And then I saw her race: Race-based expectations affect infants’ word processing

Drew Weatherhead*, Katherine S. White

Department of Psychology, University of Waterloo, Ontario, Canada

ARTICLE INFO

Keywords:
Speech perception
Infant word recognition
Infant social cognition
Sociolinguistic variation
Accent processing

ABSTRACT

How do our expectations about speakers shape speech perception? Adults’ speech perception is influenced by social properties of the speaker (e.g., race). When in development do these influences begin? In the current study, 16-month-olds heard familiar words produced in their native accent (e.g., “dog”) and in an unfamiliar accent involving a vowel shift (e.g., “dag”), in the context of an image of either a same-race speaker or an other-race speaker. Infants’ interpretation of the words depended on the speaker’s race. For the same-race speaker, infants only recognized words produced in the familiar accent; for the other-race speaker, infants recognized both versions of the words. Two additional experiments showed that infants only recognized an other-race speaker’s atypical pronunciations when they differed systematically from the native accent. These results provide the first evidence that expectations driven by unspoken properties of speakers, such as race, influence infants’ speech processing.

1. Introduction

Speech perception is often thought of as a bottom-up process, relying only on acoustic information in the speech signal. However, there are many “unspoken” properties of speakers that impact our perception of their speech, such as gender (Strand & Johnson, 1996), sexual orientation (Munson, Jefferson, & McDonald, 2006), age (Drager, 2011), and nationality (Hay, Nolan, & Drager, 2006; Niedzielski, 1996). In the current study, we investigate for the first time whether infants’ word recognition is shaped by expectations linked to unspoken properties of speakers. In particular, we ask whether infants have expectations about how a speaker will talk based on their race.

Adult speakers invoke their knowledge of socially linked variation during language processing. The same speech sequence can be interpreted differently, and be better or worse understood, based on known or inferred properties of the speaker. These effects are seen even when listeners are primed very subtly. For example, Hay and Drager (2010) exposed New Zealander participants to either a stuffed kangaroo (associated with Australia) or a stuffed kiwi bird (associated with New Zealand) prior to completing a vowel perception task. They found that participants’ vowel perception shifted as a function of the exposure toy, such that those who saw the kangaroo were more likely to classify vowels as Australian-like than those who saw stuffed kiwis. Similarly, visual cues like race can impact the way speech is perceived. For example, American listeners understand native-accented English better when it is paired with a picture of a Caucasian face than with a Chinese face (Kang & Rubin, 2009; Rubin, 1992; consistent with Babel & Russell, 2015). Likewise, Mandarin-accented English is better understood when it is paired with a Chinese face than with a Caucasian face (McGowan, 2015). Together, these studies suggest that adult listeners form associations between properties of social groups and linguistic variation, which in turn lead to expectations that affect speech processing.

Although there is evidence that speech perception in adults is affected by top-down knowledge about the speaker, there is no research addressing whether this type of knowledge affects infants’ speech perception. It may seem unlikely that it would, given that infants do not have the specific, learned associations that adults have. But even so, infants do have language experience that could potentially shape their expectations in more general ways. Because infants are exquisitely sensitive to race, a natural place to start this investigation is by exploring the effect of a speaker’s race on infants’ word recognition.

By 3 months, infants prefer to attend to familiar-race faces over unfamiliar-race faces (Bar-Haim, Ziv, Lamy, & Hodes, 2006; Kelly et al., 2005). With age, infants become less capable of discriminating or recognizing unfamiliar-race faces; by 9 months, infants categorize faces by race, and are significantly better at recognizing individual familiar-race faces (Anzures, Quinn, Pascalis, Slater, & Lee, 2010; Kelly et al., 2007; Kelly et al., 2009), and their scanning patterns for familiar-race and unfamiliar-race faces differ (Wheeler et al., 2011).

* Corresponding author at: Department of Psychology, University of Waterloo, 200 University Ave W, Waterloo, ON N2L 3G1, Canada.
E-mail address: deweathe@uwaterloo.ca (D. Weatherhead).

https://doi.org/10.1016/j.cognition.2018.04.004
Received 30 March 2017; Received in revised form 13 March 2018; Accepted 2 April 2018
Available online 12 April 2018
0010-0277/ © 2018 Elsevier B.V. All rights reserved.
Although no work has asked whether a speaker’s race affects infants’ word processing, infants do make some very general assumptions about the relationship between a speaker’s productions and their physical characteristics. For example, 5-month-old infants match the vocalizations of humans and monkeys to the appropriate faces (Vouloumanos, Druhen, Hauser, & Huizink, 2009). Six-month-olds also match other-race faces with non-native languages, though not with backwards speech (Uttley et al., 2013). Thus, infants not only have species-level associations, but also appear to have some understanding of same-race and other-race speakers as separate groups, and different beliefs about how these groups speak—in particular, that same-race individuals speak in a familiar way, and other-race individuals speak in a novel way. However, these studies do not indicate whether infants’ speech processing is affected by factors such as speaker race.

In the current study, we explore how infants’ race impacts infants’ recognition of words that are pronounced in a familiar or unfamiliar accent. Young language learners often have difficulty recognizing words when they are produced in an unfamiliar accent, at least in the absence of a learning period (Best, Tyler, Gooding, Orlando, & Quann, 2009; Van Heugten & Johnson, 2014; Van Heugten, Krieger, & Johnson, 2015; White & Aslin, 2011). For example, without prior exposure to an unfamiliar accent, 15-month-olds do not look preferentially at a target object when its label is produced in that accent (Mulak, Best, Tyler, Kitamura, & Irwin, 2013). This difficulty with unfamiliar pronunciations is perhaps unsurprising, given the narrow range of speaker variation that infants have been exposed to in their input. If speech processing is constrained by a listener’s prior experience (Kleinschmidt & Jaeger, 2015), and infants have little experience with variability in general, then they should initially expect new speakers to talk in familiar ways. Encountering an unfamiliar accent should lead to processing difficulty because it violates these expectations.

