What’s a Foo? Toddlers Are Not Tolerant of Other Children’s Mispronunciations

Dana E. Bernier and Katherine S. White

1. Introduction

Recent research has suggested that toddlers both represent familiar words in considerable detail, and efficiently recognize them when the words are spoken by adults (Fernald, Perfors, & Marchman, 2006). The precision of toddlers’ processing is reflected in the fact that they detect even slight mispronunciations of familiar words (e.g., Swingley & Aslin, 2000). For example, toddlers look to an image of a baby less, and more slowly, when it is labeled as vaby, compared to when it is labelled correctly.

Not only do toddlers show processing costs for mispronunciations, but they also show graded sensitivity to the degree of mispronunciation (Mani & Plunkett, 2011; Ren & Morgan, 2011; Tamási et al., 2017; White & Morgan, 2008). In looking paradigms, toddlers progressively decrease their looking to the target object following labelling as the degree of mispronunciation increases (White & Morgan, 2008). For example, there may be a relatively small cost associated with a 1-feature phonetic change (e.g., the word shoe pronounced as foo), but a larger cost associated with a 2-feature change (e.g., shoe pronounced as voo), and an even larger cost for a 3-feature change (e.g., shoe pronounced as goo). This graded sensitivity has also been demonstrated using pupillometry, where the larger the degree of mispronunciation, the greater the change in pupil diameter (Tamási et al., 2017).

However, all of these studies assessed toddlers’ recognition of words produced by adult speakers, and adults are not the only ones speaking in children’s environments. Toddlers are also exposed to the speech of other children, and in some environments, in amounts similar to that of adult speech (Bernier & Soderstrom, 2016). Yet we know very little about how well young language learners process this kind of speech (but see Dodd, 1975). Given that other children’s speech likely represents a substantial source of input for at least some children (e.g., those with siblings or attending daycare), this represents a significant gap in our understanding of early language processing.

Nevertheless, we do know something about what children’s speech sounds like. Very young children’s productions are characterized by a number of

* Dana E. Bernier, University of Waterloo, dbernier@uwaterloo.ca. Katherine S. White, University of Waterloo, white@uwaterloo.ca. This work was supported by an operating grant from the Natural Sciences and Engineering Research Council to KSW.

phonological deviations from adult target pronunciations. These changes include (among others) substitutions of one sound for another (e.g., *fumb* for *thumb*), and omissions of sounds (or syllables) altogether (e.g., *nake* for *snake*), leading to low intelligibility for naïve adult listeners (Flipsen, 2006; Hodson & Padden, 1981). By the age of 4, children’s productions are (on average) quite intelligible for even naïve adult listeners, although those with more experience listening to children are more accurate and reliable in their judgments of individual sounds produced by children (Munson, Johnson, & Edwards, 2012). However, even once children are producing sounds that adults judge to be in the correct phonological category, children continue to show less accuracy and more variability in their productions than adults. For example, school-aged children produce less accurate /l/ (Lin, Inkelas, McConnaughey, & Dohn, 2016), less distinguishable /s/ and /ʃ/ (Nissan & Fox, 2005; Nittrouer, 1995), and have a larger vowel space than adults (Hillenbrand, Getty, Clark, & Wheeler, 1995; Lee, Potamianos, & Narayanan, 1999). All of this suggests that child speech will be more difficult for a young language learner to process.

Given this probable difficulty, it is relevant to consider how toddlers do with other forms of non-canonical speech, such as accented speech. Studies of toddlers’ perception of accented speech have found that, while there are processing costs associated with initial exposure to an accent (e.g., van Heugten, Kreiger, & Johnson, 2015; White & Aslin, 2011), by 19 months toddlers are able to recognize intended referents after only brief exposure to the accent (Schmale, Cristia, & Seidl, 2012; van Heugten & Johnson, 2014; White & Aslin, 2011). These studies suggest that if child speech, like accented speech, is initially difficult for toddlers to process, experience hearing other children speak may help.

