in the current work, and in related studies.

Weak lexical representations

Another possibility is that effects of lexical familiarity arise from the architecture of the lexical processing system. Models of spoken and visual word recognition posit that less familiar words are represented weakly in memory – either because they have low resting activation levels (Marslen-Wilson, 1990; McClelland & Elman, 1986; McClelland & Rumelhart, 1981), are instantiated by weak lexical attractors (Morgan, 2002), or have weak connections (Gaskell & Marslen-Wilson, 1997).

Reduced sensitivity to onset mismatches in weakly represented words could be explained by the time course of spoken word recognition. If a word is weakly represented, a greater amount of acoustic input will be processed before the lexical representation is activated enough to be recognized (consistent with observations that low frequency words take longer to identify than high frequency words; Grosjean, 1980; Van Petten, Coulson, Rubin, Plante, & Parks, 1999). As a result of the additional input required to recognize weakly represented words, the overall similarity between a phonologically similar mispronunciation and the

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

stored form might exert a strong influence on processing (e.g., if, to access the stored form *carrot*, the entire word form must be heard, then its overall similarity to the acoustic input "*parrot*" would be high). In contrast, for strongly represented words, less acoustic input is needed to recognize the word, increasing the relative importance of a word's onset. In this case, listeners should be more disrupted by initial mismatches, even if the overall similarity between the mispronunciation and the stored form is still high (e.g., if the stored form *carrot* can be accessed via hearing only "*car*", the relative importance of the onset is increased, making a mismatch between that onset and the onset of "*parrot*", more disruptive). This is consistent with the results of Magnuson et al. (2003), who observed that rhyme competition is stronger among less familiar words than among more familiar words.

The idea that the relative importance of a word's onset is reduced when lexical representations are weak is also consistent with an explanation proposed for the lexical processing deficits observed in a different population: Broca's aphasics. Milberg et al. (1988) have hypothesized that Broca's aphasics' lexical processing deficits are due to abnormally low levels of lexical activation. And one possibility is that this produces a reduction in the relative importance of a word's onset (as described above), leading to the abnormally large rhyme competitor effects observed in this population (Yee, 2005). It is also consistent with the observation that adults show greater activation of rhyme competitors in their L2 than in their (presumably more strongly represented) native language (Ruschemeyer, Nojack, & Limbach, 2008). Older listeners, who require more input to identify words (Craig, 1992), also show a similar pattern (Ben-David et al., 2011).

In addition to requiring additional input for recognition, weakly represented words may also be less able to inhibit lexical competitors (i.e., may exert less lateral inhibition), again increasing the influence of rhyme competitors. This would be consistent with TRACE simulations (Mayor & Plunkett, submitted for publication), which suggest that the graded sensitivity to mispronunciations of familiar words exhibited by toddlers (White & Morgan, 2008) emerges only when lateral inhibition among lexical competitors is turned off. In our data, we also observed a graded pattern, in that larger deviations had a greater effect on behavior. This suggests that rather than being a binary parameter that is "switched on" as the vocabulary grows during childhood, some other mechanism may influence the strength of lateral inhibition in the lexicon. For instance, lateral inhibition from weakly represented (or weakly activated) words may always be weak (e.g., there may be a threshold of activation that a word must reach before it is able to inhibit lexical competitors), or inhibition may be globally up- or down-regulated depending upon the familiarity or clarity of the input. An intriguing suggestion that lateral inhibition *can* be dynamically modulated comes from evidence that when uncertainty is increased, through the addition of noise in non-target words (therefore, not comprising the clarity of the targets), young adults show increased looks to rhyme competitors (i.e., less evidence of lateral inhibition; McQueen & Huettig, 2012).

Findings from disparate populations (Broca's aphasics, older listeners, and normal adults learning new words) thus all converge to suggest that when lexical activation is relatively weak, rhyme competition increases. Increased rhyme competition, in turn, may lead to the appearance of reduced phonetic sensitivity. Thus, reduced phonetic sensitivity early in development might be due, not to a restructuring of lexical representations, but to a combination of weak lexical representations and reduced lateral inhibition, both of which cause the entire word to exert more of an influence on processing.

Are children's representations qualitatively different?