But importantly, just as infants have only heard words produced in a narrow range of pronunciations, most have only heard words produced by a narrow range of (in many cases, same-race) people. If infants’ beliefs about word pronunciations are tied to the types of speakers who say them, then an expectation that words should be pronounced in familiar ways should not necessarily extend to unfamiliar, other-race speakers. Instead, infants may wait for evidence from speakers to determine the specific accent.

In the present study, we tested infants’ comprehension of unfamiliar words produced in their familiar native accent (e.g. “dog”) and in an unfamiliar accent involving a vowel shift (e.g. “dag”) using the inter-modal preferential looking procedure (in which infants are presented with objects on a screen and hear corresponding audio). These words were presented following an image of either a same-race speaker or an other-race speaker (for brevity, we hereafter refer to the familiar-accented words as “unaccented” and the unfamiliar-accented words as “accented”, although we recognize that there is no such thing as “unaccented” speech). If infants’ word processing is affected by expectations about speakers based on race, then infants should interpret the two types of words differently depending on the speaker’s identity.

2. Experiment 1

2.1. Method

2.1.1. Participants

Forty 16-month-old infants were tested (23 females; mean age: 16 months, 0 days; age range: 15;16–16;16). Nine additional participants were tested, but not included due to non-completion (3), failure to attend to both objects during the baseline period for at least half of each trial type in each block of trials (3), or an overall difference score exceeding 2.5 standard deviations from the mean for either trial type (3).

Infants were randomly assigned to one of two conditions: Same-race speaker or Other-race speaker. Participants in both conditions were monolingual English-learners and Caucasian. Overall, participants had very minimal exposure to people who spoke a foreign language, had an accent, or were of a different race (average exposure per week was 2.6%, 7.2%, and 7.3%, respectively, as indicated by parental reports; by condition: Same-race Condition = 3.1%, 7.2%, and 7.2%, respectively; Other-race Condition = 2.1%, 6.5%, and 7.5%, respectively).

2.1.2. Stimuli

2.1.2.1. Audio stimuli. The test words were six words highly familiar to 16-month-olds, all containing the same vowel, /a/: “ball”, “block”, “bottle”, “car”, “dog”, and “sock”. All of these words are comprehended by 67–95% of children by 15-months of age, according to the MacArthur Communicative Development Inventories (Dale & Fenson, 1996). Additionally, parental reports in the current study indicate that for each test word individually, 78–97% of children had “seen the object before and understand the word very well”, and across all words, the average was 88.5%. A female native speaker of English (from the same geographic region as the participants) produced each word four times, twice unaccented and twice accented, in which the /a/ vowel was shifted to /æ/ (i.e., “bottle” to “battle”, “sock” to “sack”, etc.).

Acoustic measurements confirmed that the /a/ and /æ/ versions were realized as intended. These measurements are provided in Appendix A. Each version was produced in each of two sentence contexts, “Do you see the X” or “Find the X”. All sentences were naturally produced in an infant-directed-manner. Importantly, the same audio stimuli were used for both conditions.

2.1.2.2. Visual stimuli. Depending on the condition, participants either saw a still image of a same-race woman or an other-race woman (Fig. 1). The same-race woman was a 22-year-old Caucasian with pale skin and long brown hair. The other-race woman was a 23-year-old other-race woman (Fig. 1). The same-race woman was a 22-year-old Caucasian with pale skin and long brown hair. The other-race woman was a 23-year-old

---

1 It should be noted that although this is an existing vowel shift in the Northern US, it is not one that participants in this study were reported to have exposure to. Additionally, a previous study (White & Aslin, 2011) using the same shift in a nearby geographic region found that toddlers did not recognize words in this accent without experimental exposure to it.
Table 1

Mean scores (and standard deviations) for the unaccented and accented trials for each speaker. Baseline scores indicate the proportion of time spent attending to the familiar object prior to labeling (3 s period). Test scores indicate the proportion of time spent attending to the familiar object after labeling (3 s period). Difference scores were calculated using the proportions for each period (test – baseline); positive difference scores indicate an increase in attention to the familiar object after labeling.

<table>
<thead>
<tr>
<th></th>
<th>Unaccented</th>
<th>Accented</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Same-Race Speaker</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experiment 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>.54 (.08)</td>
<td>.51 (.07)</td>
</tr>
<tr>
<td>Test</td>
<td>.62 (.12)</td>
<td>.48 (.08)</td>
</tr>
<tr>
<td>Difference Score</td>
<td>.08 (.09)</td>
<td>-.03 (.07)</td>
</tr>
<tr>
<td><strong>Other-Race Speaker</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experiment 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>.51 (.06)</td>
<td>.49 (.07)</td>
</tr>
<tr>
<td>Test</td>
<td>.59 (.13)</td>
<td>.59 (.09)</td>
</tr>
<tr>
<td>Difference Score</td>
<td>.07 (.12)</td>
<td>.10 (.08)</td>
</tr>
<tr>
<td>Experiment 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>.53 (.07)</td>
<td>.53 (.06)</td>
</tr>
<tr>
<td>Test</td>
<td>.61 (.06)</td>
<td>.55 (.08)</td>
</tr>
<tr>
<td>Difference Score</td>
<td>.08 (.09)</td>
<td>.02 (.07)</td>
</tr>
<tr>
<td><strong>Systematic Accent Condition</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>.54 (.07)</td>
<td>.49 (.07)</td>
</tr>
<tr>
<td>Test</td>
<td>.59 (.08)</td>
<td>.57 (.1)</td>
</tr>
<tr>
<td>Difference Score</td>
<td>.05 (.08)</td>
<td>.08 (.11)</td>
</tr>
<tr>
<td><strong>Random Pronunciations Condition</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>.52 (.05)</td>
<td>.54 (.08)</td>
</tr>
<tr>
<td>Test</td>
<td>.59 (.10)</td>
<td>.51 (.10)</td>
</tr>
<tr>
<td>Difference Score</td>
<td>.08 (.09)</td>
<td>-.03 (.09)</td>
</tr>
</tbody>
</table>