Experience hearing other children’s speech could facilitate processing in a number of ways. One way that experience may help is by allowing listeners to learn something about the sound categories children produce. For example, experience hearing more variable /s/ productions could help listeners learn about children’s category boundaries, similar to the way that exposure to ambiguous sounds embedded in words alters adults’ judgments of phoneme category boundaries (Norris, McQueen, & Cutler, 2003). Experience with the types of phonological changes common in child speech could also allow listeners to remap deviant sounds to their appropriate categories, so that, e.g., *bwush* is understood to be *brush*. This would be similar to the way in which toddlers (and adults) are thought to alter the mapping between phonological categories and the lexicon following exposure to speech containing systematic phoneme shifts (as in Maye, Aslin, & Tanenhaus, 2008; White & Aslin, 2011).

Alternatively, experience may help listeners to realize that for successful comprehension, they should pay less attention to the specifics of child speakers’ productions, and rely more on top-down knowledge about their referential intent (in other words, be more tolerant of deviations). It is clear that, with very young children’s productions (when they are not very intelligible), listeners rely heavily on context for comprehension. The same may be true for slightly older children’s speech as well. Indeed, recent work has shown that even young children modulate
their use of top-down information during word processing based on the reliability of the acoustic input (Yurovsky, Case & Frank, 2016). This type of tolerance when faced with child speech may require knowledge that children misarticulate sounds. Thus, toddlers who have experience with children’s speech may adopt a strategy of relying more heavily on context than phonetic content when processing children’s pronunciations.

In the current study, we were interested in toddlers’ perceptions of other children’s speech. In particular, we asked about their processing of both correctly pronounced and mispronounced words, and how this might be influenced by experience listening to other children. We presented toddlers with visual displays containing one familiar and one novel object and recorded their looking behaviour in response to instructions from a female, first-grade speaker directing them to look at one of the objects. The speaker either pronounced the labels for the familiar objects correctly, or with an onset mispronunciation of 1 to 3 phonetic features. Our motivation for presenting mispronunciations was two-fold. First, previous studies have documented toddlers’ sensitivity to the degree of mismatch between a pronunciation and the stored representation of a familiar word. This work provides a benchmark for looking at their sensitivity to phonetic detail in child productions. Second, the presence of phonetic mismatch in the context of an unfamiliar referent makes this a difficult task. In a task with only familiar referents and correct pronunciations, toddlers might successfully recognize words even if they do not pay much attention to phonetic detail (through a more general matching strategy). In contrast, in the present task, toddlers must determine whether a mispronunciation that differs minimally from the canonical representation of a familiar word is a variant of the familiar word or a new word. It is unclear whether toddlers apply the same criteria for adult and child speech, as the latter may be inherently more distant from the canonical even for correct pronunciations.

We envisioned at least three possible outcomes for toddlers as a group. The first was that, as a group, toddlers would have the same pattern of responses for a child speaker as they do for an adult speaker. In this case, they should show the same type of graded sensitivity to the degree of mispronunciation as they do for adult pronunciations, and no more tolerance for deviations than they have for an adult speaker. A second possible outcome was that toddlers would accept anything close to a correct label. Such an outcome might arise due to an inability to resolve differences between correct and mispronounced labels, or to a higher level of tolerance for deviations in child speech. A third possible outcome was that toddlers would have difficulty processing child speech in general, and would not only have more difficulty recognizing correct and novel labels than with adult speakers, but also have difficulty distinguishing the mispronunciations.

Given that experience is likely a mediating factor in toddlers’ perception of child speech, we not only looked at the overall group averages, but also considered separately toddlers with limited experience with other children (less than 5 hours per week on average) and toddlers with more experience (10 or more hours per week on average). Based on studies with accented adult speech, we expected that
toddlers with Low Experience might have difficulty processing child speech, while those with High Experience might not.

2. Method
2.1. Participants

A total of 28 monolingual English-learning toddlers (mean age 21.7 months; range 20.9 – 23.2) were recruited from the Kitchener-Waterloo area of Ontario, Canada. Ten additional toddlers were tested but not included due to technical issues with equipment (n = 3), fussiness/crying (n = 2), completing less than half of the trials successfully (n = 2), parental interference (n = 1), not successfully completing at least 2 correct label trials (n = 1), and performance on correct label trials more than 3.5 SD below the mean (n = 1).

