Real-time coordination of visual and linguistic processes in novice readers

Yi Ting Huang

Department of Hearing and Speech Sciences, University of Maryland College Park, College Park, MD 20742, USA

Abstract

Skilled reading requires coordinating real-time visual fixations, orthographic analyses, and phonological encoding across multiple words in sentences. These procedures are well studied in experienced readers, but less is known about their status during development. To investigate how visual properties influence the origins of coordinated processing, the current study combined rapid automatized naming (RAN) with an eye-tracking paradigm and compared the timing of fixations and vocalizations in typically developing adults and 6-year-old children. Within RAN displays, sequences varied visual features of items (i.e., similar such as \( p-q \) vs. dissimilar such as \( p-t \)) and their locations in rows (i.e., row-initial vs. row-medial positions). Adults and children accessed parafoveal preview of subsequent items when fixating on current items, leading to longer latency to speak for similar items compared with dissimilar ones. Both groups also vocalized previous items while fixating on current items, leading to longer eye–voice overlap for row-medial items compared with row-initial ones. Yet, relative to adults, children exhibited exaggerated delays in latency to speak from parafoveal preview and reduced eye–voice overlap due to row transitions. Together, this suggests that coordinated processing is present at the earliest points of development but that greater inexperience increases susceptibility to momentary visual hurdles. Relationships to previous work on real-time RAN performance in dyslexic adults and children are discussed.

© 2018 Elsevier Inc. All rights reserved.
Introduction

Skilled reading forms the basis for education and communication, but its ubiquity masks its underlying complexity. To extract meaning from text, readers must visually analyze and linguistically retrieve properties of multiple words in sentences. As such, they must negotiate the need to move forward to not forget prior material with the potential for confusion when too much information is concurrently available. These decisions occur seamlessly in adults, but less is known about how they unfold in children. To examine coordinated processing in less experienced populations, past studies have relied on the rapid automatized naming (RAN) task, which asks children to name letter or number displays as quickly and accurately as possible. Similar to reading, RAN performance requires targeting fixations, encoding orthographic properties, and accessing phonological features across simultaneously presented items (Fig. 1) (Breznitz, 2005; Gordon & Hoedemaker, 2016; Jones, Obregon, Kelly, & Branigan, 2008; Kuperman, Van Dyke, & Henry, 2016; Protopapas, Altani, & Georgiou, 2013a). Total response times are associated with current reading ability (Gordon & Hoedemaker, 2016; Wagner & Torgesen, 1987), future aptitude (Compton, 2003; Lervag & Hulme, 2009; Wagner et al., 1997), and impairment risk (Denckla & Rudel, 1976; Georgiou, Parrila, Manolitsis, & Kirby, 2011).

Importantly, recent advancements in eye-tracking have provided a window into the real-time dynamics of RAN performance. Much of this work investigates the processing correlates of reading impairments. For example, although all adults engage in parafoveal preview of subsequent items (n + 1) when fixating on current items (n) (Jones, Ashby, & Branigan, 2013; Jones, Snowling, & Moll, 2016; Jones et al., 2008), those with dyslexia experience greater interference from visually similar items (e.g., p–q) relative to non-dyslexic peers (Al Dahhan et al., 2014; Moll & Jones, 2013). Nevertheless, evidence from development remains mixed. Relative to non-dyslexic peers, dyslexic 10-year-olds demonstrate delayed naming rates and increased error for visually similar items (Al Dahhan, Kirby, Brien, & Munoz, 2017). This suggests access to parafoveal preview, much like that among adults. Yet, unlike non-dyslexic peers, dyslexic 10-year-olds show limited improvements in fixation duration for simultaneously presented items compared with individually presented ones (Yan, Pan, Laubrock, Kliegl, & Shu, 2013). This suggests less parafoveal preview compared with non-dyslexic peers. Similarly, recent work suggests that impairment status generates processing distinctions that go beyond developmental delays. Although total response times and error rates in dyslexic 10-year-olds are similar to those in non-dyslexic 7-year-olds, saccade and regression counts for visually similar items remain exaggerated in impaired populations (Al Dahhan et al., 2017).

However, prior focus on reading impairments leaves open questions of how coordinated processing unfolds during typical development. This creates challenges in interpreting population differences.