Displays were static images containing one familiar (corresponding to one of the six test words) item paired with one unfamiliar item. All participants received the same six familiar–unfamiliar object pairings.

2.1.3. Procedure

The participant sat on his/her parent’s lap approximately 1.5 ft. from a 36 × 21-inch plasma screen television in a sound-treated testing room. A camera under the television recorded the child’s looking behavior for the entirety of the session. The camera was linked to a monitor and recording device in the lab area adjacent to the testing room for the experimenter’s viewing purposes and for later off-line coding. Stimuli were played at approximately 65 dB and presented in PsyScope X (Cohen, MacWhinney, Flatt, & Provost, 1993). Parents were instructed not to interact with their infants during the session and wore noise-cancelling headphones playing instrumental music to mask the audio being played to the infant.

Infants first viewed a silent 8-s introductory video of the speaker smiling and waving to reinforce the identity of the speaker. Infants then completed a total of 24 test trials, in two consecutive blocks of 12 trials. In each block, each of the test words occurred twice, once unaccented and once accented (whether the unaccented or accented version occurred first was counterbalanced across words and participants). Each trial was 10 s in length. At the start of each trial, a static image of the speaker’s face and shoulders appeared at the top center of the screen for two seconds, with two black outlined boxes appearing centrally on either side of the screen. Following this, the speaker’s face disappeared, and an object appeared in each of the two outlined boxes. One object corresponded to the test word (i.e., the target object), and the other object was a novel distractor (i.e., the distractor object).

These two objects stayed on the screen for eight seconds, the first three seconds of which was a silent baseline period, followed by an audio recording of the test word in the naming phrase (either “Do you see the X” or “Find the X”). Each block was pseudo-randomized such that the target object was never on the same side for more than three trials in a row, the same sentence context did not occur more than two trials in a row, no more than three unaccented or accented trials occurred in a row, and the same word did not occur fewer than four trials apart.

2.1.4. Coding of looking times

Looking time was coded off-line using in house software, frame-by-frame (1 frame = 33 ms). Looking proportions to the objects were determined for the baseline period and for the test period, which began 300 ms after the onset of the test word to account for the time necessary to program an eye movement. Both the baseline and test period were a priori chosen to be 3 s in length. This length test phase is consistent with other studies on toddlers’ word recognition with displays that include novel objects (Bion, Borovsky, & Fernald, 2013; White & Aslin, 2011; White & Morgan, 2008).

2.2. Results

For both the baseline and test periods, the proportion of time infants spent looking at each of the objects was computed (out of the total time looking at either object during that 3-s period). These proportions are provided in Table 1. Trials in which infants did not look at both objects during the baseline period (or at either object during the test period) were not included in the analyses. Across conditions, there was no difference in the percentage of discarded trials, t(38) = .14, p = .707, and no difference in the proportion of time that infants spent looking at the familiar object during baseline, t(38) = .48, p = .144 (.53 in the Same-race condition and .50 in the Other-race condition). Additionally, infants in the two conditions paid an equivalent amount of attention to the speaker overall during the 2-s speaker presentation prior to the objects’ appearance, t(38) = .92, p = .362.

Detail about the time course of infants’ looking over the test period is provided in Fig. 2. This figure plots looking over time in terms of the difference from infants’ average baseline preference. Differences between the conditions are evident throughout the test phase. In the Same-race condition, the differentiation of the unaccented and accented pronunciations occurs early. In contrast, in the Other-race condition, there is no clear differentiation of the looking curves at any point during the test phase.2

For the primary analysis assessing infants’ recognition of the words, a difference score was calculated for each trial using the overall looking proportions for each period (proportion target objecttest-proportion target objectbaseline). This measure indicates the change in looking toward the target object after labeling. Note that a difference score of zero (no change following labeling) indicates a failure to recognize the pronunciation as an instance of the target word. Difference scores were averaged across trials for each word type. These difference scores are presented in Fig. 3.

A mixed measures ANOVA with the within-subject factor of word type (Unaccented vs. Accented) and the between-subject factor of condition (Same-race vs. Other-race) found a main effect of condition, F (1, 38) = 11.14, p = .002, η2 = .227, and a significant word type X condition interaction, F(1, 38) = 9.49, p = .004, η2 = .200, but no effect of word type, F(1, 38) = 3.65, p = .064, η2 = .088. Thus, infants

---

2 To get a further sense of infants’ looking during test, we also conducted analyses of the latency to shift to the familiar target. Considering only those trials in which infants were not already looking at the target at word onset, a repeated-measures ANOVA in- cluding condition and word type found a significant effect of word type, F(1, 37) = 8.12, p = .007, η2 = .180, and a significant condition x word type interaction, F(1, 37) = 4.22, p = .047, η2 = .102. There was no main effect of condition, F(1, 37) = .87, p = .37. Infants in the Same-race condition were significantly faster to shift to the target for unaccented pronunciations than accented pronunciations (t(19) = 3.3, p = .004, d = .941). In contrast, infants in the Other-race condition showed the same latency to shift for the two types of pronunciations (t(18) = .6, p = .56).
interpreted the same words differently depending on which speaker they saw.