2.2. Procedure

Toddlers were tested using the Intermodal Preferential Looking Procedure in a sound attenuated room. The child sat on the parent’s lap while the parent listened to music over circumaural headphones. In front of them was a 42-inch widescreen television and two hidden speakers located at the base of the television; both were connected to a computer in an adjacent room. The participants were monitored over closed-circuit video feed that was recorded for later off-line coding; the camera was centrally located beneath the television and hidden behind a black curtain.

Each trial consisted of images of 1 familiar object and 1 novel object presented on the left and right sides of the screen. Each trial began with the objects shown alone for 3 seconds; this was used as the baseline phase. Following the baseline phase, the audio stimulus (e.g., “Where’s the shoe”?) began to play. The images remained up for an additional 5 seconds from the start of the audio (for a total trial length of 8 seconds). We defined a 3-second naming phase that commenced 267ms after the start of the target word. The dependent measure was the change in the proportion of time toddlers spent looking at each object from the baseline to the naming phase.

Following the testing session, the parent completed a questionnaire on their child’s comprehension and production of the study items (both familiar and novel), and the amount of time their child interacted with other children (siblings, daycare, and playgroups) on a weekly basis.

2.3. Design

There were a total of 20 test trials, each with a unique familiar object – novel object pair. Of these 20 trials, the familiar object was labelled correctly in 4 trials, the novel object was labelled with a novel word (e.g., tibble) in 4 trials, and in the remaining 12 trials the familiar object was labelled with one of three types of mispronunciations (4 trials for each type). These mispronunciations consisted of
onset consonant changes of 1 – 3 features that resulted in either a non-word or a word toddlers are unlikely to know (see Table 1 for an example).

Each participant heard 1 type of pronunciation (correct, 1-, 2-, 3-feature change) for each familiar target object. The assignment of pronunciations to familiar objects was counterbalanced across participants. Novel label trials used different familiar object – novel object pairs, and were the same across all participants.

Table 1: Stimulus conditions with sample target stimuli

<table>
<thead>
<tr>
<th>Condition</th>
<th>Sample Target</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correct /ʃu/ shoe</td>
<td></td>
</tr>
<tr>
<td>1-feature Change /fu/ foo</td>
<td></td>
</tr>
<tr>
<td>2-feature Change /vu/ voo</td>
<td></td>
</tr>
<tr>
<td>3-feature Change /gu/ goo</td>
<td></td>
</tr>
</tbody>
</table>

2.4. Stimuli

Stimuli were recorded in a sound-attenuated room using a Sennheiser e945 microphone connected to a laptop via a blue icicle USB adaptor and recorded into PRAAT (Boersma & Weenink, 2015). All stimuli were produced by a female first grade student (aged 7;5). To elicit these productions, the child saw an image of the target object on a laptop and repeated the productions of a female adult speaking in an infant-directed speech register. A minimum of three productions were elicited for each target word and the best token was selected. Stimuli were later edited in PRAAT, and adjusted so that all tokens were of equal intensity.

2.5. Analysis

Looking behaviour was coded off-line using customized software at a rate of 30 frames per second (~33.33ms/frame). For each trial, the proportion of time toddlers spent looking at the familiar object (out of the total time spent looking at the two objects) was calculated for both baseline and naming phases. A difference score was then computed (Naming-minus-Baseline), which indicates how much toddlers changed their looking to the familiar object from baseline to naming. A positive score indicates that they increased their looking to the familiar object following naming, while a negative score indicates that they decreased their looking to the familiar object following naming. These difference scores were then averaged across trials within each condition (correct, 1-feature, 2-feature, 3-feature, novel), resulting in 5 data points for each participant.

For a trial to be included in the analysis, the participant needed to look at each of the objects for a minimum of 267ms during the baseline phase. This criterion resulted in 18.75% of trials being discarded (20.8% for the High Experience Group; 15.0% for the Low Experience Group). Participants also needed to attend to the objects for a minimum of 1 second during the naming phase. This criterion
resulted in 1.25% of trials being discarded (1.6% for the High Experience Group; 1.0% for the Low Experience Group).

