Toddlers recognize words in an unfamiliar accent after brief exposure

Rachel Schmale\textsuperscript{1,+}, Alejandrina Cristia\textsuperscript{2,+} and Amanda Seidl\textsuperscript{3}

\textsuperscript{1}. Department of Psychology, North Park University, USA
\textsuperscript{2}. Department of Neurobiology of Language, Max Planck Institute for Psycholinguistics, Germany
\textsuperscript{3}. Department of Speech, Language, and Hearing Sciences, Purdue University, USA

Abstract

Both subjective impressions and previous research with monolingual listeners suggest that a foreign accent interferes with word recognition in infants, young children, and adults. However, because being exposed to multiple accents is likely to be an everyday occurrence in many societies, it is unexpected that such non-standard pronunciations would significantly impede language processing once the listener has experience with the relevant accent. Indeed, we report that 24-month-olds successfully accommodate an unfamiliar accent in rapid word learning after less than 2 minutes of accent exposure. These results underline the robustness of our speech perception mechanisms, which allow listeners to adapt even in the absence of extensive lexical knowledge and clear known-word referents.

Introduction

Unfamiliar accents incur a processing cost for listeners, ultimately compromising word recognition for both adults (Clarke & Garrett, 2004) and infants (Schmale & Seidl, 2009). Because early phonological and lexical representations are shaped by language experience (e.g. Jusczyk, 1997), it is not unexpected that listeners may initially find it difficult to understand speakers who talk with an unfamiliar accent. However, in most societies, children are bound to come across speakers who do not talk exactly like their parents do, including people from different economic backgrounds (Labov, 2001), different ethnic identities (Thomas, 2007), individuals with a different first language (Ghurczek, Newheimer & Dovidio, 2011), children who may not be able to produce target phonemes, and even adults with speech disorders. In all of these cases, the phonetic/phonological form of words will likely deviate from the forms the child is accustomed to hearing from caregivers, to a greater or lesser extent. Thus, in instances where children are exposed to pronunciation differences, do unfamiliar accents completely impede word recognition?

The struggle with unfamiliar accents has been documented throughout development, with most work concentrating on unfamiliar foreign and dialectal accents. For example, 9-month-old infants fail to generalize newly heard words in continuous speech across native and unfamiliar accents (dialect: Schmale, Cristià, Seidl & Johnson, 2010; foreign: Schmale & Seidl, 2009), 15-month-olds show a preference for high frequency familiar words in their native dialectal accent, but not an unfamiliar one (Best, Tyler, Gooding, Orlando & Quann, 2009), and 24-month-olds cannot recognize a recently learned word when spoken in an unfamiliar foreign accent (Schmale, Hollich & Seidl, 2011; for work on systematic mispronunciations of familiar words, see e.g. Swingley & Aslin, 2000, 2002). In fact, even older children (e.g. Floccia, Butler, Girard & Goslin, 2009a) and adults (Munro & Derwing, 1995; Van Wijngaarden, 2001) experience degraded word recollection accuracy and slowed processing speed for foreign-accented speech.

However, much work suggests that this processing disadvantage is rapidly modulated by experience (Bradlow & Bent, 2008; Clarke & Garrett, 2004; Gass & Varonis, 1984). For instance, foreign-accent processing costs in adults are dramatically reduced with as little as 1 minute of prior exposure (Clarke & Garrett, 2004; although see Floccia, Butler, Goslin & Ellis, 2009b). Moreover, exposure helps adults develop more general representations of the previously unfamiliar accent (Bradlow & Bent, 2008). Further, laboratory-learning studies suggest that unfamiliar accents are encoded as constraints on the mapping of sounds, which are induced from mismatches between the accented and stored lexical forms, and can later be applied to novel words (e.g. Kraljic & Samuel, 2005, 2007; Maye, Aslin & Tanenhaus, 2008). For example, in Maye \textit{et al.} (2008), listeners heard...
passages from the Wizard of Oz story where all tokens of one sound were produced as another target sound (e.g. ‘the weckud wetch of the wast’). In the context of this familiar phrase, these vowel deviations clearly indicate that the speaker has a different accent. To resolve the perceptual discord, adult listeners (implicitly) abstracted a generalization such as [e] maps to /æ/, and extended it to untrained words.

