Infants adapt to speaking rate differences in word segmentation

Yuanyuan Wang
Department of Otolaryngology–Head and Neck Surgery, The Ohio State University, Columbus, Ohio 43202, USA

Fernando Llanos
Department of Languages and Cultures, Purdue University, West Lafayette, Indiana 47906, USA

Amanda Seidl
Speech, Language and Hearing Sciences, Purdue University, West Lafayette, Indiana 47906, USA

(Received 19 July 2016; revised 1 March 2017; accepted 22 March 2017; published online 11 April 2017)

Throughout their development, infants are exposed to varying speaking rates. Thus, it is important to determine whether they are able to adapt to speech at varying rates and recognize target words from continuous speech despite speaking rate differences. To address this question, a series of four experiments were conducted to test whether infants can recognize words in continuous speech when rate is variable. In addition, the underlying mechanisms that infants may use to cope with variations induced by different speaking rates were also examined. Specifically, using the Headturn Preference procedure [Jusczyk and Aslin (1995), Cognitive Psychol. 29, 1–23], infants were familiarized with normal-rate passages containing two trisyllabic target words (e.g., elephants and dinosaurs), and tested with familiar (elephants and dinosaurs) and unfamiliar (crocodiles and platypus) words embedded in normal-rate (experiment 1), fast-rate (experiments 2 and 3), or slow-rate passages (experiment 4). The results indicate that 14-month-olds, but not 11-month-olds, recognized target words in passages with a fast speaking rate. In addition, findings suggest that infants used context to normalize speech across different speaking rates. © 2017 Acoustical Society of America. [http://dx.doi.org/10.1121/1.4979704]

I. INTRODUCTION

Speaking rate varies across registers (Fernald and Simon, 1984; Papousek et al., 1985; Fernald, 1992; Bergeson et al., 2006; Van de Weijer, 1997), the lifespan (Kowal et al., 1975; Chermak and Schneiderman, 1985), gender (Lutz and Mallard, 1986), as well as dialects (Robb et al., 2003). Even during normal conversation, talkers often produce large variations in the rate at which they speak (Miller et al., 1984b). Given the variability of speaking rate in the input, in order to become competent language processors, infants will need to come to recognize words across different rates at some point in life. Although there has been considerable interest in examining infants’ and young children’s ability to cope with variations in other aspects of the speech signal, such as gender, age, dialect, affect, and speech style, etc. (e.g., Houston and Jusczyk, 2000; Schmale et al., 2012; Singh et al., 2002, 2008), very little is known about the timeline of infants’ recognition of words at different rates. In this study, we take the first step to explore the question of when and how infants might recognize words produced at different speaking rates. While it is well known that adults compensate for rate differences using cues (e.g., Miller and Liberman, 1979), both available (e.g., acoustic cues) and unavailable (e.g., semantic cues) to the infants, we do not know whether infants will even be able to use the cues that are available to them. However, there is some evidence that infants are sensitive to temporal information when discriminating speech tokens (Eimas and Miller, 1980; Miller and Eimas, 1983). Therefore, in this study we also explore the perceptual mechanism of adaptation potentially used by infants to compensate for rate differences.

As mature language learners, adults are able to adapt to the differences induced by speaking rate; however, prior studies do show that variability in speaking rate influences speech perception and word recognition in adults with normal hearing and with hearing impairment (Miller and Liberman, 1979; Miller and Volaitis, 1989; Volaitis and Miller, 1992; Sommers et al., 1994; Bradlow et al., 1999; Uchanski et al., 1996). For example, listeners show decreased word identification scores when they are presented with word lists with mixed speaking rates relative to the word lists with a constant speaking rate (Sommers et al., 1994). Similarly, adults are more accurate in recognizing a word when it is repeated at the same speaking rate than at a different rate (Bradlow et al., 1999). In addition, slow speaking rates have been found to be associated with higher speech intelligibility and this in turn leads to better speech processing of slower speech. In contrast, fast speaking rates have a negative impact on listeners’ processing of speech, especially when the speech stimuli are heavily compressed (Bradlow et al., 1999; Dupoux and Green, 1997; Uchanski et al., 1996). For example, listeners show an increase in the number of correctly recalled content words when the words are compressed at a lower compression rate (38%) than at a higher compression rate (45%) (Dupoux and Green, 1997).
Taken together, these findings suggest that the temporal information, as instantiated by speaking rate, is crucial for accurate speech comprehension and word recognition. Although adults are able to cope with speaking rate variations in the speech, their speech processing is affected by speaking rate. Specifically, fast-rate speech in general leads to poorer intelligibility and thus poorer recognition and comprehension, while slow-rate speech may facilitate speech comprehension.