If a child did not have a minimum of 2 trials in a condition that met the above criteria, they did not receive a score for that condition. This resulted in the loss of 5 data points across all participants: 2 in the 3-feature condition, and 3 in the novel condition (all 5 from toddlers in the High Experience Group). No participant was missing more than 1 data point.

3. Results
3.1. Overall

We first assessed baseline preferences across all trials. A one-sample $t$-test against chance (50%) showed that toddlers spent longer looking at the familiar objects during baseline ($59.5\% \ [SD = 5.5\%]$), $t(27) = 9.12$, $p < .001$. We next looked at performance on trials where the familiar object was labeled correctly. A one-sample $t$-test evaluating toddlers’ difference scores against 0 (no change from baseline) revealed that toddlers recognized these words, increasing their looking to the familiar object by $16.6\% \ (SD = 15.6\%)$ from baseline, $t(27) = 5.61$, $p < .001$ (Figure 1).

Second, to determine whether there was an effect of pronunciation type, we conducted a one-way ANOVA across the 4 conditions in which the familiar object was labeled (i.e., excluding novel label trials). This analysis revealed that toddlers’ responses were not the same across the different pronunciations, $F(3,75) = 10.65$, $p < .001$. In contrast to their performance with correctly pronounced labels, toddlers increased their looking to the familiar object by only $5.6\% \ (SD = 16.0\%)$ for a 1-feature change, $t(27) = 1.85$, $p = .076$, and decreased their looking to the familiar object by $3.0\% \ (SD = 17.3\%)$ for a 2-feature change, $t(27) = -0.92$, $p = .367$, and by $6.8\% \ (SD = 18.5\%)$ for a 3-feature change, $t(25) = -1.88$, $p =$
.072. Planned comparisons revealed that all 3 levels of change were significantly different from the correct condition, $t(27) = 2.43, p = .022$, $t(27) = 4.61, p < .001$, $t(25) = 5.54, p < .001$, respectively.

Third, we looked at whether toddlers were sensitive to the degree of mispronunciation via a trend analysis across the 3 levels of mispronunciation. This analysis revealed a significant linear trend, $F(1,25) = 8.08, p = .009$, indicating that toddlers are sensitive to the degree of mispronunciation in child speech.

Finally, we found that on novel label trials, toddlers reduced their looking to the familiar object from baseline by 5.1% ($SD = 21.0%$), however, this difference was not significant, $t(24) = -1.22, p = .235$. We next turn to a breakdown of the data based on toddlers’ experience with other children.

### 3.2. By Experience Group

Participants were divided into 2 groups based on parental estimates of their exposure to other children. Toddlers in the Low Experience Group (n = 10) were those who spent less than 5 hours weekly with other children (mean of 3.2 hours/week [$SD = 1.5$] in playgroups; no siblings or daycare). Toddlers in the High Experience Group (n = 18) were those who spent 10 or more hours weekly with other children; 13 attended daycare at least 8 hours/week (mean of 32.4 hours/week [$SD = 12.2$] in daycare and playgroups; 5 also had siblings), 2 only attended playgroups (mean of 12.5 hours/week [$SD = 3.5$]; no siblings), and 3 were primarily exposed to siblings (with an additional mean of 5.3 hours/week [$SD = 1.5$] with other children).

#### 3.2.1. High Experience Group

As in the overall analysis, we first assessed baseline preferences across all trials. A one-sample $t$-test against chance (50%) showed that toddlers spent longer looking at the familiar objects during baseline (61.0% [$SD = 5.1%$]), $t(17) = 9.13, p < .001$. We next looked at performance on correctly pronounced labels, and found that the High Experience Group significantly increased their looking to the familiar object by 15.9% ($SD = 13.9%$) from baseline, $t(17) = 4.87, p < .001$ (Figure 2, left side), indicating that they recognized the correctly pronounced labels.