If children can implement the same strategies, we might expect them to also rapidly compensate for unfamiliar accents. Recent work by White and Aslin (2011) suggests that toddlers can utilize top-down knowledge to guide accommodation of mispronunciations of familiar words. Specifically, 18- to 20-month-olds were trained on three familiar label–object pairings (approximately 24 repetitions), all of which evidenced a single sound change in the label (e.g. they heard /dæg/ while a picture of a dog loomed on a screen; they heard /bæl/ while seeing a ball, etc.). At test, toddlers generalized this sound change to an untrained familiar object. For example, they looked at a sock when hearing /sæk/ (‘sack’) but not when hearing the word /slk/ (‘sick’); interestingly, they also accepted the pronunciation /sk/ (‘sec’). Thus, when given ample evidence, toddlers can learn rather (but not completely) specific patterns, and apply them to highly familiar words. While these toddlers were able to use top-down information to guide their accent accommodation strategy, there are considerable differences between the controlled accent training provided in White and Aslin (2011) and the accent exposure that children are likely to encounter in their natural environments. In reality, toddlers may more often encounter accented talkers who display several accentual features, rather than a single phonetic change. These talkers may provide examples of their speech spread out over a sentence, with only a few repetitions of each phonetic target and without necessarily brandishing the object whose label is being produced. In such fluent situations, implementing a top-down strategy is challenging, as it requires sufficient world knowledge, vocabulary, and processing resources to bootstrap from relatively few and imperfect matches in a complex speech stream. Thus, in the present work, we sought to face toddlers with exposure to a natural accent in the context of fluent speech, without providing access to explicit, top-down lexical bootstrapping cues.

To investigate whether and how toddlers can learn to accommodate unfamiliar accents, we tested 24-month-olds’ ability to recognize a newly learned word when spoken in a foreign accent (Spanish-accented English) after different types of exposure to variable speech. The word-learning task used in our study was based on recent work on the perception of Spanish-accented English by toddlers raised in the Midwest (Schmale et al., 2011). Spanish-accented English differs substantially from the children’s North Midland American (NMA) English dialect on several phonological levels (e.g. Jongman & Wade, 2007), and is therefore likely to pose a considerable challenge to toddlers unaccustomed to hearing it.

Indeed, 24-month-olds trained on a novel word by an NMA speaker failed to recognize this newly learned word when spoken by a Spanish-accented speaker (Schmale et al., 2011). To assess what types of experience might improve toddlers’ performance, we preceded this word-learning task with a 2-minute Exposure phase, during which toddlers heard four passages by single or multiple, NMA (‘local’) or Spanish-accented (‘foreign’) speakers. These exposure passages involved some word repetition and only a few known content words, but no anchors or targets that directly mapped on to the words to be learned. Thus, the Exposure phase was designed to approximate naturalistic exposure to native and Spanish-accented English. Participants were randomly assigned to one of four different Exposure conditions: Single Local, Multiple Local, Single Foreign, Multiple Foreign, as illustrated in Figure 1 in the Methods section. In short, the Single Local condition is a replication of Schmale et al. (2011) with the addition of the Exposure phase spoken by the same talker who produces the subsequent word-learning task. In the three other Exposure conditions, the speaker(s) are not the same as the one in the subsequent word-learning task; they all prepare toddlers for a speaker change, thereby reducing the ‘surprise’ effect during subsequent testing. It was predicted that exposure to the foreign accent (Single Foreign, Multiple Foreign) would have a positive effect for learning (Clarke & Garrett, 2004). In addition, the number of speakers in each Exposure condition was manipulated since multiple talkers are likely to be more variable than a single talker, which could positively impact performance by promoting more abstract representations (Bradlow & Bent, 2008; Rost & McMurray, 2010), and increasing attention by virtue of being more interesting.

Method

Participants

Twenty-two monolingual English-learning 24-month-olds were included in each of four conditions (n = 88). Participants’ age, sex, and productive vocabulary size, as estimated by the short form A of the MacArthur-Bates Communicative Developmental Inventory: Level II Vocabulary Checklist (CDI; Fenson, Pethick, Renda, Cox, Dale & Reznick, 2000), are reported in Table 1; a CDI was not available for one child. An additional 19 children were not included because of the following reasons: crying, being overly restless, or failing to finish the study (13), equipment or experimenter error (six), or foreign language exposure (one).