Changes in speaking rate have a dramatic impact on the acoustic cues along the temporal dimension, such as voice onset time, vowel duration, and transition duration, etc. (Miller et al., 1984a, 1986; Summerfield, 1981). Previous studies have demonstrated that adult listeners are able to alter the perceptual criteria to adapt to the changes in different rates of speech to help them decode degraded speech as well as recognize target words (e.g., Dilley and Pitt, 2010; Gordon, 1988; Wade and Holt, 2005; Miller and Liberman, 1979; Miller et al., 1984a; Green et al., 1994). For example, Dilley and Pitt (2010) tested adult listeners with sentences containing a target function word; they found that slowing the speaking rate of the carrier sentence with the target word caused listeners to perceive sentences as lacking the target words (e.g., leisure or time was perceived as leisure time). In contrast, when listeners were tested with sentences lacking function words, speeding the speaking rate of the carrier sentence caused the listeners to perceive a function word that was never spoken around the region where it was supposed to occur (e.g., leisure time was perceived as leisure time). These findings suggest that adult listeners are sensitive to the speaking rate of the carrier phrase such that variation in the rate at which the speech is presented can induce a change in how listeners perceive the same acoustic signal.

Despite the large number of studies examining the effects of speaking rate on speech processing in adults as well as the underlying mechanisms adults may use to cope with speaking rate variability, there has been very little research so far on how speaking rate may impact speech processing in infants and young children. A small number of studies have demonstrated that infants, similar to adults, are better at processing slow-rate speech relative to fast-rate speech (Nelson, 1976; Zangl et al., 2005). In addition, children’s ability to compensate for speaking rate differences shows a developmental pattern. For example, Zangl et al. (2005) examined 12- to 32-month-old children’s word recognition by measuring the accuracy and reaction time to unaltered stimuli and time-compressed stimuli in an online comprehension task. Results showed that children recognized the target words more accurately when listening to the unaltered speech stimuli. In a similar vein, Nelson (1976) found that fast speaking rate (4.9 syllables per second) hindered sentence comprehension in 5- to 9-year-old children as compared to a slower speaking rate (2.5 syllables per second); however, older children (9- to 16-year-olds) were less affected by the fast-rate speech. These studies suggest that the ability to process fast-rate speech is enhanced with the development of perceptual and/or cognitive ability. However, the literature is limited as to whether/how temporal information is used by infants.

Therefore, in this paper we primarily ask whether infants are able to cope with variations of speaking rate in a word segmentation task. Specifically, we ask whether infants are able to recognize target words introduced and segmented from continuous speech at a normal-rate in fast-rate speech presented later. In order to address this question, we manipulated one simple variable that changed between normal- and fast-rate speech, specifically the speed with which syllables were produced (experiments 1–3). Although natural fast-rate speech varies with normal-rate speech along more than this one single dimension, this question is nonetheless important because it has implications for understanding the impact of temporal variation of speech within a developmental framework, as well as the cues that infants use to recognize and segment words from continuous speech, an issue that is a long-standing puzzle in language acquisition literature (e.g., White and Aslin, 2011; Seidl et al., 2014; van Heugten and Johnson, 2014).

The secondary research question we explore in this study focuses on the mechanisms infants might employ in processing speech at different rates. In order to recognize target words across different speaking rates, the perceptual system may use different mechanisms to compensate for variations in the input. We know that adults use contextual cues to facilitate speaking rate normalization. However, infants’ speech processing strategies may differ from adults due to their limited linguistic experience and working memory. For example, given limited working memory it is possible that infants may rely more on information carried by smaller linguistic units, such as words, instead of being able to make use of context for normalization of rate. Therefore, we will explore whether infants use contextual cues to facilitate their recognition of words introduced at normal-rate in fast-rate speech.

In summary, the purpose of the current study was twofold. The primary goal was to explore whether infants are able to adjust to compressed speech and show recognition of target words from normal to fast-rate speech. The secondary goal was to determine whether infants rely on the context to help them adapt to speaking rate differences. To address these goals we used the modified Headturn Preference procedure (HPP; Jusczyk and Aslin, 1995), and tested 11- and 14-month-old English-learning infants in a series of four experiments. Experiment 1 was a baseline condition in which we tested 11-month-olds’ ability to segment trisyllabic words from normal-rate speech in a passage-to-passage design. In experiments 2 and 3, we familiarized 11- and 14-month-olds with target words in normal-rate passages and then tested them with familiar and unfamiliar words in fast-rate passages. In experiment 4, in order to explore infants’ use of contextual information for normalization, we familiarized 14-month-olds with target words in normal-rate passages and then tested them with fast-rate target words embedded in slow-rate passages.

