Central Lancashire Online Knowledge (CLoK)

Title	Effects of aging and text-stimulus quality on the word-frequency effect during Chinese reading.
Type	Article
URL	https://clok.uclan.ac.uk/26464/
DOI	https://doi.org/10.1037/pag0000259
Date	2018
Citation	Wang, Jingxin, Li, Lin, Li, Sha, Xie, Fang, Liversedge, Simon Paul and Paterson, Kevin B. (2018) Effects of aging and text-stimulus quality on the word-frequency effect during Chinese reading. Psychology and Aging, 33 (4). pp. 693-712. ISSN 0882-7974
Creators	Wang, Jingxin, Li, Lin, Li, Sha, Xie, Fang, Liversedge, Simon Paul and Paterson, Kevin B.

Effects of Aging and Text Stimulus Quality on the Word Frequency Effect during Chinese Reading

Jingxin Wang[1]

Lin Li[1]

Sha Li[1]

Fang Xie[1]

Simon P. Liversedge[2]

Kevin B. Paterson[3]

1. Academy of Psychology and Behavior, Tianjin Normal University, Tianjin, China.

2. School of Psychology, University of Southampton, Southampton, UK.

3. School of Psychology, University of Leicester, Leicester, UK.

Running head: Aging Effects in Chinese reading

Abstract

Age-related reading difficulty is well-established for alphabetic languages. Compared to young adults (18-30 years), older adults (65+ years) read more slowly, make more and longer fixations, more regressions, and produce larger word frequency effects. However, whether similar effects are observed for non-alphabetic languages like Chinese remains to be determined. In particular, recent research suggests Chinese readers experience age-related reading difficulty but do not produce age differences in the word frequency effect. This might represent an important qualitative difference in aging effects, so we investigated this further by presenting young and older adult Chinese readers with sentences that included high or low frequency target words. Additionally, to test theories which suggest reductions in text stimulus quality differentially affect lexical processing by adult age groups, either the target words (Experiment 1) or all characters in sentences (Experiment 2) were presented normally or with stimulus quality reduced. Analyses based on mean eye movement parameters and distributional analyses of fixation times for target words showed typical age-related reading difficulty. We also observed age differences in the word frequency effect, predominantly in the tails of fixation time distributions, consistent with an aging effect on the processing of high and low frequency words. Reducing stimulus quality disrupted eye movements more for the older readers, but the influence of stimulus quality on the word frequency effect did not differ across age groups. This suggests Chinese older readers' lexical processing is resilient to reductions in stimulus quality, perhaps due to greater experience recognising words from impoverished visual input.

Key Words: Aging, Chinese, Eye Movements during Reading, Word Frequency, Text Stimulus Quality

Compared to young adults (aged 18-30 years), healthy older adults (aged 65+) experience greater reading difficulty and so often read more slowly despite normal levels of comprehension (e.g., Rayner, Reichle, Stroud, Williams, & Pollatsek, 2006; Stine-Morrow, Loveless, & Soederberg, 1996; for a review, see Gordon, Lowder, & Hoedemaker, 2015). In particular, eye movement studies show that older adults make more and longer fixations, and more regressions (Kliegl, Grabner, Rolfs, & Engbert, 2004; McGowan, White, Jordan, & Paterson, 2014; Paterson, McGowan, & Jordan, 2013a,b,c; Rayner, Castelhano, & Yang, 2009; Rayner et al., 2006; Rayner, Yang, Castelhano, & Liversedge, 2011; Rayner, Yang, Schuett, & Slattery, 2013; Stine-Morrow, Shake, Miles, Lee, Gao, & McConkie, 2010; see also Choi, Lowder, Ferreira, Swaab, & Henderson, 2017; Whitford & Titone, 2016, 2017). Some studies also show that older adults produce larger word frequency effects than young adults, by fixating for longer on words that have a lower frequency of written usage (Kliegl et al., 2004; McGowan et al., 2014; Rayner et al., 2006, 2013; Whitford & Titone, 2017). Such findings are consistent with older adults experiencing greater difficulty identifying words during reading due to slower lexical processing in older adulthood.

Most of these studies have been conducted in alphabetic languages (e.g., English, German). However, several recent studies extended this work by investigating aging effects for Chinese (Wang et al., 2018; Zang, Zhang, Bai, Yan, Paterson, & Liversedge, 2016). Written Chinese has very different visual and linguistic characteristics compared to alphabetic languages, and so such studies are important to establish if aging effects are similar across different writing systems. In particular, Chinese uses a logographic writing system in which text is printed as an unspaced sequence of often visually-complex pictorial characters (Hoosain, 1991, 1992). Moreover, while individual characters sometimes correspond to a word, words generally comprise two or more characters, although boundaries between words are not marked by spaces or other visual cues (Li, Zang, Liversedge & Pollatsek, 2015; Zang, Liversedge, Bai, & Yan, 2011). Accordingly, given the specific visual and linguistic demands of Chinese reading, it seemed possible that older adults might experience

particular difficulty reading this language.

Consistent with previous investigations, the studies showed clear age-related reading difficulty (i.e., slower reading times, more and longer fixations, and more regressions for older adults), although the effects were larger than reported for alphabetic languages. However, an unexpected finding was that the word frequency effect appeared similar across adult age groups (Wang et al., 2018), in contrast to the larger word frequency effects reported for older readers of alphabetic languages. This apparently divergent finding may represent an important qualitative difference in aging effects for Chinese compared to alphabetic languages and so clearly requires further investigation. It may also have implications for current computational models of eye movement control during reading, such as E-Z Reader and SWIFT (Engbert, Nuthmann, Richter, & Kliegl, 2005; Reichle, Rayner, & Pollatsek, 2003) and similar models of Chinese reading (Rayner, Li, & Pollatsek, 2007). Aging effects in alphabetic languages are simulated in these models by adjusting key parameters, including slowing the rate of lexical processing, so that the models predict larger word frequency effects for older readers (Laubrock, Kliegl, & Engbert, 2006; Rayner et al., 2006). Consequently, if such effects are not observed in Chinese, this might indicate an important limitation to these models. We therefore investigated this issue further.

Text stimulus quality

A further concern was to assess age differences in the effects of text stimulus quality. Studies with young adult readers show longer reading times for faint compared to normally presented text, both for entire sentences (Hohenstein & Kliegl, 2014; Liu, Li & Han, 2015; Jainta, Nikolova, & Liversedge, 2017; White & Staub, 2012) and words within sentences (Drieghe, 2008; Glaholt, Rayner, & Reingold, 2014; Reingold & Rayner, 2006; Sheridan & Reingold, 2013; Wang & Inhoff, 2010; White & Staub, 2012). However, few studies have examined effects of stimulus quality for older readers (Mitzner & Rogers, 2006, see also Madden, 1988; Speranza, Daneman, & Schneider, 2000), and even fewer have investigated effects in Chinese (Liu et al., 2015). Importantly, however,

older adults typically experience reductions in visual sensitivity (Elliott, Yang, & Whitaker, 1995; Owsley, Sekuler, Siemsen, 1983; see Owsley, 2011, 2016) and so may require higher stimulus quality to reading efficiently. Consequently, as text encountered in everyday reading can vary substantially in stimulus quality (e.g., due to print or display quality) and reading often occurs in sub-optimal luminance (Charness & Dijkstra, 1999), this may have important practical consequences for older adults' reading performance. Such effects may be especially severe for Chinese reading due to its specific visual and linguistic characteristics. In particular, studies of text legibility show Chinese characters often require high visual sensitivity to be recognized efficiently (Zhang, Zhang, Xue, & Yu, 2007; Zhang, Zhang, Xue, Liu, & Yu, 2009; see also Wang, He, & Legge, 2014), and this may pose particular problems for older readers, especially if stimulus quality is reduced. Accordingly, the present research also assessed the effects of reducing text stimulus quality on the eye movements of young and older Chinese readers.

Lexical processing

Effects of word frequency on eye movements have long been taken to provide a measure of lexical processing difficulty (Inhoff & Rayner, 1986; Rayner, Sereno, & Raney, 1996). Such studies show that words with a lower frequency of written usage have longer fixation times. As noted above, older adults often produce larger word frequency effects, by making disproportionately longer fixation times on lower frequency words compared to young adults. We therefore investigated age differences in the word frequency effect in Chinese reading. However, the design of the study also allowed us to examine if text stimulus quality modulates these effects, and if this influence of stimulus quality on the word frequency effect differs across adult age groups.

These issues relate to a longstanding interest in how manipulations of stimulus quality affect word recognition. A particular concern has been to establish if reductions in stimulus quality affect only the early encoding of visual and orthographic features or also the lexical processing of words (e.g. Becker & Killion, 1977). Additive effects of stimulus quality and word frequency might

indicate these variables influence separate processing stages, while interactive effects might suggest

they influence a common stage (Sternberg, 1969; but see McClelland, 1979). Effects of stimulus

quality on lexical processing by young adults have been shown to depend on specific task

requirements (Yap & Balota, 2007; Yap, Balota, Tse & Besner, 2008). Additive effects are shown in

lexical decision tasks (Balota & Abrams, 1995; Becker & Killion, 1977; Plourde & Besner, 1997;

Stanners, Jastrzembski, & Westbrook, 1975; Yap & Balota, 2007, but see Balota, Aschenbrenner, &

Yap, 2013) while interactive effects are observed in naming and semantic categorisation tasks

(O'Malley, Reynolds, & Besner, 2007; Yap & Balota, 2007). Studies of eye movements are

informative about the influence of such variables during natural reading and have shown interactive

patterns of results when the stimulus quality and lexical frequency of a critical word in a sentence is

manipulated. For instance, Sheridan and Reingold (2013) showed initial fixation durations on critical

words produced an interactive pattern of effects, with larger word frequency effects for faint

compared to normally-presented words. Such findings suggest that stimulus quality can influence

lexical processing as well as the early encoding of visual and orthographic features during natural

reading. However additive effects have been shown in eye movement studies in German when the

stimulus quality of the entire sentence was manipulated (Jainta et al., 2017) and in Chinese when

stimulus quality was manipulated for the entire sentence or a single word (Liu et al., 2015). Effects

of stimulus quality on lexical processing therefore may vary depending on the reading situation and

even differ across languages. However, relatively few studies have been conducted on this issue in

languages other than English. Accordingly, we investigated this issue further by examining the

influence of reduced stimulus quality on the word frequency effect in our experiments.

The present research is also relevant to accounts which argue that age differences in lexical

processing can be affected by text stimulus quality (e.g., Gao, Stine-Morrow, Noh & Eskew Jr.,

2011; Gao, Levinthal, & Stine-Morrow, 2012). Studies show that degrading text by adding visual

noise can disrupt reading more for older adults (see also Speranza, Daneman, & Schneider, 2000).

Text degradation also produces larger word frequency effects in self-paced reading times for words (as well as increased processing costs associated with semantic integration) for older than younger adults (Gao et al., 2011, 2012). Such findings are interpreted in terms of the flexible allocation of processing resources (see Rabbitt, 1968; Wingfield, Tun, & McCoy, 2005), following the logic that age-related sensory declines might lead older adults to allocate more resources to decoding the surface form of text at the expense of higher-level processes, including word identification. Consequently, when the surface form is more difficult to decode (due to degradation), this may impair higher-level processes more for older readers. Accordingly, the present research also investigated if stimulus quality mediates aging effects on the word frequency effect.

We addressed the above issues by presenting high or low frequency target words in each sentence normally or faint and all other text normally (Experiment 1), or the entire sentence normally or faint (Experiment 2). The reasoning behind the second experiment was that manipulating a specific word in the first experiment might cue readers to the nature of the manipulation and induce an abnormal reading strategy in which attention is focused on that word (Glaholt et al., 2014; Liu et al., 2015; White & Staub, 2012). Therefore, to provide a more general indication of stimulus quality effects, the second experiment examined effects when all characters were changed. The experiments allowed us to assess if Chinese readers show standard patterns of age-related reading difficulty and, crucially, if there are age differences in the word frequency effect. The results are also informative about the influence of stimulus quality on lexical processing by revealing if reduced stimulus quality modulates the word frequency effect, and if these effects differ across adult age groups.

General Method

The research was approved by the research ethics committee in the Academy of Psychology and Behavior at Tianjin Normal University and conducted in accordance with the principles of the Declaration of Helsinki.

Participants. Participants were 80 young adults from Tianjin Normal University and 64 older

adults from a residential home in Tianjin. Of these, 40 young and 32 older adults participated in

Experiment 1 and a different 40 young and 32 older adults participated in Experiment 2. All were

native Chinese readers. In each experiment, the age groups were closely matched on years of formal

education and all participants reported reading for at least several hours per week. All participants

were also screened for high-contrast acuity (>20/40 in Snellen values) using a Tumbling E eye-chart

(Taylor, 1978). Acuity was higher for young than older adult groups, as is typical (Elliott et al.,

1995). Table 1 shows mean participant characteristics in each experiment.