Stimuli and procedure

All toddlers were tested using the Preferential Looking Procedure (Fagan, 1971; Spelke, 1979), in which the toddler sits on a caregiver’s lap and watches images
projected onto a video screen while an experimenter videotapes their looking patterns. The experiment began with the Exposure phase followed by a word-learning task involving two repetitions of the same Training-Test block. Since the Training-Test blocks were presented two times sequentially, they are referred to as first and second block. Thus, the second block is an exact replica of the first block. This timeline is represented in Figure 1.

Toddlers were randomly assigned to two experimental orders that were counterbalanced for test trial order, presentation side, and label–object pairings. During the Exposure phase, all children heard the same four passages drawn from work on infant word segmentation and identical to those in Schmale and Seidl (2009; see Table 2). None of these passages contained words used in the subsequent word-learning phase. To reduce attrition, the passages were accompanied by an unrelated visual stimulus: a (silent) Curious George cartoon. The speaker(s) that produced the passages differed across four Exposure conditions to which participants were assigned (see Figure 1). In the Single Local condition, one native speaker of NMA English produced the passages, who was the same speaker used in the Training of the word-learning phase. In the Single Foreign condition, one speaker of Spanish-accented English produced the passages, who was the same speaker used in the Test of the word-learning phase. Multiple measures were used to determine that the voices of the speakers used in our Single Local and Single Foreign conditions were highly similar (see Schmale & Seidl, 2009). In the Multiple conditions, different speakers produced each one of the four passages: either native speakers of NMA English (Multiple Local), or speakers of Spanish-accented English (Multiple Foreign), but these four

Table 1 Characteristics of participants in each exposure condition: Age mean (and range), Number of males/females, and productive vocabulary mean (and range) according to the short form of the CDI

<table>
<thead>
<tr>
<th>Condition</th>
<th>Age (M–F)</th>
<th>Sex (M/F)</th>
<th>Vocabulary Mean (and range)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single Local</td>
<td>23.93 (23.36–24.47)</td>
<td>18/4</td>
<td>49.36 (14–100)</td>
</tr>
<tr>
<td>Single Foreign</td>
<td>24.05 (23.55–24.54)</td>
<td>16/6</td>
<td>57.05 (16–100)</td>
</tr>
<tr>
<td>Multiple Local</td>
<td>23.98 (23.49–24.74)</td>
<td>11/11</td>
<td>46.81 (6–93)</td>
</tr>
<tr>
<td>Multiple Foreign</td>
<td>23.96 (23.49–24.77)</td>
<td>16/6</td>
<td>44.05 (2–85)</td>
</tr>
</tbody>
</table>

Figure 1 Each toddler heard only one of the four possible Exposure passages accompanied by an unrelated cartoon. The Exposure conditions differed in accent and on how many talkers produced the passages. After Exposure, toddlers were tested on a word-learning task with two repetitions of the same Training-Test block (only one block is shown here). Notice that all toddlers, regardless of whether they heard one or multiple talkers in Exposure, were trained with the same North Midland American talker and tested with the same Spanish-accented talker. The objects, side of presentation, and labels in the Training-Test block were counterbalanced across toddlers within each Exposure condition.
The candle in the kitchen was almost melted. So Annie bought another candle at the stationary store. She came home and put away the old candle. Fran gave that candle to you later. Then she made a place for the new big candle. Your candle is very pretty and smells nice too.

Your hamlet lies just over the hill. Far away from here near the sea is an old hamlet. People from the hamlet like to fish. Another hamlet is in the country. People from that hamlet really like to farm. They grow so much that theirs is a very big hamlet.

The candle in the kitchen was almost melted. So Annie bought another candle at the stationary store. She came home and put away the old candle. Fran gave that candle to you later. Then she made a place for the new big candle. Your candle is very pretty and smells nice too.