We started with the observation that children's lack of sensitivity to phonetic changes, particularly in less familiar words, has been used to argue for a restructuring in the format of

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

lexical representations over development. An additional observation is that toddlers sometimes continue to look at target pictures at above chance levels when hearing single-feature mispronunciations, even when there is another possible referent available (Mani & Plunkett, 2011; Swingley & Aslin, 2000, 2002; White & Morgan, 2008; and in simulations of toddler performance by Mayor & Plunkett, submitted for publication). Some researchers have claimed that this pattern of performance by young learners reflects a failure to interpret phonetic detail in line with the phonological system of the native language (Ramon-Casas, Swingley, Sebastian-Galles, & Bosch, 2009; Swingley & Aslin, 2007).

Thus, in addition to asking whether adults detect phonetic deviations in newly learned words, we also asked what their behavior indicates about how they interpret these deviations. We found that, for both types of mispronunciations at low levels of familiarity, and for single-feature mispronunciations even with more training, adults looked significantly more at the familiar object than the untrained object. This is despite the fact that the phonology of English (where these sorts of changes should signal lexical contrast) indicates that these pronunciations should be interpreted as novel words and mapped onto the untrained objects. The fact that adults' behavior with newly learned words parallels toddlers' raises questions regarding assumptions about what a mature response looks like in these tasks. Rather than being driven by interpretation failures, this type of familiarity bias could also be interpreted as normal lexical (i.e., rhyme) competition.[4]

Conclusion

The present findings suggest that effects of lexical familiarity on phonetic sensitivity – which in the past have been attributed to immaturities in young children's lexical representation and processing – reflect mechanisms that operate throughout development. Parallel findings from adults and infants suggest that the architecture underlying spoken word representation and processing may be constant across development – in other words, the differences between the populations are differences of degree, not kind. If this claim is correct, it has both methodological and theoretical implications for the conduct of research on spoken word recognition. Methodologically, assumptions of continuity mean that questions that are difficult to pursue in one population may be informed by study of the other population. From a theoretical point of view, this means that, rather than existing independently, accounts of adult processing must be constrained by developmental considerations and that accounts of language development must be constrained by observations of processing in adults.

Acknowledgments

This research was supported by NIH Grant F 31 DC 007541-01 to K.S.W., Brown University research funds to S.E.B., and NSF IGERT training Grant 9870676 to the Department of Cognitive & Linguistic Sciences at Brown University. We thank Andrew Duchon for help with analysis, James Magnuson for providing visual stimuli templates, and Cara Misiurski for helpful input in the experimental design.

[4]The above discussion brings up an interesting point: greater than chance looking at an object in the intermodal preferential looking paradigm is typically thought to indicate that a child has interpreted that object as a referent for the label. However, this may not always be the case. In our data, we found that adults looked significantly more to a trained object (than an untrained competitor) when they heard a one-feature mispronunciation, even at higher levels of exposure. Despite this, at higher level of exposure, adults did not show a bias for selecting the familiar object (they were at chance in selecting between the two objects). Thus, higher looking at the familiar object did not always translate into selection of that object. This dissociation suggests that it will be important to consider the role of processing factors (such as rhyme effects) when interpreting looking behavior, in both infants and adults.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

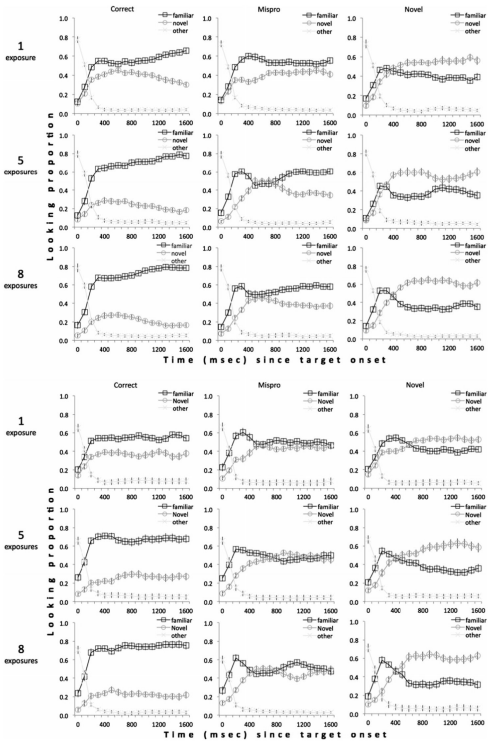
A: Experiment 1 auditory stimuli

(Note that novel labels are listed in arbitrary order, as the pairing of novel and training labels differed across participant groups. p = place mispronunciations, v = voicing mispronunciations.)