II. EXPERIMENT 1

Our first experiment investigated English-learning 11-month-olds’ ability to segment trisyllabic words in a passage-to-passage word segmentation task. We tested infants on this baseline condition because using passage-to-passage design is crucial to addressing our two research questions: (1) whether
infants are able to recognize target words despite speaking rate differences, and (2) whether they rely on context speaking rate to normalize speech. Without the passage-to-context design we would not be able to manipulate the context without manipulating the target word (or vice versa). Previous word segmentation studies with infants have involved either passage-to-word or word-to-passage segmentation, but not passage-to-passage segmentation (e.g., Jusczyk et al., 1999; Johnson and Seidl, 2008; Jusczyk and Aslin, 1995; Houston and Jusczyk, 2000). Although these studies showed that as young as 7.5 months of age, infants recognize target words from continuous speech (Jusczyk et al., 1999), we do not know whether they are also able to do so by using a possibly more challenging passage-to-passage design with trisyllabic words. Therefore, in this baseline experiment, we familiarized and tested infants with normal-rate passages, with normal referring to speaking rate that is typical of American English IDS. We started with 11-month-old infants because previous studies, using a similar paradigm and task, suggest that infants at this age are able to adapt to differences along the dimensions of accent and gender. For example, by 10.5 months, infants can generalize different instances of the same word across talkers of different genders (Houston and Jusczyk, 2000). In addition, 12-month-olds, but not 9-month-olds, recognize familiar words across different accents (Schmale et al., 2010).

3. Design

The experiment consisted of a familiarization and a test phase. During the familiarization phase, infants heard two passages with the target words (e.g., elephants and dinosaurs). During the test phase, all the infants heard three blocks of four randomly ordered passages, with two containing the familiar words (elephants and dinosaurs) and the other two containing the unfamiliar words (crocodiles and platypus). The passages used in the familiarization and test phases were different even when the target words were familiar.

4. Apparatus and procedure

A modified version of the HPP (Jusczyk and Aslin, 1995) was used. Each infant was seated on a caregiver’s lap in the center of a single-walled sound booth. The experimenter who sat outside of the booth observed the infant through a video camera which recorded the infant’s head orientation during the experiment. The caregiver wore headphones (3M Peltor Aviation headset 7050, Indianapolis, IN) which played continuous music mixed with white noise to mask the stimuli. The booth was quiet and comfortable and consisted of three panels: a center panel with a green light and two side panels each with a red light. An overhead light was dimmed to make the panel lights more salient. Each trial began with the blinking of the green light on the center panel. When the infant looked at the green light, the light was extinguished and one of the two red lights would begin to blink. A computer program randomly chose which red light to trigger. When the infant oriented at least 30° in the direction of the red light, the stimuli for that trial began to play. The red light blinked until either the infant looked away for 2 consecutive seconds or the auditory stimuli file was complete. At this point, the red light was extinguished and then the center green light began to blink in preparation for the next trial. During the familiarization phase, only the light, but not sounds, were contingent on the infants’ orientation. In other words, the passage was played to completion even if the infants turned away from the red light before it completed. However, during the test phase, both light and sounds were contingent on infants’ head orientation. The

<table>
<thead>
<tr>
<th>Table I. Durational information [Mean, SD, Minimum (Min), Maximum (Max)] for the passages and target words across the normal-, fast-, and slow-rate speech.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Passage</strong></td>
</tr>
<tr>
<td>Type</td>
</tr>
<tr>
<td>Normal</td>
</tr>
<tr>
<td>Fast</td>
</tr>
<tr>
<td>Slow</td>
</tr>
</tbody>
</table>

J. Acoust. Soc. Am. 141 (4), April 2017

Wang et al. 2571
computer recorded the amount of time the infant oriented to the red light while the stimuli played. Orientation time was defined as the amount of time the infant spent looking at the target red light. If the infant turned away from the red light by 30° for less than 2 s, that time was not included in the orientation time, although the light did not extinguish.

Each experimental session began with the familiarization phase which was immediately followed by a test phase. During the familiarization phase, infants listened to two normal-rate passages containing two trisyllabic target words (e.g., elephants and dinosaurs) on alternating trials until they accumulated 30 s of listening time to each passage. The passages were presented continuously and simultaneously from both side speakers. Sound files were initiated by the initial orientation toward one of the flashing side lights and played continuously until infants listened to the stimuli for 30 s.