Fig. 1. Real-time RAN performance can be measured along multiple dimensions. Vocal duration assesses the time from the onset of articulating an item (e.g., B) to the offset of articulating the same item (e.g., B). Latency to speak assesses the time from the onset of fixating an item (e.g., N) to the onset of articulating the same item (e.g., N). Eye–voice overlap assesses the time from the onset of fixating a current item (e.g., N) to the offset of articulating the previous item (e.g., B). Total response time increases with vocal duration and latency to speak, but it decreases with eye–voice overlap.
which may reflect when coordination strategies are acquired or what strategies are attained. Moreover, because previous studies examined children with one or more years of classroom instruction, the developmental origins of coordinated processing remain unclear. To address this gap in knowledge, the current study compared real-time RAN performance in typically developing adults and 6-year-olds along three dimensions. First, children’s vocal duration for current items (n) is highly related to their total response time (Clarke, Hulme, & Snowling, 2005; Georgiou, Parrila, & Kirby, 2006; Georgiou, Parrila, Kirby, & Stephenson, 2008). Thus, varying syllable length provides a baseline for developmental delays in item-level processing speed. Second, adults’ latency to speak for (n) is influenced by parafoveal preview of (n + 1) (Jones et al., 2008, 2013, 2016). Thus, varying the visual similarity of adjacent items reveals the extent to which such effects are also present in novice readers. Finally, prior work demonstrates that unimpaired adults (Gordon & Hoedemaker, 2016) and 10-year-olds (Pan, Yan, Laubrock, Shu, & Kliegl, 2013) often initiate fixation to (n) before articulation of (n – 1) is complete. This parallel strategy is traditionally indexed by the number of items where the eyes lead the voice. Eye–voice spans are longer with highly practiced numbers compared with less practiced letters (Gordon & Hoedemaker, 2016), but this asymmetry is smaller in impaired adults (Hogan-Brown, Hoedemaker, Gordon, & Losh, 2014) and children (Pan et al., 2013). However, because the eyes often lead the voice by only a single item, categorical spans may be inadequate for isolating processing variation in novice readers. Moreover, categorical spans have focused on between-display effects; thus, they do not address the impacts of within-display features such as row transitions. To investigate these dynamics, the current study adopted a continuous measure of parallel processing called eye–voice overlap. Relative to row-medial items, fixations from (n/C0) to (n) involve greater distances for row-initial items. This may shorten eye–voice overlap by delaying the millisecond onset of fixating on (n) relative to the millisecond offset of articulating (n – 1). Comparisons between adults and novice readers will reveal the extent to which common visual hurdles generate disproportionate coordination challenges during development.

Method

Participants

A total of 46 6- and 7-year-old children were recruited from first-grade classrooms in local schools. Because data collection occurred during the fall semester, children had 1–4 months of literacy instruction. A parental survey indicated no known impairments in vision, hearing, or language. Reading readiness was assessed through the Woodcock Reading Mastery Tests (Woodcock, 2011). Of the 46 child participants, 1 was excluded due to scores that were more than 2 standard deviations below age means, 3 were excluded due to excessive track loss, and 2 were excluded due to problems with vocalization recordings. This resulted in a final sample of 40 children (15 girls and 25 boys) (Table 1). They were compared with 22 undergraduates who participated for course credit. All reported normal or corrected-to-normal vision and no hearing or language impairments. Of the 22 adult participants, 2 were excluded due to excessive track loss, resulting in a final sample of 20 adults (16 women and 4 men). Their performance on the Author Recognition Task (M accuracy = 15, SD = 5) revealed reading scores consistent with documented college-aged samples (Acheson, Wells, & MacDonald, 2008; Moore & Gordon, 2015).

Table 1
Age and reading measures for the current sample of novice readers.

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Standard deviation</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years; months)</td>
<td>6; 7</td>
<td>0; 5</td>
<td>6; 3–7; 6</td>
</tr>
<tr>
<td>WRMT-III</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phonological awareness</td>
<td>99</td>
<td>14</td>
<td>75–124</td>
</tr>
<tr>
<td>Word identification</td>
<td>106</td>
<td>19</td>
<td>83–106</td>
</tr>
<tr>
<td>Word attack</td>
<td>106</td>
<td>18</td>
<td>87–106</td>
</tr>
</tbody>
</table>

**Procedures and materials**

Participants were told that they would see letters or numbers on a screen and needed to produce their names as quickly and accurately as possible, from left to right and from top to bottom. They first took part in practice displays with sequentially ordered letters and numbers. After feedback, they moved onto critical trials. Each trial began with a fixation point marking the location of the first item in the upper left-hand corner. Once participants fixated on this point, the experimenter initiated the display, which remained until vocalization of the last item. Displays were presented approximately 20 in. away from participants’ eyes using a 17-in. Dell monitor with 1280 × 1024 resolution and a 60-Hz refresh rate. While displays were viewed with both eyes, right-eye fixations were sampled at a rate of 500 Hz using an EyeLink 1000 eye-tracker (SR Research, Ottawa, Ontario, Canada). Speech production was recorded using a table-mounted Shure SM58 cable microphone that was paired with an M-Audio amplifier and ASIO sound card.