B: Experiment 2 auditory stimuli

(Note that novel labels are listed in arbitrary order, as the pairing of novel and training labels differed across participant groups.)

c: Raw looking proportions (Experiments 1 and 2)



References

Allopenna PD, Magnuson JS, Tanenhaus MK. Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. Journal of Memory and Language. 1998; 38(4):419–439.

Altmann, GTM.; Kamide, Y. Now you see it, now you don't: Mediating the mapping between language and the visual world. In: Henderson, J.; Ferreira, F., editors. The interface of language, vision, and action: Eye movements and the visual world. Psychology Press; 2004. p. 347-386.

Anderson JL, Morgan JL, White KS. A statistical basis for speech sound discrimination. Language and Speech. 2003; 46(2–3):155–182. [PubMed: 14748443]

Andruski JE, Blumstein SE, Burton M. The effect of subphonetic differences on lexical access. Cognition. 1994; 52:163–187. [PubMed: 7956004]

Bailey TD, Plunkett K. Phonological specificity in early words. Cognitive Development. 2002; 17(2): 1265–1282.

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

Ballem KD, Plunkett K. Phonological specificity in children at 1;2. Journal of Child Language. 2005; 32:159–173. [PubMed: 15779881]

Barton D. Phonemic discrimination and the knowledge of words in children under 3 years. Papers and Reports on Child Language Development. 1976; 11:61–68.

Barton, D. Phonemic perception in children. In: Yeni-Komshian, G.; Kavanaugh, J.; Ferguson, C., editors. Child phonology: Perception. Vol. Vol. 2. Academic Press; New York: 1980. p. 97-116.

Bates TC, Oliveiro L. Psyscript: A Macintosh application for scripting experiments. Behavior Research Methods, Instruments, & Computers. 2003; 35(4):565–576.

Ben-David BM, Chambers CG, Daneman M, Pichora-Fuller MK, Reingold EM, Schneider BA. Effects of aging and noise on real-time spoken word recognition: Evidence from eye movements. Journal of Speech, Language, and Hearing Research. 2011; 54:243–262.

Connine CM, Titone D, Deelman T, Blasko D. Similarity mapping in spoken word recognition. Journal of Memory and Language. 1997; 37:463–480.

Craig CH. Effects of aging on time-gated isolated word-recognition performance. Journal of Speech and Hearing Research. 1992; 35:234–238. [PubMed: 1735973]

Dahan D, Magnuson JS, Tanenhaus MK. Time course of frequency effects in spoken-word recognition: Evidence from eye movements. Cognitive Psychology. 2001; 42:317–367. [PubMed: 11368527]

Eilers RE, Oller DK. The role of speech discrimination in developmental sound substitutions. Journal of Child Language. 1976; 3:319–329.

Eimas P. Auditory and linguistic processing of cues for place of articulation by infants. Perception and Psychophysics. 1974; 16:513–521.

Eimas PD, Siqueland ER, Jusczyk P, Vigorito J. Speech perception in infants. Science. 1971; 171:303–306. [PubMed: 5538846]

Fennell CT. Object familiarity enhances infants' use of phonetic detail in novel words. Infancy. 2012; 17:339–353.

Fennell CT, Werker JF. Early word learners' ability to access phonetic detail in well-known words. Language and Speech. 2003; 46(2):245–264. [PubMed: 14748446]

Fowler, AE. How early phonological development might set the stage for phoneme awareness. In: Brady, SA.; Shankweiler, DP., editors. Phonological processes in literacy: A tribute to Isabelle Y. Liberman. Lawrence Erlbaum; Mahwah, NJ: 1991. p. 97-117.

Garlock VM, Walley AC, Metsala JL. Age-of-acquisition, word frequency, and neighborhood density effects on spoken word recognition by children and adults. Journal of Memory and Language. 2001; 45:468–492.

