1. Introduction

Learning a language is a long, complex process, yet in the first few years of life humans are able to become masters of their native language with relative ease. One particularly complicated factor of language learning is the variability inherent to the speech environment, especially as toddlers are starting to recognize more and more words. The interaction between the nature of young learners’ phonological knowledge and phonological variability, such as that in mispronunciations, affects toddlers’ ability to recognize known words and learn new ones.

Previous research has demonstrated that toddlers are sensitive to even single-feature consonantal mispronunciations of familiar labels and that such sensitivity is graded as the number of features involved increases. For example, Ballem and Plunkett (2005) and Swingley and Aslin (2002) showed that infants as young as 14-months-old can detect changes in word-initial consonants of familiar words, suggesting that consonant categories are highly salient during lexical access. Similarly, White and Morgan (2008) found that toddlers are sensitive to increasing severity of mispronunciations in word onset consonants. In this study, toddlers saw a familiar image and an unfamiliar image and heard the target word, like “shoe” or a mispronounced word, like “foo” (/fu/). Mispronunciations could depart from target mispronunciations by 1, 2, or 3 phonological features. The proportion of time that the toddlers looked at the familiar image was measured, and looking time decreased linearly with number of features altered, demonstrating graded sensitivity to increasing severity of mispronunciation (White & Morgan, 2008, see Figure 1). In fact, this phenomenon holds with both adults and toddlers in word-initial and word-final consonants along the same dimensions of phonological features (Ren & Morgan, 2011; White, Yee, Blumstein, & Morgan, 2013; Gasdaska, 2016; Tin, 2013). This effect may be a result of discrete categorical perception of stop consonants, a phenomenon which has been widely studied (Liberman et al., 1967; Eimas, 1963; Kronrod, Coppess, & Feldman, 2016). Indeed, it has been shown that stop consonants on either side of acoustic boundaries are perceived strongly as being members of one category, despite acoustic difference from a prototypical exemplar (Eimas et al., 1971).
Conversely, perception of vowels has often shown to be continuous in nature, suggesting the possibility that vowels may be perceived differently during lexical access.

![IPLP Data from White & Morgan (2008)](image)

**Figure 1:** Toddlers’ proportion of time looking to target averaged over the entire trial and corrected for baseline looking preferences. Data from White & Morgan’s (2008) onset consonant mispronunciation study.

For infants, there are contrasting accounts on the role of vowels in word recognition tasks. Some studies have shown that infants rely on phonological differences in vowels during lexical access (Stager & Werker, 1997; Mani & Plunkett, 2007; Curtin, Fennell, & Escudero, 2009), while others have suggested that although vowels do contribute to lexical identity, they may be less important than consonants (Nespor et al., 2003; Nazzi, 2005). Similarly, adults privileged consonants over vowels during lexical access in reading tasks (New, Araujo, & Nazzi, 2009) and word reconstruction tasks (Van Ooijen, 1996; Cutler et al., 2000), but in phonological pattern generalization tasks, adults were able to generalize patterns of vowels but not consonants (Toro, Nespor, Mehler, & Bonatti, 2009). Therefore, listeners’ sensitivity to changes in vowels during tasks associated with lexical information suggests a complicated relationship between the lexicon and vocalic information.

In a series of studies investigating word-medial vowels, Mani and colleagues (Mani & Plunkett, 2007; Mani, Coleman & Plunkett 2008) found that toddlers from fifteen to twenty-four months displayed different patterns of gaze after
hearing correct or mispronounced versions of familiar labels. More specifically, Mani et al. (2008) found that 18-month-olds gaze time declined when hearing mispronunciations, suggesting that toddlers are similarly sensitive to consonant and vowel mispronunciations.