--Table 1--

We also assessed vocabulary and working memory capabilities for 48 young and 48 older

adults from the participant sample, using the Vocabulary Knowledge Test from the WAIS-III

Chinese version (Wechsler, Chen, & Chen, 2002) and the WAIS-III Digit Span subtest (Wechsler,

1997). Vocabulary scores were higher for older adults (M=15.7, SD=0.8) than young adults (M=15.0,

SD=1.4; t(94)=2.85, p<.01), consistent with a vocabulary advantage for older adults (Ben-David,

Erel, Goy, & Schneider, 2015; Keuleers, Stevens, Mandera, & Brysbaert, 2015). Scores for the digit

span task were lower for older adults (M=12.0, SD=1.8) than young adults (M=14.0, SD=2.3;

t(94)=4.70, p<.001), also typical for these age groups (Ryan, Sattler, & Lopez, 2000).[1]

Materials and Design. Stimuli consisted of 60 sentence pairs that included a 2-character target

word that had a high or low frequency of written usage (according to the Dictionary of Chinese

Character Information, Li & Liu, 1988; high frequency, M=75 counts/million, low frequency, M=6

counts/million; t=2.03, p<.05). These were from a study by Wang J. et al. (2014) that produced

robust word frequency effects. High and low frequency words were matched for complexity (number

of character strokes) and were equally plausible and predictable in each sentence. Sentences

averaged 19 characters (range=15-23 characters) and target word always appeared towards the

middle of sentences. In Experiment 1, target words were presented normally as black text (RGB 0 0

0) on a white background (RGB 255 255 255) or with contrast reduced to 15% of normal as light

gray text (RGB 243 243 243) on a white background, and all other characters displayed normally. Experiment 2 used the same stimuli, except that all characters in each sentence were displayed normally or faint. The full stimulus set is provided as Supplementary Materials. See Figure 1 for example stimuli.

--Figure 1--

Both experiments used a 2(age group: young or older adults) x 2(word frequency: high or low) x 2(text stimulus quality: normal or faint) mixed design. In each, sentences were pseudo-randomly sampled so each participant saw a sentence containing a high or low frequency target word once and an equal number of high and low frequency target words were shown normally or faint for each age group. Each participant saw 60 experimental and 20 filler sentences, presented in random order, so that sentences containing normal or faint text were inter-mixed, preceded by 8 practice sentences.

Apparatus and Procedure. An SR Research EyeLink 1000 eye-tracker sampled right eye-movements each millisecond during binocular viewing. Sentences were presented on a high-resolution 19-inch monitor in Song font. Each character subtended approximately 1.2°.

Participants were instructed to read normally and for comprehension. At the start of the experiment, participants completed a three-point horizontal calibration procedure (ensuring spatial accuracy of .30° or better), and calibration was checked before each trial and the eye-tracker re-calibrated as necessary. At the start of each trial, a fixation cross was presented on the left of the screen. Once the participant fixated this cross, the sentence was presented across a single line with the first character replacing the cross. The participant pressed a response key to terminate the display once they finished reading. The sentence was then replaced by a comprehension question on 40% of trials. Participants responded by pressing a response key, and their response was recorded.

Comprehension accuracy was higher for the young adults in Experiment 1 (93% vs. 81%, $z=1.33$, $SE=.003$, $t=44.3$) and Experiment 2 (93% vs. 77%, $z=0.17$, $SE=0.003$, $t=60.6$), indicating that older adults comprehended the sentences less well. Mean comprehension accuracy for the older

adults was below the conventional cut-off in many experiments (<80%). This was due, in part at least, to the experimental naivety of many older participants, who frequently could provide accurate verbal responses to comprehension questions but made inaccurate manual responses. To determine if manual response accuracy was related to older adults' reading behavior, we computed correlations between their comprehension scores and eye movement behaviour. For both experiments, we observed no significant correlations for any eye movement measure (all rs=<.4, all ps=>.07), and correlations only approached significant for sentence reading times in Experiment 1 ($r(32)$=-.33, p=.07) and length of forward saccades in Experiment 2 ($r(32)$=-.32, p=.08). Differences in the older adults' response accuracy therefore did not correlate with reading performance. We also examined the pattern of effects in eye movement measures by older adults with response accuracy above and below the conventional 80% cut-off. Both groups differed from the young adults in important respects but produced similar patterns to each other, providing a further indication that variation in older adults' comprehension did not produce qualitative differences in eye movement behavior.[2] The experiments lasted 20-30 minutes for each participant.

Analyses. Two sets of analyses based on mean eye movement parameters were computed for each experiment. Sentence-level analyses examined sentence reading times and eye movements using standard measures: sentence reading time, number of fixations, average fixation duration, number of regressions (backward eye movements), and progressive saccade amplitude (length, in characters, of forward eye movements). Additional measures were computed for the target words. These included measures sensitive to effects during first-pass reading (the initial reading of a word prior to a fixation to its right): word-skipping (probability of not fixating a word), first-fixation duration (FFD; duration of the first fixation on a word), single-fixation duration (SFD; duration of the first fixation on a word receiving only one first-pass fixation), gaze duration (GD; sum of all first-pass fixations), and regressions-out (probability of a backward eye movement from a word). We also examined total reading times (TRT; sum of all fixations on a word) and regressions-in

(probability of a backward eye movement to a word) as measures of later processing.

Data were analyzed by Linear Mixed-Effects Models (LMEMs, Baayen, 2008; Baayen, Davidson, & Bates, 2008) using R (R Development Core Team, 2016) and the lme4 package (Bates, Maechler, & Bolker, 2011). A maximal random effects structure was used for continuous variables, and for generalised linear models conducted for dichotomous variables, with participants and sentence/word stimuli as crossed random effects (Barr, Levy, Scheepers, & Tily, 2013). For sentence-level measures, age group and stimulus quality were fixed factors, and for target word measures, age group, stimulus quality, and word frequency were fixed factors. Pair-wise comparisons of main effects and interactions were defined using the MASS package (Venables & Ripley, 2002). Following convention, t/z>1.96 were considered significant.

In addition to these analyses, we performed distributional analyses that provide insights into aging effects on the processing of words that extend beyond comparisons focused on the mean (Balota, Yap, Cortese, & Watson, 2008; Balota & Yap, 2011; Payne & Stine-Morrow, 2014; Staub, 2011; Staub, White, Drieghe, Hollway, & Rayner, 2010; White & Staub, 2012; see also Luce, 1986; Ratcliff, 1979). In these, an ex-Gaussian distribution is fitted to the data. This is a convolution of the normal distribution and exponential distribution and has been shown to provide a good fit to reaction time data (Heathcote, Brown, & Cousineau, 2004; Heathcote, Brown, & Mewhort, 2002). The resulting distributions can be described by three parameters: the central tendency of the Guassian component (μ), the variance of this component (σ, the standard deviation), and the rate parameter for the exponential component (τ) which describes the degree of slowing in the tail. As an additional nonparametric means of assessing distributional effects, we constructed vincentile plots of the data (Ratcliff, 1979; Vincent, 1912; see Andrews & Heathcote, 2001; Balota et al., 2008; Yap & Balota, 2007). These were created by dividing observations for each participant in each condition into the shortest 10% (vincentile 1), next shortest 10% (vincentile 2), and so on, and then calculating the mean of the observations in each vincentile. The means of the individual participant values in each

vincentile was then calculated and displayed with vincentile number on the x-axis and fixation time on the y-axis. The resulting plots show how fixation time changes across the distribution. The steepness of the curve on the right indicates the degree of right skew. A difference between two conditions due to a shift in the distribution with no change in skew produces parallel curves, with the slowest on top. A difference due to increased skew in one condition compared to another will show little difference in the curves on the left of the graph and a larger difference on the right. Finally, a difference due to a shift in distributions and a change in skew will manifest as a vertical separation in the curves across vincentiles that is larger on the right than the left. We also simulated fixation time distributions by randomly sampling distributions based on the computed ex-Gaussian parameters and using this simulated data to predict vincentile values. We included these predicted vincentile values on the plots as a visual depiction of the goodness-of-fit of the ex-Gaussian parameters with the obtained data (for further discussion, see, e.g., Staub et al., 2010; White & Staub, 2012)

Eye movement studies have used this approach to assess if variables (e.g., age group, word frequency, stimulus quality) produce a shift in the distribution and so affect μ; increase variability, and so affect σ; or influence the skew in the tails of distributions, and so affect τ. Staub et al. (2010) showed word frequency produces a shift and a longer tail in fixation time distributions, while White& Staub (2012) showed that reduced stimulus quality can produce a shift in fixation time distributions for young adult readers. Other studies show adult age differences in all three parameters (Rayner et al., 2013), and that sentence-level factors which produce adult age differences in reading performance also affect all three parameters (Payne & Stine-Morrow, 2014; Rayner et al., 2013). We therefore used this approach to further investigate how adult age group, word frequency, and stimulus quality might influence these components of fixation time distributions for target words in the present experiments.

Experiment 1

Experiment 1 compared the eye movements of young and older adult Chinese readers for

sentences containing a high or low frequency target word that was displayed normally or faint while other characters in sentences were displayed normally.

Results

Following standard procedures, fixations <80ms and >1200ms were excluded. Trials were also excluded due to track loss or error (affecting <2% of trials).

Sentence-Level Analyses. Mean sentence-level eye movement measures are shown in Table 2 and statistical analyses summarized in Table 3.

Tables 2 and 3

Main effects of age group were obtained in all measures. Compared to young adults, the older adults read almost twice as slowly (young adults=4094ms, older adults=8129ms), made more and longer fixations (number of fixations, young adults=14.5, older adults=24.0; average fixation duration, young adults=234ms, older adults=282ms), shorter forward saccades (young adults=3.5 characters, older adults=2.3 characters), and more regressions (young adults=3.8, older adults=6.2). The findings are consistent with previous Chinese findings showing age-related reading difficulty (Wang et al., 2018; Zang et al., 2016). Main effects of stimulus quality for sentence reading time, average fixation duration, and number of fixations and regressions were qualified by interactions with age group. Pairwise comparisons showed older adults had longer reading times, longer fixation durations and made more regressions for sentences that included a faint word than sentences displayed normally, but no effect of stimulus quality for the young adults.

Target Word Analyses. Eye movement measures for the target words are shown in Table 4 and statistical analyses summarized in Table 5.

Tables 4 and 5

Main effects of age group and stimulus quality and interactions between these variables in FFD, SFD, GD, and TRT were due to larger effects of stimulus quality for the older adults (FFD, young adults=25ms effect, older adults=64ms effect; SFD, young adults=28ms effect, older

adults=62ms effect; GD, young adults=22ms effect, older adults=274ms effect; TRT, young adults=47ms effect, older adults=571ms effect). Main effects of word frequency in GD and TRT were qualified by interactions between age group and word frequency, and stimulus quality and word frequency. Interactions between age group and word frequency were due to larger word frequency effects for the older adults (GD, young adults=17ms effect, older adults=71ms effect; older adults=114ms effect; TRT, young adults=42ms effect, older adults=146ms effect). Interactions between stimulus quality and word frequency were due to larger word frequency effects for faint than normal words (GD, normal=16ms effect, faint=72ms effect; TRT, normal=15ms effect, faint=173ms effect). A marginal 3-way interaction in TRT was due to a potentially larger influence of stimulus quality on the word frequency effect for older than younger adults.

A main effect of stimulus quality for regressions-in was due to more regressions to faint than normal words (7% difference), and a main effect of age group on regressions-out due to more regressions from target words by older adults (4% difference). Main effects of age group and stimulus quality in word-skipping were qualified by an interaction. The older adults skipped normal words less than young adults (6% difference), and both groups skipped faint words rarely (1% difference).

Distributional Analyses. Following Staub (2010) we focused on FFD and GD when computing distributional analyses. While for the young adults, most observations in GD will include only one fixation and so are likely to produce effects similar to FFD, older adults are more likely to make multiple first-pass fixations on a word, and so a measure that includes these additional first-pass fixations may be informative with respect to age differences. We conducted two analyses for each measure, one collapsed over word frequency to examine age differences in stimulus quality effects, and another collapsed over stimulus quality to examine age differences in word frequency effects. This approach allowed for a maximum of 30 observations per participant in each analysis. The actual number of observations was fewer, however, due to trials missing because of tracking error or word-

skipping during first-pass reading. We therefore examined the number of usable fixations (which

was the same for FFD and GD). For age group x stimulus quality, the number of trials with usable

fixations averaged 27.4 for normal words and 26.7 for faint words (range=16-30 trials). For age

group x word frequency, this averaged 27.1 for high frequency words and 27.0 for low frequency

words (range=15-30 trials). The number of observations per condition in both analyses is fewer than

regarded as capable of delivering reliable parameter estimates (i.e., ~40; Heathcote et al., 2002,

Speckman & Rouder, 2004). Accordingly, while the models converged successfully for all

participants, the number of observations per participant suggest the findings should be treated with

caution.

<div align="center">Table 6 and Figure 2</div>

The ex-Gaussian distribution was fit to the data from each participant, in each condition, using

the quantile maximum probability estimator (QMPE v2.18) software developed by Heathcote et al.

(2004). This divides the distribution of duration values into quantiles (each containing an equal

number of observations) and uses a maximum likelihood estimation to determine the distributional

parameters that best estimate quantile boundaries. This approach yielded μ, σ, and τ parameters for

each subject in each condition (see Table 6). We conducted separate analyses of these parameters

using LMEMs with participants as a random factor and age group and stimulus quality or age group

and word frequency as fixed factors.

The age group x stimulus quality analysis showed μ and σ were influenced by age group and

stimulus quality in FFD and GD, and an interaction between these factors for μ in GD and σ in FFD

and GD. The τ parameter was influenced by age group, stimulus quality and an interaction between

these factors in FFD and GD. Vincentile plots (Figure 2) show age differences in shift and skew, due

to older adults making generally longer as well as more long fixations times. The plots also show

that stimulus quality caused a shift and skew in FFD and GD for the older but not young adults, due

to older adults making generally longer as well as more long fixation times for faint words. The age

group x word frequency analysis showed age-group effects on μ, σ, and τ for FFD and GD, but only a word frequency effect on μ for GD. Vincentile plots confirmed this limited effect of word frequency, except that these plots suggest word frequency may produce age differences in skew, due to older adults making more long fixation times on lower than higher frequency words.