Garnica, OK. The development of phonemic speech perception. In: Moore, TE., editor. Cognitive development and the acquisition of language. Academic Press; New York: 1973. p. 215-222.

Gaskell MG, Marslen-Wilson WD. Integrating form and meaning: A distributed model of speech perception. Language and Cognitive Processes. 1997; 12:631–656.

Goldinger SD, Luce PA, Pisoni DB. Priming lexical neighbors of spoken words: Effects of competition and inhibition. Journal of Memory and Language. 1989; 28(5):501–518.

Grosjean F. Spoken word recognition and the gating paradigm. Perception & Psychophysics. 1980; 28:267–283. [PubMed: 7465310]

Jusczyk PW, Aslin RN. Infants' detection of sound patterns of words in fluent speech. Cognitive Psychology. 1995; 29(1):1–23. [PubMed: 7641524]

Kay-Raining Bird E, Chapman RS. Partial representation and phonological selectivity in the comprehension of 13- to 16-month olds. First Language. 1998; 18:105–127.

Kuhl PK, Williams KA, Lacerda F, Stevens KN, Lindblom B. Linguistic experience alters phonetic perception in infants by 6 months of age. Science. 1992; 255:606–608. [PubMed: 1736364]

Luce PA, Pisoni DB. Recognizing spoken words: The neighborhood activation model. Ear and Hearing. 1998; 19:1–36. [PubMed: 9504270]

Magnuson JS, Tanenhaus MK, Aslin RN, Dahan D. The time course of spoken word recognition and learning: Studies with artificial lexicons. Journal of Experimental Psychology: General. 2003; 132(2):202–227. [PubMed: 12825637]

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

Magnuson JS, Dixon JA, Tanenhaus MK, Aslin RN. The dynamics of lexical competition during spoken word recognition. Cognitive Science. 2007; 31:1–24. [PubMed: 21635286]

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

Mani N, Plunkett K. Does size matter? Subsegmental cues to vowel mispronunciation detection. Journal of Child Language. 2011; 38:606–627.

Marslen-Wilson WD. Functional parallelism in spoken word recognition. Cognition. 1987; 25:71–102. [PubMed: 3581730]

Marslen-Wilson, W. Activation, competition, and frequency in lexical access. In: Altmann, GTM., editor. Cognitive models of speech processing: Psycholinguistic and computational perspectives. MIT Press.; Cambridge, MA: 1990. p. 148-172.

Mayor, J.; Plunkett, K. Infant word recognition: Insights from TRACE simulations. submitted for publication

McClelland JL, Rumelhart DE. An interactive activation model of context effects in letter perception: I. An account of basic findings. Psychological Review. 1981; 88:375–407.

McClelland JL, Elman JL. The TRACE model of speech perception. Cognitive Psychology. 1986; 18(1):1–86. [PubMed: 3753912]

McMurray B, Tanenhaus MK, Aslin RN, Spivey MJ. Probabilistic constraint satisfaction at the lexical/phonetic interface: Evidence for gradient effects of within-category VOT on lexical access. Journal of Psycholinguistic Research. 2003; 32(1):77–97. [PubMed: 12647564]

McQueen JM, Huettig F. Changing only the probability that spoken words will be distorted changes how they are recognized. Journal of the Acoustical Society of America. 2012; 131:509–517. [PubMed: 22280612]

Metsala, JL.; Walley, AC. Spoken vocabulary growth and the segmental restructuring of lexical representations: Precursors to phonemic awareness and early reading ability. In: Metsala, JL.; Ehri, LC., editors. Word recognition in beginning literacy. Lawrence Erlbaum Associates; Mahwah, NJ: 1998. p. 89-120.

Milberg W, Blumstein SE, Dworetzky B. Phonological factors in lexical access: Evidence from an auditory lexical decision task. Bulletin of the Psychonomic Society. 1988; 26:305–308.

Miller JL, Eimas PD. Studies on the categorization of speech by infants. Cognition. 1983; 13:135–165. [PubMed: 6682742]

Morgan, JL. Word recognition and phonetic structure acquisition: Some possible relations. Paper presented at the meeting of acoustical society of America; Pittsburgh, PA. 2002.