For infants in the Same-race speaker condition, paired sample t-tests comparing unaccented pronunciations to the accented pronunciations revealed that infants interpreted the two types of pronunciations differently, $t(19) = 3.74, p = .001, d = .823$. One-sample t-tests against chance (zero change) revealed that for the unaccented pronunciations, infants’ looking increased significantly to the target object, $t(19) = 3.77, p = .001, d = .843$. In contrast, for the accented words, infants’ looking increased to the distractor object, $t(19) = 2.09, p = .05, d = .467$. Thus, for the same-race speaker, infants recognized only the unaccented pronunciations. The pattern of results for the same-race speaker is consistent with previous demonstrations that toddlers fail to recognize accented words in the absence of exposure when no information about the speaker is present (e.g., Mulak et al., 2013; White & Aslin, 2011).

For infants in the Other-race speaker condition, however, paired sample t-tests comparing unaccented pronunciations to the accented pronunciations revealed that infants did not interpret these two types of words differently, $t(19) = .79, p = .442$. One-sample t-tests against chance revealed that infants’ looking increased significantly toward the target object for both the unaccented words, $t(19) = 2.71, p = .014, d = .606$, and the accented words, $t(19) = 5.42, p < .001, d = 1.212$. Thus, for the other-race speaker, infants mapped both the accented and unaccented pronunciations to the familiar objects.

Because previous studies (White & Aslin, 2011) have found that toddlers may learn about a speaker’s accent during the test phase, we also conducted planned analyses of the test-baseline difference scores with test block as a factor. The first block is more representative of infants’ initial interpretation of the words, whereas the second block indicates what they learned after some exposure to the speaker.

A mixed measures ANOVA with the within-subject factors block and word type, and the between-subject factor condition, revealed a main effect of block, $F(1, 38) = 6.73, p = .013, \eta^2 = .150$, condition, $F(1, 38) = 12.89, p = .001, \eta^2 = .253$, and, crucially, the significant condition X word type interaction, $F(1, 38) = 10.11, p = .003, \eta^2 = .21$. No other effects were significant, $p$s > .092. The lack of a 3-way block x word type X condition interaction indicates that the infants’ differential treatment of the pronunciations between the two speaker conditions was present in both blocks of testing. Consistent with this, the critical word type X condition interaction found in the overall analysis was found for each block separately: for block 1, there was a significant condition X word type interaction, $F(1, 38) = 6.34, p = .016, \eta^2 = .143$, but no main effect of condition, $F(1, 38) = 3.23, p = .080, \eta^2 = .078$, or word type, $F(1, 38) = 3.48, p = .070, \eta^2 = .084$. For the second block, there was a significant condition X word type interaction, $F(1, 38) = 4.78, p = .035, \eta^2 = .112$, and a main effect of condition, $F(1, 38) = 7.43, p = .010, \eta^2 = .164$, but no effect of word type, $F(1, 38) = .26, p = .614$. Thus, for both blocks individually, infants interpreted the same words differently depending on which speaker they saw.

We then considered each speaker condition separately. A repeated measures ANOVA with the within-subject factors block and word type found that for the same-race speaker, there was a main effect of word
type, $F(1, 19) = 12.93, p = .002, n^2 = .405$, no effect of block, $F(1, 19) = 1.14, p = .300$, and no block X word type interaction, $F(1, 19) = 2.09, p = .165$. In block 1 infants in the Same-race speaker condition interpreted unaccented and accented words differently, $t(19) = 3.26, p = .004, d = .713$, but by block 2 this difference diminished, $t(19) = 1.87, p = .077, d = .416$. However, in both blocks their looking increased to the target object for unaccented words (block 1: $t(19) = 2.86, p = .010, d = .640$; block 2: $t(19) = 2.03, p = .057, d = .454$), but did not for the accented pronunciations (in block 1, there was a significant increase in looking to the distractor object, $t(19) = 2.59, p = .018, d = .579$; in block 2, there was no change from baseline, $t(19) = .101, p = .921$). In other words, infants in this condition recognized only the unaccented pronunciations.

For the other-race speaker, a repeated measures ANOVA with the within-subject factors block and word type revealed a main effect of block, $F(1, 19) = 6.30, p = .021, n^2 = .249$, but no main effect of word type, $F(1, 19) = .99, p = .333$, and no block X word type interaction, $F(1, 19) = 1.20, p = .732$. In block 1, infants showed no difference between the unaccented and accented words, $t(19) = .44, p = .665$. However, one-sample t-tests suggest that while infants did not recognize unaccented pronunciations, $t(19) = .82, p = .421$, they did recognize accented pronunciations, $t(19) = 2.01, p = .059, d = .449$. By block 2, infants’ looking increased significantly toward the target object for both the unaccented words, $t(19) = 2.69, p = .015, d = .602$, and accented pronunciations, $t(19) = 6.69, p < .001, d = 1.495$, with no differences between the two word types, $t(19) = 1.21, p = .241$.

These findings demonstrate that infants’ word recognition is significantly affected by the race of the speaker. Overall, although infants in the Same-race condition recognized only unaccented words, infants in the Other-race condition recognized both types of words as labels for the familiar objects. Moreover, the block analyses show the changes over time: when infants first encountered the other-race speaker, they recognized only accented pronunciations. However, after some experience with the speaker, they recognized her pronunciations as instances of the target words, regardless of how they were pronounced. Note that infants who saw the same-race speaker also changed their processing over time: they initially interpreted the accented words as referring to the distractor object, but by the second block, they no longer did so. Therefore, although infants in the two groups differed in their initial treatment of the accented words, both groups started to learn the accented words with exposure.