The vincentile plots include predicted vincentiles based on the best-fitting ex-Gaussian parameters for each condition (diamonds for high frequency / faint words, triangles for low frequency / normally presented words). These were generated by randomly sampling the ex-Gaussian distributions. For each condition, we took 20,000 random samples, each created by summing a sample from a normal distribution created using μ and σ and a sample from an exponential distribution based on the rate parameter $1/\tau$. For each condition, the random sample were used to generate predicted vincentiles. The predicted vincentiles in Figure 2 generally are close the observed values, with the exception that the predicted 10th vincentiles in each condition in each plot tend to undershoot the observed value. This is common for highly skewed distributions and observed in other studies that include less modest numbers of observations per participant (e.g., Staub et al., 2010; White & Staub, 2012). The pattern of the predicted vincentiles therefore suggests that the best-fitting Gaussian parameters successfully capture the shape of the observed distributions despite our reservations about the number of observations per participant.

<div align="center">Discussion</div>

Experiment 1 showed clear age differences in mean eye movement behavior. Compared to young adults, older adults read more slowly, made more and longer fixations and more regressions, and fixated target words for longer. These aging effects were similar to those in previous research (Wang et al., 2018; Zang et al., 2016), and so provide further evidence that age-related reading difficulty is found for both alphabetic languages and non-alphabetic languages like Chinese. Reducing the stimulus quality of a specific target word had little effect on sentence-level eye movement behavior by young adults, but caused substantial disruption for older adults, who read

almost twice as slowly. Word-level effects of stimulus quality were seen for both age groups but were greater for older adults. These emerged early in the eye movement record, indicating that stimulus quality influenced an early stage of the processing of words. The indication, therefore, is that reduced stimulus quality for a single word in a sentence substantially disrupts the normal reading behaviour of older adults but produces only, short-lived and localized disruption for young adults.

An interaction between stimulus quality and word frequency in GD and TRT (larger word frequency effects for faint compared to normal words) showed stimulus quality can affect lexical processing. This was consistent with research showing an interaction between these factors (Sheridan & Reingold, 2013) but contrasts with other research that shows additive effects (Jainta et al., 2017; Liu et al., 2015). Crucially, the indication from these findings is that stimulus quality may have a similar influence on lexical processing in non-alphabetic languages as observed previously for alphabetic languages. We discuss this further in the General Discussion. We also observed larger word frequency effects for older than younger adults in GD and TRT. This contrasts with Chinese findings that reported no age difference (Wang et al., 2018), but is consistent with findings from alphabetic languages (Kliegl et al., 2004; McGowan et al., 2015; Rayner et al., 2006, 2013; Whitford & Titone, 2017). Again, this finding is important as it suggests that aging effects on the word frequency effect are similar across alphabetic and non-alphabetic languages. A marginal 3-way interaction in total reading times showed this influence was observed for both age groups but might be greater for older readers. Such an interaction would be consistent with a flexible resource allocation account of aging effects (Gao et al., 2011, 2012). However, as the effect was marginal, it will be important to establish if this effect can be replicated in Experiment. 2.

Distributional analyses of fixation times on the target words allowed us to examine effects of variables beyond comparisons based on mean behavior. A combination of ex-Gaussian parameters that describe the shape of fixation time distributions, and vincentile plots that illustrate shifts and skews in these distributions, showed clear effects of age group and stimulus quality, and weaker

effects of word frequency. Effects of age group were due to generally longer fixation times and more very long fixation times by older adults, consistent with previous findings (Rayner et al., 2013). Reduced stimulus quality primarily affected the older adults' eye fixation times. This produced a shift and skew in distributions due to older adults producing both generally longer fixations and more very long fixation times for faint words than words presented normally as compared to the young adults. The effects for young adults were weaker than in previous research that has used distributional methods to investigated effects of text stimulus quality (White & Staub, 2012), presumably because the present study used a subtler manipulation. Although effects of word frequency were weak, there was some evidence they produced a shift in the GD distribution, and vincentile plots showed a skew due to larger word frequency effects for the older than younger adults. If the effects of aging on the word frequency effect are principally found in the tails of fixation time distributions, this would suggest a non-linearity in aging effects such that older adults have particular difficulty processing some lower frequency words. This may reflect age differences in lexical processing, perhaps due to older adults accessing lexical representations more slowly for words that are especially different to recognize. Alternatively, it may be a post-lexical effect, due to older adults having particular difficulty integrating the meaning of more unfamiliar words into the sentence context (see, e.g., White, Drieghe, Liversedge, & Staub, 2018).

In Experiment 1, the contrast of a specific target word was manipulated. This isolated change may have been detectable prior to a reader encountering the target word and so cued readers to the nature of the manipulation and induced an abnormal reading strategy in which attention was focused on that word (Glaholt et al., 2014; White & Staub, 2012). Therefore, to explore whether effects arose due to readers focusing on a single target word with unique visual characteristics, and to investigate effects of reduced text stimulus quality more generally, Experiment 2 examined eye movements for sentences where the target word and all other characters were normal or faint.

Experiment 2

Experiment 2 compared the eye movements of young and older adult readers when sentences containing high or low frequency target words were entirely faint or normal.

Results

Following the same procedures as Experiment 1, short and long fixations were deleted as were trials with track loss or error (affecting <2% of trials).

Sentence-Level Analyses. Mean sentence-level eye movement measures are shown in Table 7 and statistical analyses summarized in Table 8.

Tables 7 and 8

Main effects of age group were significant in all measures. Compared to young adults, the older adults read more slowly (young adults=4411ms, older adults=6940ms), made more and longer fixations (number of fixations, young adults=15.4, older adults=21.6; average fixation duration, young adults=230ms, older adults=271ms), more regressions (young adults=4.3, older adults=5.5), and shorter forward saccades (young adults=2.7 characters, older adults=2.1 characters).

Stimulus quality affected sentence reading time and average fixations durations, qualified by interactions with age group, due to larger effects for older adults (sentence reading time, young adults=25ms effect, older adults=862ms effect; average fixation duration, young adults=10ms effect, older adults=40ms effect). Stimulus quality did not affect numbers of fixations or regressions and so disrupted reading primarily by increasing the duration rather than number of fixations.

Target Word Analyses. Eye movement measures for the target words are shown in Table 9 and statistical analyses summarized in Table 10.

Tables 9 and 10

Main effects of age group and stimulus quality, qualified by interactions, were obtained in fixation time measures, due to larger effects for older adults (FFD, young adults=6ms effect, older adults=53ms effect; SFD, young adults=9ms effect, older adults=52ms effect; GD, young

adults=9ms effect, older adults=98ms effect; TRT, young adults=10ms effect, older adults=110ms

effect). We also observed a word frequency effect, marginal in FFD, reliable in GD and qualified by

an interaction with age group in TRT. There was also an interaction between stimulus quality and

word frequency in SFD. This was due to a larger word frequency effect in faint (high

frequency=408ms, low frequency=499ms) than normal (high frequency=258ms, low

frequency=253ms) displays, most likely because longer single-fixations when text was faint

provided greater time for lexical effects to emerge. An interaction between age group and word

frequency in TRT revealed a larger word frequency effect for older (122ms effect) than younger

(45ms effect) adults, although this emerged later in the eye movement record than in Experiment 1.

The effect was also larger for older adults than Experiment 1. This could reflect differences between

the older adults participants in the two experiment. Alternatively, it may be due greater post-lexical

checking of words in Experiment 2 by the older adults, which may reflect the general difficulty they

experience in environments containing a large proportion of sentences in which all characters are

faint. An effect of word frequency for regressions-out was due to more regressions from lower

frequency words (3% effect), and an interaction between stimulus quality and word frequency for

regressions-in was due to more regressions to low than high frequency words when text was faint

(7% effect) than normal (-1% effect). Finally, effects of age group and stimulus quality on word-

skipping were qualified by an interaction. As in Experiment 1, both age groups skipped faint words

rarely (4% effect), and young adults skipped normal words more than older adults (9% effect).

    Distributional Analyses. We fitted ex-Gaussian distribution to each subjects' data using the

same procedures as Experiment 1 to yield μ, σ, and τ parameters for age group x stimulus quality and

age group x word frequency in FFD and GD (see Table 11). We also followed the same procedures

as in Experiment 1 to create vincentile plots (see Figure 3), which included estimates of predicted

vincentiles based on the ex-Gaussian parameters we obtained.

<p align="center">Table 11 and Figure 3</p>

As in Experiment 1, there was a maximum of 30 observations per participant in each analysis, although the number of usable observations was fewer (averaging 26.2, range=12-30 observations, for normal words and 25.6, range=13-30 trials, for faint words per participant in the age group x text stimulus quality analysis, and 25.5, range=10-30 observations, for high frequency words and 26.3, range=15-30 observations, for low frequency words in the age group x word frequency analysis). The relatively low number of observations per participant suggest the findings should be treated with caution (Heathcote et al., 2002, Speckman & Rouder, 2004).

Following the same approach as Experiment 1, we analysed the distribution parameters separately using LMEs with participants as a random factor and age group and stimulus quality or age group and word frequency as fixed factors. For age group x stimulus quality, age group affected μ and τ, stimulus quality affected μ, and age group and stimulus quality interacted to affect τ for FFD; while age group and stimulus quality interacted to affect μ and τ for GD. Vincentile plots show that age group produces a shift and skew in FFD and GD distributions. They also show larger effects for older adults, due to reduced stimulus quality producing a larger shift and greater skew in the fixation time distributions for the older than younger adults. The findings therefore resonate with those from Experiment 1. Age group x word frequency showed effects of age group on μ and τ for FFD, and μ, σ and τ for GD, while word frequency affected only τ for FFD, and interactions between age group and word frequency affected τ for FFD and μ for GD. Vincentile plots showed greater skew in FFD and GD due to larger word frequency effects for older adults. Consequently, as in Experiment, it appears that age differences in word frequency effects are restricted to the tails of fixation time distributions. As in Experiment 1, predicted vincentiles (diamonds for high frequency / faint words, triangles for low frequency / normally presented words) were close to the observed values; although, the predicted 10th vincentiles in each condition in each plot tended to undershoot the observed value (likely due to highly skewed distributions, e.g., Staub et al., 2010; White & Staub, 2012). Therefore, as in Experiment 1, the best-fitting Gaussian parameters appeared to successfully

captured the shape of the observed distributions despite low numbers of observations per participant.

Discussion

Experiment 2 replicated the age-related reading difficulty in Experiment 1. Moreover, as in Experiment 1, reduced stimulus quality disrupted eye movements more for older readers. This effect emerged early in the eye movement record and affected a range of fixation time measures, indicating that text stimulus quality has an early and pervasive influence on the processing of words during reading. Effects of word frequency, by comparison, emerged later, in SFD and GD. Interactions between age group and word frequency also appeared later than in Experiment 1, in TRT rather than GD for target words. This delay in the appearance of the word frequency effect may be due to the different manipulation of stimulus quality across the two experiments, but is also consistent with relatively slow lexical processing during Chinese reading, perhaps due to the need to first identify characters before combining these to form words. The TRT effect was larger than in Experiment 1, perhaps due to this delay in its appearance. The finding is nevertheless consistent with an age difference in lexical processing and so provides a further indication, in addition to the Experiment 1 findings, that aging effects on the word frequency effect are similar across alphabetic and non-alphabetic languages. Interactions between stimulus quality and word frequency also appeared weaker than in Experiment 1. This intersection was observed in SFD, and also in regressions-in, where a larger word frequency effect when stimulus quality was reduced suggested readers were more likely to re-read less familiar words when text was faint, presumably because they initially misidentified these words more often. Consequently, while this finding was consistent with the Experiment 1 finding that stimulus quality has a similar modulating effect on lexical processing in Chinese as reported previously for alphabetic languages, the effect was weaker when the stimulus quality of all characters in sentences was reduced. Contrary to the flexible resource allocation account (Gao et al., 2011, 2012), we did not replicate the marginal 3-way interaction in TRT in Experiment 1. Rather, the findings are compatible with reduced stimulus quality impairing early

visual processing more for older adults and disrupting lexical processing similarly across age groups, although this may be due to the generally weaker effects of word frequency in this experiment.

Distributional analyses of fixation times on the target words complemented these findings. These showed age group differences due to larger shift and skew in FFD and GD distributions for the older than younger adults. Moreover, reduced stimulus quality produced a larger shift and skew in these distributions for the older than younger adults. The indication therefore is that both aging effects and effects of reduced stimulus quality have a pervasive influence on fixation times on words, and that both disrupt the processing of words more for older than younger adult readers. Word frequency, by comparison, showed limited effects, although vincentile plots suggest it produced a greater skew in fixation times for the older adults. As in Experiment 1, this skew may be due either to older readers taking exceptionally long to recognize more difficult words or experiencing post-lexical difficulty integrating these words into the sentence context. Effects of age group, stimulus quality and word frequency were therefore broadly similar to Experiment 1.

General Discussion

The experiments reported here are among the first to investigate adult age differences in eye movements during Chinese reading. Both showed clear effects of age-related reading difficulty, consistent with findings from recent studies (Wang et al., 2018; Zang et al., 2016). Distributional analyses of fixation times (FFD and GD) on target words (e.g., Staub et al., 2010; White & Staub, 2012) additionally revealed that age-group differences produced a shift and rightward skew in these distributions. Vincentile plots suggested these effects are due to older adults making both generally longer and more very long fixation times on words than young adults, consistent with previous research using these methods (Rayner et al., 2013; see also Payne & Stine-Morrow, 2014). The findings therefore show typical patterns of age-related reading difficulty in Chinese reading and a pervasive influence on eye-fixation durations on words similar to that observed in studies with alphabetic languages.