The videos of toddlers’ looking patterns to objects on the screen were digitized at 30 frames per second and coded offline by a highly trained coder. Looking times to each object in each trial type were measured over a 2-second period, starting 367 ms after the onset of the label (e.g. Swingley & Aslin, 2000). If children reliably recognize the recently learned words, they should exhibit longer looking times to the trained object than to the novel object when hearing the trained label, but longer looking times to the novel object than the trained one when hearing a novel label. Thus, in order to demonstrate successful word learning in this demanding task, toddlers must identify the appropriate trained label–object pairing in Trained Test and use mutual exclusivity to infer a novel label–object pairing in Novel Test, on the fly, in order to map a novel label to a novel object. They must do this twice within each Training-Test block, for a total of four times over the whole experiment.

Results

We calculated a measure of difference in looking times (LT) as LTTrainedObject – LTNovelObject. If toddlers successfully recognize the recently learned words, this average should be reliably above zero for Trained trials (where the label corresponds to the trained object), and reliably below zero for Novel trials (where the label corresponds to a novel object, not the trained object). A repeated measures Analysis of Variance (ANOVA) with Speaker Number (Single, Multiple) and Speaker Accent (Local, Foreign) as between-subjects factors and Block (1, 2) and Trial Type (Trained, Novel) as within-subject factors revealed a main effect of Type [F(1, 84) = 31.06, p < .001] and an interaction of Type and Accent [F(1, 84) = 4.02, p < .05]. Given the lack of effects or interactions involving Speaker Number and Block, we collapsed across these factors in subsequent analyses and explored the interaction with ANOVAs within each condition. In the Local condition, there was a main effect of Trial Type [F(1, 43) = 7.48, p < .01] due to toddlers looking longer at the trained object than the novel object upon hearing the Trained label, but showing the opposite preference when hearing a Novel label. This pattern of preference was even stronger in the Foreign

Table 2  Exposure passages

| The candle in the kitchen was almost melted. So Annie bought another candle at the stationary store. She came home and put away the old candle. Fran gave that candle to you later. Then she made a place for the new big candle. Your candle is very pretty and smells nice too.
| Your hamlet lies just over the hill. Far away from here near the sea is an old hamlet. People from the hamlet like to fish. Another hamlet is in the country. People from that hamlet really like to farm. They grow so much that theirs is a very big hamlet.
| The candle in the kitchen was almost melted. So Annie bought another candle at the stationary store. She came home and put away the old candle. Fran gave that candle to you later. Then she made a place for the new big candle. Your candle is very pretty and smells nice too.
| The videos of toddlers’ looking patterns to objects on the screen were digitized at 30 frames per second and coded offline by a highly trained coder. Looking times to each object in each trial type were measured over a 2-second period, starting 367 ms after the onset of the label (e.g. Swingley & Aslin, 2000). If children reliably recognize the recently learned words, they should exhibit longer looking times to the trained object than to the novel object when hearing the trained label, but longer looking times to the novel object than the trained one when hearing a novel label. Thus, in order to demonstrate successful word learning in this demanding task, toddlers must identify the appropriate trained label–object pairing in Trained Test and use mutual exclusivity to infer a novel label–object pairing in Novel Test, on the fly, in order to map a novel label to a novel object. They must do this twice within each Training-Test block, for a total of four times over the whole experiment.

Results

We calculated a measure of difference in looking times (LT) as LTTrainedObject – LTNovelObject. If toddlers successfully recognize the recently learned words, this average should be reliably above zero for Trained trials (where the label corresponds to the trained object), and reliably below zero for Novel trials (where the label corresponds to a novel object, not the trained object). A repeated measures Analysis of Variance (ANOVA) with Speaker Number (Single, Multiple) and Speaker Accent (Local, Foreign) as between-subjects factors and Block (1, 2) and Trial Type (Trained, Novel) as within-subject factors revealed a main effect of Type [F(1, 84) = 31.06, p < .001] and an interaction of Type and Accent [F(1, 84) = 4.02, p < .05]. Given the lack of effects or interactions involving Speaker Number and Block, we collapsed across these factors in subsequent analyses and explored the interaction with ANOVAs within each condition. In the Local condition, there was a main effect of Trial Type [F(1, 43) = 7.48, p < .01] due to toddlers looking longer at the trained object than the novel object upon hearing the Trained label, but showing the opposite preference when hearing a Novel label. This pattern of preference was even stronger in the Foreign