During the test phase, infants were tested with four new normal-rate six-sentence passages. There were two kinds of passages: two familiar and two unfamiliar. Familiar passages contained the two target words (e.g., elephants and dinosaurs) embedded in new passages and unfamiliar passages contained two novel words not heard during familiarization (e.g., crocodiles and platypus). The test trials were blocked in groups of four so that each passage occurred once per block. Each infant received three blocks of test passages with the order of familiar and unfamiliar passages randomized within a block. The order of each of the sentences within a passage was fixed such that all the infants received passages with the same order of sentences. We also counterbalanced the passages presented in the familiarization and test phases such that half the infants were familiarized with passage 1 and tested with passage 2, while the other half were familiarized with passage 2 and tested with passage 1 (see the passages in the Appendix). In addition, the target words were also counterbalanced such that half the infants were familiarized with elephants and dinosaurs whereas the other half with crocodiles and platypus. The dependent measure was the average orientation time across trials to each stimulus type (familiar, unfamiliar) at test. A Macintosh computer controlled the presentation of the stimuli and recorded the experimenter’s coding of infants’ orientation via a button box. The audio output was fed to two Cambridge Soundworks Ensemble II speakers.

We predict that if infants are able to recognize the target words, they would show significantly different looking times to the familiar versus unfamiliar passages. It should be noted that either a familiarity or a novelty preference at the group level would suggest that infants have recognized differences in the two stimulus sets. Given the design of the study in which one group hears one set of words as familiar and another hears the opposite set of words as familiar, either such effect (familiarity, novelty) would suggest recognition. Nonetheless, these different responses are interpretable based on previous literature examining the factors that lead to familiarity or novelty responses from infants. A familiarity preference typically occurs when a task is more challenging for a group of infants. On the other hand, a novelty preference typically occurs when a task is easier for a group of infants. Previous studies examining the mechanisms leading to these different preference patterns suggest for example that novelty preferences may be found when familiar stimuli have been completely processed and infants can then devote their attentional resources to the unfamiliar stimuli, but familiarity preferences may occur when representation of familiar stimuli are still tenuous (Hunter and Ames, 1988; Rose et al., 1982; Roder et al., 2000). This ease/difficulty is not just related to processing ease, but also to infant age, though the two are surely highly correlated. For example, Colombo and Bundy (1983) showed that the direction of preference changed as a function of infant age. Specifically, 2-month-olds showed a familiarity preference after receiving 18 familiarization trials, whereas 4-month-olds, who were cognitively more advanced, showed a novelty preference.

B. Results and discussion

In all experiments reported in this paper, no interactions were found between differences in orientation times to the different types of passages (familiar, unfamiliar) with Block (1,2,3) [experiment 1: F(2, 60) = 0.41, p = 0.526, block data were missing for one child; experiment 2: F(2, 62) = 1.14, p = 0.327; experiment 3: F(2, 62) = 0.77, p = 0.469; experiment 4: F(2, 62) = 0.71, p = 0.490]. Thus, within the results discussed below mean orientation times to the familiar and unfamiliar passages were calculated across the three test blocks in all experiments.

The average orientation times to familiar and unfamiliar trials were 6.32 s (SD = 3.10 s) and 7.40 s (SD = 3.67 s), respectively. A paired-samples t-test revealed a significant novelty preference: Infants looked longer to the unfamiliar than to the familiar passages, t(31) = 2.06, p = 0.048, $\eta^2 = 0.120$ (all the reported t-tests were two-tailed), see Fig. 1. This novelty preference may seem surprising, because the passage-to-passage segmentation task involving trisyllabic target words in our study was expected to be challenging and novelty preferences are typically observed when tasks are relatively easy (Colombo and Horowitz, 1986). However, it should be noted that our stimuli, task, and infant age differed from previous studies showing familiarity preferences (e.g., Houston and Jusczyk, 2000; Singh et al., 2002) and these differences may have all contributed to the novelty preference observed in our study. Nonetheless, the important thing to note here is that a consistent preference direction, either novel or familiar, suggests that word recognition did occur at test.

This baseline passage-to-passage segmentation experiment thus revealed that 11-month-olds were able to recognize target trisyllabic words in continuous speech. Thus, in the next three experiments, we manipulated the speaking rate of the test stimuli to investigate infants’ ability to recognize words across different speaking rates.

III. EXPERIMENT 2

Experiment 2 aimed to explore whether 11-month-old infants are able to recognize familiar words in continuous speech across different speaking rates. Specifically, after being familiarized with normal-rate speech, infants were tested on passages in which context and target words were produced in fast-rate speech. If 11-month-olds are able to
recognize and segment words from speech stream with different rates, we expect that their orientation times to familiar and unfamiliar passages would be different. On the contrary, if they are not able to recognize words across a rate change from normal-rate to fast-rate speech, then we expect the orientation times to the familiar and unfamiliar passages would be similar.