It was of particular concern for the present research to assess age differences in the word frequency effect in Chinese reading. Previous research with alphabetic languages shows larger word frequency effects (i.e., longer fixation times for lower frequency words) for older than younger adults (Kliegl et al., 2004; McGowan et al., 2015; Rayner et al., 2006, 2013; Whitford & Titone, 2017). Moreover, simulations of aging effects in models of eye movement control attribute these larger word frequency effects to slower lexical processing in older age (Laubrock et al., 2006; Rayner et al., 2006). The apparent absence of a corresponding age difference in the word frequency effect in Chinese reading (Wang et al., 2018) might therefore represent an important qualitative difference in aging effects for non-alphabetic compared to alphabetic languages and so this warranted further investigation. Crucially, both present experiments showed age differences in the word frequency effect in mean fixation times on target words. These emerged in GD and TRT in Experiment 1 but only TRT in Experiment 2, most likely due to differences in the manipulation of text stimulus quality across the two experiments. Distribution analyses shed further light on these aging effects. These showed a generally weak influence of word frequency on FFD and GD for target words. But, crucially, vincentile plots of fixation time distributions showed age differences in the size of the word frequency effect in very long fixation times (in the 10th vincentile). This may indicate that age differences in the word frequency effect (certainly during Chinese reading) primarily affect the tail of fixation time distributions, due to older but not younger adults producing very long fixation times for low rather than high frequency words.

The possibility that age differences in the word frequency effect during Chinese reading may be restricted to the tails of fixation time distributions is important as it might explain why such effects were not observed in previous experiments (Wang et al., 2018). In particular, while this previous research used similar manipulations of word frequency as the present study, there were crucial differences in the statistical methods used that might affect the likelihood of detecting an age difference in the word frequency effect that primarily affects long fixation times on words. Both

studies used LMEMs to analyze mean fixation times. But, crucially, and in contrast with the present experiments, Wang et al. first log-transformed the fixation time data before performing these analyses. Log-transformation is widely advocated to ensure data are normally distributed and meet the requirements of LMEM (Baayen, 2008; Baayen et al., 2008). Importantly, this data normalization is also likely to minify or remove the tail of very long fixation times from fixation time distributions, with the consequence that effects that reside in the tail may go unnoticed, even though these may be theoretically meaningful (see Balota et al., 2013). Accordingly, to test if Wang et al. failed to observe age differences in the word frequency effect because they had normalized their data, we reanalyzed these data with and without log-transformation[4]. Log-transformed data produced the same non-significant interaction as Wang et al. reported, confirming these findings. However, untransformed data produced statistically reliable interactions in GD and TRT due to a larger word frequency effect for older adults. The pattern of effects was essentially the same as in Experiment 1 and so it appears the studies show similar age differences in the word frequency effect when data are not normalized.

These findings are important in two respects. First, they suggests that age differences in word frequency effects in Chinese reading may be similar to those observed for alphabetic languages (Kliegl et al., 2004; McGowan et al., 2014; Rayner et al., 2006, 2013; Whitford & Titone, 2017), and in line with the predictions of models of eye movement control (Engbert et al., 2005; Reichle et al., 2003). Second, the findings suggest that the aging effects we observed arise from a small proportion of long fixation times on words. As we noted earlier, this non-linearity in ageing effects might be due to older readers accessing lexical representations more slowly for words that were especially difficult to recognise. Alternatively, it might a post-lexical effect, due to older readers having greater difficulty integrating the meaning of some lower frequency words into the sentence context, perhaps because these words are unfamiliar (e.g., White et al., 2018). The distributional analyses we conducted that support this possibility were based on fewer observations than typically regarded as capable of providing reliable estimates of distributional parameters (Heathcote et al., 2002,

Speckman & Rouder, 2004), although predicted vincentiles based on the ex-Gaussian parameters we obtained suggest these provide a good fit with the data. Accordingly, given the capacity for these methods to advance our understanding of aging effects on eye movements during reading, it will be important for future research to further assess the influence of aging effects on the word frequency effect in fixation times during reading using larger stimulus sets. Moreover, while the present research is potentially informative about effects during Chinese reading, studies to date have also not investigated such effects for alphabetic languages, although this clearly will be required to more fully understand the nature of aging effects across different writing systems. It will also be important to establish if these effects can be explained by models of eye movement control. As noted above, these aging effects currently are accounted for in models of alphabetic reading by slowing the rate of lexical processing (e.g., Laubrock et al., 2006; Rayner et al., 2006). It will be important to establish if aging effects in the present research can be accounted for using similar mechanisms in models of Chinese reading (e.g., Rayner e al., 2007). Alternatively, if aging effects on the word frequency effect are observed only in very long fixations, such effects might be better explained in terms the integration of words into context, and so better accounted for by varying parameters associated with integration difficulty in more recent models (Reichle, Warren, & McConnell, 2009).

Text stimulus quality

It was also of concern for the present research to assess age differences in effects of text stimulus quality. Substantial research shows that reducing text stimulus quality by making words faint can disrupt eye movements (Drieghe, 2008; Glaholt et al., 2014; Hohenstein & Kliegl, 2014; Liu et al., 2015; Reingold & Rayner, 2006; Sheridan & Reingold, 2013; Wang & Inhoff, 2010; White & Staub, 2012). However, studies have not examined age differences in this effect, although these seemed likely given age-related visual declines (Owsley, 2011, 2016). Moreover, most research has used alphabetic languages, and it also seemed likely that such effects might be greater

for Chinese given the specific visual and cognitive demands of this writing system (e.g., Zhang et al., 2007, 2009; Wang, H. et al., 2014). We therefore investigated this issue.

Consistent with previous findings, mean eye movement measures showed that presenting a single target word or an entire sentence as faint disrupted normal reading behavior. Crucially, however, in line with our expectation of an age difference in this effect, the older adults experienced much greater disruption and read sentences in which either a single word or all characters were faint almost twice as slowly compared to the young adults. Distributional analyses resonated with these findings and showed reduced stimulus quality produced a shift and greater skew in FFD and GD distributions for target words for older compared to younger adults. These findings are important as they reveal that older readers are highly vulnerable to changes in text stimulus quality. This is theoretically important as it reveals that visual declines that occur naturally with older age (see Owsley, 2011, 2016) have a major influence on the processing of text. This finding is also of considerable practical importance as text encountered in everyday reading can vary substantially in stimulus quality (Charness & Dijkstra, 1999). Moreover, the difficulties this can cause older readers may impact considerably on their visual quality of life (e.g., Kosnik, Winslow, Kline, Rasinski, & Sekuler, 1988; Rubin, Roche, Prasada-Rao, & Fried, 1994). However, these effects of reduced stimulus quality may be especially severe for non-alphabetic languages like Chinese, due to the visual demands associated with recognizing its visually complex characters (e.g., Wang, H. et al., 2014; Zhang et al., 2007, 2009). It therefore will be important to establish if effects are similar for alphabetic languages like English.

Lexical processing

It was of further concern to establish if stimulus quality affects only the early encoding of visual and orthographic features or also the lexical processing of words. Following previous eye movement studies, this was investigated by assessing effects of stimulus quality on the word

frequency effect (Jainta et al., 2017; Liu et al., 2015; Sheridan & Reingold, 2013). Interactive effects of these variables were observed in GD and TRT for target words in Experiment 1, and regressions back to target words in Experiment 2. The fixation time effects showed readers had more difficulty identifying lower frequency words when faint; while the regression effect suggests readers were more likely to misidentify low frequency words when entire sentences were faint. Such findings are important as they suggest that effects of stimulus quality and word frequency are interactive during Chinese reading and therefore affect a common stage of processing. The fixation time effects we observed, in particular, resonate with the interaction between stimulus quality and word frequency reported by Sheridan and Reingold (2013), although this effect emerged early in the eye movement record in their study, in the first fixation on target words rather than later measures such as GD or TRT. The precise reasons for this difference are yet to be established, although it may reflect the slower time-course of word identification during Chinese reading compared to alphabetic languages. In particular, because word identification in Chinese depends on readers first identifying characters that might combine to form a word (Li et al., 2015; Zang et al., 2011), word frequency effects may emerge more slowly in this language, and this may explain the apparent delay in the effect of stimulus quality on lexical processing. The absence of this effect in fixation times in Experiment 2 may also reflect slower lexical processing in Chinese, which may be compounded when all characters in sentences are faint and therefore potentially more difficult to identify.

These findings contrast with those from studies that have reported additive effects of stimulus quality and word frequency (Jainta et al., 2017; Liu et al., 2015). Jainta et al. employed German text, for which critical words could be capitalised (as is standard for German nouns). Capitalisation has been shown to influence how words are processed, perhaps due to the visual salience of the initial letter (Hohenstein & Kliegl, 2013; Rayner & Schotter, 2014). Accordingly, it could be that visually salient capitalisation cues in Jainta et al.'s study facilitated orthographic processing, mitigating the effects of reduced stimulus quality. Moreover, while Liu et al. (2014) found additive effects on initial

fixation on words during Chinese reading, they observed interactive effects in later measures. This may also be because the interactive effect emerges more slowly in Chinese due to its specific orthography. The findings from these studies therefore may show that stimulus quality can affect both the visual encoding and the lexical processing of words during natural reading but with variation in the timing of effects across languages depending the specific orthographic and lexical characteristics of the writing system. Such findings will be important for models of word recognition in these languages. Crucially, however, we found little evidence that the interaction between stimulus quality and word frequency effect differed across age groups, as the flexible resource allocation account of aging effects suggests (Gao et al., 2011, 2012). In particular, while a marginal interaction consistent with this effect was observed in Experiment 1, it was not corroborated in Experiment 2. It therefore appears that even though older adults have greater difficulty visually processing text when stimulus quality is reduced, this does not substantially impact on the efficiency with which they can recognize words. This resilience in their word recognition capabilities may be a consequence of their greater experience of recognizing words when visual input is impoverished due to the sensory declines that older readers typically experience. However, this lack of support for the flexible resource allocation account may also reflect differences in lexical processing for Chinese compared to non-alphabetic languages, and so clearly requires further investigation.

In sum, the present experiments provide fresh insights into aging effects on eye movements during reading by confirming that Chinese readers experience age-related reading difficulty and age differences in the word frequency effect similar to those reported for alphabetic languages. Consequently, there appear to be important commonalities in aging effects on eye movement control when reading alphabetic and non-alphabetic languages. Our findings also reveal that, although text stimulus quality can influence both the visual encoding and lexical processing of words during Chinese reading, age differences in these effects are observed only during the visual but not lexical processing of words. We attribute this resilience in Chinese older adults' word recognition to their

greater experience of recognizing words under conditions in which lexical input is impoverished. Finally, distributional analyses shed further light on these aging effects by revealing that age differences in effects of text stimulus quality emerge early in the eye movement record, whereas age differences in the word frequency effect emerge late and only in long fixation times. Crucially, this may explain why aging effects on the word frequency effect are only sometimes observed in studies that make comparisons based on the mean. Our findings therefore highlight the value of using these distributional methods to better understand aging effects on eye movements during reading.

Footnotes

1. Note that these are raw scores in the task, not the size of the digit span.

2. LMEMs (and generalized linear models for dichotomous variables) were used to compare the eye movements of young adult readers and low- (<80% comprehension accuracy) and high- (>80% comprehension accuracy) scoring older adult groups. These were computed by comparing the young adults with the high- and low-scoring older adult groups separately, then comparing the two older adult groups. All statistical models converged except word-skipping. Effects were considered significant when t>1.96. Compared to the older adult groups, the young adults had shorter reading times, fewer and shorter fixations, and longer fixation times on words. Comparisons between the two older adult groups revealed few group differences in sentence-level measures, except that lower text stimulus quality slowed reading and increased regression rates more for the low scorers. Target word-level measures showed similar patterns of effects for low and high scoring older adults, except that lower text stimulus quality produced longer GD for high scorers and longer TRT and increased regressions for low scorers. For Experiment 2, sentence-level analyses showed similar effects for high and low scorers, except that reduced contrast slowed reading and increased fixations and regressions more for the low scorers, and low scorers made shorter progressive saccades than high scorers. Target word-level analyses showed that reduced contrast produced larger increases in GD and TRT for low than high scoring older adults but no other group differences. These analyses reveal few qualitative differences between the low and high scorers and show that both groups performed differently compared to the young adults. The findings therefore show no evidence of a differences in the reading strategy of low- and high-scoring older adults.

3. We are grateful to Vicky McGowan for reanalyzing these data. Analyses of untransformed data showed no interaction of age group and word frequency in FFD, $B=3.19$, $SE=7.26$, $t=.44$, but significant interactions in GD, $B=37.03$, $SE=16.39$, $t=2.26$, and TRT, $B=98.37$, $SE=29.74$, $t=3.31$. In both cases, the interaction was due to a larger word frequency effect for older than younger adults.

References

Andrews, S., & Heathcote, A. (2001). Distinguishing common and task specific processes in word identification: A matter of some moment? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*, 514-544.

Baayen, R.H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R.* Cambridge University Press; Cambridge, UK.

Baayen, R.H., Davidson, D.J., & Bates, D.M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*, 390-412.

Balota, D.A., & Abrams, R.A. (1995). Mental chronometry: Beyond onset latencies in the lexical decision task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 1289-1302.

Balota, D.A., Aschenbrenner, A.J., & Yap, M. J. (2013). Additive effects of word frequency and stimulus quality: The influence of trial history and data transformations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*, 1563-1571.

Balota, D.A., & Yap, M.J. (2011). Moving beyond the mean in studies of mental chronometry: The power of response time distributional analyses. *Current Directions in Psychological Science*, *20*, 160-166.

Balota, D.A., Yap, M.J., Cortese, M.J., & Watson, J.M. (2008). Beyond mean response latency: Response time distributional analyses of semantic priming. *Journal of Memory & Language*, *59*, 495–523.

Bates, D., Maechler M., & Bolker, B. (2011). LME4: Linear mixed-effects models using S4 classes. R Package Version 0.999375–39. Retrieved from http://CRAN.R-project.org/package=lme4

Barr, D.J., Levy, R., Scheepers, C., & Tily, H.J. (2013). Random effects structure for confirmatory hypothesis testing: keep it maximal. *Journal of Memory and Language*, *68*, 255-278.

Becker, C.A., & Killion, T.H. (1977). Interaction of visual and cognitive effects in word recognition.

