Task Demands Modulate the Effects of Speech on Text Processing

Zhu Meng[1, 2]

Zebo Lan[1]

Guoli Yan[1]

John E. Marsh[2,3]

Simon P. Liversedge[2]

[1]Academy of Psychology and Behavior, Tianjin Normal University, China

[2]School of Psychology, University of Central Lancashire, United Kingdom

[3]Department of Building Engineering, Energy Systems, and Sustainability Science, University of Gävle, Gävle, Sweden

Corresponding author: SPLiversedge@uclan.ac.uk (ORCID iD: 0000-0002-8579-8546)

Word Count: 9052

Abstract

Task-irrelevant background sound can disrupt performance of visually-based cognitive tasks. The cross-modal breakdown of attentional selectivity in the context of reading was addressed using analyses of eye-movements. Moreover, the study addressed whether task-sensitivity to distraction via background speech on reading was modulated by the cognitive demands of the focal task. Two randomly-assigned groups of native-Chinese participants read the same set of Chinese experimental sentences while being exposed to meaningful speech, meaningless (foreign) speech, or silence. For one group, participants were instructed to judge whether the sentences made sense (i.e., semantic acceptability task); for another, participants were instructed to detect whether the sentences contained a non-character (i.e., non-character detection task). Results showed no significant effect across sound conditions for the non-character detection task. For the semantic acceptability task, however, there was a substantial disruptive effect of the meaningfulness of the speech. Compared with reading with meaningless speech or reading in silence, the meaningful speech increased numbers of fixations, regressions, regression path and total reading times. These results suggest that the disruption of reading by background speech is jointly dependent on the nature of the speech and the task-process deployed, thereby favouring an Interference-by-Process account over Interference-by-Content and Attentional Diversion accounts of distraction to reading by background sound.

Keywords: auditory distraction, reading, eye-movements, process-oriented

Word count: 206

Other than in specific circumstances, the living or working environment is replete with background sounds (e.g., traffic noise, instrumental music, speech). Typically, empirical findings demonstrate that speech in particular is a significant distracter, resulting in impaired performance of tasks requiring cognitive processing (e.g., Cauchard, Cane, & Weger, 2012; Martin, Wogalter, & Forlano, 1988; Sörqvist, 2010). The interference with visual information processing by background speech is a phenomenon that has been termed the irrelevant speech effect[1]. Although the influence of speech on cognitive processing has been long investigated (e.g., Colle & Welsh, 1976), the precise nature of such effects, especially for complex tasks such as reading, remains controversial (see Jones, 1995; Vasilev, Kirkby, & Angele, 2018). Thus, a goal of the present study was to clarify understanding of the irrelevant speech effect in reading through the use of eye-tracking techniques.

Phonological-Interference-by-Content

One assumption is that Working Memory (Baddeley, 1979; Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001) is involved in natural reading (Gathercole & Baddeley, 1993), particularly for the identification and representation of read words (Rayner, 1998; 2009). Within the Working Memory model, the phonological loop comprises a phonological store and a subvocal rehearsal device (the articulatory loop). Entry to the phonological store is either direct for auditory input or indirect for visual input whereby written words are converted into phonological form via a grapheme-phoneme process undertaken by the articulatory loop (Baddeley, 1986). A second function of the articulatory loop is to revivify decaying traces of stimuli within the phonological store (Baddeley, 1986). One hypothesis stemming from the purported role of Working Memory in reading is that disruption due to

[1] The ISE was originally termed the “irrelevant speech effect”, but it has also been referred to as the “unattended speech effect” and the “irrelevant sound effect”. The latter label (Beaman & Jones, 1997) recognizes that irrelevant sound need not necessarily be speech to produce marked disruption of cognitive processing.

auditory distraction occurs because representations in the phonological store become corrupted by automatically encoded auditory speech input (Salamé & Baddeley, 1982; but see Page & Norris, 2003). It is assumed that susceptibility to disruption arises to the extent that there is an overlap in phonemes between the written text and background speech for cases in which a grapheme-phoneme conversion process is used to identify words. In other words, according to the Phonological-Interference-by-Content account, all speech-like sounds and only speech-like sounds could cause disruption.