3. Experiment 2

Infants had different expectations for the same-race and other-race speakers in Experiment 1. In particular, they expected the same-race speaker to use familiar pronunciations, but did not have the same expectation for the other-race speaker – for this speaker, they initially showed somewhat more robust recognition of the accented words. Ultimately, however, they recognized both unaccented and accented pronunciations from this speaker. One possibility is that, over time, infants simply began to recognize anything “close” from the unfamiliar-looking speaker, perhaps because they paid less attention to the specifics of the speaker’s productions or had a higher tolerance for variability from this speaker. In Experiment 2 we tested this possibility, by mispronouncing the words in random ways. If infants simply disregard the unfamiliar-looking speaker’s vowels, then they should similarly recognize the random mispronunciations.

3.1. Method

3.1.1. Participants

Twenty 16-month-old infants were tested (10 females; mean age: 15 months 27 days; age range: 15;10–16;15 days). Four additional participants were tested, but not included due to non-completion (3), and failure to attend to both objects during the baseline period for at least half of each trial type in each block of trials (1). As in Experiment 1, participants had minimal exposure to people who spoke a foreign language, had an accent, or were of a different race (average exposure per week was 4.4%, 8.3%, and 9.1%, respectively, as indicated by parental reports).

3.1.2. Stimuli

3.1.2.1. Audio stimuli. The same six highly familiar words from Experiment 1 were used. Recall, all six words contain the same vowel (/a/). Unlike the prior experiment, here there was no systematicity to the new pronunciations; a random vowel change was assigned to each word. For example, “bottle” was produced as “bootive”, “sock” as “seck”, “block” as “blick”, etc. A female native speaker of English (from a region of central New York with merged /a/ and /ɔ/) produced two versions of each word, one unaccented and one with a random mispronunciation. Acoustic analyses of the unaccented versions are provided in the supplementary text. Once again, each version was produced naturally in infant-directed speech in the context of two sentences, “Do you see the X’” or “Find the X’”.

3.1.2.2. Visual stimuli. Visual stimuli were identical to the Other-race speaker condition of Experiment 1.

3.1.3. Procedure

Same procedure used in Experiment 1.
3.1. Coding of looking times

Same procedure used in Experiment 1.

3.2. Results

As in Experiment 1, trials in which infants did not look at both objects during the baseline period (or at either object during the test period) were not included in the analyses. There was no difference in the percentage of discarded trials across Experiments 1 and 2, $F(2, 57) = .07, p = .937$. In addition, the proportion of time that infants spent looking at the familiar object during baseline in Experiment 2 (.54) was equivalent to the proportions found in Experiment 1, $F(2, 57) = 1.15, p = .258$.

Fig. 4 displays the time course of infants’ looking during the test period.

To explore infants’ recognition of the words, a difference score was again calculated for each trial (overall proportion looking target object$_{test}$-overall proportion looking target object$_{baseline}$). These difference scores are displayed in Fig. 5. A paired sample $t$-test revealed a significant difference in how infants interpreted the unaccented and random pronunciations, $t(19) = 2.34, p = .030, d = .561$. Infants’ looking increased significantly toward the target object for the unaccented pronunciations, $t(19) = 3.97, p = .001, d = .888$, but was at chance levels for the random pronunciations, $t(19) = 1.41, p = .176$.

As in Experiment 1, we also conducted analyses with test block as a factor to explore changes over the experiment. A repeated measures ANOVA with within-subjects factors block and word type revealed a main effect of word type, $F(1, 19) = 8.23, p = .010, n^2 = .302$. No block X word type interaction, $F(1, 19) = 1.51, p = .235$, or main effect of block, $F(1, 19) = .22, p = .642$, was found.

For the first block, a paired sample $t$-test revealed no difference between the unaccented and random pronunciations, $t(19) = .66, p = .517$. Infants did not show a significant change in looking to the target for either the unaccented pronunciations, $t(19) = 1.91, p = .072, d = .427$, or the random pronunciations, $t(19) = 1.27, p = .221$. In the second block, a paired sample $t$-test revealed a significant difference in how infants interpreted the unaccented and random pronunciations, $t(19) = 2.59, p = .018, d = .543$. Infants’ looking increased significantly toward the target object for the unaccented pronunciations, $t(19) = 4.26, p < .001, d = .953$, but was at chance levels for the random pronunciations, $t(19) = 1.9, p = .850$.

To determine whether infants’ behavior in this experiment was different from the behavior of infants in the Other-race speaker condition of Experiment 1, we conducted a mixed-measures ANOVA, with the within-subject factors block and word type, and the between-subject
factor experiment (only the Other-race condition was included for Experiment 1). This ANOVA revealed a main effect of block, \( F(1, 38) = 5.37, p = .026, n^2 = .124 \) and a significant word type X experiment interaction, \( F(1, 38) = 6.01, p = .019, n^2 = .137 \). No other effects were significant, \( ps > .083 \). When each block was considered separately, for block 1, we found no statistical differences between experiments, \( ps > .458 \). However, for block 2, we found a main effect of experiment, \( F(1, 38) = 5.28, p = .027, n^2 = .122 \), and a significant experiment X word type interaction, \( F(1, 38) = 7.53, p = .009, n^2 = .165 \).

Therefore, as in the Other-race speaker condition of Experiment 1, infants in Experiment 2 did not show significant recognition of either type of pronunciation in the first block of testing. However, in contrast to Experiment 1, infants in Experiment 2 who heard random mispronunciations also did not recognize the pronunciations after exposure.