A. Methods

1. Participants

Data were analyzed from 32 infants (13 female) between the ages of 10.1 and 12.37 months ($M = 11.15, SD = 0.73$). An additional 12 infants were tested whose data were not included due to fussiness (9) and being 2.5 SD off the mean (3).

2. Stimuli

The familiarization stimuli were the same as those used in experiment 1. However, infants were tested on fast-rate passages. To create the fast-rate test stimuli, we first spliced out the medial unstressed syllables from the target words (le from elephants, no from dinosaurs, co from crocodiles, and ty from platypus) at upward zero crossings. Then the normal-rate passages were time-compressed by a factor of 0.6 by using a MATLAB implementation of the Synchronous-OverLap-Add (SOLA) algorithm, see waveform in Fig. 2. This SOLA manipulation kept intact the crucial spectral information of the speech and retained normal rhythmic and prosodic information, thus the manipulated sentences sounded quite natural. All stimuli were then amplitude-normalized to 70 dB. The durational information for the passages and target words is shown in Table I.

Our motivations for deleting the unstressed medial syllables of the target words were twofold. First, in fast-rate natural speech, unstressed syllables in trisyllabic words undergo substantial reduction (Fokes and Bond, 1993). For example, the word probably could be produced as proly in fast-rate speech. Thus, deleting the unstressed syllable from the target word produces a fair representation of the properties observed in real life fast-rate speech with trisyllabic words. Second, deletion of medial unstressed syllables allows us to tap into our second research question of interest—specifically, whether language-learning infants can recognize words produced at different speaking rates in the same the way adults would such that they would recognize words even when deletion occurs in the context of fast-rate speech (Dilley et al., 2014) rather than treating words with these omitted syllables as novel. Note that the reason we only removed word-medial unstressed syllables from the target, but not context words, was because the words in the context were either monosyllabic or bisyllabic, in which the unstressed syllables would less often be deleted (Johnson, 2004; Fokes and Bond, 1993).

![FIG. 1. (Color online) Averaged orientation time to familiar and unfamiliar passages. Error bars indicate mean standard errors. * indicates that groups are statistically significant.](image1)

![FIG. 2. Waveforms of a sample stimulus across different speaking rates. The sections of the waveform with background shading correspond to the context speech while the sections without shading represent target words.](image2)

J. Acoust. Soc. Am. 141 (4), April 2017 Wang et al. 2573
3. Design

The design was the same as in experiment 1 except for the substitution of the fast-rate test passages for the normal-rate test passages.

4. Apparatus and procedure

The apparatus and procedure were the same as in experiment 1.

B. Results and discussion

Infants’ orientation times to the familiar and unfamiliar passages were 4.41 s (SD = 1.86 s) and 4.46 s (SD = 1.65 s), respectively. A paired-samples t-test showed no difference in infants’ looking times to the two types of passages \( t(31) = 0.16, p = 0.878 \), indicating that 11-month-olds did not distinguish between passages containing familiar and unfamiliar words. These findings suggest that the task may be too challenging for 11-month-olds, thus in experiment 3, we tested an older group, 14-month-olds, using the same design.2

IV. EXPERIMENT 3

In experiment 3, we tested whether 14-month-olds are able to recognize target words across normal-rate and fast-rate speech.

A. Methods

1. Participants

Thirty-two 14-month-old infants (14 female) were tested. They were between the ages of 13.19 and 15.1 month \( (M = 14.08, SD = 0.59) \). An additional 8 infants were tested whose data were not included due to fussing, crying (5), or experimenter error (3).

2. Stimuli

The stimuli were same as in experiment 2.

3. Design

The design was same as in experiment 2.

4. Apparatus and procedure

The apparatus and procedure were the same as in experiment 2.