Journal of Experimental Psychology: Human Perception and Performance, 3, 389-401.

Ben-David, B.M., Erel, H., Goy, H., & Schneider, B.A. (2015). "Older is always better": Age-related differences in vocabulary scores across 16 years. *Psychology and Aging*, *30*, 856-862.

Brown, B. (1981). Reading performance in low vision patients: relation to contrast and contrast sensitivity. American Journal of Optometry and Physiological Optics, 58, 218-226.

Charness, N., Dijkstra, K. (1999). Age, luminance, and print legibility in homes, offices, and public places. *Human Factors, 41,* 173-193.

Choi, W., Lowder, M.W., Ferreira, F., Swaab, T.Y., & Henderson, J.M. (2017). Effects of word predictability and preview lexicality on eye movements during reading: A comparison between young and older adults. *Psychology and Aging*, *32*, 232–242.

Drieghe, D. (2008). Foveal processing and word skipping during reading. *Psychonomic Bulletin & Review*, *15*, 856-860.

Elliott, D.B., Yang, K.C., & Whitaker, D. (1995). Visual acuity changes throughout adulthood in normal, healthy eyes: Seeing beyond 6/6. *Optometry and Vision Science*, *72*, 186-191.

Engbert, R., Nuthmann, A., Richter, E.D., & Kliegl, R. (2005). SWIFT: A dynamical model of saccade generation during reading. *Psychological Review*, *112*, 777-813.

Gao, X., Levinthal, B.R., & Stine-Morrow, E.A.L. (2012) The effects of ageing and visual noise on conceptual integration during sentence reading. *The Quarterly Journal of Experimental Psychology*, *65*, 1833-1847.

Gao, X., Stine-Morrow, E.A.L., Noh, S.R., & Eskew Jr., R.T. (2011). Visual noise disrupts conceptual integration in reading. *Psychonomic Bulletin & Review*, *18*, 83-88.

Glaholt, M.G., Rayner, K., & Reingold, E.M. (2014). A rapid effect of visual quality on the durations of individual fixations during reading. *Visual Cognition*, *22*, 377-389.

Gordon, P.C., Lowder, M.W., & Hoedemaker, R.S. (2015). Reading in normally aging adults. In Wright H.H. (Ed.), *Cognitive-linguistic processes and aging* (pp. 165-191). Amsterdam,

Netherlands: John Benjamins Publishing.

Heathcote, A., Brown, S., & Cousineau, D. (2004). QMPE: estimating Lognormal, Wald, and Weibull RT distributions with a parameter-dependent lower bound. *Behavior Research Methods, Instruments and Computers*, *36*, 277-290.

Heathcote, A., Brown, S., & Mewhort, D.J. (2002). Quantile maximum likelihood estimation of response time distributions. *Psychonomic Bulletin & Review*, *9*, 394-401.

Hohenstein S., & Kliegl R. (2014). Semantic preview benefit during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40,* 166-190.

Hoosain, R. (1991). *Psycholinguistic implications for linguistic relativity: A case study of Chinese*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Hoosain, R. (1992). Psychological reality of the word in Chinese. In H.-C. Chen & O.J.L. Tzeng (Eds.), *Language processing in Chinese* (pp. 111-130). Amsterdam, Netherlands: North-Holland.

Jainta, S., Nikolova, M., & Liversedge, S.P. (2017). Does text contrast mediate binocular advantages in reading? *Journal of Experimental Psychology: Human Perception and Performance*, *43*, 55-68.

Keuleers, E., Stevens, M., Mandera, P., & Brysbaert, M. (2015). Word knowledge in the crowd: Measuring vocabulary size and word prevalence in a massive online experiment. *The Quarterly Journal of Experimental Psychology*, *68*, 1665–1692.

Kliegl, R., Grabner, E., Rolfs, M., & Engbert, R. (2004). Length, frequency, and predictability effects of words on eye movements in reading. *European Journal of Cognitive Psychology*, *16*, 262-284.

Kosnik, W., Winslow, L., Kline, D., Rasinski, K., & Sekuler, R. (1988). Visual changes in daily life throughout adulthood. *Journal of Gerontology: Psychological Sciences*, *43*, 63-70.

Laubrock, J., Kliegl, R., & Engbert, R. (2006). SWIFT explorations of age differences in eye

movements during reading. *Neuroscience and Biobehavioral Reviews*, *30*, 872-884.

Li, G.Y., & Liu, R.S. (1988). *A dictionary of Chinese character information*. Beijing: Science Press. (pp. 1228-1662)

Li, X., Zang, C., Liversedge, S.P., & Pollatsek, A. (2015). The role of words in Chinese reading. In A. Pollatsek, & T. Rebecca (Eds.), *The Oxford Handbook of Reading*. New York, NY: Oxford University Press.

Liu, P., Li, X., & Han, B. (2015). Additive effects of stimulus quality and word frequency on eye movements during Chinese reading. *Reading and Writing*, *28*, 199-215.

Lott, L.A., Schneck, M.E., Haegerström-Portnoy, G., Brabyn, J.A., Gildengorin, G.L., & West, C.G. (2001). Reading performance in older adults with good acuity, *Optometry and Vision Science*, *78*, 316-324.

Luce, R.D. (1986). *Response times: Their role in inferring elementary mental organization*. Oxford University Press; Oxford.

Madden, D.J. (1988). Adult age differences in the effects of sentence context and stimulus degradation during visual word recognition. *Psychology and Aging, 3*, 167-172.

McClelland, J.L. (1979). On the time relations of mental processes: An examination of systems of processes in cascade. *Psychological Review, 86*, 287-330.

McGowan, V.A., White, S.J., Jordan, T.R., & Paterson, K.B. (2014). Aging and the use of interword spaces during reading: Evidence from eye movements. *Psychonomic Bulletin & Review*, *21*, 740-747.

Mitzner, T.L., & Rogers, W.A. (2006). Reading in the dark: Effects of age and contrast on reading speed and comprehension. *Human Factors*, *48*, 229-240.

O'Malley, S., Reynolds, M., & Besner, D. (2007). Qualitative differences between the joint effects of visual quality and word frequency in lexical decision and reading aloud: Extensions to Yap and Balota. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33,* 451-458.

Owsley, C. (2011). Aging and vision. *Vision Research*, *51*, 1610-1622.

Owsley, C. (2016). Vision and aging. *Annual Review of Vision Science*, *14*, 255-271.

Owsley, C., Sekuler, R., & Siemsen, D. (1983). Contrast sensitivity throughout adulthood. *Vision Research*, *23*, 689-699.

Payne, B.R., & Stine-Morrow, E.A.L. (2014). Adult age differences in wrap-up during sentence comprehension: Evidence from ex-Gaussian distributional analyses of reading time. *Psychology and Aging*, *29*, 213-228.

Paterson, K.B., McGowan, V.A., & Jordan, T.R. (2013a). Filtered text reveals adult age differences in reading: Evidence from eye movements. *Psychology and Aging*, *28*, 352-364.

Paterson, K.B., McGowan, V.A., & Jordan, T.R. (2013b). Effects of adult aging on reading filtered text: Evidence from eye movements. *PeerJ* 1:e63.

Paterson, K.B., McGowan, V.A., & Jordan, T.R. (2013c). Aging and the control of binocular fixations during reading. *Psychology and Aging*, *28*, 789-795.

Plourde, C.E., & Besner, D. (1997). On the locus of the word frequency effect in visual word recognition. *Canadian Journal of Experimental Psychology, 51,* 181-194.

R Development Core Team (2016). R: A language and environment for statistical computing. R foundation for statistical computing, Vienna, Austria. <http://www.R-project.org>.

Rabbitt, P.M.A. (1968). Channel capacity, intelligibility and immediate memory. *The Quarterly Journal of Experimental Psychology*, *20*, 241-248.

Ratcliff, R. (1979). Group reaction time distributions and an analysis of distribution statistics. *Psychological Bulletin*, *86*, 446-461.

Rayner, K., Castelhano, M. S., & Yang, J. (2009). Eye movements and the perceptual span in older and younger readers. *Psychology and Aging*, *24*, 755-760.

Rayner, K., Li, X, & Pollatsek, A. (2007). Extending the E-Z Reader model of eye movement control to Chinese readers. *Cognitive Science*, *31*, 1021-1033.

Rayner, K., Reichle, E., Stroud, M., Williams, C., & Pollatsek, A. (2006). The effect of word frequency, word predictability, and font difficulty on the eye movements of young and older readers. *Psychology and Aging*, *21*, 448-465.

Rayner, K., Sereno, S.C., & Raney, G.E. (1996). Eye movement control in reading: a comparison of two types of models. *Journal of Experimental Psychology: Human Perception and Performance*, 22, 1188-1200.

Rayner, K., & Schotter, E.R. (2014). Semantic preview benefit in reading English: The effect of initial letter capitalisation. *Journal of Experimental Psychology: Human Perception and Performance, 40,* 1617-1628.

Rayner, K., Yang, J., Castelhano, M.S., & Liversedge, S.P. (2011). Eye movements of older and younger readers when reading disappearing text. *Psychology and Aging*, *26*, 214-223.

Rayner, K., Yang, J., Schuett, S., Slattery, T.J. (2013). Eye movements of older and younger readers when reading unspaced text. *Experimental Psychology*, *60*, 354-361.

Reichle, E.D., Rayner, K., & Pollatsek, A. (2003). The E-Z Reader model of eye movement control in reading: Comparisons to other models. *Behavioral and Brain Sciences*, *26*, 445-476.

Reichle, E.D., Warren, T., & McConnell, K. (2009). Using E-Z reader to model the effects of higher level language processing on eye movements during reading. *Psychonomic Bulletin & Review*, *16*, 1-21.

Reingold, E.M., & Rayner, K. (2006). Examining the word identification stages hypothesized by the E-Z Reader model. *Psychological Science*, *17*, 742-746.

Rubin, G.S., Roche, K.B., Prasada-Rao, P., & Fried, L.P. (1994). Visual impairment and disability in older adults. *Optometry & Vision Science*, *71*, 750-760.

Ryan, J.J., Sattler, J.M., & Lopez S.J. (2000). Age effects on Wechsler Adult Intelligence Scale-III subtests. *Archives of Clinical Neuropsychology*, *15*, 311-317.

Sheridan, H., & Reingold, E.M. (2013). A further examination of the lexical processing stages

hypothesized by the E-Z Reader model. *Attention, Perception, & Psychophysics*, *75*, 407-414.

Speckman, P.L., & Rouder, J.N. (2004). A comment on Heathcote, Brown, and Mewhort's QMLE method for response time distributions. *Psychonomic Bulletin & Review*, *11*, 574–576.

Speranza, F., Daneman, M., & Schneider, B.A. (2000). How aging affects the reading of words in noisy back- grounds. *Psychology and Aging*, *15*, 253-258.

Stanners, R.F., Jastrzembski, J.E., & Westbrook, A. (1975). Frequency and visual quality in a word-nonword classification task. *Journal of Verbal Learning and Verbal Behavior, 14,* 259-264.

Sternberg, S. (1969). The discovery of processing stages: Extensions of Donders' method. In W.G. Koster (Ed.), *Attention and Performance II. Acta Psychologica, 30,* 276-315.

Staub, A. (2011). The effects of lexical predictability on distributions of eye fixation durations. *Psychonomic Bulletin & Review*, *18*, 37-376.

Staub, A., White, S.J., Drieghe, D., Hollway, E.C., & Rayner, K. (2010). Distributional effects of word frequency on eye fixation durations. *Journal of Experimental Psychology: Human Perception and Performance*, *36*, 1280-1293.

Stine-Morrow, E.A.L., Loveless, M.K., & Soederberg, L.M. (1996). Resource allocation in on-line reading by younger and older adults. *Psychology and Aging*, *11*, 475-486.

Stine-Morrow, E.A.L., Shake, M.C., Miles, J.R., Lee, K., Gao, X., & McConkie, G. (2010). Pay now or pay later: Aging and the role of boundary salience in self-regulation of conceptual integration in sentence processing. *Psychology and Aging*, *25*, 168-176.

Vincent, S.B. (1912). The function of the viborissae in the behavior of the white rat. *Behavioral Monographs*, 1, No. 5.

Venables, W.N., & Ripley, B.D. (2002). *Modern Applied Statistics with S* (4th ed.). New York, NY: Springer.

Wang, C.-A., & Inhoff, A. W. (2010). The influence of visual contrast and case changes on parafoveal preview benefits during reading. *Quarterly Journal of Experimental Psychology*,

63, 805-817.

Wang, H., He, X., & Legge, G.E. (2014). Effect of pattern complexity on the visual span for Chinese and alphabet characters. *Journal of Vision*, *14*, 6.

Wang, J., Li, L., Li, S., Xie, F., Chang, M., Paterson, K.B., White, S.J., & McGowan, V.A. (2018). Adult age differences in eye movements during reading: The evidence from Chinese. *Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, in press.

Wang, J., Tian, J., Han, W., Liversedge, S.P., & Paterson, K.B. (2014). Inhibitory stroke neighbour priming in character recognition and reading in Chinese. *Quarterly Journal of Experimental Psychology*, *67*, 2149-2179.

Wechsler, D. (1997). *WAIS-III administration and scoring manual*. San Antonio, TX: The Psychological Corporation.

Wechsler, D., Chen, Y.H., & Chen, X.Y. (2002). *WAIS-III Chinese Version Technical Manual*. San Antonio, TX: The Psychological Corporation.

Whitford, V., & Titone, D. (2016). Eye movements and the perceptual span during first- and second-language sentence reading in bilingual older adults. *Psychology and Aging, 31*, 58-70.

Whitford, V., & Titone, D. (2017). The effects of word frequency and word predictability during first- and second-language paragraph reading in bilingual older and younger adults. *Psychology and Aging*, *32*, 158-177.