Ramon-Casas M, Swingley D, Sebastian-Galles N, Bosch L. Vowel categorization during word recognition in bilingual toddlers. Cognitive Psychology. 2009; 59:96–121. [PubMed: 19338984]

Ruschemeyer S-A, Nojack A, Limbach M. A mouse with a roof? Effects of phonological neighbors on processing of words in sentences in a non-native language. Brain and Language. 2008; 104:132–144. [PubMed: 17391746]

Schvachkin, NK. The development of phonemic speech perception in early childhood. In: Ferguson, C.; Slobin, D., editors. Studies of child language development. Holt, Rinehart, and Winston; New York: 1973. p. 91-127.(originally published in 1948)

Stager CL, Werker JF. Infants listen for more phonetic detail in speech perception than in word learning tasks. Nature. 1997; 388(6640):381–382. [PubMed: 9237755]

Storkel HL. Restructuring similarity neighbourhoods in the developing mental lexicon. Journal of Child Language. 2002; 29:251–274. [PubMed: 12109371]

Swingley D, Aslin RN. Spoken word recognition and lexical representation in very young children. Cognition. 2000; 76:147–166. [PubMed: 10856741]

Swingley D, Aslin RN. Lexical neighbourhoods and the word-form representations of 14-month-olds. Psychological Science. 2002; 13(5):480–484. [PubMed: 12219818]

Swingley D, Aslin RN. Lexical competition in young children's word learning. Cognitive Psychology. 2007; 54:99–132. [PubMed: 17054932]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Van Petten C, Coulson S, Rubin S, Plante E, Parks M. Time course of word identification and semantic integration in spoken language. Journal of Experimental Psychology: Learning, Memory, and Cognition. 1999; 25:394–417.

Walley AC, Metsala JL. The growth of lexical constraints on spoken word recognition. Perception & Psychophysics. 1990; 47(3):267–280. [PubMed: 2326149]

Werker JF, Curtin S. PRIMIR: A developmental model of speech processing. Language Learning and Development. 2005; 1(2):197–234.

Werker JF, Fennell CT, Corcoran K, Stager CL. Infants' ability to learn phonetically similar words: Effects of age and vocabulary size. Infancy. 2002; 3:1–30.

Werker JF, Tees RC. Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. Infant Behavior and Development. 1984; 7:49–63.

White KS, Morgan JL. Sub-segmental detail in early lexical representations. Journal of Memory and Language. 2008; 59:114–132.

Yee, EJ. The time course of lexical activation during spoken word recognition: Evidence from unimpaired and aphasic individuals. Brown University; Providence, R.I.: 2005. Unpublished doctoral dissertation

Yoshida KA, Fennell CT, Swingley D, Werker JF. Fourteen-month-old infants learn similar-sounding words. Developmental Science. 2009; 12:412–418. [PubMed: 19371365]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

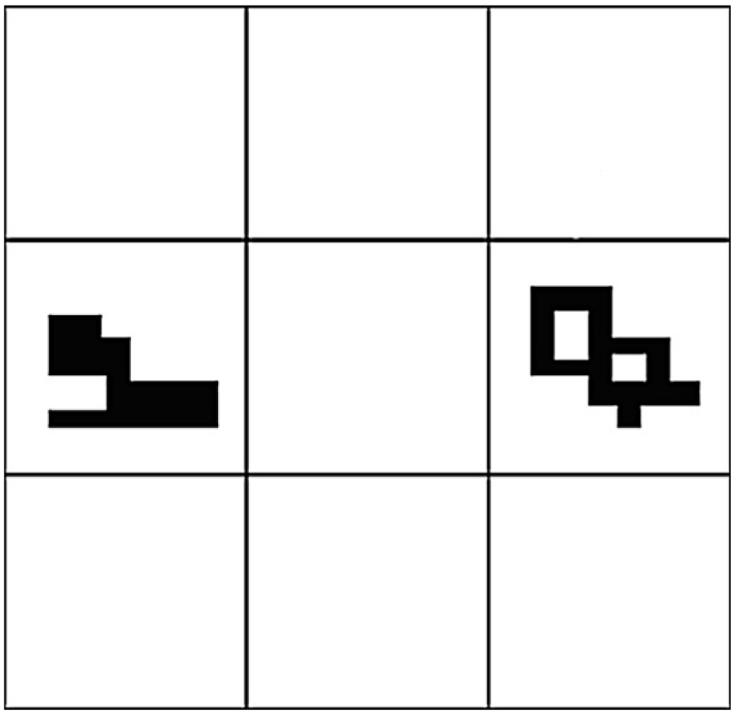


Fig. 1.
Sample test display.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