Despite this claim, a number of studies concerning the irrelevant sound effect in the context of focal tasks requiring semantic processing report findings that run counter to the Phonological-Interference-by-Content account. Specifically, non-speech sounds, without phonemes, produce similar disruptive effects to those of speech (Banbury & Berry, 1998, Experiment 2; Jones & Macken, 1993; Marsh, Hughes, & Jones, 2009, Experiment 1; Murphy, Bailey, Pearson, & Albert, 2018) and speech comprising phonological content (e.g., foreign speech) often fails to impair performance in complex tasks that require semantic processing (Boyle & Coltheart, 1996; Jones, Miles, & Page, 1990; Martin et al., 1988).

Thus, the Phonological-Interference-by-Content account does not appear to offer sufficient flexibility to comprehensively explain results obtained with complex tasks (e.g., reading). It predicts that the phonological properties of background speech drive the disruption of performance but this is not observed (e.g., Boyle & Coltheart, 1996; Martin et al. 1988). Specifically, then the phonological properties of background speech are not determining factors for the occurrence of auditory distraction in the context of reading.

Semantic-Interference-by-Content

While Phonological-Interference-by-Content appears to be an improbable account of auditory distraction in reading, it is possible that disruption to reading from meaningful background speech could occur at the level of lexical or semantic processing of a word. For

example, if meaningful background sound activates lexical entries of words then conceivably semantic interference by content may arise between semantic representations derived from auditory and visual origins (Anderson, 1983; Oberauer & Lange, 2008). This theoretical perspective, the Semantic-Interference-by-Content account, holds that the disruptive effect of irrelevant speech is due to the semantic properties of the speech (Hyönä & Ekholm, 2016; Vasilev et al., 2018).

In the case of reading, the activation of words within the speech stream can generate interference with the semantic meanings of the words in the text being read when represented in a common memory system (e.g., Working Memory). The importance to reading of a capacity within working memory for "semantic storage" (compared to "phonological storage") has previously been outlined (e.g., Haarmann, Davelaar, & Usher, 2003; see also Just & Carpenter, 1992). Thus, the Semantic-Interference-by-Content account proposes that task-irrelevant speech conveying words in a language a participant understands (meaningful speech) should produce disruption compared to background speech comprising words in a language the participant does not understand, which in turn should produce no disruption relative to quiet. Furthermore, the degree to which irrelevant speech interferes with reading should depend on the degree of similarity in semantic content between the background speech and the visually-attended text, that is, the "between-sequence semantic similarity effect" (Marsh, Hughes, & Jones, 2008).

At odds with this account is that between-sequence semantic similarity does not, in fact, determine distraction in reading (Hyönä & Ekholm, 2016; see also Martin et al. 1988; Marsh, Perham, Sörqvist, & Jones, 2014, for related findings). Furthermore, several studies have shown that the meaningfulness of background speech fails to disrupt reading (Boyle & Coltheart, 1996; Hyönä & Ekholm, 2016, Experiment 1; Johansson, Holmqvist, Mossberg, & Lindgren, 2012). In addition, Martin et al. (1988) showed that a background of random

spoken words produced as much disruption to written comprehension as continuous meaningful irrelevant speech. These findings suggest that background speech produces disruption due to the processing of individual word meanings within the irrelevant speech, but there is no evidence that background speech impairs identification of individual words within visually-attended text (Vasilev, Liversedge, Rowan, Kirkby, & Angele, 2019).

Semantic-Interference-by-Content approaches assume that background speech should interfere with reading when the speech signal conveys semantic content and that such disruption should be exacerbated the greater the similarity in semantic content between the background speech and visually-attended text. This pattern of findings is not consistently observed. Thus, the findings presented above undermine the Semantic-Interference-by-Content account in its most simplistic form and suggest a more dynamic process-oriented account of the permeability of selective attention to auditory distracters may be a necessity.