White, S.J., & Staub, A. (2012). The distribution of fixation durations during reading: Effects of visual quality. *Journal of Experimental Psychology: Human Perception and Performance*, *38*, 603-617.

White, S.J., Drieghe, D., Liversedge, S.P., & Staub, A. (2018). The word frequency effect during sentence reading: A linear or nonlinear effect of log frequency? *The Quarterly Journal of Experimental Psychology*, *71*, 46-55.

Wingfield, A., Tun, P.A., & McCoy, S.L. (2005). Hearing loss in older adults: What it is and how it

interacts with cognitive performance. *Current Directions in Psychological Science*, *14*, 144-148.

Wotschack, C., & Kliegl, R. (2013). Reading strategy modulates parafoveal-on-foveal effects in sentence reading. *The Quarterly Journal of Experimental Psychology*, *66*, 548-562.

Yap, M.J., & Balota, D.A. (2007). Additive and interactive effects on response time distributions in visual word recognition. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 33,* 274-296.

Yap, M.J., Balota, D.A., Tse, C.S., Besner, D. (2008). On the additive effects of stimulus quality and word frequency in lexical decisions: Evidence for opposing additive influences revealed by RT distributional analyses. *Journal of Experimental Psychology: Learning, Memory & Cognition, 34,* 495-513.

Zang, C., Liversedge, S.P., Bai, X., & Yan, G. (2011). Eye movements during Chinese reading. In, I. Gilchrist & S. Everling (Eds.), *The Oxford Handbook of Eye Movements* (pp. 961-978). Oxford, UK: Oxford University Press.

Zang, C., Zhang, M., Bai, X., Yan, G., Paterson, K.B., & Liversedge, S.P. (2016). Effects of word frequency and visual complexity on eye movements of young and older Chinese readers. *The Quarterly Journal of Experimental Psychology*, *69*, 1409-1425.

Zhang, J-T, Zhang, T., Xue, F., & Yu, C. (2007). Legibility variations of Chinese characters and implications for visual acuity measurement in Chinese reading population. *Investigative and Ophthalmology and Visual Science*, *48*, 2383-2390.

Zhang, J.-Y., Zhang, T., Xue, F., Liu, L., & Yu, C. (2009). Legibility of Chinese characters in peripheral vision and the top-down influences on crowding. *Vision Research*, *49*, 44-53.

Table 1a. Mean Participant Characteristics for Experiments 1 and 2

	Age Group	Age (Years)	Visual Acuity (Log Mar)	Formal Education (Years)
Expt. 1	Young Adult	24 (2.2)	.14 (.07)	16 (1.7)
	Older Adults	78 (6.2)	.18 (.05)	16 (3.1)
Expt. 2	Young Adult	22 (2.2)	.14 (.07)	16 (1.9)
	Older Adult	80 (6.4)	0.18 (.07)	17 (2.8)

The Standard Error of the Mean is shown in parentheses.

Table 2: Mean Sentence-Level Measures for Experiment 1

Measure		Text Stimulus Quality	
		Normal	Faint
Sentence Reading Time (ms)	Young	4109 (54)	4079 (52)
	Older	7598 (106)	8659 (145)
Average Fixation Duration (ms)	Young	233 (.98)	234 (1)
	Older	275 (1.98)	289 (2.21)
Number of Fixations	Young	14.5 (.16)	14.5 (.16)
	Older	23.3 (.29)	24.7 (.34)
Progressive Saccade Amplitude (characters)	Young	2.5 (.02)	2.5 (.02)
	Older	1.9 (.02)	1.9 (.02)
Number of Regressions	Young	3.7 (.07)	3.8 (.08)
	Older	5.8 (.12)	6.6 (.14)

The Standard Error of the Mean is shown in parentheses.

Table 3: Statistical Results for Sentence-Level Measures for Experiment 1

	Source of variance	*Estimate*	*SE*	*t*
Sentence Reading Time	Intercept	6098.9	248.3	24.56
	Age Group	4059.2	459	8.84*
	Text Stimulus Quality	526.8	125.5	4.20*
	Age Group x Text Stimulus Quality	1075.5	249.5	4.31*
Average Fixation Duration	Intercept	257.87	5.17	49.89
	Age Group	48.37	10.21	4.74*
	Text Stimulus Quality	7.86	1.56	5.04*
	Age Group x Text Stimulus Quality	12.93	3.12	4.15*
Number of Fixations	Intercept	19.25	.61	31.59
	Age Group	9.47	1.08	8.70*
	Text Stimulus Quality	.71	.19	3.72*
	Age Group x Text Stimulus Quality	1.38	.38	3.64*
Progressive Saccade Amplitude	Intercept	2.22	.07	34.04
	Age Group	.65	.13	5.04*
	Text Stimulus Quality	.02	.02	.99
	Age Group x Text Stimulus Quality	.03	.03	.86
Number of Regressions	Intercept	4.92	.21	23.19
	Age Group	2.54	.39	6.51*
	Text Stimulus Quality	.48	.11	4.56*
	Age Group x Text Stimulus Quality	.58	.19	3.02*

Note. Asterisks are used to indicate statistically significant fixed-factor effects ($t/z > 1.96$).

Table 4: Mean Target Word Measures for Experiment 1

		Normal		Faint	
		High	**Low**	**High**	**Low**
First fixation duration (ms)	Young	244 (4)	245 (4)	262 (4)	275 (5)
	Older	311 (6)	324 (7)	371 (9)	389 (11)
Single fixation duration (ms)	Young	247 (5)	248 (4)	266 (5)	280 (6)
	Older	313 (9)	321 (11)	376 (14)	390 (17)
Gaze duration (ms)	Young	315 (8)	321 (7)	324 (7)	353 (8)
	Older	496 (14)	523 (17)	725 (26)	835 (32)
Total reading time (ms)	Young	446 (13)	457 (14)	459 (12)	533 (16)
	Older	774 (28)	789 (25)	1220 (56)	1487 (69)
Regressions in (%)	Young	24 (2)	26 (2)	30 (2)	31 (2)
	Older	28 (2)	28 (2)	36 (2)	36 (2)
Regressions out (%)	Young	11 (1)	9 (1)	13 (1)	15 (1)
	Older	14 (2)	17 (2)	18 (2)	20 (2)
Word skipping (%)	Young	10 (1)	12 (1)	11 (1)	11 (1)
	Older	5 (1)	4 (1)	9 (1)	10 (1)

The Standard Error of the Mean is shown in parentheses.

Table 5: Statistical Results for Target Word Measures for Experiment 1

	Source of variance	*Estimate*	*SE*	*t/z*
First Fixation Duration	Intercept	301.84	8.43	35.79
	Age Group	92.06	16.59	5.55*
	Text Stimulus Quality	43.80	7.59	5.77*
	Word Frequency	10.55	5.31	1.99[*]
	Age Group x Text Stimulus Quality	40.02	14.58	2.74*
	Age Group x Word Frequency	7.72	9.28	.83
	Text Stimulus Quality x Word Frequency	8.50	8.29	1.03
	Age Group x Text Stimulus Quality x Word Frequency	6.63	15.61	.42
Single Fixation Duration	Intercept	307.17	8.62	35.62
	Age Group	91.36	16.79	5.44*
	Text Stimulus Quality	46.57	8.01	5.81*
	Word Frequency	9.56	7.28	1.31
	Age Group x Text Stimulus Quality	38.46	15.72	2.45*
	Age Group x Word Frequency	6.23	12.96	.48
	Text Stimulus Quality x Word Frequency	6.56	10.93	.60
	Age Group x Text Stimulus Quality x Word Frequency	5.68	20.33	.28
Gaze Duration	Intercept	482.61	19.55	24.69
	Age Group	318.50	35.85	8.88*
	Text Stimulus Quality	144.77	22.42	6.46*
	Word Frequency	43.35	17.00	2.55*
	Age Group x Text Stimulus Quality	252.41	43.50	5.80*
	Age Group x Word Frequency	52.48	26.11	2.01*
	Text Stimulus Quality x Word Frequency	55.94	24.56	2.28*
	Age Group x Text Stimulus	61.71	42.83	1.44

	Quality x Word Frequency			
Total Reading Time	Intercept	766.19	45.26	16.93
	Age Group	598.27	83.29	7.18*
	Text Stimulus Quality	304.67	58.00	5.25*
	Word Frequency	92.76	31.99	2.90*
	Age Group x Text Stimulus Quality	524.92	113.67	4.62*
	Age Group x Word Frequency	106.47	46.66	2.28*
	Text Stimulus Quality x Word Frequency	156.27	55.63	2.81*
	Age Group x Text Stimulus Quality x Word Frequency	198.21	100.37	1.98*
Regressions-In	Intercept	.95	.09	10.48
	Age Group	.25	.17	1.50
	Text Stimulus Quality	.33	.09	3.83*
	Word Frequency	.05	.07	.74
	Age Group x Text Stimulus Quality	.11	.17	.69
	Age Group x Word Frequency	.09	.14	.60
	Text Stimulus Quality x Word Frequency	.01	.14	.10
	Age Group x Text Stimulus Quality x Word Frequency	.01	.29	.04
Regressions-Out	Intercept	2.02	.11	17.71
	Age Group	.50	.20	2.51*
	Text Stimulus Quality	.35	.09	3.80*
	Word Frequency	.07	.09	.79
	Age Group x Text Stimulus Quality	.19	.18	1.04
	Age Group x Word Frequency	.20	.18	1.07
	Text Stimulus Quality x Word Frequency	.11	.18	.60
	Age Group x Text Stimulus Quality x Word Frequency	.38	.37	1.02

Word skipping	Intercept	2.82	.15	18.40
	Age Group	.77	.30	2.57*
	Text Stimulus Quality	.46	.12	3.89*
	Word Frequency	.02	.12	.20
	Age Group x Text Stimulus Quality	.81	.24	3.45*
	Age Group x Word Frequency	.24	.24	1.00
	Text Stimulus Quality x Word Frequency	.12	.24	.51
	Age Group x Text Stimulus Quality x Word Frequency	.81	.47	1.71

Note. Asterisks are used to indicate statistically significant fixed-factor effects (t/z>1.96).

Table 6. Mean Ex-Guassian Parameters for FFD and GD for (a) Age Group x Text Stimulus Quality and (b) Age Group x Word Frequency Effects, and Summaries of Statistical Analyses for (c) Age Group x Text Stimulus Quality and (d) Age Group x Word Frequency Effects for Experiment 1.

(a)

		Text Stimulus Quality	μ	σ	τ
FFD	Young Adults	Normal	182 (3.82)	38 (3.55)	61 (4.49)
		Faint	207 (5.86)	53 (4.68)	62 (5.56)
		Effect	25	15	1
	Older Adults	Normal	229 (9.28)	44 (5.68)	91 (8.59)
		Faint	254 (16.08)	102 (10.42)	122 (10.51)
		Effect	25	58	31
GD	Young Adults	Normal	195 (6.89)	61 (7.38)	113 (7.56)
		Faint	224 (9.03)	70 (7.28)	105 (9.81)
		Effect	29	9	-8
	Older Adults	Normal	295 (16.97)	125 (14.18)	198 (17.11)
		Faint	421 (30.41)	220 (28.59)	336 (32.94)
		Effect	126	95	138

(b)

		Word Frequency	μ	σ	τ
FFD	Young Adults	High	196 (5.40)	46 (4.10)	55 (4.78)
		Low	193 (4.85)	45 (4.39)	68 (5.21)
		Effect	-3	-1	13
	Older Adults	High	233 (10.81)	66 (7.78)	110 (10.04)
		Low	250 (15.19)	80 (10.26)	103 (9.51)
		Effect	17	14	-7
GD	Young Adults	High	205 (8.42)	60 (7.26)	106 (8.83)
		Low	214 (7.94)	71 (7.38)	112 (8.69)
		Effect	9	11	6
	Older Adults	High	332 (20.93)	165 (22.42)	258 (24.72)
		Low	383 (29.69)	180 (24.17)	276 (30.25)
		Effect	51	15	18

(c)

Fixed effects		FFD			GD		
		μ	σ	τ	μ	σ	τ
（Intercept）	*Estimate*	218.37	59.67	84.47	284.18	119.57	188.47
	SE	6.58	3.89	4.29	11.53	8.91	11.79
	t	33.15	15.33	19.67	24.65	13.40	15.98
Age Group	*Estimate*	46.58	27.45	45.05	148.52	106.81	157.96
	SE	13.17	7.78	8.58	23.05	17.83	23.58
	t	3.54*	3.52*	5.24*	6.44*	5.98*	6.69*
Text Stimulus Quality	*Estimate*	24.80	36.86	16.32	78.18	51.69	64.78
	SE	7.41	5.65	6.78	14.18	14.51	15.72
	t	3.35*	6.52*	2.40*	5.51*	3.56*	4.12*
Age Group x Text Stimulus Quality	*Estimate*	0.68	42.06	28.84	96.91	84.75	146.12
	SE	14.83	11.30	13.56	28.37	29.02	31.44
	t	0.05	3.72*	2.12*	3.41*	2.92*	4.64*

(d)

Fixed effects		FFD			GD		
		μ	δ	τ	μ	δ	τ
（Intercept）	*Estimate*	218.38	59.68	84.48	284.18	119.58	188.47
	SE	6.59	3.89	4.29	11.53	8.92	11.79
	t	33.15	15.33	19.68	24.66	13.41	15.99
Age Group	*Estimate*	46.59	27.45	45.06	148.52	106.81	157.96
	SE	13.18	7.78	8.59	23.05	17.84	23.58
	t	3.54*	3.53*	5.25*	6.44*	5.99*	6.7*
Word Frequency	*Estimate*	6.92	6.34	3.26	30.34	13.2	11.82
	SE	7.57	6.26	6.89	15.22	15.12	16.93
	t	0.91	1.01	0.47	1.99*	0.87	0.7
Age Group x Word Frequency	*Estimate*	19.52	15.48	20.07	41.54	3.49	11.56
	SE	15.13	12.52	13.77	30.43	30.24	33.85
	t	1.29	1.24	1.46	1.37	0.12	0.34

Table 7: Mean Sentence-Level Measures for Experiment 2

Measure		Text Stimulus Quality	
		Normal	Faint
Sentence Reading Time (ms)	Young	4428 (61)	4461 (62)
	Older	6439 (93)	7450 (128)
Average Fixation Duration (ms)	Young	226 (1.01)	236 (1.15)
	Older	252 (1.63)	295 (2.35)
Number of Fixations	Young	15.6 (.19)	15.4 (.19)
	Older	21.2 (.26)	21.3 (.28)
Progressive Saccade Amplitude (degrees)	Young	3.2 (.02)	3.2 (.03)
	Older	2.4 (.02)	2.5 (.03)
Number of Regressions	Young	4.4 (.09)	4.3 (.09)
	Older	5.4 (.10)	5.4 (.11)

The Standard Error of the Mean is shown in parentheses.