Critical trials involved eight RAN displays (four letters and four numbers), each featuring 48 total items (8 unique items repeated six times) equally spaced in eight columns and six rows. Number displays manipulated length such that \((n)\) had fewer syllables in short displays \((20, 12)\) and more syllables in long displays \((21, 11)\). Additional numbers served as baseline and spillover items \((3, 4, 5, 10)\). Letter displays manipulated similarity such that \((n)\) and \((n+1)\) were confusable in similar displays \((q–p, b–d)\) and not confusable in different displays \((t–p, f–d)\). Other letters were used as baseline items \((s, n)\). Within displays, location was manipulated such that row-initial sequences were in first, second, and third columns, whereas row-medial sequences were in fifth, sixth, and seventh columns. For length and similarity manipulations, four critical sequences were located in medial columns to minimize inaccurate saccades. Across conditions, \((n – 1)\) and \((n + 1)\) were identical. For the location manipulation, eight sequences were located in short-number and different-letter displays to minimize length and similarity effects. Moreover, to avoid task acclimation or fatigue, critical sequences never occurred in first or last rows. Items were printed in black Calibri font on white background with a 1.1° × 0.7° visual angle for letters and single-digit numbers and a 1.1° × 1.5° angle for double-digit numbers. To minimize fatigue during the study, each RAN display was followed by four filler trials involving single-object naming (e.g., ball, dog, spoon, foot). The order of display presentation was randomized across participants.

**Results**

Areas of interest (AOIs) were defined from the center of an item to halfway to the adjacent item and featured a 3° × 3° visual angle. Onsets and offsets for fixating AOIs were calculated in Data Viewer (SR Research). Onsets and offsets for vocalizing items were coded in Praat (Boersma, 2001). Approximately 4.2% of items in adults and 11.2% of items in children were excluded from subsequent analyses due to no fixations (e.g., track loss, skips). Another 0.1% of items in adults and 0.9% of items in children were excluded due to vocalization errors (e.g., name substitutions, disfluencies). The remaining data were used to calculate four measures:

1. **Total response time:** For each display, the time from the onset of fixating on the first item in a display until the offset of vocalizing the last item. Larger values correspond to slower processing of the entire display.
2. **Vocal duration:** For critical sequences, the time from the onset of vocalizing an item \((n)\) to the offset of vocalizing the same item. Larger values correspond to increased challenges articulating an item.
3. **Latency to speak:** For critical sequences, the time from the onset of fixating on an item \((n)\) to the onset of vocalizing the same item. Larger values correspond to increased challenges due to parafoveal preview.
4. **Eye–voice overlap:** For critical sequences, the time from the onset of fixating on an item \((n)\) to the offset of vocalizing the previous item \((n – 1)\). Larger values correspond to increased parallel processing across adjacent items.
To minimize the impacts of outliers, we first log-transformed raw values and then excluded values more than 3 standard deviations from age means. In adults, this omitted 0.6% of vocal duration, 2.8% of latency to speak, and 1.0% of eye–voice overlap. In children, this omitted 1.6% of vocal duration, 6.7% of latency to speak, and 3.9% of eye–voice overlap. Table 2 presents descriptive statistics for the remaining data. Log values were analyzed in linear mixed-effects models using the lme4 software package in R (Bates, Maechler, Bolker, & Walker, 2014). Maximal models included both random slopes and intercepts for participants and items, but simpler models were adopted with random intercepts only when maximal models failed to converge (Barr, Levy, Scheepers, & Tily, 2013). Fixed-effects variables included condition (length, similarity, or location), display (number or letter), and age (adults or children). Deviation coding compared condition means with grand means. Parameter-specific p values were estimated through normal approximation of t statistics. All data sets and analysis codes can be found at https://osf.io/dwt82.