B. Results and discussion

Infants’ orientation times to the familiar and unfamiliar passages were 4.41 s (SD = 1.86 s) and 4.46 s (SD = 1.65 s), respectively. A paired-samples t-test showed no difference in infants’ looking times to the two types of passages \( t(31) = 0.16, p = 0.878 \), indicating that 11-month-olds did not distinguish between passages containing familiar and unfamiliar words. These findings suggest that the task may be too challenging for 11-month-olds, thus in experiment 3, we tested an older group, 14-month-olds, using the same design.2

VI. EXPERIMENT 4

The findings from experiments 1–3 show that 11-month-olds recognize target trisyllabic words from normal-rate speech in a passage-to-passage segmentation task; however, only 14-month-olds, but not 11-month-olds, recognize words across different (fast and normal) speaking rates. In experiment 4, we explore possible mechanisms that may have allowed the 14-month-old infants to recognize target words in fast-rate speech. As mentioned in Sec. I, it is possible that 14-month-olds succeeded in word recognition across different speaking rates because they were able to normalize the speaking rate of the target words based on the rate of the contextual speech (in this case fast), using mechanisms similar to that used by adults (Dilley and Pitt, 2010; Gordon, 1988; Wade and Holt, 2005). Alternatively, it is possible that infant listeners did not use context for normalization at all, but instead exploited some other more local mechanisms. To address the question of whether infants use the rate of the contextual speech for rate normalization, we tested infants on their ability to accommodate fast-rate familiar words (with the deletion of the middle unstressed syllable) embedded in slow-rate carrier sentences. If infants in this experiment orient equally long to the familiar and unfamiliar passages, it may be that they fail to recognize fast-rate words in slow-rate carrier sentences, suggesting that infants may be at least somewhat reliant on context to normalize different speaking rates. On the contrary, if they look longer to the familiar than to the unfamiliar passages, then this may suggest that they are not reliant on context to normalize speaking rate.

A. Methods

1. Participants

Data were analyzed from 32 infants (15 female) between the ages of 13.39 and 14.93 months \( (M = 13.94, SD = 0.38) \). An additional 11 infants were tested whose data were not included due to fussiness.

2. Stimuli

The stimuli was the same as in experiments 2 and 3, except that the context was slowed through time expansion with a SOLA parameter of 1.6 from the original normal-rate speech, see waveform in Fig. 2. All stimuli were then amplitude-normalized to 70 dB. The durational information for the passages and target words is shown in Table I.
3. Design

The design was the same as in experiments 2 and 3.

4. Apparatus and procedure

The apparatus and procedure were the same as in experiments 1–3.

B. Results and discussion

Infants’ orientation times to the familiar and unfamiliar passages were 10.2 s (SD = 5.96 s) and 9.51 s (SD = 5.68 s), respectively. A paired-samples t-test showed that infants’ orientation times to the familiar and unfamiliar passages were similar, t(31) = 0.70, p = 0.492, see Fig. 1. These results showed that when the context and the target differ in speaking rate, 14-month-olds failed to recognize the target words. This suggests that infants may have succeeded in experiment 3 because they use context to normalize speaking rate differences.

VI. GENERAL DISCUSSION

Despite the acoustic variability of speech signals adult listeners are able to perceptually accommodate speech across different talkers, accents, and speaking rates (Bradlow and Bent, 2008; Samuel and Kraljic, 2009; Toscano and McMurray, 2015). However, the ability of young infants to process speech under conditions of variability seems to be more limited and dependent on language experience and maturation. For example, young infants show more difficulty in recognizing familiar words across different accents and talkers than their older peers (Houston and Jusczyk, 2000; Best et al., 2009; Schmale et al., 2011). One source of variability, speaking rate, has not been explored up to this point. We took a first step toward filling this gap by asking whether infants can recognize words across different speaking rates. As a secondary research question, we explored one mechanism infants might use to recognize words across different speaking rates. It was predicted that given their lesser degree of top-down knowledge, working memory, and/or proficiency in using phonetic cues, they may utilize different mechanisms from adults and may fail to use contextual rate for normalization. Our results show that 14-month-olds, but not 11-month-olds, were able to cope with variability in speaking rate and succeeded in recognizing target words across different rates. In addition, infants seemed to use a contextually driven rate-specific mechanism to normalize speech across different speaking rates, much the way adults do.

In sum, our results are the first to show that infants are capable of handling variation in speaking rate from very early on; specifically, by the age of 14 months, they recognize familiar words from an unfamiliar context in which speech is time-compressed. The findings that 14-month-olds, but not 11-month-olds, succeed in this task supports the view that success in word recognition under taxing conditions may depend on a complex interplay of linguistic knowledge and cognitive factors, both of which may be involved in the rapid recognition of target words from speech. The timeline of this development is broadly consistent with what we know about the impact of age/experience on infants’ accommodation of other types of variability in speech, such as unfamiliar accents, gender, or idiolects. For example, only infants at 12 months, but not 9 months, recognize instances of familiar words across different dialects/accents (Schmale et al., 2010). Our results, together with these findings, suggest that infants only gradually learn to disregard linguistically irrelevant acoustic variation in order to achieve fast and reliable speech processing.