Table 8: Statistical Results for Sentence-Level Measures for Experiment 2

	Source of variance	*Estimate*	*SE*	*t*
Sentence Reading Time	Intercept	5651.4	248.7	22.73
	Age Group	2496.1	469.9	5.31*
	Text Stimulus Quality	491.1	112.6	4.36*
	Age Group x Text Stimulus Quality	952.0	224.1	4.24*
Average Fixation Duration	Intercept	252.04	5.03	50.16
	Age Group	43.23	9.97	4.34*
	Text Stimulus Quality	26.50	2.44	10.87*
	Age Group x Text Stimulus Quality	33.23	4.82	6.89*
Number of Fixations	Intercept	18.28	.63	29.11
	Age Group	5.81	1.15	5.06*
	Text Stimulus Quality	.15	.19	.75
	Age Group x Text Stimulus Quality	.26	.40	.66
Progressive Saccade Amplitude	Intercept	2.83	.07	40.31
	Age Group	.73	.13	5.39*
	Text Stimulus Quality	.02	.03	.70
	Age Group x Text Stimulus Quality	.02	.05	.41
Number of Regressions	Intercept	4.79	.23	20.87
	Age Group	1.07	.42	2.52*
	Text Stimulus Quality	.06	.09	.70
	Age Group x Text Stimulus Quality	.09	.17	.50

Note. Asterisks are used to indicate statistically significant fixed-factor effects ($t/z > 2$).

Table 9: Mean Target Word Measures for Experiment 2

		Text Stimulus Quality			
		Normal		Faint	
		Word Frequency			
		High	Low	High	Low
First fixation duration (ms)	Young	242 (4)	241 (4)	243 (4)	252 (4)
	Older	267 (5)	278 (5)	322 (7)	330 (8)
Single fixation duration (ms)	Young	240 (4)	236 (4)	244 (4)	254 (5)
	Older	276 (7)	270 (8)	318 (10)	336 (11)
Gaze duration (ms)	Young	290 (7)	307 (8)	304 (8)	317 (7)
	Older	440 (13)	469 (14)	531 (17)	591 (19)
Total reading time (ms)	Young	439 (13)	465 (13)	436 (15)	494 (15)
	Older	615 (21)	704 (23)	706 (25)	853 (35)
Regressions in (%)	Young	29 (2)	27 (2)	24 (2)	32 (2)
	Older	19 (2)	21 (2)	18 (2)	24 (2)
Regressions out (%)	Young	10 (1)	11 (1)	8 (1)	10 (1)
	Older	12 (2)	16 (2)	11 (2)	16 (2)
Word skipping (%)	Young	17 (1)	15 (1)	17 (1)	13 (1)
	Older	6 (1)	8 (1)	12 (1)	10 (1)

The Standard Error of the Mean is shown in parentheses.

Table 10: Statistical Results for Target Word Measures for Experiment 2

	Source of variance	*Estimate*	*SE*	*t/z*
First Fixation Duration	Intercept	270.56	6.46	41.91
	Age Group	55.65	12.63	4.41*
	Text Stimulus Quality	29.03	4.35	6.67*
	Word Frequency	7.72	3.40	1.93[+]
	Age Group x Text Stimulus Quality	47.45	8.49	5.59*
	Age Group x Word Frequency	6.16	6.37	.97
	Text Stimulus Quality x Word Frequency	5.24	6.75	.78
	Age Group x Text Stimulus Quality x Word Frequency	15.86	12.73	1.25
Single Fixation Duration	Intercept	272.77	6.88	39.67
	Age Group	59.52	13.40	4.44*
	Text Stimulus Quality	31.58	5.03	6.28*
	Word Frequency	6.05	4.71	1.29
	Age Group x Text Stimulus Quality	42.38	9.73	4.35*
	Age Group x Word Frequency	4.25	7.51	.57
	Text Stimulus Quality x Word Frequency	23.15	8.91	2.60*
	Age Group x Text Stimulus Quality x Word Frequency	3.52	15.97	.22
Gaze Duration	Intercept	398.13	15.00	26.55
	Age Group	198.98	27.33	7.28*
	Text Stimulus Quality	53.08	11.13	4.77*
	Word Frequency	32.60	11.92	2.74*
	Age Group x Text Stimulus Quality	90.07	20.56	4.38*
	Age Group x Word Frequency	31.96	16.35	1.95[+]
	Text Stimulus Quality x Word Frequency	19.10	17.54	1.09
	Age Group x Text Stimulus	37.19	29.68	1.25

	Quality x Word Frequency			
Total Reading Time	Intercept	577.86	26.81	21.55
	Age Group	257.77	47.24	5.46*
	Text Stimulus Quality	59.96	17.01	3.52*
	Word Frequency	83.90	23.31	3.60*
	Age Group x Text Stimulus Quality	101.00	32.54	3.10*
	Age Group x Word Frequency	76.25	30.39	2.51*
	Text Stimulus Quality x Word Frequency	50.33	34.83	1.45
	Age Group x Text Stimulus Quality x Word Frequency	28.43	55.54	.51
Regressions-In	Intercept	1.28	.09	13.55
	Age Group	.43	.16	2.73*
	Text Stimulus Quality	.04	.08	.49
	Word Frequency	.21	.08	2.67*
	Age Group x Text Stimulus Quality	.04	.16	.26
	Age Group x Word Frequency	.13	.16	.84
	Text Stimulus Quality x Word Frequency	.38	.16	2.43*
	Age Group x Text Stimulus Quality x Word Frequency	.38	.31	1.22
Regressions-Out	Intercept	2.22	.10	21.40
	Age Group	.34	.19	1.82^{+}
	Text Stimulus Quality	.14	.10	1.38
	Word Frequency	.29	.10	2.89*
	Age Group x Text Stimulus Quality	.13	.20	.65
	Age Group x Word Frequency	.17	.20	.85
	Text Stimulus Quality x Word Frequency	.07	.20	.34
	Age Group x Text Stimulus Quality x Word Frequency	.08	.40	.21

Word skipping	Intercept	2.53	.16	15.78
	Age Group	1.08	.31	3.49*
	Text Stimulus Quality	.30	.11	2.78*
	Word Frequency	.11	.11	1.03
	Age Group x Text Stimulus Quality	.71	.22	3.30*
	Age Group x Word Frequency	.31	.22	1.45
	Text Stimulus Quality x Word Frequency	.42	.22	1.99+
	Age Group x Text Stimulus Quality x Word Frequency	.58	.43	1.35

Note. Asterisks are used to indicate statistically significant fixed-factor effects ($t/z > 2$).

Table 11. Mean Ex-Guassian Parameters for FFD and GD for (a) Age Group x Text Stimulus Quality and (b) Age Group x Word Frequency Effects, and Summaries of Statistical Analyses for (c) Age Group x Text Stimulus Quality and (d) Age Group x Word Frequency Effects for Experiment 1.

(a)

		Text Stimulus Quality	μ	σ	τ
FFD	Young Adults	Normal	186 (4.7)	41 (3.7)	54 (4.15)
		Faint	199 (4.8)	40 (4.14)	48 (4.48)
		Effect	13	-1	-6
	Older Adults	Normal	210 (6.1)	47 (5.20)	64 (7.12)
		Faint	233 (7.6)	47 (6.24)	98 (10.90)
		Effect	23	0	34
GD	Young Adults	Normal	180 (5.9)	46 (6.05)	113 (7.44)
		Faint	195 (5.8)	50 (6.46)	106 (7.95)
		Effect	15	4	-7
	Older Adults	Normal	263 (13.8)	97 (12.38)	176 (14.01)
		Faint	320 (19.7)	127 (17.82)	226 (20.67)
		Effect	57	30	50

(b)

		Word Frequency	μ	σ	τ
FFD	Young Adults	High	191 (4.2)	42 (4.05)	51 (4.44)
		Low	194 (5.4)	39 (3.75)	50 (4.21)
		Effect	3	-3	-1
	Older Adults	High	227 (6.7)	47 (5.34)	69 (8.40)
		Low	216 (7.2)	46 (6.06)	92 (10.05)
		Effect	-11	-1	23
GD	Young Adults	High	190 (5.8)	49 (6.29)	99 (6.30)
		Low	185 (6.1)	47 (6.23)	120 (8.71)
		Effect	-5	-2	21
	Older Adults	High	275 (15.4)	97 (13.43)	199 (18.34)
		Low	308 (18.7)	127 (16.79)	203 (17.33)
		Effect	33	30	4

(c)

Fixed effects		FFD			GD		
		μ	σ	τ	μ	σ	τ
(Intercept)	*Estimate*	207.37	44.09	66.46	240.04	80.51	155.53
	SE	4.22	2.87	4.77	8.93	6.87	9.06
	t	49.2	15.34	13.93	26.89	11.72	17.17
Age Group	*Estimate*	29.25	6.19	50.59	104.14	64.27	90.55
	SE	8.43	5.75	11.02	17.85	13.74	18.12
	t	3.47*	1.08	4.59*	5.83	4.68	5
Text Stimulus Quality	*Estimate*	17.87	0.52	9.78	36.22	17.3	20.71
	SE	4.5	4.38	10.95	8.58	9.65	10.19
	t	3.97*	0.12	0.89	4.22	1.79	2.03
Age Group x Text Stimulus Quality	*Estimate*	10.51	1.32	14.33	41.06	25.24	55.14
	SE	9	8.76	5.45	17.17	19.31	20.37
	t	1.17	0.15	2.63*	2.39*	1.31	2.71*

(d)

Fixed effects		FFD			GD		
		μ	σ	τ	μ	σ	τ
(Intercept）	*Estimate*	207.28	44.1	66.2	239.58	80.23	155.26
	SE	4.22	2.87	4.77	8.94	6.88	9.08
	t	49.14	15.35	13.89	26.8	11.66	17.09
Age Group	*Estimate*	29.19	6.18	29.61	103.31	63.74	90.11
	SE	8.44	5.75	9.53	17.88	13.76	18.17
	t	3.46*	1.08	3.11*	5.78*	4.63*	4.96*
Word Frequency	*Estimate*	3.46	1.28	11.01	13.87	14.43	13.25
	SE	4.64	4.38	5.59	8.87	9.65	10.36
	t	0.74	0.29	1.97*	1.56	1.5	1.28
Age Group x Word Frequency	*Estimate*	13.64	1.75	24.12	39.3	31.9	15.37
	SE	9.28	8.76	11.17	17.73	19.3	20.73
	t	1.47	0.2	2.16*	2.22*	1.65	0.74