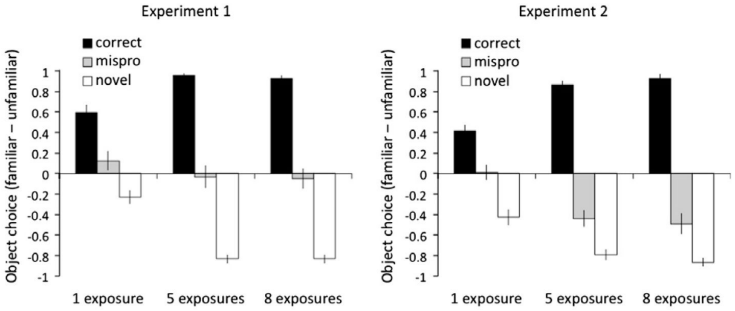


Fig. 2.

Familiar–unfamiliar object selection from Experiments 1 and 2. Leftmost bars represent 1-exposure conditions, middle bars 5-exposure conditions, and rightmost bars 8-exposure conditions. Vertical bars indicate standard errors.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

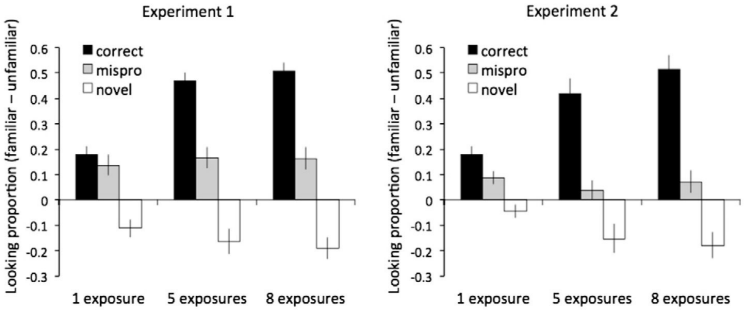


Fig. 3.

Familiar–unfamiliar looking proportion (averaged across the trial) from Experiments 1 and 2. Leftmost bars represent 1-exposure conditions, middle bars 5-exposure conditions, and rightmost bars 8-exposure conditions. Vertical bars indicate standard errors.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

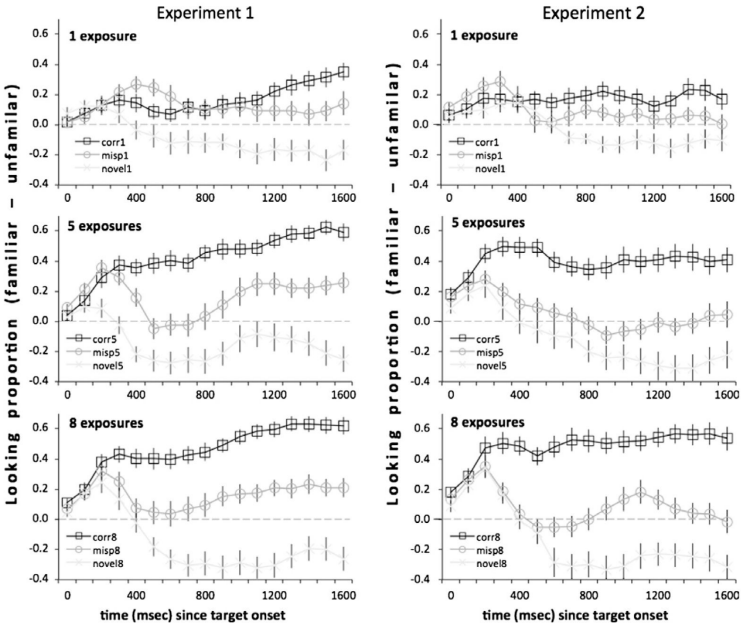


Fig. 4.