4. Experiment 3

Because this is the first demonstration of such visually mediated effects on toddlers’ word processing, in Experiment 3 we replicate our findings from the Other-race speaker conditions of Experiment 1 and Experiment 2, using a different other-race speaker. This time, we chose a speaker whose non-Caucasian ethnicity was more visually salient.

4.1. Method

4.1.1. Participants

Forty 16-month-old infants were tested (20 females; mean age: 16 months 1 day; age range: 15;18–16;17 days). Three additional participants were tested, but not included due to non-completion (1), and failure to attend to both objects during the baseline period for at least half of each trial type in each block of trials (2). As in Experiments 1 and 2, participants had minimal exposure to people who spoke a foreign language, had an accent, or were of a different race, and the amount of exposure was similar across conditions (Systematic accent condition: 4.3%, 7.2%, and 6.0%, respectively; Random pronunciations condition: 4.5%, 8.4%, and 8.7%, respectively).

4.1.2. Stimuli

4.1.2.1. Audio stimuli. The same six highly familiar words from the previous two experiments were used. Half the infants heard the words with a systematic vowel shift (identical to that of Experiment 1), while the other half heard random pronunciations (identical to that of Experiment 2). A female native speaker of English (same speaker as in Experiment 2) produced all test stimuli. Once again, each version was produced naturally in infant-directed speech in the context of two sentences, “Do you see the X” or “Find the X”.

4.1.2.2. Visual stimuli. The test trials were the same as in the previous two experiments with the substitution of a new other-race speaker (see Fig. 1).

4.1.3. Procedure

Same procedure used in Experiment 1 and 2.

4.1.4. Coding of looking times

Same procedure used in Experiment 1 and 2.

4.2. Results

As in previous experiments, trials in which infants did not look at both objects during the baseline period (or at either object during the test period) were not included in the analyses. Across conditions, there was no difference in the percentage of discarded trials, \( t(38) = .16, p = .874 \), and no difference in the proportion of time that infants spent looking at the familiar object during baseline, \( t(38) = .81, p = .423 \). Fig. 6 displays the time course of infants’ looking during the test period.
In the Systematic condition, there is no clear differentiation of the looking curves at any point during the test phase, as in the Other-race condition of Experiment 1. In the Random condition, infants differentiate between the two pronunciation types very early on and this remains stable throughout test.

Test-baseline difference scores are displayed in Fig. 7. A mixed measures ANOVA with the within-subject factor of word type (Unaccented vs. Accented) and a between-subject factor of condition (Systematic accent vs. Random pronunciations) found a main effect of condition, $F(1, 38) = 5.17, p = .029, \eta^2 = .120$, and a significant word type $X$ condition interaction, $F(1, 38) = 9.22, p = .004, \eta^2 = .195$, but no effect of word type, $F(1, 38) = 3.12, p = .087, \eta^2 = .076$.

For each condition, paired sample $t$-tests comparing unaccented pronunciations to the accented pronunciations were run. In the Systematic accent condition, there was no difference across word types, $t(19) = .83, p = .416$; however, in the Random pronunciations condition, there was a significant difference between the unaccented and randomly pronounced words, $t(19) = 3.72, p = .001, d = .853$. One-sample $t$-tests against chance (zero change) showed that in the Systematic accent condition, infants’ looking increased significantly to the familiar object for both word types (unaccented: $t(19) = 2.90, p = .009, d = .684$; accented: $t(19) = 3.32, p = .004, d = .742$). In the Random pronunciations condition, infants increased their looking to the familiar object for the unaccented words, $t(19) = 3.85, p = .001, d = .861$, but not the randomly pronounced words, $t(19) = 1.48, p = .154$.

As in the previous experiments, we also conducted analyses including test block as a factor to explore changes over the experiment. For the Systematic accent condition, a repeated measures ANOVA with within-subjects factors block and word type revealed a main effect of block, $F(1, 19) = 7.28, p = .014, \eta^2 = .277$. No other effects were significant, $ps > .462$. For the first block, a paired sample $t$-test revealed no difference between the unaccented and accented pronunciations, $t(19) = .76, p = .456$. Infants did not show a significant change in looking to the target for either the unaccented pronunciations, $t(19) = .06, p = .950$, or the accented pronunciations, $t(19) = 1.08, p = .294$. In the second block, a paired sample $t$-test also revealed no difference between the unaccented and accented pronunciations, $t(19) = .16, p = .879$. However, in this block, infants’ looking increased significantly toward the target object for the unaccented pronunciations, $t(19) = 3.43, p = .003, d = .767$, and the accented pronunciations, $t(19) = 3.54, p = .002, d = .791$.

For the Random pronunciations condition, a repeated measures ANOVA with within-subjects factors block and word type revealed a main effect of word type, $F(1, 19) = 12.13, p = .002, \eta^2 = .390$, and a main effect of block, $F(1, 19) = 13.16, p = .002, \eta^2 = .409$, but no word type $X$ block interaction, $F(1, 19) = 1.28, p = .272$. For the first block, a paired sample $t$-test revealed no difference between the unaccented and random pronunciations, $t(19) = 1.52, p = .146$. Infants did not show a significant change in looking to the target for the unaccented pronunciations, $t(19) = .265, p = .794$, but for the random pronunciations looking increased to the distractor, $t(19) = 2.16, p = .044, d = .483$. In the second block, a paired sample $t$-test revealed a significant difference in how infants interpreted the unaccented and random pronunciations, $t(19) = 3.20, p = .005, d = .699$. Infants’ looking increased significantly toward the target object for the unaccented pronunciations, $t(19) = 5.23, p < .001, d = 1.169$, but was at chance levels for the random pronunciations, $t(19) = .26, p = .801$.