1 While analyzing looking time within the time window of interest we found an error in the SuperCoder macro used for these analyses. After notification of the error, this macro is accompanied with a disclaimer in the SuperCoder site. We hence designed our own R-based analysis of the data for the time window of interest, which is available together with the data as supplementary material.
condition \( F(1, 43) = 25.65, p < .001 \). Thus, the interaction between Type and Accent clearly occurred because performance in the Foreign conditions was better than performance in the Local conditions. In addition, given that there were some imbalances in the samples included in each condition (e.g. number of males and females; see Table 1) and to evaluate a possible effect of vocabulary size, we carried out a linear regression that also incorporated age, sex, and number of words reported in the CDI. This linear regression confirmed the effect of Trial Type and the interaction with Accent even when these additional factors were incorporated, none of which was a significant predictor.\(^7\) Data and analyses routines are available for download from https://sites.google.com/site/toddlersaccent/home. Difference in LTs by Trial Type and Accent are shown in Figure 2.

Discussion

The present work examined 24-month-olds’ ability to cope with an unfamiliar foreign accent in word learning when provided with brief native or foreign accent exposure. Results suggest that toddlers exhibit remarkable plasticity, especially once they have experience with the foreign accent. In addition, other variables that we explored did not significantly affect toddlers’ performance: Because there was no effect or interaction involving the number of talkers in Exposure, this suggests that the Single and Multiple conditions (within their respective accents) were comparable. Therefore, talker variability (found in the two Multiple conditions) was not as helpful as foreign-accent exposure. Thus, we confirm that unfamiliar accents do not constitute an insurmountable roadblock for young word learners, particularly when provided with appropriate exposure. Together with previous work, the current results show that while unfamiliar accents incur processing costs for listeners, relevant exposure reduces (and potentially eliminates) this negative impact.

Taking these findings together with other work with children and adults, we would like to propose that listeners could, in fact, employ two distinct strategies in accent accommodation. First, and as discussed in the Introduction, listeners may use top-down knowledge to guide accent accommodation; we will refer to this as a lexically based specific expansion strategy. For example, hearing a strange sound in the context of *croco_ile* as opposed to the context of *luna_ić* shifts adult listeners’ perception of the boundary between /d/ and /V/ (Kraljic & Samuel, 2007); and toddlers trained with *dag* as a label for *dog* shifted their vowel category to accept ‘sack’ for ‘sock’. This strategy is linguistically informed, and crucially depends on listeners having clear evidence for each and every type of pronunciation change. The problem becomes challenging for natural, foreign accents, which typically vary along a number of dimensions, often including several sound category changes in addition to differences in suprasegmental properties. When faced with such widespread and radical changes, a lexically based strategy may prove overly complex for toddlers, who would have to learn a myriad of rewrite rules or constraints for the unfamiliar accent, with rather scarce evidence for each of them.

In such circumstances, language users could also employ a second accent accommodation strategy, which we will call general expansion. It is possible that, when faced with speech that differs dramatically from native input, listeners may generally expand or relax their phonemic categories to accept a certain degree of deviation from native pronunciation norms. The advantage of this strategy is that it could be applied any time the novel talker deviates from expected standards, and thus may facilitate accommodation when dealing with unfamiliar and/or highly diverse accents. However, a general expansion strategy would carry an important processing cost, since listeners would not constrain subsequent lexical access through newly built expectations. This strategy may also present an implicit danger, whereby toddlers might not *a priori* limit the acceptable sound changes. As a result, two patterns of errors are expected: First, listeners might accept changes in pronunciation that span phonemic boundaries when faced with accented speech, something that they are not keen to do in other situations. For example, toddlers were slower to fixate on a picture of a baby when hearing the mispronounced word *vaby* in a native speech stream, than when hearing the correct *baby* pronunciation (Swingley & Aslin, 2000). If toddlers applied the general expansion strategy to

\[ \text{LT}_{\text{Trained Object}} - \text{LT}_{\text{Novel Object}} \]

Figure 2 Difference looking times (LT) \( \text{LT}_{\text{Trained Object}} - \text{LT}_{\text{Novel Object}} \) by Trial Type and Accent of the speaker(s) during initial exposure. Error bars represent 2 Standard Errors. If toddlers learn, it is expected that this difference will be positive in Trained trials (they look reliably longer to the trained than the novel object when hearing the trained label) and negative in Novel trials (they look reliably longer to the novel than the trained object when hearing the novel label).

\(^7\) The estimate for Trial Type was \( \beta = 1.53, t(341) = 6.15, p < .05 \); Accent \( \beta = .74, t(341) = 2.98, p < .05 \); their interaction \( \beta = -.85, t(341) = 2.41, p < .05 \). Non-significant estimates were found for Sex \( \beta = -.23, t(341) = 1.19, p = .23 \), age \( \beta = -.05, t(341) = 0.2, p = .85 \), and CDI \( \beta = .006, t(341) = 1.82 p = .07 \).
foreign-accented speech, they might be as accepting of 
\textit{vaby} as they are of \textit{baby}, a process that may be modu-
lated by the presence of lexical competitors, the spectral
similarity between the variant and target, as well as,
possibly, the child’s vocabulary development. In fact, an
effect akin to this one predicted by the general expansion
has already been documented: In White and Aslin (2011),
toddlers trained repetitively on a sound change
from /a/ to /æ/ also accepted /æ/ for the same /æ/ target,
showing inappropriate relaxation of phonemic boundaries.
The second predicted pattern of error builds on the fact that the
general expansion strategy is a fall-
back plan when insufficient evidence is provided for
precise changes. As a result, one predicts that listeners
should actually accept changes for which they have had
no evidence. For example, when faced with a talker with
an unfamiliar accent who pronounces ‘dog’ as ‘dak’,
toddlers may also accept ‘beg’ for ‘peg’, and perhaps
even ‘sit’ for ‘seat’, although the talker has not provided
evidence of mispronouncing any of those categories. We
believe this is an understudied, but plausible strategy,
worthy of further exploration. The general expansion
strategy would be of most use to young children, whose
limited top-down knowledge and reduced processing
abilities may limit the utility of the lexically based spe-
cific expansion strategy. Nonetheless, even adults could
profit from the general expansion strategy when facing
radically different, unfamiliar accents. In fact, some very
recent evidence points in that direction: Eisner, Weber
and Melinger (2010) found that listeners extrapolate
mispronunciations from syllable-final to syllable-initial
stops, for which they had had no evidence, and
Witteman, Weber and McQueen (2011) report that
adults accommodate for even large vowel deviations
without any specific training.

Thus, an additional question for further research is the
modulation of this expansion strategy on the basis of
experience, and in particular lexical advancement. Not
only does vocabulary development predict performance
in word recognition (Fernald, Perfors & Marchman,
2006) and word learning (Weker, Fennell, Corcoran &
Stager, 2002), but also recent work has documented an
impact of vocabulary development on accommodation
for unfamiliar accents. Best et al. (2009) show that 19-
month-olds, but not 15-month-olds, prefer highly fre-
quent to infrequent words uttered in an unfamiliar dia-
lect. They propose that this change could be triggered by
the vocabulary expansion that takes place between 15
and 19 months: As children’s vocabulary expands, so
does their ability to retrieve the abstract phonological
shape of the word. Recent word recognition studies have
provided some support for this hypothesis (Best, Tyler,
Kitamura & Bundgaard-Nielsen, 2010; Mulak, Best
et al., 2010). Although our data provide no strong evi-
ence of an association between vocabulary size and
performance, the children tested here are much older,
and thus the variance in vocabulary size may not be as
informative of their linguistic development. Nonetheless,

References

Best, C.T., Tyler, M.D., Gooding, T.N., Orlando, C.B., &
Quann, C.A. (2009). Development of phonological con-
stancy: toddlers’ perception of native- and Jamaican-

Best, C.T., Tyler, M.D., Kitamura, C., & Bundgaard-Nielsen,
R.L. (2010). Vocabulary size at 17 months and the emergence
of phonological constancy in word recognition across native
and nonnative dialects. Poster presented at the 2010 Inter-
national Conference on Infant Studies, Baltimore, MD.


Received: 9 August 2011
Accepted: 10 February 2012