Next we conducted a one-way ANOVA across the 4 different pronunciations of the familiar object (i.e., excluding novel label trials), and found that toddlers’ responses were not the same across pronunciations, $F(3,45) = 7.27, p < .001$. In contrast to correctly pronounced labels, toddlers increased their looking to the familiar object by only 3.0% ($SD = 19.0%$) for a 1-feature change, $t(17) = 0.68, p = .504$, and decreased their looking to the familiar object by 3.7% ($SD = 19.0%$) for a 2-feature change, $t(17) = -0.83, p = .420$, and by 9.9% ($SD = 17.2%$) for a 3-feature change, $t(15) = -2.32, p = .035$. Planned comparisons revealed that all 3 levels of mispronunciation were significantly different from the correct
condition, \( t(17) = 2.13, p = .049, \) \( t(17) = 4.01, p = .001, \) \( t(15) = 4.71, p < .001, \) respectively. There was also a significant linear trend across the 3 levels of mispronunciation, \( F(1,15) = 5.26, p = .037. \)

And finally, toddlers in the High Experience Group decreased their looking to the familiar object by 8.2% \( (SD = 24.7) \) when they heard a novel label, though this difference was not significant, \( t(14) = -1.28, p = .221. \)

![Figure 2: Toddlers’ change in looking to the familiar object (naming-minus-baseline) by experience group; error bars represent standard error.](image)

### 3.2.2. Low Experience Group

As with previous analyses, we first assessed baseline preferences across all trials. A one-sample \( t \)-test against chance (50%) showed that the Low Experience Group also spent longer looking at the familiar objects during baseline (56.9\% \( [SD = 5.5\%] \)), \( t(9) = 3.96, p = .003. \) This bias for the familiar object was marginally less than that of the High Experience Group, \( t(26) = 1.98, p = .058. \)

We next looked at performance on correctly pronounced labels, and found that toddlers in the Low Experience Group significantly increased their looking to the familiar object by 17.7\% \( (SD = 19.1\% \) from baseline, \( t(9) = 2.93, p = .017 \) (Figure 2, right side), indicating that they also recognized the correctly pronounced labels.

Next we conducted a one-way ANOVA across the different pronunciations for the familiar object (i.e., excluding novel label trials), and found again that toddlers’ responses were not the same across pronunciations, \( F(3,27) = 2.47, p = .030. \) However, their pattern of responses differed from that of the High Experience Group. Toddlers in the Low Exposure Group significantly increased their looking to the familiar object by 10.1\% \( (SD = 7.2\% \) for a 1-feature change, \( t(9) = 4.44, p = .002, \) and decreased their looking to the familiar object by 1.8\% \( (SD = 14.7\% \) for a 2-feature change, \( t(9) = -0.38, p = .715, \) and by 1.8\% \( (SD = 20.4\% \) for a 3-feature change, \( t(9) = -0.28, p = .785. \) Planned comparisons
revealed that the 1-feature change condition did not differ from the correct condition, $t(9) = 1.126$, $p = .289$, while larger changes did, $t(9) = 2.33$, $p = .045$, $t(9) = 2.89$, $p = .018$, respectively. Additionally, there was neither a linear, $F(1,9) = 2.67$, $p = .137$, nor a quadratic, $F(1,9) = 0.92$, $p = .092$, trend across the 3 levels of mispronunciation.

And finally, toddlers in the Low Experience Group did not change their looking from baseline after hearing a novel label ($M = -0.5\%; SD = 13.8\%$), $t(9) = -0.12$, $p = .904$.

4. Discussion

Overall, toddlers successfully mapped correctly pronounced labels to the appropriate referents. When the labels were mispronounced, toddlers’ looking behaviour did not differ from chance, showing that they were sensitive to each type of change (even the 1-feature changes). Further, toddlers showed graded sensitivity to the degree of change such that their looking for the small 1-feature changes trended towards the familiar object, while their looking for the larger 3-feature changes trended towards the novel object. These patterns are quite similar to those previously found for toddlers’ processing of adult speech (White & Morgan, 2008), with the caveat that toddlers showed a more significant disruption for a 1-feature change in the present study. Thus, overall, toddlers appear quite sensitive to phonetic content of children’s speech.

However, the pattern of looking for mispronounced labels differed according to toddlers’ experience interacting with other children. Toddlers who spent more than 10 hours a week with other children showed a pattern similar to the one described above, with the exception that the 3-feature mispronunciations were mapped to the novel object. Therefore, toddlers who have experience interacting with other children show fine sensitivity to the properties of child speech.