Figure Legend

Figure 1. Example Stimuli for the Experiments

Figure 2. Vincentile Plots of Age Group x Text Stimulus Quality for (a) FFD and (b) GD, and for Age Group x Word Frequency for (a) FFD and (b) GD in Experiment 1. Vincentiles based on the observed data are represented by unfilled squares and circles (for high versus low frequency words and faint versus normal words) linked by solid lines for the older adults, and unfilled triangles and diamonds (for high versus low frequency words and faint versus normal words) linked by solid lines for the young adults. For each age group, the predicted vincentiles based on ex-Gaussian parameters are represented by filled shapes. Error bars represent standard error of the mean.
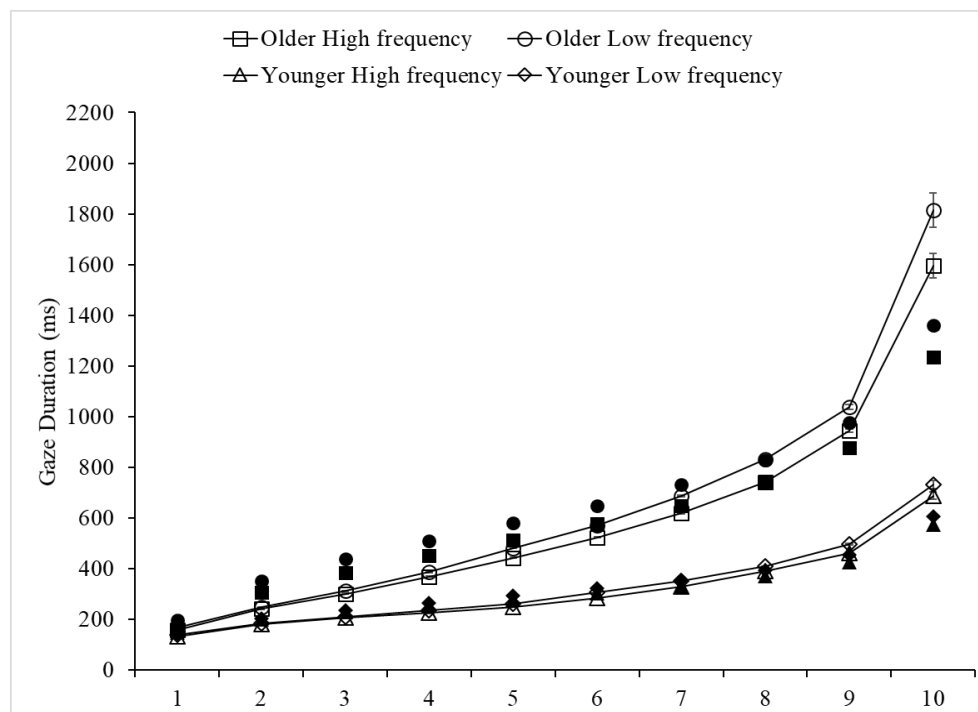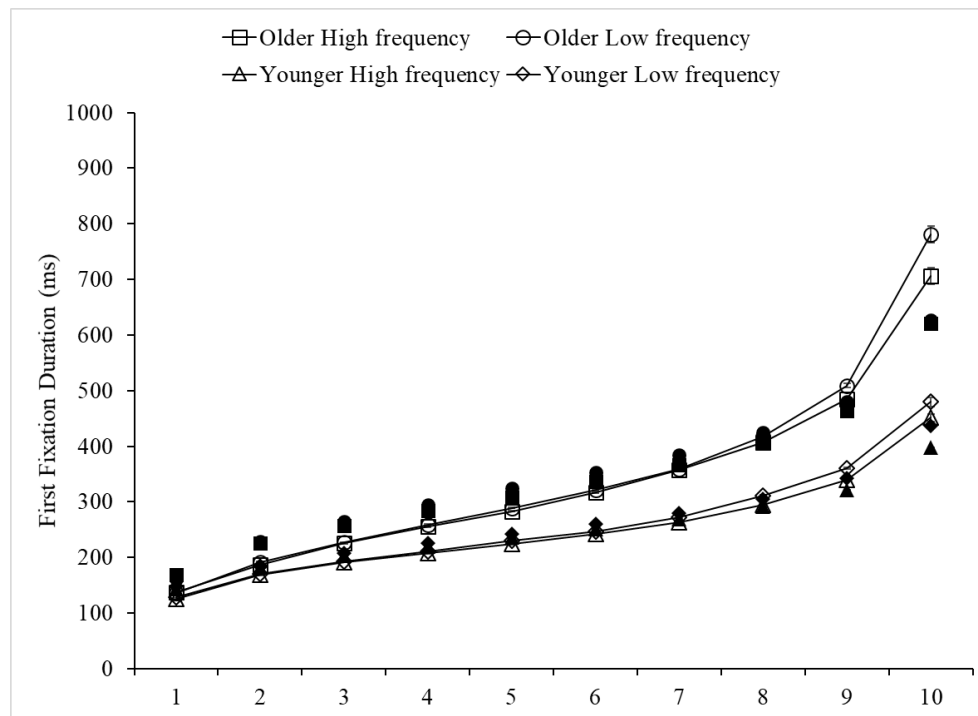
Figure 3. Vincentile Plots of Age Group x Text Stimulus Quality for (a) FFD and (b) GD, and for Age Group x Word Frequency for (a) FFD and (b) GD in Experiment 2. Vincentiles based on the observed data are represented by unfilled squares and circles (for high versus low frequency words and faint versus normal words) linked by solid lines for the older adults, and unfilled triangles and diamonds (for high versus low frequency words and faint versus normal words) linked by solid lines for the young adults. For each age group, the predicted vincentiles based on ex-Gaussian parameters are represented by filled shapes. Error bars represent standard error of the mean.

Figure 1

Condition	Sentence
Normal_HF	张大爷看着小王手提猪肉匆匆走回了村子。
	Uncle Zhang was looking at Xiao Wang hurrying back to their village with pork in his hand.
Normal_LF	张大爷看着小王手提猎物匆匆走回了村子。
	Uncle Zhang was looking at Xiao Wang hurrying back to their village with prey in his hand.
Experiment 1	
Faint-target_HF	张大爷看着小王手提猪肉匆匆走回了村子。
	Uncle Zhang was looking at Xiao Wang hurrying back to their village with pork in his hand.
Faint-target_LF	张大爷看着小王手提猎物匆匆走回了村子。
	Uncle Zhang was looking at Xiao Wang hurrying back to their village with prey in his hand.
Experiment 2	
Faint_HF	张大爷看着小王手提猪肉匆匆走回了村子。
	Uncle Zhang was looking at Xiao Wang hurrying back to their village with pork in his hand.
Faint_LF	张大爷看着小王手提猎物匆匆走回了村子。
	Uncle Zhang was looking at Xiao Wang hurrying back to their village with prey in his hand.
Question	小王是否走回了村子？
	Did Xiao Wang go back to their village?

Note. Target words appear underlined (but not during the experiment).

HF=high-frequency targets; LF = low-frequency targets.
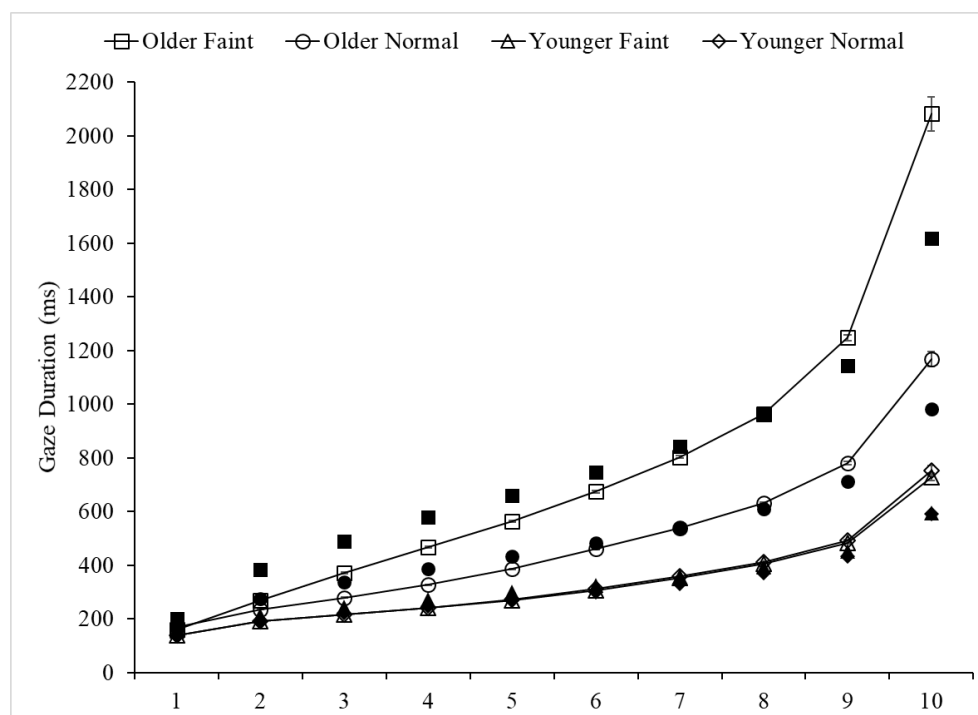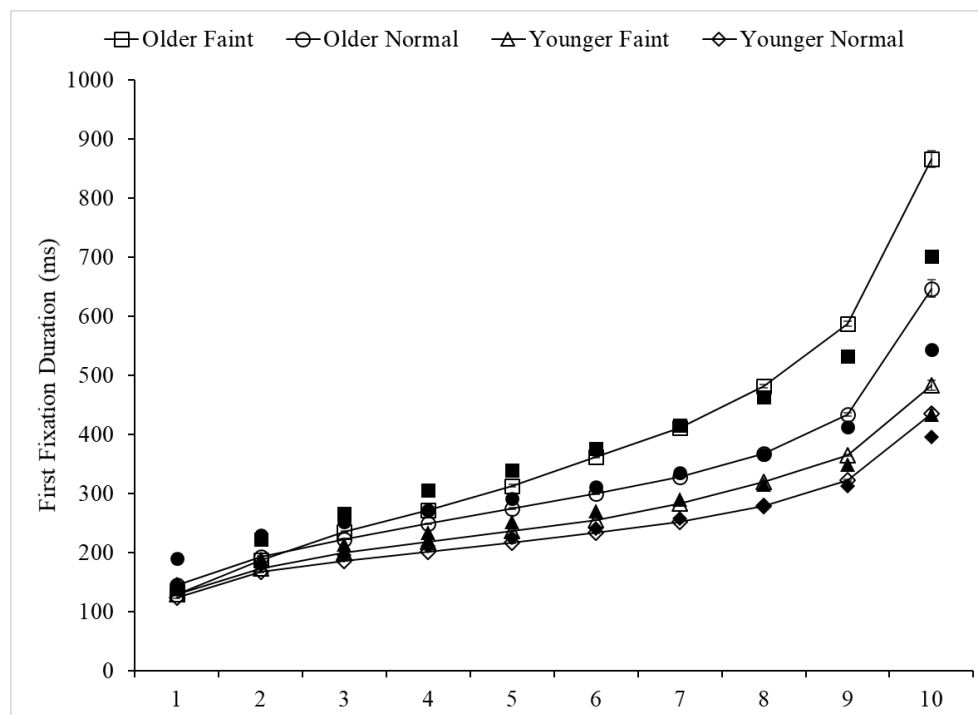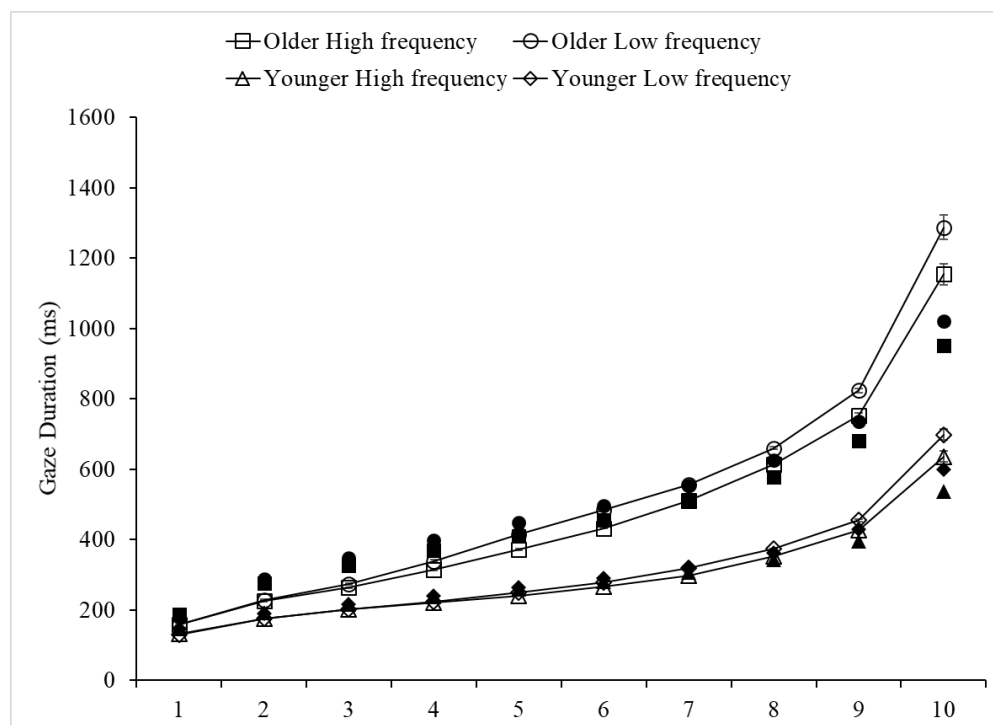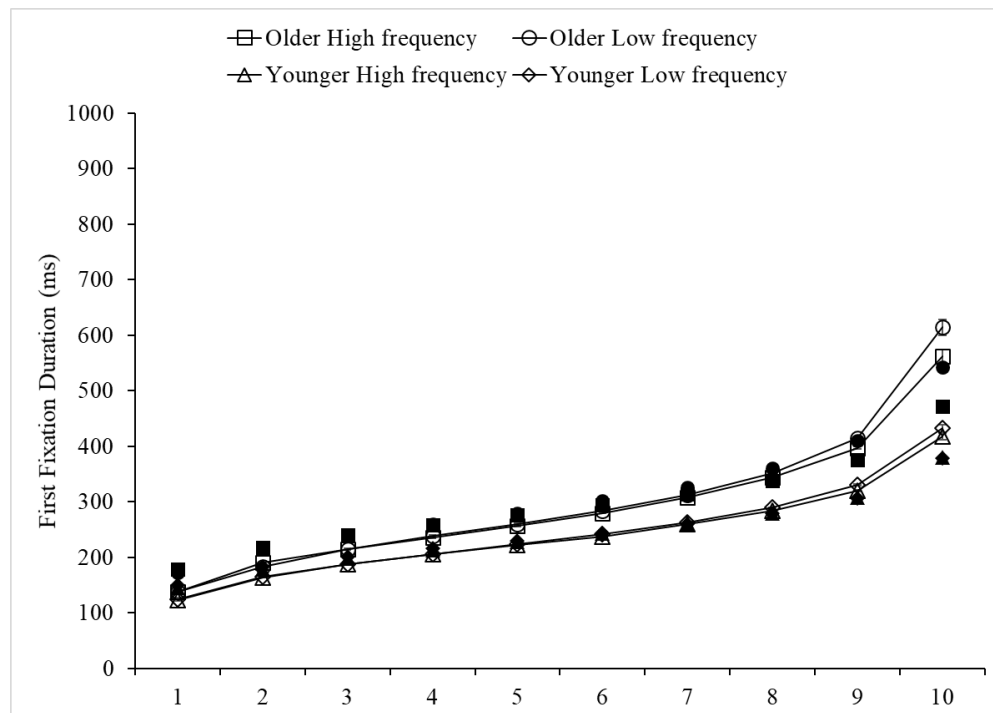
Figure 2

(a)

(b)

Figure 3

(a)

(b)