However, it is important to point out that infants’ ability to recognize words in the face of variability may largely depend on the nature of the task or the acoustic complexity of the stimuli. To begin with, infants may need more language experience to accommodate unfamiliar speech in tasks that involve word learning, because word learning requires infants to encode both visual and auditory signals within a very short time period. Indeed, only 30-, but not 24-month-olds, are able to generalize learned words across unfamiliar accents (Schmale et al., 2011). In addition, the task of passage-to-passage segmentation may be more demanding than a passage-to-word or word-to-passage segmentation task, and therefore, only by 14 months of age could infants normalize for speaking rate differences in our studies, whereas by 10.5 and 12 months of age, infants were able to recognize words across gender or accent (Houston and Jusczyk, 2000; Schmale et al., 2010). Second, the ability to successfully accommodate variable speech may also be influenced by the degree of acoustic distortion that is under examination. From an acoustic perspective, it is a more challenging task for infants to accommodate accentual differences, relative to speech differences in rate, because accents tend to involve acoustic changes along both indexical and linguistically-relevant dimensions. This could explain why, in our study, infants succeed at a fairly young age (14 months), but why it took until 19 months for English-learning infants to accommodate Jamaican English (Best et al., 2009). This is perhaps because, unlike accentual variability, speech-rate distortions are more homogeneously distributed across all sounds independent of their phonetic nature, and therefore are potentially more predictable than those more phonetically distorted changes. In a nutshell, the complexity of the stimuli, coupled with task demands, could account for some of the developmental differences reported in the literature for coping with different types of speech variability, but only a systematic study of these different kinds of variability while holding task-related demands the same could help to address this question.

Our second research question explored whether infants use rate-related contextual information to accommodate speech-rate variability. Our results revealed that, when the speaking rate of the context did not match the speaking rate of the target words, 14-month-olds were not able to recognize the target words, suggesting that they rely on the rate of the contextual speech to facilitate their processing of speech with different rates. In light of these findings, it seems that infants’ word recognition skills are modulated by speech context. When contextual variability displays specific rate
parameters, infants and young toddlers are able to rapidly adjust their expectations for the pronunciation of target words in relation to the rate at which the contextual speech is produced. Therefore, when target words deviate from their rapidly formed expectations of what they should sound like in terms of rate, infants fail to identify the target words from the speech.

Note that the fact that infants in our study at least partially relied upon a contextual-rate mechanism to normalize speech does not necessarily argue against the use of other mechanisms at any point in processing. First, adjustment to a new speaking rate may occur gradually, with recognition improving over a relatively longer period of experience. Indeed, adult performance improves significantly with increased exposure to compressed speech (Dupoux and Green, 1997; Voor and Miller, 1965). Like adults, toddlers accept non-standard pronunciations when variability is expected after sufficient exposure. Specifically, in three theoretically connected studies involving recognition of Spanish-accented words previously learned in an American English accent (Schmale et al., 2012, 2015, 2011), 24-month-old toddlers succeed if, and only if, they were previously exposed to different sources of variability, such as English or Spanish-accented passages recorded from different speakers, or presented with different pictures of human faces varying in age and gender. Thus, it is possible that infants may also use other strategies after they have been exposed to a higher degree of speech variability for a longer period of time, which may assist them in forming more flexible expectations on how the target words should be pronounced. One important implication of this interpretation is that if infants were pre-exposed to speech with different speaking rates for a sufficient amount of time, they might be able to recognize target words from speech even if the speaking rate of the context speech does not match the speaking rate of the target words. However, this research does not allow us to directly test this hypothesis and we encourage future studies to address this question.

A question that remains unanswered is whether infants rely on the speech context to a greater extent than adults do. Adult literature has demonstrated that in addition to relying on speech context for speech processing (Holt et al., 2000; McMurray and Jongman, 2011; Sawusch and Pisoni, 1974; Toscano and McMurray, 2015), adults also use lexical knowledge to recognize words in the presence of contextual rate discontinuities (Bradlow et al., 1996). Adult listeners tend to recruit more top-down lexical processing in their perceptual accommodation of unfamiliar speech, as compared to the accommodation of familiar speech (Lev-Ari, 2014), and adjust their short-term representation of unfamiliar words to achieve accurate processing. For example, Norris et al. (2003) found that adult listeners have difficulty in calibrating non-native pronunciations of native sounds when these pronunciations are provided in a non-native lexical environment. Like adults, infants may also have difficulties in accommodating speech variation in the absence of lexical expectations providing evidence for how sounds should be pronounced in native speech. Indeed, research has demonstrated an improvement in infants’ accommodation of unfamiliar speech after being exposed to speech samples that are lexically familiar to them (van Heugten and Johnson, 2012; White and Aslin, 2011). Therefore, the fact that infants in our study rely only on contextual rate to normalize speech is not surprising, given that their very limited lexical experience may prevent them from using top-down lexical knowledge to process this speech.