In contrast, toddlers who spent less than 5 hours a week with other children had a different pattern of responses. First, they showed no significant cost for the 1-feature changes. Second, overall, they did not show a graded pattern of sensitivity to the size of the mispronunciation. It remains to be seen whether these findings will hold with a larger sample size (testing is ongoing in this group).

With respect to the possible roles of experience, these results suggest that toddlers with high levels of experience hearing child speech do not accept greater deviations from a child speaker than they would from an adult speaker. In other words, they do not seem to rely on top-down knowledge at the expense of phonetic detail. Additionally, because the target sounds in children’s speech are more variable (acoustic analyses of the present stimuli are ongoing), toddlers’ fine sensitivity to the contrasts we presented suggests that they have learned something about the variability of children’s speech categories (as in adults’ perceptual learning of accents and other atypical speech; Norris, McQueen, & Cutler, 2003). In contrast, the failure of the low experience group to distinguish correct and 1-feature mispronunciations, together with the lack of a graded effect, suggests that they had more difficulty processing the child speech.
A secondary finding was that toddlers who had little exposure to other children did not change their looking from baseline after hearing a novel label. In other words, this group did not map novel labels to novel objects, a disambiguation response typically seen in monolingual toddlers as young as 17-months when a novel label is presented by an adult (Byers-Heinlein & Werker, 2009; Halberda, 2003; White & Morgan, 2008). It is not clear whether this is due to processing difficulties or to toddlers placing less trust in information conveyed by child speakers (Jaswal & Neely, 2006), however a similar failure to learn words from a child speaker was found by Wang and Seidl (2015). We have also replicated this finding in another study that used the same correct and novel label recordings as the current study, but did not include any mispronunciations (Figure 3).

![Figure 3: Toddlers' change in looking to the familiar object (naming-minus-baseline) for familiar-novel object pairs in a separate study that used the same correct and novel audio recordings as the current study, but with no mispronunciations.](image)

In that study, when 21- to 23-month-olds were presented with 1 familiar object and 1 novel object, all experience groups successfully mapped familiar labels to familiar objects. However, when they heard a novel label, performance differed depending on their level of experience with other children. As in the current study, toddlers with low experience (i.e., no siblings or daycare and limited playgroup activities) did not change their looking from baseline. Interestingly, toddlers whose high experience came primarily from exposure to siblings also did not change their looking from baseline. Only those whose experience came from a variety of speakers (e.g., daycare) successfully mapped novel labels to novel objects. This highlights the potential role of speaker...
variability in toddlers’ processing of novel words (Bergmann, ten Bosch, Fikkert, & Boves, 2015). It is also possible that other factors correlated with high vs. low experience with child speakers are responsible for this difference in novel word processing. For example, it could be that toddlers who differ in their exposure to other children also differ in vocabulary size or in the amount of child-directed speech they hear, factors that have been shown to affect the use of a disambiguation strategy (Bion, Borovsky, & Fernald, 2013).

In conclusion, toddlers are not only exposed to adult speakers. The current study demonstrates that, even for speech produced by a relatively clear 7-year-old child speaker, toddlers’ ability to process this speech depends on their previous experience with children. A full picture of the impact this has on toddlers’ language development will require additional research, as some studies of toddlers in other cultures (such as the Mayan Yucatec) indicate that their vocabulary size is best predicted by the quantity of speech directed at them from adults, and not from the children they spend most of their time with (Shneidman & Goldin-Meadow, 2012).

Even so, this study highlights the role of experience in how toddlers process the speech of an unfamiliar child, and raises important questions about what specific aspects of experience matter. Perhaps it is not the pure amount of experience with other children, but the number of different speakers encountered. Or perhaps there is some other factor correlated with experience that shapes processing. In future research, we will continue to explore the role of experience in toddlers’ processing of child speech.

References


Norris, Dennis, McQueen, James M., & Cutler, Anne (2003). Perceptual learning in speech. *Cognitive Psychology, 47*, 204-238.