Therefore, in both conditions, infants did not recognize either type of pronunciation in the first block of testing. However, infants who heard systematic pronunciations (but not random mispronunciations) did recognize the pronunciations after some exposure. Thus, this experiment replicates the findings of the previous two experiments, confirming that infants initially fail to recognize even familiar pronunciations in the presence of an other-race speaker. After exposure to the speaker, however, infants recognize words produced in a systematic accent (whether familiar or not).

5. Discussion

Do infants have race-based expectations about how speakers produce words? In three experiments we explored how speaker race impacts infants’ recognition of words produced in a familiar or unfamiliar accent. In particular, we explored infants’ initial expectations for same-race and other-race speakers’ word pronunciations, and what they learned about those speakers’ pronunciations over time. Overall, we found that infants interpreted the same words differently depending on the speaker’s race. In Experiment 1, infants who viewed a same-race speaker recognized only familiar versions of words. In contrast, infants who viewed an other-race speaker recognized the words both when they were produced in a familiar and unfamiliar accent. Experiment 2 further demonstrated that, following exposure, infants did not simply accept any similar-sounding variant of a word from an other-race speaker, but rather, only recognized words produced with a systematic accent. Experiment 3 replicated the findings of Experiment 1 and 2 with a different other-race speaker. These results provide the first evidence that unspoken properties of speakers, such as race, influence infants’ speech processing.

The fact that infants in the same-race condition recognized only familiar pronunciations of words replicates the findings of multiple studies in which no social information about the speaker was provided (Mulak et al., 2013; Van Heugten & Johnson, 2014; White & Aslin, 2011). In fact, in this condition, infants’ initial bias was to assume that the novel pronunciations referred to new objects, consistent with a large body of research that has demonstrated that young language
learners interpret novel wordforms as labels for novel objects (e.g., Golinoff, Mervis, & Hirsh-Pasek, 1994; Halberda, 2003; Markman, 1989; Markman, 1990; Merriman & Bowman, 1989). It is important to note that in block 2, infants no longer showed this kind of disambiguation response, suggesting that over time they were learning the accentuated pronunciations from the same-race speaker. We return to this point below.

In striking contrast, infants were initially unsure about how to interpret words from other-race speakers. Across experiments, infants were variable in their interpretation of the accentuated versions of the words in block 1, and they never recognized the familiar versions in block 1. The fact that infants did not recognize familiar pronunciations is particularly interesting, given that infants this age reliably map known words to target objects when there is no information about the speaker’s appearance. The contrast between the patterns for the same- and other-race speakers in block 1 has the intriguing implication that infants’ default expectations in the absence of any visual speaker information correspond to their expectations about a same-race speaker.

Where do these expectations come from? Recent work on listeners’ adaptation to new speakers has explored how adults integrate prior beliefs with new data (i.e., speakers’ productions) during learning (Kleinschmidt & Jaeger, 2015). Our data provide critical insights from the youngest language users into the role of experience in generating these prior beliefs about speakers. We suggest that infants’ word representations are tied to the speakers who say them (consistent with proposals that linguistic and social information are linked in adult speech perception (e.g., Hay & Drager, 2007; Sumner, Kim, King, & McGowan, 2014). The infants in our sample had very little exposure to other races and accents and, therefore, had mostly heard these familiar words produced in a particular way by same-race speakers. As a result, they appear to have linked those pronunciations with same-race speakers, leading to an expectation that new same-race speakers would produce words in the same way. In contrast, encountering an other-race speaker appeared to trigger a different process. With little experience to draw on, infants did not recognize the familiar pronunciations, suggesting that their expectations for familiar-race individuals did not extend to unfamiliar-race individuals. If infants do indeed link social information, like race, to linguistic representations, we predict that different experiences may lead to different outcomes. For example, monolingual/dialectal biracial infants may be less likely to use race as a linguistic marker, as their experience has been that physically different types of people speak in the same way.

Importantly, infants’ responses for the other-race speaker across experiments demonstrates that they paid attention to the vowels produced by the other-race speaker (contrary to claims that adult listeners may simply perceive speech less veridically for other-race speakers, e.g., Kang & Rubin, 2009; Kang & Rubin, 2014). If they had not, the same pattern of results would have been found across experiments. Instead, infants eventually recognized the atypical pronunciations only when the vowels differed systematically from the familiar accent. One possibility consistent with these findings is that infants have the expectation that other-race speakers’ productions be linguistically plausible (i.e., not deviate randomly from the native accent). Indeed, in cases of real-world sociolinguistic variation, phonetic differences across accents are often systematic across multiple (though not necessarily all) lexical items. A second possibility is that, although they did not recognize the words with random phonemic changes, infants did not expect or learn a systematic shift either. Instead, they may have expanded their /æ/ category boundaries to accommodate the /æ/ pronunciations (Schmale, Cristia, & Seidl, 2012; Schmale, Seidl, & Cristia, 2015). Finally, it is also possible that, even though infants were unable to learn the random pronunciations over the course of this experiment, they would be able to learn these random pronunciations given sufficient exposure. Regardless, the difference between the conditions show that (a) infants attended to the specific vowels for both speakers and (b) infants have different expectations about whether familiar-race and unfamiliar-race speakers will produce words in ways they have heard them in the past.