The current work explores toddlers’ sensitivities to vowel mispronunciations in greater detail and compares these to adults’ sensitivities. Recent work has shown that for adults, vowels and consonants do not play symmetric roles: as revealed by time-course analyses of looking in a visual world procedure, adults appear to detect vowel mispronunciations earlier than consonant mispronunciations, but vowel mispronunciations have less effect on lexical access (Franklin 2016; see Figure 2). If toddlers’ sensitivity to mispronunciations of vowels in familiar words is similar to that of adults, it is likely that, at least for familiar words, toddlers have detailed but flexible phonological representations. Adult-like representations of vowels in lexical items at this young of an age have implications on our knowledge of the types of information children use in early stages of word learning, such as distributions and frequencies of phonological units in the child’s native language environment.

![Figure 2: Adult time course of looking for onset consonant, coda consonant, and vowel mispronunciation experiments (Tin 2013, Gasdaska 2016, Franklin 2016, respectively).](image)

2. Methods
2.1. Subjects

Fifty-nine toddlers age 17.5 to 19.5 months were recruited from Providence, Rhode Island and surrounding areas to participate in the current study. Of the children who participated, eleven were excluded (7 due to eye-tracker calibration issues, 3 due to fussiness, and 1 due to bilingual exposure). The remaining 48 toddlers (22 female, mean age 572.7 days; 26 male, mean age 565.3 days) provided the data reported here.
2.2. Stimuli

The stimuli comprised a set of 12 familiar American English test words with 3 mispronunciations of each word for a total of 48 test items. In the mispronounced conditions, the vowel in the mispronounced token differed from the familiar word along one, two, or three phonological dimensions typically associated with vowels: height, backness, and roundedness. For example, \textit{ball} (\textipa{bol/}) was manipulated by one phonological feature, height, to form \textipa{bol/}, two features, height and roundedness, to form \textipa{bol/}, and three features, height, backness, and roundedness, to form \textipa{bæl/}. Half of the familiar words contained front vowels and half contained back vowels, all of which were lax. Therefore, there were two patterns of change along phonological dimensions: \textit{Height} > \textit{Height + Backness} > \textit{Height + Backness + Roundedness} or \textit{Height} > \textit{Height + Roundedness} > \textit{Height + Roundedness + Backness}. Additionally, there were 4 unfamiliar fillers (e.g. \textit{hive}) and 2 familiar fillers (e.g. \textit{bunny}). Individuals items were chosen according to the Cross Linguistic Lexical Norms (CLEX) such that at 17 months, at least 50\% of children had the familiar words in their receptive vocabulary and were easily imageable (Dale & Fenson, 1996; Fenson et al., 2007).

Audio stimuli was produced by a 22-year-old female native speaker of American English from the Southern New England area. Test words were produced in positive-affected infant-directed speech in carrier phrases such as “Where is the X?” and “Look at the X!” The sound files were recorded using Audacity, with at least 8 recordings made for each item and the best single token chosen for each (Audacity Team, 2014). These tokens were confirmed independently by members of the Metcalf Infant Research Lab to contain vowels perceptually typical of their intended category. All sound files were digitally edited with Praat software to equate intensity, but otherwise they were not digitally manipulated (Boersma & Weenink, 2018). Additionally, information on sentence and word duration and formant frequencies at vowel midpoints was collected.

Each word was paired with two images, one familiar item (in test and familiar fillers, this image was a prototypical picture of the target), and one unfamiliar item that the infants would not know a name for. In all trials, one image was familiar and one was novel, and no images were used for more than one trial in each condition. The presentation of familiar and unfamiliar images was based on previous studies such as White and Morgan (2008) that have shown that participants may perceive a mispronounced token as a novel label for an unfamiliar (and therefore nameless) item; moreover, this avoids the possibility found in studies that have used pairs of familiar items that participants may look at the target image because the mispronunciation was a poor label for the distractor image. Familiar images were determined by members of the Metcalf Infant Research Lab to be typical exemplars of the word referents they matched. All images were of real items, and familiar and unfamiliar images were impressionistically matched for visual interest. Altogether, each toddler saw a total of 18 trials.
2.3. Procedure