VII. CONCLUSIONS AND FUTURE DIRECTIONS

In summary, the current study presents an important step in our understanding of how infants cope with speaking rate variability. We show that by 14 months of age, infants can recognize target words in the face of variability in speaking rate. In addition, we show that infants are sufficiently reliant on the rate of contextual speech for normalization such that when there is a mismatch between contextual rate and target word rate there is no recognition of segmented wordforms. These findings have implications for our understanding of how infants identify words in connected speech, as well as how the prosodic properties and temporal properties of speech context influence speech segmentation and word recognition. Since infants grow to become competent language processors, there is a genuine need for future research on the transition children make from using rate-specific mechanisms to a variety of different mechanisms (e.g., lexical) in processing speech at different rates.

ACKNOWLEDGMENTS

This research was partly supported by Purdue Research Foundation Grant. We would like to thank Rana Abu Zhaya and Katie Courtaney for assistance recruiting and testing infants; and the infants and their parents who participated in our study. We would also like to thank Dr. Derek Houston, Dr. Jongmin Jung, and anonymous reviewers for helpful comments on an earlier version of this paper.

APPENDIX

Dinosaurs passage 1

Dinosaurs are large beasts that eat a lot. Your dinosaurs are much bigger than mine! Long ago the Earth was covered with dinosaurs. Dinosaurs can eat meat or plants like these. I wish that I had dinosaurs like yours! Look between those trees and you will see dinosaurs.

Dinosaurs passage 2

Dinosaurs appeared in the Triassic period. All the dinosaurs died millions of years ago. Birds are the descendants of dinosaurs. Dinosaurs can be herbivores or carnivores. The word dinosaurs means “terrible lizard.” T-Rexs are very scary dinosaurs.

Elephants passage 1

Elephants are large animals that eat a lot. His elephants are much bigger than yours! The African plains are covered with elephants. Elephants can charge if they are angry. I wish that I could see elephants like you did! Go for a walk in Kenya to see elephants.
Elephants passage 2
Elephants are big mammals that live in Africa. Asian elephants are found in India. Mastodons are the ancestors of elephants. Elephants eat lots and lots of grasses. Young elephants live with their families. Low frequencies are perceptible to elephants.

Crocodiles passage 1
Crocodiles have webbing on their toes to help them swim. Some crocodiles live in America. Alligators are not the same as crocodiles. Crocodiles live in rivers, lakes, or wetlands. If you see crocodiles you should run away. It would not be fun to be bitten by crocodiles.

Crocodiles passage 2
Crocodiles can swim very fast after prey. Some crocodiles like to hunt in the dark. The fastest ones are Australian crocodiles. Crocodiles can grow as long as 17 feet! The jaw of crocodiles is very strong. Humans should be careful to avoid crocodiles.

Platypus passage 1
Platypus venom can kill a dog or cat. Duckbilled platypus can be found in Australia. Snakes and water rats can eat platypus. Platypuses eat worms and shrimp for dinner. Some platypus sleep for 14 h a day. It is important to protect the platypus.

Platypus passage 2
Platypus fur is thick, brown, and waterproof. The platypus tail is made mostly of fat. Digging for food is important for the platypus. Platypus nurse their young like humans do! I would not want a platypus for a pet. Argentina has animals like platypus.

1We time-compressed the normal-rate passages using a range of parameters and tested these passages on six native speakers of English. Specifically, listeners were asked to rate the naturalness of each passage. The passages created by using the parameter of 0.6 were voted as the most natural fast-rate speech by these listeners.

2A reader raised an interesting question regarding our interpretation of these results. He pointed out that we were not able to tell whether (1) fast-rate speech was just more difficult, or (2) 11-month-olds detected the difference between the fast-rate words with the deletion of the middle unstressed syllable (e.g., ephant) and the target words (e.g., elephant). It is not possible to tease apart these two factors using our current design. However, it would be very interesting for future studies to address this question. In addition, in an unreported experiment conducted in our lab, we tested 11-month-olds’ ability to recognize target words from slow-rate speech after they were familiarized with normal-rate speech. The design was similar to experiment 2 reported in this paper, with the exception that the test speech rate was slow and the unstressed syllables in target words were maintained. Infants in this experiment failed to recognize target words. Thus it is more likely that the null results in experiment 2 were due to the fact that different rate/non-normal rate speech is more difficult to process for infants of this age.