As noted above, infants in the familiar-race condition did not show a disambiguation response (mapping the accented labels onto the novel objects) in the 2nd block. The infants in the unfamiliar-race condition (with the exception of block 1 in the random condition of Experiment 3) did not show one at all. Previous work has demonstrated that infants and children tend to map novel labels to novel objects (Halberda, 2003; Markman, 1989; Markman, 1990; Merriman & Bowman, 1989; Shukla, White, & Aslin, 2011). However, infants and children do not show a disambiguation response across languages or accents (Au & Glusman, 1990; Weatherhead & White, 2016). This suggests that children’s typical disambiguation response may be due to a more nuanced understanding of the mapping between words and referents. In particular, infants may have an implicit understanding that new words refer to new referents only within a language community (Clark, 1990; Clark, 2007). In the current study, the unfamiliar-race face may have signaled to infants that the speaker was not a member of their language community, allowing them to rapidly learn the mapping between the accentuated words and familiar referents. The familiar-race speaker, in contrast, may have been initially treated as part of the infants’ language community (leading to a strong disambiguation response for accentuated pronunciations in block 1). However, over time, infants began to learn the accentuated pronunciations from this speaker as well (and perhaps, implicitly, that this speaker was a member of a different language community).

Finally, although we chose the age of 16-months in the current study, we believe that infants may very well develop these links between social and linguistic variation earlier. Young infants are very sensitive to race (Anzures et al., 2010; Kelly et al., 2007; Kelly et al., 2009) and speech processing and face processing are intertwined (e.g., Kuhl & Meltzoff, 1982; Lewkowicz & Hansen-Tift, 2012; Teinonen, Aslin, Alku, & Caihra, 2008). We chose to use race because of infants’ early sensitivity to this property of individuals. However, we do not believe that these effects are specific to race. Infants also appear to be able to reason about two individuals’ social relationships (e.g., Kuhlmeier, Wynn, & Bloom, 2003; Spokes & Spelke, 2016; Spokes & Spelke, 2017). More abstract social properties such as these are also closely linked to linguistic variation (Hay & Drager, 2007; Niedzielski, 1996). Future work should explore whether the use of these more abstract social properties during language processing emerges on the same time course as race. It is possible that only some, more perceptually salient properties, like race and gender, are considered relevant dimensions of variation early in development, with other, more abstract properties emerging later (Foulkes & Docherty, 2006).

Adults have very strong associations between accent and race (McGowan, 2015; Rubin, 1992). We demonstrate for the first time that infants also possess some expectations about accent and race: (1) familiar-race speakers will pronounce words in familiar ways, and (2) unfamiliar-race speakers may not pronounce words in familiar ways. These basic assumptions may set the stage for learning the associations between specific unspoken speaker characteristics and speech properties that are observed in adults.

---

3 It is also important to note that, in fact, infants ultimately recognized both the unaccented and accented pronunciations from the unfamiliar-race speaker. This in itself constitutes a failure to disambiguate distinct wordforms, even for a single speaker. The same pattern of results was found in White and Aslin (2011) and in Maye, Aslin, and Tanenhaus (2008), where, after training on a novel accent, toddlers and children recognized both familiar pronunciations of those words and pronunciations consistent with the accent (again, from a single speaker). In the present case, where infants receive the same amount of evidence for both types of pronunciations, they learn both. This suggests that infants may not have trouble learning multiple forms for a single object when provided with sufficient evidence.
Acknowledgments

The authors would like to thank Carolyn Baer, Jasmine Taylor, and Luxsi Sribaskaran for providing stimuli, the research assistants who helped with coding the data, and all of the families and infants who participated. This work was funded by an operating grant from the Natural Sciences and Engineering Research Council of Canada awarded to K.S.W.

Appendix A

A.1. Acoustic measurements of stimuli

Note that for Experiments 2 and 3, the randomly pronounced words are not included (Unaccented refers to /a/ productions and Accented refers to /æ/ productions).

<table>
<thead>
<tr>
<th>Word</th>
<th>Word Type</th>
<th>Speaker 1 (Exp. 1)</th>
<th>Speaker 2 (Exp. 2 &amp; 3)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>F1</td>
<td>F2</td>
</tr>
<tr>
<td>Ball</td>
<td>Unaccented</td>
<td>840</td>
<td>1087.5</td>
</tr>
<tr>
<td></td>
<td>Accented</td>
<td>1077</td>
<td>1721</td>
</tr>
<tr>
<td>Bottle</td>
<td>Unaccented</td>
<td>938.5</td>
<td>1182.5</td>
</tr>
<tr>
<td></td>
<td>Accented</td>
<td>1179.5</td>
<td>1906</td>
</tr>
<tr>
<td>Block</td>
<td>Unaccented</td>
<td>1003.5</td>
<td>1397.5</td>
</tr>
<tr>
<td></td>
<td>Accented</td>
<td>1255</td>
<td>1752.5</td>
</tr>
<tr>
<td>Car</td>
<td>Unaccented</td>
<td>1010</td>
<td>1347.5</td>
</tr>
<tr>
<td></td>
<td>Accented</td>
<td>1269.5</td>
<td>1774</td>
</tr>
<tr>
<td>Dog</td>
<td>Unaccented</td>
<td>869</td>
<td>1282</td>
</tr>
<tr>
<td></td>
<td>Accented</td>
<td>1094</td>
<td>1901</td>
</tr>
<tr>
<td>Sock</td>
<td>Unaccented</td>
<td>967.5</td>
<td>1579.5</td>
</tr>
<tr>
<td></td>
<td>Accented</td>
<td>1200.5</td>
<td>1901</td>
</tr>
<tr>
<td>MEAN</td>
<td>Unaccented</td>
<td>938</td>
<td>1213</td>
</tr>
<tr>
<td></td>
<td>Accented</td>
<td>1179</td>
<td>1826</td>
</tr>
</tbody>
</table>

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.cognition.2018.04.004.

References