Modeled after White and Morgan (2008), the present study tested 17.5-19.5 month olds using the Intermodal Preferential Looking Paradigm (IPLP) with eye tracking. Infants were tested using an SMI remote binocular eye tracker paired with ExperimentCenter and iViewX software. The study was conducted in a dimmed, sound-attenuated booth, where the toddler sat on a caregiver’s lap. The caregiver was asked to wear headphones playing music and avert gaze from the testing monitor in order to reduce the possibility of caregiver influence. In each trial, the child first saw a center fixation paired with a non-speech sound. Once the child fixated in the center of the screen for 500ms, the target and distractor images appeared on the screen for 4000ms in silence in order to measure the toddler’s baseline preference for each image. After this, toddlers say a different center fixation, and after fixating for 500ms, a blank screen appeared and toddlers heard the test word in a carrier sentence, such as “Where is the ball?” At the offset of the test word, the target and distractor images appeared once again and after 1000ms, infants heard the test word a second time in a carrier sentence such as “Look at the ball!” After the initial utterance of the test word, images were on the screen for a total of 5000ms (see Figure 3).

![Figure 3: Stimulus sequence in an example trial.](image)

Monocular (left eye) gaze was recorded at 120 Hz and downsampled to 60 Hz. Raw gaze data were preliminarily analyzed using in-house software to ascertain looking time to each of the images in each trial, as well as longest gaze fixation on each image, latency to each image, and number of gaze switches in each trial.
3. Results

For each trial, proportion of time looking to target (PLT = Time looking at the target/Time looking at any location on the screen) was calculated by subtracting the PLT during salience from the PLT during test. The resulting PLTs were then averaged for each condition. Positive mean difference scores indicate greater looking to the target, while negative mean difference scores indicate less looking to the target. Chance looking is at a PLT of zero. Mean PLT values for test trials were Correct: 0.064 (SD = 0.271), 1-Feature Mispronunciation: 0.054 (SD = 0.236), 2-Feature Mispronunciation: 0.014 (SD = 0.284), 3-Feature Mispronunciation: 0.003 (SD = 0.271) (see Figure 4).

![Figure 4: Toddlers’ proportion of time looking to target for correctly pronounced tokens and tokens in which vowels were mispronounced along one, two, or three phonological dimensions. PLT was averaged over the entire trial and corrected for baseline looking preferences.](image)

3.1. ANOVA and pairwise comparisons

Trials in which the parent reported the toddler did not know the target word were excluded from analysis. An ANOVA with PLT and condition revealed a main effect of condition, $F(3, 412) = 2.634, p = 0.045$. Follow up t-tests showed that PLT on correct trials was significantly different from PLT on 2-feature and 3-feature mispronunciation trials, $t(201) = 2.2007, p = 0.029$ and $t(201) = 2.2812, p = 0.024$, respectively. This finding suggests that toddlers may be sensitive only
to mispronunciations that diverge greatly from the correct pronunciation and that toddlers’ vowel representations within lexical items are not as detailed as those of consonants.

3.2. Time course analysis

In addition to PLT computed over the entire test trial, data from the eye tracker also allows for an analysis of the time course across the trial. Analysis was cut off at peak looking time, approximately 1500ms after the offset of the target word. The overall curves were modeled with second-order orthogonal polynomials, fixed effects of condition (number of feature changes) on all time terms, and random effects of participant. The correct condition (zero feature changes) was treated as the baseline and parameters were estimated for the other three conditions. The fixed effects of condition were added for each model, and their effects on model fit were evaluated with a $\chi^2$ test using the lme4 package in R (Bates et al., 2015; R Core Team, 2013).