Familiar–unfamiliar looking over time for Experiment 1 (left panels) and Experiment 2 (right panels). Dotted line at 0 indicates no bias. Vertical bars indicate standard errors.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Training labels	Mispronunciations	Novel labels
bæv	gæv^{p}	tʃæg
biθ	giθ^{p}	flɪm
blɛm	glɛm^{p}	slud
boð	goɪð^{p}	dʒɪk
bos	pos^{v}	ʃɛn
buʒ	puʒ^{v}	kaʊt
baʊdʒ	paʊdʒ^{v}	faɪs
daʃ	taʃ^{v}	san
dɛtʃ	bɛtʃ^{p}	vɛd
ðɛk	ʒɛk^{p}	taɪf
dɪv	bɪv^{p}	vis
dɹeb	gɹeb^{p}	daɪg
fɪdʒ	vɪdʒ^{v}	pɹæk
faʊm	vaʊm^{v}	stʌt
gʌdʒ	kʌdʒ^{v}	læt
gɹol	bɹol^{p}	gaʊk
gaɪn	baɪn^{p}	tɛz
gef	kef^{v}	tʃum
kadʒ	padʒ^{p}	tʃʌb
kɛl	gɛl^{v}	paɪtʃ
kɪʒ	gɪʒ^{v}	dʒid
gʊz	dʊz^{p}	fæl
klop	glop^{v}	voʒ
pʌv	bʌv^{v}	fitʃ
pluk	kluk^{p}	
pʊθ	tʊθ^{p}	
sɛp	fɛp^{p}	
sig	ʃig^{p}	
soɪk	zoɪk^{v}	
sot	zot^{v}	
ʃub	ʒub^{v}	
sʊd	zʊd^{v}	
saɪp	ʃaɪp^{p}	
tæs	kæs^{p}	
θʌp	fʌp^{p}	
teg	peg^{p}	
tib	dib^{v}	
tof	dof^{v}	
toɪp	doɪp^{v}	

Training labels	Mispronunciations	Novel labels
tɹam	dɹam^{v}	
tuv	kuv^{p}	
vad	zad^{p}	
voɪn	ʒoɪn^{p}	
vup	zup^{p}	
zʌl	sʌl^{v}	
zæb	sæb^{v}	
zed	sed^{v}	
ʒɪb	ʃɪb^{v}	

Phonotactic probability characteristics of training and mispronunciation sets, respectively. Mean 1st biphone = .003, Mean 1st biphone = .0024. Mean 1st segment = .0507, Mean 1st segment = .0455.

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

Training labels	Mispronunciations	Novel labels
bʌv	tʌv	tʃæg
bluk	kluk	flɪm
ʒub	sub	slud
dæs	kæs	dʒɪk
ðʌp	fʌp	ʃɛn
kʌdʒ	dʌdʒ	kaʊt
kʊz	dʊz	faɪs
doɪp	koɪp	san
dɹam	kɹam	vɛd
zoɪk	ʃoɪk	taɪf
fad	zad	vis
foɪn	ʒoɪn	daɪg
dof	kof	pɹæk
gadʒ	padʒ	stʌt
sed	ved	læt
piθ	giθ	gaʊk
glop	plop	tɛz
deg	peg	tʃum
tɪv	bɪv	tʃʌb
kɹol	bɹol	paɪtʃ
fup	zup	dʒid
kaɪn	baɪn	fæl
pæv	gæv	voʒ
gɪʒ	pɪʒ	fitʃ
plɛm	glɛm	
sʌl	vʌl	
pos	gos	
zaɪp	ʃaɪp	
paʊdʒ	daʊdʒ	
sæb	væb	
poɪð	goɪð	
gɛl	pɛl	
ʃɪb	zɪb	
zɛp	fɛp	
tɛtʃ	bɛtʃ	
θɛk	ʒɛk	
kef	bef	
tɹeb	gɹeb	
vɪdʒ	sɪdʒ	
taʃ	baʃ	
vaʊm	saʊm	

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

Training labels	Mispronunciations	Novel labels
zɪg	ʃig	
duv	kuv	
zot	fot	
bʊθ	tʊθ	
zʊd	fʊd	
puʒ	guʒ	
dib	kib	

Phonotactic probability characteristics of training and mispronunciation sets, respectively. Mean 1st biphone = .0024, Mean 1st biphone = .0027. Mean 1st segment = .0496, Mean 1st segment = .0489.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript