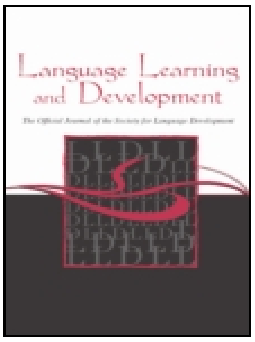

Routledge
Taylor & Francis Group

Language Learning and Development

ISSN: 1547-5441 (Print) 1547-3341 (Online) Journal homepage: http://www.tandfonline.com/loi/hlld20

He Says Potato, She Says Potahto: Young Infants Track Talker-Specific Accents

Drew Weatherhead & Katherine S. White

To cite this article: Drew Weatherhead & Katherine S. White (2016) He Says Potato, She Says Potahto: Young Infants Track Talker-Specific Accents, Language Learning and Development, 12:1, 92-103, DOI: 10.1080/15475441.2015.1024835

To link to this article: http://dx.doi.org/10.1080/15475441.2015.1024835

Published online: 09 Nov 2015.

Submit your article to this journal

Article views: 65

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
http://www.tandfonline.com/action/journalInformation?journalCode=hlld20

Download by: [University of Waterloo] **Date:** 11 April 2016, At: 06:24

LANGUAGE LEARNING AND DEVELOPMENT
2016, VOL. 12, NO. 1, 92–103
http://dx.doi.org/10.1080/15475441.2015.1024835

Routledge
Taylor & Francis Group

He Says Potato, She Says Potahto: Young Infants Track Talker-Specific Accents

Drew Weatherhead and Katherine S. White

Department of Psychology, University of Waterloo

ABSTRACT

One of the most fundamental aspects of learning a language is determining the mappings between words and referents. An often-overlooked complication is that infants interact with multiple individuals who may not produce words in the same way. In the present study, we explored whether 10- to 12-month-olds can use talker-specific knowledge to infer the intended referents of novel labels. During exposure, infants heard two talkers whose front vowels differed; one talker trained them on a word-referent mapping. At test, infants saw the trained object and a novel object; they heard a single novel label from both talkers. When the label had a front vowel (Experiment 1), infants responded differently as a function of talker, but when it had a back vowel (Experiment 2), they did not, mapping the novel label to the novel object for both talkers. These results suggest that infants can track the phonetic properties of two simultaneously presented talkers and use information about each talker's previous productions to guide their referential interpretations.

Downloaded by [University of Waterloo] at 06:24 11 April 2016

Introduction

One of the most fundamental aspects of learning a language is determining the mappings between words and referents. However, learning even the words themselves presents a formidable categorization problem due to rampant variability in the speech signal. In addition to within-speaker variability, infants interact with multiple individuals, who may not produce words in the same way. This can be due to physical or idiosyncratic differences across speakers as well as to systematic language-based differences. For example, consider two individuals from different regions of the United States: Sarah says "bag" to refer to a sack, while John says "beg" for the same object. More confusingly, this word, "beg," is similar to Sarah's word for pleading. How do we determine when the same phonetic form (e.g., Sarah's and John's "beg") maps onto different word categories (and, therefore, meanings) and when different phonetic forms (e.g., Sarah's "bag" vs. John's "beg") map onto the same word category (and meaning)? One source of information is context (e.g., if the speaker's attention is directed toward a sack). However, contextual information is not always available or unambiguous. For adults, another important source of information is knowledge about the speaker's language background. This can come in the form of general knowledge about the speaker's language community, which can activate stored information about that community's accent (Hay, Nolan, & Drager, 2006), or from direct observation of a speaker's productions.

Encoding such "talker-specific" information not only allows adults to understand the particular words someone has produced before, but also to make inferences about other words. If Sarah has previously heard John say "beg" for a sack, she might infer that "teg" is his pronunciation of her "tag." However, if she knows nothing about him, she might treat "teg" as a new word, because

CONTACT Drew Weatherhead deweathe@uwaterloo.ca Department of Psychology, University of Waterloo, 200 University Avenue West, Waterloo, ON, Canada, N2L 3G1.

© 2016 Taylor & Francis

listeners have a bias to assume one-to-one mappings between phonetic form and meaning. This bias likely contributes to the difficulty that listeners initially have in understanding a talker with an unfamiliar accent (e.g., Bradlow & Bent, 2008; Clarke & Garrett, 2004).

The problems posed by this type of accent variability are potentially much more significant for young language learners. One reason for this is that young learners adhere more strongly than adults to the assumption that novel phonetic forms should be mapped to novel referents—an adaptive assumption, given how often new words occur in their environments. A large body of research has demonstrated that learners as young as 6 months interpret novel wordforms as labels for novel objects, although the mechanism underlying this mapping preference may change across development (e.g., Golinkoff, Mervis, & Hirsh-Pasek; Halberda, 2003; Markman, 1989, 1990; Merriman, Bowman, & MacWhinney, 1989; Shukla, White, & Aslin, 2011). Whether this mapping bias results from exclusion reasoning (objects have only one label) or a mapping of novelty (of the label) to novelty (of the referent), accented speech poses a challenge: if an accented pronunciation is judged to be different from known words, it will be treated as a new word. In other words, learners will posit wordform-meaning mappings that do not exist, potentially slowing lexical development. Indeed, children sometimes map mispronunciations of familiar words to novel referents (Mani & Plunkett, 2011; Merriman & Schuster, 1991; White & Morgan, 2008).

Recent work demonstrates that young learners do have some difficulty processing accented words. Infants are unable to recognize familiarized wordforms across accents until the end of the first year (Schmale & Seidl, 2009; Schmale, Cristia, Seidl, & Johnson, 2010) and have difficulty recognizing accented versions of known words even at later ages (Best, Tyler, Gooding, Orlando, & Quann, 2009), unless they are given sufficient exposure to the accent (van Heugten & Johnson, 2014). Similarly, it is not until 19 months that toddlers recognize, under some conditions, that accented pronunciations map onto familiar referents (Mulak, Best, Tyler, & Kitamura, 2013; Schmale, Hollich, & Seidl, 2011; Schmale, Cristia, & Seidl, 2012; White & Aslin, 2011). Even 4-year-olds have difficulties recognizing familiar words in an unfamiliar dialect (Nathan, Wells, & Donlan, 1998).

Here, we explore whether infants can encode the phonetic properties of particular individuals, and use this talker-specific knowledge to overcome the problems posed by accent variability. Just as children realize that the one-to-one mapping assumption does not operate across languages (Au & Glusman, 1990), they might also realize that what counts as a novel label depends on a speaker's accent. Therefore, if learners can track the differences between two talkers' accents, they may understand that whether two phonetic forms refer to the same or different referents depends on the speaker.

Infants have the ability to link certain types of vocal properties and speakers, at both a group and an individual level. Infants can match a familiar accent to a face of a familiar race and an unfamiliar accent to a face of an unfamiliar race (Uttley et al., 2013). Infants are also capable of remembering links between talkers and the global properties of their speech: they prefer to look at an individual who previously spoke their native language over one who previously spoke a foreign language (Kinzler, Dupoux, & Spelke, 2007). However, in the case of native vs. foreign speech, it is enough to recognize simply that one speaker sounds familiar and the other sounds unfamiliar. Tracking specific phonetic properties across speakers where there is no familiarity difference is a more challenging task.

We asked whether 10- to 12-month-olds could learn about the specific properties of two talkers' accents and, in the absence of any contextual information, use that talker-specific information to determine the intended referent of novel words. Previous research on early accent processing has considered whether infants can map a novel accent to their own accent, but not whether they can track speaker-specific phonetic information. In addition, although it has been shown that exposure improves infants' and toddlers' recognition of accented speech, which properties of the accent are learned has remained virtually untested (but see White & Aslin, 2011). We chose to focus on 10- to 12-month-olds because, by the end of the first year, infants have started tuning to the relevant sound properties of their native language (Houston & Juscyzk, 2000; Singh, White, & Morgan, 2008;

Downloaded by [University of Waterloo] at 06:24 11 April 2016

Werker & Tees, 1984), making the kind of accent variability used in the present study (which involves phonemic category changes) disruptive.

We presented infants with two talkers whose productions systematically differed in the height of their front vowels – a "Training" Speaker and an "Extension" Speaker. The Extension Speaker's front vowels were higher than the Training Speaker's. We chose to manipulate vowel height across speakers, as accent differences commonly involve vowels, and English-learning infants are sensitive to various types of vowel contrasts, including subtle distinctions like [i] vs. [I] (Swoboda, Morse, & Leavitt, 1976), [a] vs. [ɔ] (Kuhl, 1983), and [e] vs. [E] (Sundara & Scutellaro, 2011). In addition, 14-month-olds are sensitive to a range of vowel changes in both familiar and newly trained words (Mani & Plunkett, 2008; Mani, Mills, & Plunkett, 2012). Following this exposure, infants learned the label for a novel object from the Training Speaker (*tEpu*), but did not hear the Extension Speaker label it. In other words, infants were not directly exposed to the Extension Speaker's label for the object. In Experiment 1, at test, infants saw this trained object and an untrained object, and heard each talker use the label *tIpu*. If infants are able to track the systematic difference between the speakers, their interpretation of the test label *tIpu* should differ as a function of the talker's identity. In Experiment 2, we changed the test label to *topu*. If infants have learned that the talkers differ specifically in their front vowels, their interpretation of the test label *topu* (with only back vowels) should not differ by talker.

Experiment 1

Participants

Thirty 10- to 12-month-olds were tested (11 females and 19 males; mean age: 324 days; age range: 297–358 days). Ten additional participants were tested but not included due to non-completion (2), parental headphone difficulties (3), failure to attend to both objects during the baseline period (3), or difference scores exceeding 2.5 standard deviations from the mean of either speaker (2).

Stimuli

Audio stimuli

The stimuli consisted of four pairs of CVCV nonsense words (see Table 1) produced by two female native speakers of English. The pronunciation of the first vowel (always a front vowel) varied by talker, but the remainder of the word (including the second, back, vowel) did not differ across talkers. Three of the word pairs (*m[I/i]to, d[E/I]lu*, and *b[I/i]mo*[1]) were presented during exposure without referents. The word *tEpu* was used by the Training Speaker during exposure to label an object. The last word, *tIpu*, was heard in test. Stimuli were recorded in a sound-treated booth at a sampling rate of 44100 Hz and were later equated for amplitude in Praat (Boersma & Weenink, 2009). See Table 2 for acoustic information. The audio stimuli for exposure were inserted into the videos described below.

Table 1. Audio stimuli used during exposure.

Word type	Training speaker	Extension speaker
Exposure Pair 1	mIto	mito
Exposure Pair 2	dɛlu	dIlu
Exposure Pair 3	bImo	bimo
Trained object label	tɛpu	

[1][I] represents the sound in "big", [i] the sound in "beep", and [o] (Experiment 2) the sound in "boat."

Downloaded by [University of Waterloo] at 06:24 11 April 2016

Table 2. Acoustic information (first and second formant of the critical vowel in Hz, mean pitch of the word in Hz, and word duration in seconds) for key tokens used in both experiments. The first column refers to the Training Speaker's pronunciation of tEpu, which appears in the object presentation phase of both experiments. The values in this column are a calculated mean across the 3 tokens used during the object presentation event. The second column refers to each speaker's pronunciation of the test word in Experiment 1 (tIpu). The third column refers to each speaker's pronunciation of the test word in Experiment 2 (topu). For the test words, each value was calculated for the single token.

	Trained label: tEpu (Experiment 1 & 2)	Test word: tIpu (Experiment 1)		Test word: topu (Experiment 2)	
	Training speaker	Training speaker	Extension speaker	Training speaker	Extension speaker
F1	914	539	618	544	508
F2	2357	2599	2620	882	1098
Mean Pitch	276	288	257	289	253
Duration	0.62	0.68	0.64	0.66	0.76

Audiovisual stimuli (exposure phase)

Both talkers, Caucasian females aged approximately 22 years, were recorded against the same backdrop. They were dressed in different colored t-shirts to provide a salient cue that they were different people. Each talker recorded three exposure videos, in which a single exposure word was repeated three times in infant-directed speech approximately two seconds apart. Each talker also recorded an object presentation event. In the Training Speaker's object presentation event, she held and waved the target object while labelling it *tEpu* three times (this object is hereafter referred to as the trained object). In the Extension Speaker's object presentation event, she held and waved the trained object, but did not label it. Infants were either trained with an unfamiliar blue object or an unfamiliar yellow object.

Downloaded by [University of Waterloo] at 06:24 11 April 2016

Procedure

The participant sat on his/her parent's lap approximately 1.5 ft. from a 36x21-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 65dB 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.

The first video pair of the exposure phase involved the object presentation events from both talkers, to signal to the infants that they were in a word-learning situation. Next, the three pairs of yoked exposure videos (e.g., *mIto-mito*) were presented in random order (see Table 1). These pairs served to highlight the front-vowel difference between the talkers. The object presentation event pair was repeated twice at the end of the exposure phase. Overall, infants heard the trained object labeled nine times by the Training Speaker. An attention getter occurred between the video pairs, with the next pair beginning when the experimenter judged that the participant was focused on the attention getter. See Figure 1 for a schematic of the exposure phase.

The test phase began immediately after the exposure. There were two test trials, one for each talker. Each trial was 10 seconds in length. At the start of each trial, the talker's face and shoulders appeared alone for 2 seconds, followed by a display with the trained object and a novel untrained object. The objects remained on the screen for 8 additional seconds, the first 3 seconds of which was a silent baseline period, followed by an audio recording of the pictured talker saying the test word (*tIpu*). The talker in the first test trial and the side on which the trained

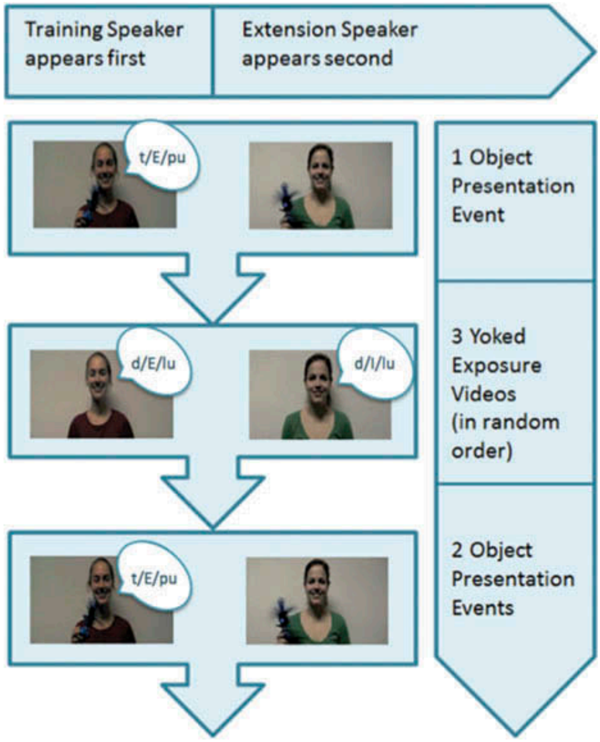

Figure 1. Schematic of the Exposure Phase: The exposure phase begins with one Object Presentation Event, followed by the three Exposure Events, followed by two more Object Presentation Events. In each event, the Training Speaker is seen first (approximately 8 seconds), followed by the Extension Speaker (approximately 8 seconds). In each event the speaker is alone on the screen. We present them together in the Figure to highlight the alternation.

object appeared were counterbalanced across participants (this side assignment remained constant for both test trials). See Figure 2 for a schematic of the test trials.

If infants learned the trained label *tEpu* from the Training Speaker during the exposure phase, then the novel label *tIpu* should be mapped to the untrained object for this talker. If, in addition, they learned that the two talkers differ in their pronunciations of front vowels and, in particular, that the Extension Speaker had higher front vowels, then they should interpret *tIpu* as the Extension Speaker's pronunciation of the trained object's label. In that case, they should look longer to the trained object for this talker.

Coding of looking times

Looking time during the test phase was coded off-line using in house software (Brown University), frame-by-frame (1 frame = 33 msec). Looking proportions for the objects were determined for the baseline period and for the test period, which began 430 msec after test word onset. This delay corresponded to the time necessary to program an eye movement in response to the first vowel in *tIpu* (shifting the analysis window at test is a common practice in word recognition studies, e.g., Bailey & Plunkett, 2002; Swingley & Aslin, 2002; White & Aslin, 2011). To equate the length of the baseline and test periods, only the first 3 seconds of the test period were analyzed.

Downloaded by [University of Waterloo] at 06:24 11 April 2016

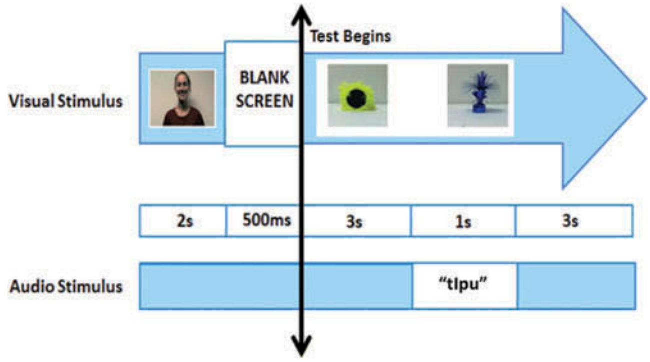

Figure 2. Schematic of a Test Trial: An image of the speaker appears alone on the screen for two seconds, followed by images of the trained and untrained object on either side of the screen. Objects are onscreen for 3 seconds before the test word is uttered (baseline period) and remain on screen for another 4 seconds post-label onset.

Downloaded by [University of Waterloo] at 06:24 11 April 2016

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 3 seconds for each phase).[2] During the baseline period, there was no difference in looking to the trained and untrained objects for the Extension Speaker (proportions of .44 and .43, respectively; $t(29) = 0.184$, *ns*)[3]; however, there was an asymmetry for the Training Speaker (.50 and .37; $t(29) = 2.0$, $p = .054$), which is addressed below. Using the proportions for each period, a difference score was calculated for each trial (proportion $\text{object}_{\text{test}}$-proportion $\text{object}_{\text{baseline}}$). This measure indicates the change in looking towards an object after labelling. Figure 3 displays the difference scores.

A repeated measures ANOVA on these difference scores with within-subjects factors of Speaker and Object and a between-subjects factor of test Order revealed no significant main effects of Speaker ($F(1,28) = 0.566$, *ns*) or Object ($F(1,28) = 0.091$, *ns*), but a significant interaction between Speaker and Object ($F(1,28) = 6.530$, $p = .016$). There was also a marginal interaction between Speaker, Object, and test Order ($F(1,28) = 4.070$, $p = .053$).

To determine the effect of labeling for each talker separately, one-sample t-tests compared difference scores for each talker and object against chance (where chance = a difference score of 0). As predicted, following labeling by the Training Speaker, looking significantly increased to the untrained object ($t(29) = 2.594$, $p = .015$). In contrast, for the Extension Speaker, looking to the trained object significantly increased ($t(29) = 2.700$, $p = .011$).[4] Thus, when the Training Speaker said

[2]Note that these proportions are out of the total duration of each phase. Thus, the proportions for each object in a trial do not necessarily sum to 1. In fact, while infants spent approximately 87% of the total time looking at the objects during the baseline phase, they spent approximately 95% of the total time looking at the objects during the test phase. The amount of time spent looking at the screen in each phase was the same for both talkers.

[3]All t-tests reported are two-tailed.

[4]To ensure that participants with more extreme baseline asymmetries did not affect the overall pattern of results, we also re-analyzed the data using a weighted difference score, in which trials with larger asymmetries carried less weight. To arrive at this weighted difference score, we first determined the difference in baseline preference for each object (degree of bias) for each trial (by participant). The actual difference scores were then multiplied by (1 - the degree of bias). Thus, the larger the bias score, the less weight the score carried in the overall mean. The pattern of results remained the same (for the Training Speaker, looking significantly increased to the untrained object $t(29) = 2.783$, $p = .016$ and for the Extension Speaker looking significantly increased to the trained object $t(29) = 3.042$, $p = .009$). In addition to this baseline correction, we also re-analyzed the data by including only trials that had less asymmetric baseline differences, equating baseline scores across the speakers for both objects. We found the same pattern of results (for the Training Speaker, looking increased to the untrained object $t(20) = 1.995$, $p = .059$, and for the Extension Speaker, looking increased to the trained object $t(25) = 2.969$, $p = .006$. Finally, note that although there was an asymmetry in the baseline for the Training speaker in Experiment 1, this asymmetry was not present in Experiment 2, where a significant increase in looking to the untrained object was also found.

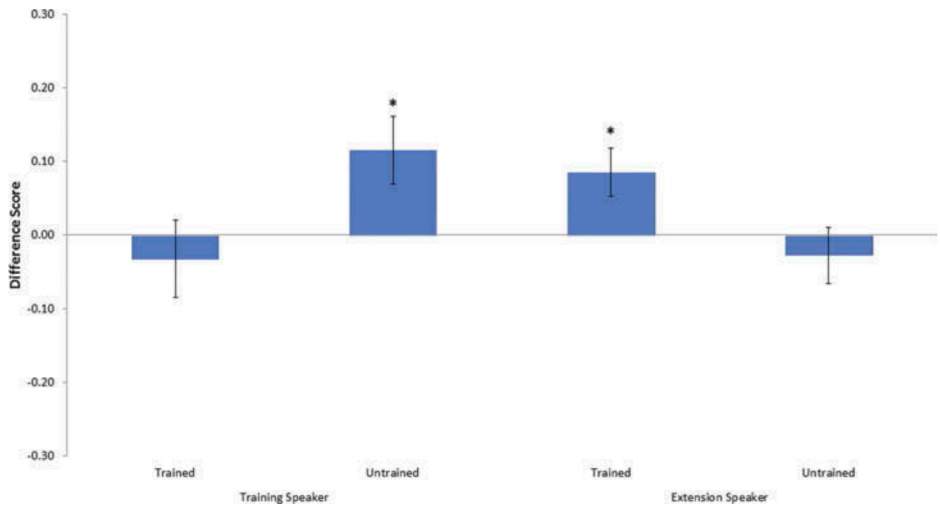

Figure 3. Difference scores for all participants in Experiment 1: looking proportions during baseline subtracted from looking proportions during the test period. Positive scores reflect an increase in looking while negative scores reflect a decrease in looking. * denotes a *p* value less than 0.05. Error bars represent the calculated standard error.

Downloaded by [University of Waterloo] at 06:24 11 April 2016

tIpu, infants increased their looking toward the untrained object, but when the Extension Speaker said *tIpu*, they increased their looking toward the trained object. In other words, infants responded differently to the same test word, depending on which talker produced it. Note from Table 2 that both talkers' pronunciations of *tIpu* were distinct from the training word *tEpu*.

This pattern of results suggests that infants learned the Training Speaker's label for the training object (*tEpu)* and when the same talker used a different label (*tIpu*), they interpreted it as a label for the untrained object. The fact that, in contrast, infants increased their looking to the trained object for the Extension Speaker suggests that they tracked the differences between the two talkers' pronunciations during the exposure phase.

However, closer analysis revealed that this was only true of infants who were first tested on the Training Speaker. For this order, the ANOVA revealed no significant effect of Speaker ($F(1,14) = .0448$, *ns*) or Object ($F(1,14) = .202$, *ns*), but a significant Speaker x Object interaction ($F(1,14) = 10.705$, $p = .006$). One-sample t-tests showed that for the Training speaker, there was a significant increase in looking to the untrained object ($t(14) = 2.654$, $p = .019$) and decreased looking to the trained object ($t(14) = -1.101$, $p = .290$). For the Extension speaker, looking significantly increased to the trained object ($t(14) = 2.918$, $p = .011$) and decreased to the untrained object ($t(14) = -1.122$, $p = .281$); see Figure 4). In contrast, those who were first tested on the Extension Speaker did not reliably change their looking behavior at test. There was no Speaker x Object order interaction ($F(1,14) = 0.141$, *ns*) and looking did not change for either speaker individually (Training Speaker untrained object: $t(14) = 0.846$, *ns*; Training Speaker trained object: $t(14) = 0.358$, *ns*; Extension Speaker untrained object: $t(14) = 0.943$, *ns*; Extension Speaker trained object: $t(14) = 0.076$, *ns*).

Given this pattern of results, it is possible that infants did not learn the systematic differences between the talkers' accents but simply learned that the two talkers pronounced words differently. If true, when the Training Speaker came first, infants could succeed by determining her intended referent and then, for the Extension Speaker trial, looking at the other referent. In the other direction, determining the Extension Speaker's intended referent would have been more difficult without the anchor provided by the Training Speaker, thus leading to poorer overall performance. To investigate whether infants were using only this type of heuristic, a

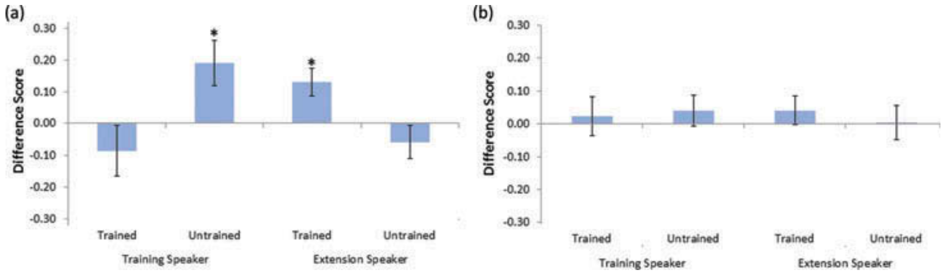

Figure 4. Experiment 1 difference scores for participants who saw the Training Speaker first at test (a), and for participants who saw the Extension Speaker first at test (b). * denotes a *p* value less than 0.05. Error bars represent the calculated standard error.

second experiment was run, in which the expected response was increased looking to the same object for both talkers.

Experiment 2

In order to determine if the participants in Experiment 1 were tracking the systematic differences between the talkers' accents, Experiment 2 used a test word, *topu*, that did not fall into the pattern learned during exposure. Recall that the difference between the talkers involved the height of front vowels; critically, the pronunciation of back vowels remained constant. If infants in Experiment 1 simply learned to respond to the two talkers differently, then infants in Experiment 2 should also look at different objects for the two talkers, even if the test word contains only back vowels. If, however, infants in Experiment 1 learned that the accent difference was specific to front vowels, then, regardless of the talker, infants in Experiment 2 should map the word *topu* to the untrained object.

Participants

Forty 10- to 12-month olds (21 females and 19 males, mean age = 334.85 days, age range = 311-363 days) took part. An additional ten infants were tested but were not included due to fussiness (1), failure to attend to both objects during familiarization (2) or the screen for the entirety of the test period (6), or a difference score exceeding 2.5 standard deviations from the mean of either speaker (1).

Stimuli and Procedure

Identical to Experiment 1, except the label *t[o]pu* was used during the test phase.

Results

Figure 5 displays the difference scores.[5] During the baseline period, there was no significant difference in looking to the trained versus untrained object for either speaker (proportions of 0.46 and 0.40[6], respectively, for the Training Speaker: $t(39) = 1.459$, $p = .15$; proportions of 0.48 and 0.43, respectively, for the Extension Speaker: $t(39) = 1.191$, $p = .24$). A Repeated-Measures ANOVA was conducted on the test-baseline difference scores with within-subjects factors of Speaker and

[5]Infants spent approximately 89% of the total time looking at the objects during the baseline phase, and approximately 92% of the total time looking at the objects during the test phase.

[6]This degree of baseline difference is on the order of those often found when a label for one object is known and for the other object is unknown (Schafer, Plunkett, & Harris, 1999; White & Morgan, 2008). In such studies, there are still reliable effects of the type of label (familiar/novel) on looking behavior.

Downloaded by [University of Waterloo] at 06:24 11 April 2016

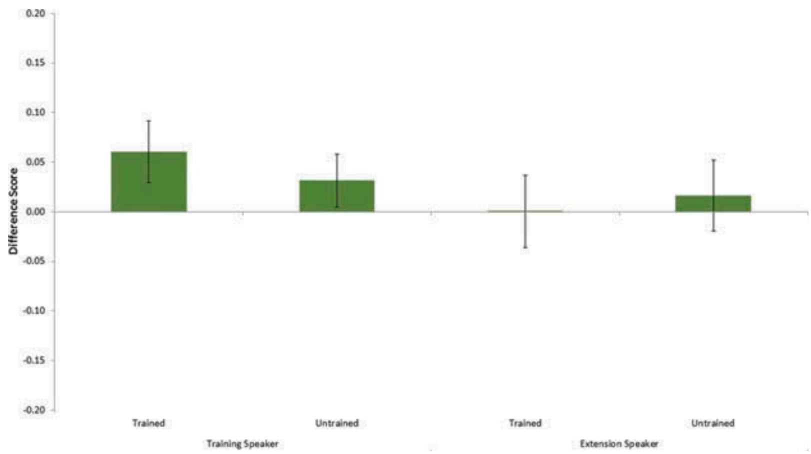

Figure 5. Experiment 2 difference scores for all participants. Error bars represent the calculated standard error.

Downloaded by [University of Waterloo] at 06:24 11 April 2016

Object and a between-subjects factor of test Order. The ANOVA revealed a significant effect of Speaker (F(1,39) = 6.123, p = .018) and a significant interaction between Object and Order (F(1,39) = 6.018, p = .019). No other effects reached significance.

Given the interaction involving Order, analyses were conducted for each of the presentation orders separately. For the participants who saw the Training Speaker first at test, a repeated-measures ANOVA found significant main effects of Speaker (F(1,19) = 5.009, p = .037) and Object (F(1,19) = 5.242, p = .034), but no interaction (F(1,19) = 1.365, p = .257). One sample t-tests against 0 showed that, as predicted, infants increased their looking to the untrained object for both the Training Speaker (t(19) = 2.141, p = .045) and Extension Speaker (t(19) = 1.951, p = .066), and decreased their looking to the trained object for both speakers (Training Speaker t(19) = -.539, *ns*; Extension Speaker t(19) = -2.223, p = .039) (see Figure 6). However, for participants who saw the Extension Speaker first, the ANOVA revealed no significant effects (Speaker: F(1,19) = 2.408, p = .137; Object: F(1,19) = 1.414, p = .249; interaction: F(1,19) = .172, *ns*). One sample t-tests showed that, unexpectedly, looking increased to the trained object for the Training Speaker (t(19) = 2.125, p = .047); there was no change for the untrained object (t(19) = -.130, *ns*). For the Extension Speaker, there were no significant changes for either object (trained: t(19) = 0.832, *ns*; untrained: t(19) = -0.674, *ns*).

Summarizing these results, infants increased their looking to the untrained object when they heard either of the two talkers say the test word *topu*, but only if they saw the Training Speaker first

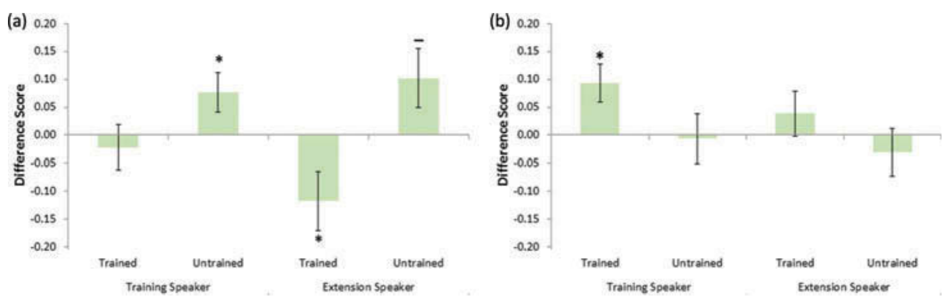

Figure 6. Experiment 2 difference scores for participants who saw the Training Speaker first at test (a), and for participants who saw the Extension Speaker first at test (b). * denotes a p value less than 0.05, — denotes a p value less than 0.10. Error bars represent the calculated standard error.

during the test phase. This suggests that infants in Experiment 1 did not learn only that the two talkers pronounced words differently. If they had, infants would have looked at different objects for each of the talkers in Experiment 2 as well. Therefore, infants must have encoded something more specific about the accent differences. We discuss the implications of these findings below.

Discussion

If infants cannot recognize the equivalence of words that are realized differently due to cross-speaker variation, they risk positing spurious word-referent mappings that could slow lexical development. We explored whether 10- to 12-month-olds could overcome the effects of talker-specific variation if given the chance to determine the relationship between the talkers' accents. Infants were first exposed to talkers whose front vowels differed. At test, they were presented with a previously unheard wordform, either *tIpu* (Experiment 1) or *topu* (Experiment 2). We predicted that, if infants were able learn the systematic vowel differences between the talkers and use this talker-specific information to make inferences about intended referents, their interpretation of *tIpu* should differ by speaker, but their interpretation of *topu* should not. Experiment 1 demonstrated that infants mapped *tIpu* to the untrained object for the Training Speaker, but to the trained object for the Extension Speaker. Experiment 2 ruled out the possibility that infants learned only a heuristic that the two talkers spoke differently: infants looked longer at the untrained object when both talkers produced the label *topu*, at least when the Training Speaker was presented first. Thus, infants appear to have learned that the difference between the talkers was specific to front vowels. The finding that infants learned about the relationship between the two talkers' productions is consistent with the fact that older toddlers can learn about the properties of accents (Schmale et al., 2012; Van Heugten & Johnson, 2014; White & Aslin, 2011). In those studies, toddlers learned the relationship between a novel accent and their own. The present work not only extends this ability to younger infants, but also shows that they can learn a phonetic relationship between two novel talkers that does not involve comparison to their own accent. This ability to track talker-specific detail parallels adults' learning of talker-specific properties for multiple speakers (such as voice-onset-time: Allen & Miller, 2004).[7]

Infants interpreted novel wordforms as a function of what they had learned about each talker's speech. This is consistent with work in other domains demonstrating that infants make person-specific attributions about certain types of information (e.g., desires, Repacholi & Gopnik, 1997; action goals, Buresh & Woodward, 2007) and can use person-specific information to guide their interactions with an individual (e.g., the person's reliability, Chow, Poulin-Dubois, & Lewis, 2008; helping and hindering behavior, Hamlin, Wynn, & Bloom, 2007; global aspects of the person's accent, Kinzler, Dupoux, & Spelke, 2007). The present results suggest that infants can also link subtle phonetic properties to particular individuals and use that information alone to infer a talker's intended referent.

In both experiments, infants succeeded only when they saw the Training Speaker first at test. This suggests that they were using their knowledge of the Training Speaker's productions to guide their behavior for the Extension Speaker. However, despite the order effects, infants' pattern of looking differed between the two experiments, demonstrating that infants were responding to the specifics of the label in each experiment. The fact that infants succeeded at all in our task is noteworthy. The task imposed a high processing load on our young participants—in order to succeed, they had to not only detect and encode the relationship between the talkers' productions, but also learn a new word in the lab. Even the latter task alone is challenging at this age; only a handful of lab studies have found

[7]As pointed out by an anonymous reviewer, an alternative possibility is that infants misattributed the accent difference to voice (treating [E] and [I] as these speakers' pronunciations of the same vowel, that is, as a within-category difference). However, under such an interpretation, it is not clear why infants would show differential treatment of 'tIpu' in Experiment 1, as the tokens of /I/ are acoustically similar for the two speakers. That said, determining how infants attribute variability to different sources is an important question for future research.

Downloaded by [University of Waterloo] at 06:24 11 April 2016

word learning in this age group from a single talker (e.g., Gogate, 2010; Shukla et al., 2011). Further, the correct interpretation in some trials was not the mapping of the trained word, but instead required that the learned mapping be used to map a novel label to the novel object. Thus, our results also demonstrate precocious use of an exclusion-based or novelty-novelty mapping strategy. In future work, we plan to further explore developmental changes in this task (e.g., at what point infants' representations are robust enough to succeed in the opposite order).

In summary, a large body of research has demonstrated that word learners have a strong bias to map novel labels to novel objects. In the present work, we find that even 10-12-month-olds do so when labels come from a single talker, but that they do not when the labels come from talkers who have different accents. Our results also suggest that, like adults, infants can track talker-specific phonetic properties, and can use information about an individual's previous language history to guide their future interactions with that individual. These findings suggest that, from a young age, infants are equipped with the tools necessary to handle the variable input around them.

Acknowledgments

The authors would like to thank the members of the Lab for Infant Development and Language for help with participant recruitment, Eiling Yee and Mohinish Shukla for helpful discussion, 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.

References

Allen, J. S., & Miller, J. L. (2004). Listener sensitivity to individual talker differences in voice-onset-time. *Journal of the Acoustical Society of America*, *115*, 3171–3183.

Au, T. K. F., & Glusman, M. (1990). The principle of mutual exclusivity in word learning: to honor or not to honor? *Child Development*, *61*(5), 1474–1490.

Bailey, T. M., & Plunkett, K. (2002). Phonological specificity in early words. *Cognitive Development*, *17*(2), 1265–1282.

Best, C. T., Tyler, M. D., Gooding, T. N., Orlando, C. B., & Quann, C. A. (2009). Development of phonological constancy: Toddlers' perception of native- and Jamaican-accented words. *Psychological Science*, *20*(5), 539–542. doi:10.1111/j.1467-9280.2009.02327.x

Boersma, P., & Weenink, D. (2009). Praat: doing phonetics by computer (Version 5.1. 05) [Computer program].

Bradlow, A. R., & Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition*, *106*(2), 707–729. doi:10.1016/j.cognition.2007.04.005

Buresh, J. S., & Woodward, A. L. (2007). Infants track action goals within and across agents. *Cognition*, *104*(2), 287–314. doi:10.1016/j.cognition.2006.07.001

Chow, V., Poulin-Dubois, D., & Lewis, J. (2008). To see or not to see: Infants prefer to follow the gaze of a reliable looker. *Developmental Science*, *11*(5), 761–770.

Clarke, C. M., & Garrett, M. F. (2004). Rapid adaptation to foreign-accented English. *The Journal of the Acoustical Society of America*, *116*(6), 3647–3658.

Cohen, J. D., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: A new graphic interactive environment for designing psychology experiments. *Behavioral Research Methods, Instruments, and Computers*, *25*(2), 257–271. Retrieved from http://psy.cns.sissa.it

Gogate, L. J. (2010). Learning of syllable–object relations by preverbal infants: The role of temporal synchrony and syllable distinctiveness. *Journal of Experimental Child Psychology*, *105*(3), 178–197.

Golinkoff, R. M., Mervis, C. B., & Hirsh-Pasek, K. (1994). Early object labels: the case for a developmental lexical principles framework. *Journal of Child Language*, *21*, 125–155.

Halberda, J. (2003). The development of a word-learning strategy. *Cognition*, *87*, B23–34.

Hay, J., Nolan, A., & Drager, K. (2006). From fush to feesh: Exemplar priming in speech perception. *The Linguistic Review*, *23*(3), 351–379.

Houston, D. M., & Jusczyk, P. W. (2000). The role of talker specific information in word segmentation by infants. *Journal of Experimental Psychology: Human Perception and Performance*, *26*, 1570–1582.

Kinzler, K. D., Dupoux, E., & Spelke, E. S. (2007). The native language of social cognition. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(30), 12577–12580. doi:10.1073/pnas.0705345104

Mani, N., & Plunkett, K. (2008). Fourteen-month-olds pay attention to vowels in novel words. *Developmental Science*, *11*, 53–59.

Downloaded by [University of Waterloo] at 06:24 11 April 2016

Mani, N., & Plunkett, K. (2011). Does size matter? Subsegmental cues to vowel mispronunciation detection. *Journal of Child Language*, *38*, 606–627.
Mani, N., Mills, D. L., & Plunkett, K. (2012). Vowels in early words: an event-related potential study. *Developmental Science*, *15*, 2–11.
Markman, E. M. (1989). *Categorization and naming in children: Problems of induction*. Cambridge, MA: MIT Press.
Markman, E. M. (1990). Constraints children place on word meanings. *Cognitive Science*, *14*(1), 57–77.
Merriman, W. E., Bowman, L. L., MacWhinney, B. (1989). The mutual exclusivity bias in children's word learning. *Monographs of the Society for Research in Child Development*, *54*(3/4), 1–129.
Merriman, W. E., & Schuster, J. M. (1991). Young children's disambiguation of object name reference. *Child Development*, *62*(6), 1288–1301.
Mulak, K. E., Best, C. T., Tyler, M. D., Kitamura, C., & Irwin, J. R. (2013). Development of phonological constancy: 19-month-olds, but not 15-month-olds, identify words in a non-native regional accent. *Child Development*, *84*(6), 2064–2078. doi:10.1111/cdev.12087
Kuhl, P. K. (1983). Perception of auditory equivalence classes for speech in early infancy. *Infant Behavior and Development*, *6*(2), 263–285.
Nathan, L., Wells, B., & Donlan, C. (1998). Children's comprehension of unfamiliar regional accents: A preliminary investigation. *Journal of Child Language*, *25*, 343–365.
Repacholi, B. M., & Gopnik, A. (1997). Early reasoning about desires: evidence from 14-and 18-month-olds. *Developmental Psychology*, *33*(1), 12–21.
Schafer, G., Plunkett, K., & Harris, P. L. (1999). What's in a name? Lexical knowledge drives infants' visual preferences in the absence of referential input. *Developmental Science*, *2*, 187–194.
Schmale, R., & Seidl, A. (2009). Accommodating variability in voice and foreign accent: Flexibility of early word representations. *Developmental Science*, *12*(4), 583–601.
Schmale, R., Cristia, A., Seidl, A., & Johnson, E. K. (2010). Developmental changes in infants' ability to cope with dialect variation in word recognition. *Infancy*, *15*(6), 650–662.
Schmale, R., Hollich, G., & Seidl, A. (2011). Contending with foreign accent in early word learning. *Journal of Child Language*, *38*(5), 1096–1108.
Schmale, R., Cristia, A., & Seidl, A. (2012). Toddlers recognize words in an unfamiliar accent after brief exposure. *Developmental Science*, *15*(6), 732–738.
Shukla, M., White, K. S., & Aslin, R. N. (2011). Prosody guides the rapid mapping of auditory word forms onto visual objects in 6-mo-old infants. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(15), 6038–6043.
Singh, L., White, K. S., & Morgan, J. L. (2008). Building a word-form lexicon in the face of variable input: Influences of pitch and amplitude on early spoken word recognition. *Language Learning and Development*, *4*, 157–178.
Sundara, M., & Scutellaro, A. (2011). Rhythmic distance between languages affects the development of speech perception in bilingual infants. *Journal of Phonetics*, *39*(4), 505–513.
Swingley, D., & Aslin, R. N. (2002). Lexical neighborhoods and the word-form representations of 14-month-olds. *Psychological Science*, *13*(5), 480–484.
Swoboda, P. J., Morse, P. A., & Leavitt, L. A. (1976). Continuous vowel discrimination in normal and at risk infants. *Child Development*, 459–465.
Uttley, L., de Boisferon, A. H., Dupierrix, E., Lee, K., Quinn, P. C., Slater, A. M., & Pascalis, O. (2013). Six-month-old infants match other-race faces with a non-native language. *International Journal of Behavioral Development*, *37*(2), 84–89. doi:10.1177/0165025412467583
Van Heugten, M., & Johnson, E. K. (2014). Learning to contend with accents in infancy: Benefits of brief speaker exposure. *Journal of Experimental Psychology: General*, *143*, 340–350.
Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, 7, 49–63.
White, K. S., & Morgan, J. L. (2008). Sub-segmental detail in early lexical representations. *Journal of Memory and Language*, *59*(1), 114–132.
White, K. S., & Aslin, R. N. (2011). Adaptation to novel accents by toddlers. *Developmental Science*, *14*(2), 372–384. doi:10.1111/j.1467-7687.2010.00986.x

Downloaded by [University of Waterloo] at 06:24 11 April 2016