The effect of condition on the intercept improved model fit significantly ($\chi^2(4) = 797.227, p < 0.001$), as did the effect of condition on the linear term ($\chi^2(4) = 37.697, p < 0.001$) and the quadratic term ($\chi^2(4) = 124.746, p < 0.001$). This indicates that conditions differed overall in the total proportion of looking to target over the range of the time course and the rate at which the participants looked to the target. Models were also fit to subsets of condition in order to investigate pairwise differences between conditions. Correct and 1-feature conditions had significantly different intercepts although slope and quadratic curve for these conditions were not significantly different, $\chi^2(1) = 52.2964, p < 0.001, \chi^2(1) = 0.1358, p = 0.71, and \chi^2(1) = 0.1626, p = 0.68$, respectively. Conversely, between 1- and 2-feature mispronunciation conditions, intercept, slope, and quadratic curve were all significantly different, $\chi^2(1) = 59.2256, p < 0.001, \chi^2(1) = 4.0769, p = 0.043, and \chi^2(1) = 5.3825, p = 0.020$, respectively. 2- and 3-feature mispronunciation conditions did not differ in intercept, but were significantly different in slope and quadratic curve, $\chi^2(1) = 0.0876, p = 0.77, \chi^2(1) = 8.1416, p = 0.004, and \chi^2(1) = 11.9248, p < 0.001$, respectively. Thus, the time course data shows subtle differences between each condition, suggesting that toddlers are indeed sensitive to vowels that are divergent from their categorical prototype, but these divergences may not have as much of an effect on lexical access as consonants (Figure 5).
Figure 5: Toddlers’ proportion of time looking to target for correctly pronounced tokens and tokens in which vowels were mispronounced along one, two, or three phonological dimensions. PLT was computed frame-by-frame, smoothed and with 95% confidence intervals (gray).

4. Discussion

The current study provides evidence that toddlers and adults are similarly sensitive to vowel mispronunciations, but in general vowel mispronunciations do not modulate lexical access in the same way as consonants. These results are most strongly supported by analyses of the time course of looking across the trial, which show subtle differences in behavior across all conditions, but a generally similar quadratic trend. Furthermore, lack of differences in sensitivity to correct pronunciations and 1-feature mispronunciations in the early part of trial point to less sensitivity to small divergences in vowels for toddlers compared to adults. During early stages of linguistic development, lexical access mechanisms seem to be adult-like to some degree, but there may be differences in the way adults and toddlers weight small divergences in vowel pronunciation. These differences may be attributed to experience with variation across the lifespan, but it is also possible that non-adult-like weighting of phonological units may aid early word-learning mechanisms. However, more research needs to be done to explore this claim fully.

Furthermore, over the entire trial, there is some graded sensitivity to mispronunciation, though not to the same magnitude as with consonants in White & Morgan (2008). This suggests that, as with adults, vowels and consonants play different roles in toddler’s lexical processes. One possibility is that due to their
longer duration and greater intensity, vowels are more robust to articulatory variation than consonants. If this were the case, we would expect to see similar results across languages, as this explanation relies on the inherent phonetic differences between vowels and consonants in natural language. However, in complementary studies, Dutch learning infants showed a vowel-bias for word-learning (Højen & Nazzi, 2016) while French learning infants at the same age showed a consonant bias in the same task (Havy & Nazzi, 2009). Findings such as these suggest that the distribution of phonological segments in a given language may play a role in such asymmetries, as Dutch is “vowel heavy” while French and English are “consonant heavy.” Nevertheless, even in infancy, learners of different languages may rely on consonantal and vocalic information to differing degrees during tasks involving the lexicon. Another possibility is that vowels may exhibit more variability in speech than consonants due to non-lexical factors, such as sociolinguistic identity and emotive tone, which are often realized in vowels. Thus listeners may be more flexible when mapping divergent vowels to items in the mental lexicon than divergent consonants due to their extralinguistic variation.

A caveat to these conclusions is that comparisons to earlier work with consonants can only be impressionistic at this time, as the measurement methods (offline coding vs. eye tracking) have inherently different granularity and consistency. Therefore, further work is currently being done to replicate Ren & Morgan’s (2011) findings of toddler’s adult-like sensitivity to mispronunciations of coda consonants with eye tracking methods in order to better explore toddler’s behavior when perceiving mispronounced consonants and vowels. Nevertheless, the current study provides interesting evidence that although infants are sensitive to mispronunciations of vowels, the nature of that sensitivity may not be completely mature and may be asymmetric to early sensitivity to consonant mispronunciations.

References


Franklin, Lauren and Morgan, James L. (2017). On the nature of vocalic representation during lexical access. 173rd Meeting of the Acoustical Society of America, June 2017 meeting [poster], Boston, MA.