Unsurprisingly, children were slower than adults across all analyses (main effect of age, ps > .001). However, to distinguish general delays from processing distinctions, we focused on interactions between age and condition. Because number displays include multisyllabic items, they generated longer total response times compared with letter displays (t = 3.26, p < .01). Nevertheless, there was no length effect in number displays (t = 0.33, p > .70), no similarity effect in letter displays (t = 0.47, p > .60), and no interactions with age (ps > .60). This confirms that total times have limited temporal resolution for isolating item-level dynamics. Next, we turned to vocal duration across critical sequences in number displays. As expected, durations were equivalent for identical items on (n – 1) (t = 1.05, p > .20) and (n + 1) (t = 1.88, p > .60). However, they were longer on (n) when items featured more syllables compared with fewer syllables (t = 6.07, p < .001). Syllable length effects were found in adults (t = 9.28, p < .001) and children (t = 3.04, p < .01), and the greater magnitude among experienced readers led to an interaction between length and age (t = 2.38, p < .05). Together, this confirms that real-time measures are sensitive to item-level RAN performance in novice readers. Moreover, although children are slower than adults in aggregated (total times) and disaggregated (vocal durations) measures, they do not experience disproportionate delays when increased syllable length makes items more difficult to produce (Fig. 2).

This contrasted with effects of visual similarity and row location. To isolate the presence of parafoveal preview, we focused on latency to speak in letter displays. As expected, no differences were found for identical items on (n – 1) (t = 0.10, p > .90) and (n + 1) (t = 1.47, p > .10). However, consistent with documented patterns, latency to speak on (n) was greater when (n + 1) was similar compared with different (t = 3.82, p < .001). Importantly, although preview effects were present in both adults (t = 3.14, p < .01) and children (t = 3.61, p < .001), an additional Age × Similarity interaction reveals that children experienced greater parafoveal interference compared with adults (t = 2.43, p < .01). Next, to examine coordinated processing across row locations, we focused on eye–voice overlap in letter and number displays. As expected, eye–voice overlap on (n – 1) was shorter for row-initial items compared with row-medial ones, leading to a location effect (t = 2.89, p < .01) and an interaction with age (t = 3.79, p < .001). Although row transitions did not affect adult processing (t = 1.42, p > .15), it reduced eye–voice overlap in children (t = 5.72, p < .001). Moreover, for both adults (t = 4.29, p < .001) and children (t = 5.15, p < .001), overlap on (n) was greater for row-initial items compared with row-medial ones, leading to a location effect (t = 5.36, p < .001) but no interaction with age (t = 1.10, p > .20). However, the impacts of row transition were resolved by (n + 1), leading to no location effect or interaction with age (ps > .40).

Finally, to investigate documented effects of item familiarity across RAN displays, we compared children’s row transitions across well-practiced numbers versus less-practiced letters. Consistent with prior findings, eye–voice overlap on (n – 1) was greater for numbers compared with letters, leading to an effect of display type (t = 3.52, p < .001) and an interaction with row location (t = 2.84, p < .001). Whereas eye–voice overlap on (n – 1) was shorter for row-initial items compared with row-medial ones, this difference was greater for letters (t = 6.99, p < .001) compared with numbers (t = 2.28, p < .05). This suggests that row transitions may pose particular challenges for maintaining parallel processing of less-practiced items. Nevertheless, these effects were short-lived. Although eye–voice overlap on (n) was greater for row-initial items compared with row-medial ones (t = 5.15, p < .001), there was no effect or interaction with display type (p > .70).
Table 2
Descriptive statistics for real-time measures in critical sequences of RAN displays.

<table>
<thead>
<tr>
<th></th>
<th>Vocal duration</th>
<th>Latency to speak</th>
<th>Eye–voice overlap</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n - 1</td>
<td>n</td>
<td>n + 1</td>
</tr>
<tr>
<td>Adults</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number short</td>
<td>383</td>
<td>72</td>
<td>421</td>
</tr>
<tr>
<td>Number long</td>
<td>358</td>
<td>70</td>
<td>533</td>
</tr>
<tr>
<td>Letter different</td>
<td>388</td>
<td>58</td>
<td>348</td>
</tr>
<tr>
<td>Letter similar</td>
<td>401</td>
<td>72</td>
<td>337</td>
</tr>
<tr>
<td>Number long</td>
<td>458</td>
<td>70</td>
<td>533</td>
</tr>
<tr>
<td>Letter different</td>
<td>401</td>
<td>72</td>
<td>337</td>
</tr>
<tr>
<td>Letter similar</td>
<td>401</td>
<td>72</td>
<td>337</td>
</tr>
<tr>
<td>Number long</td>
<td>458</td>
<td>70</td>
<td>533</td>
</tr>
<tr>
<td>Letter different</td>
<td>401</td>
<td>72</td>
<td>337</td>
</tr>
<tr>
<td>Letter similar</td>
<td>401</td>
<td>72</td>
<td>337</td>
</tr>
<tr>
<td>Number long</td>
<td>458</td>
<td>70</td>
<td>533</td>
</tr>
<tr>
<td>Letter different</td>
<td>401</td>
<td>72</td>
<td>337</td>
</tr>
<tr>
<td>Letter similar</td>
<td>401</td>
<td>72</td>
<td>337</td>
</tr>
</tbody>
</table>

The current study examined how novice readers coordinate processing across simultaneously presented items during real-time RAN performance. Similar to adults (Jones et al., 2008, 2013, 2016), typically developing 6-year-old children engage in parafoveal preview of adjacent items. Moreover, like adults (Gordon & Hoedemaker, 2016; Hogan-Brown et al., 2014), they are more likely to engage in parallel processing of well-practiced numbers compared with less-practiced letters. Nevertheless, developmental differences were also present. Parafoveal preview of visually similar items exaggerated delays in latency to speak in children compared with adults. Row transitions also pose challenges for maintaining parallel processing in children, decreasing eye–voice overlap for row-initial items relative to row-medial ones. These findings suggest that coordinated processing in novice readers is more susceptible to commonplace hurdles imposed by linguistic (e.g., letter form) and visual (e.g., text location) contexts. Notably, these age-related interactions occur independent of item-level effects on vocal duration, suggesting that developmental changes in RAN performance are likely motivated by improved coordination strategies rather than general increases in processing speed (see also Clarke et al., 2005; Georgiou et al., 2006, 2008; Protopapas et al., 2013b).

Although the current study focused on coordinated processing in typical development, these findings inform our understanding of documented patterns in reading impairments. Previous research demonstrates that dyslexic adults experience more interference from parafoveal preview compared with unimpaired peers (Jones et al., 2008, 2013, 2016; Moll & Jones, 2013). In contrast, dyslexic 10-year-olds exhibit smaller preview effects compared with unimpaired peers (Yan et al., 2013). However, on its own, this difference leaves open whether impaired and unimpaired readers converge on the same coordination strategies with developmental delays or whether impairment status promotes acquisition of different procedures altogether. Importantly, evidence of parafoveal preview in unimpaired 6-year-olds demonstrates that this ability emerges with limited experience. Moreover, although unimpaired novice readers exhibit greater preview interference relative to adults, these delays were short-lived. In contrast, visual similarity generates sustained delays across multiple items.

Fig. 2. In critical sequences: syllable length effects on vocal duration (A), visual similarity effects on latency to speak (B), and row location effects on eye–voice overlap (C).
in dyslexic adults (Jones et al., 2008). The presence of display-type effects provides additional support that deficits across populations are not the same as differences across development. Whereas impaired adults (Hogan-Brown et al., 2014) and 10-year-olds (Pan et al., 2013) show no advantages with well-practiced items compared with less-practiced ones, unimpaired 6-year-olds reveal more parallel processing for numbers compared with letters. Taken together, the precocity of coordinated processing in novice readers suggests that a little experience goes a long way in typical development and that distinctions among impaired readers likely reflect deviations in acquired strategies. Nevertheless, additional research is needed to test this hypothesis by directly assessing real-time RAN performance across ages and populations.

Finally, despite methodological differences, the current findings from real-time RAN performance are consistent with prior research on sentence reading in 7- to 12-year-olds (Blythe & Joseph, 2011; Reichle et al., 2013). This work reveals that children are sensitive to visual properties of text and generate more fixations and longer durations to words of increasing letter length (Hyönä & Olson, 1995; Joseph, Liversedge, Blythe, White, & Rayner, 2009). Comparisons of 8- to 10-year-olds suggest that age-related changes are driven by improvements in linguistic experience rather than oculomotor skill (Huestegge, Radach, Corbic, & Huestegge, 2009). Future work mapping the trajectory of typical development will enhance our understanding of reading impairments. Because disparate outcomes reflect the culmination of past learning experiences, isolating how this input is filtered through the processing strategies of impaired and unimpaired readers will reveal not only how they differ but why distinctions emerge in the first place.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.jecp.2018.02.010.

References