Feature-Based Interference-by-Content

Thus far, we have considered Phonological-Interference-by-Content and Semantic-Interference-by-Content independently, yet it is clear from Feature-Based Models (Neath, 2000; Oberauer & Lange, 2008) that Interference-by-Content can occur on multiple levels. On this approach all representations are built from features, regardless of whether those representations are meaningful or meaningless and/or occur in the visual or auditory modality. Representations of to-be-attended stimuli are disrupted by background speech due to a feature-adoption processes. Here, modality-independent features of to-be-attended stimuli can be overwritten by corresponding, mismatching features of background speech stimuli. This has the result of impairing memory traces of to-be-attended items, thereafter impairing focal task performance. As a consequence, Feature-Based Interference-by-Content accounts predict a graded disruption from background speech whereby some interference—via feature adoption—will occur from meaningless background speech to a reading task,

regardless of the nature of that task. Additional disruption will arise from meaningful background speech whereby the quantity of non-shared modality independent features—and therefore opportunity for feature adoption—is greater (i.e., the order of disruption should be: meaningful background speech > meaningless background speech > quiet).

This prediction has not been the focus of previous work but extant findings fail to provide evidence for graded disruption. For example, existing studies show that impairment due to meaningful background speech is significantly greater than that from meaningless background speech which is not significantly more disruptive than a quiet baseline (Jones, Marsh, & Hughes, 2012, Experiments 1 and 2; Marsh et al., 2008, Experiment 1; Martin et al., 1988). This pattern of non-graded disruption (meaningful background speech > meaningless background speech = silence) offers a significant challenge to Feature-Based Interference-by-Content accounts.

Attentional Diversion

The Attentional Diversion account assumes that an orienting response (or attentional capture) away from the focal task is produced by auditory changes within background sound (Bell, Röer, Dentale, & Buchner, 2012; Cowan, 1995, 1999). To explain disruptive effects attributable to the meaningfulness of background speech on focal task processing entails the further assumption that attentional capture can be triggered by its semantic features. However, on this account it is unclear why disruption is acutely sensitive to the nature of the prevailing task activity (Marsh et al., 2009).

Interference-by-Process

Generally, the Interference-by-Process account (Jones & Tremblay, 2000; Marsh et al., 2009) acknowledges the importance of the nature of processing in the focal task in relation to the nature of irrelevant speech stimuli. According to this account, the impact of background sound on cognitive processing occurs due to the extent that the background

sound and visual stimuli draw on similar processes (see Hyönä & Ekholm, 2016; Marsh et al., 2008, 2009; Sörqvist, Nöstl, & Halin, 2012). In the context of serial short-term memory the preattentive processing of changes in acoustically varying (changing-state) stimuli generate order cues that interfere with processing the serial order of visually to-be-remembered items through serial rehearsal using covert speech. This explains why between-stream phonological, or semantic, similarity fails to modulate the disruption of serial recall; why non-speech and speech sounds produce qualitatively similar effects (Jones & Macken, 1993) and why sounds that contains little, or no, acoustic variation (steady-state) produce much less disruption of serial recall than changing-state sounds (Jones, Madden, & Miles, 1992).

In the context of short-term memory, the Interference-by-Process view asserts that serial recall task is invulnerable to disruption via the meaning of background speech because the task itself does not necessitate semantic processing and the items are often arbitrary and relatively sparse with respect to their semantic properties. Thus, even though there is good evidence for the semantic processing of background speech during serial recall (Röer, Körner, Buchner, & Bell, 2017), those semantic properties do not interfere because there is no clash with any concurrent semantic processing in the focal task. However, the principle of Interference-by-Process extends beyond the bounds of short-term memory (Marsh et al., 2008; 2009). For cases in which semantic processing of visual information is a necessary component of the task, a clash occurs with concurrent involuntary semantic processing of background speech (Beaman, 2004; Marsh et al., 2008; 2009; Neely & LeCompte, 1999). Crucially then, presentation rates, background speech and visual memoranda may be held constant and still, through the manipulation of the nature of focal task processing, qualitative differences in the nature of auditory distraction may be observed. Thus, disruption is dictated by the nature of the focal task process and not by the meaningfulness of the background

speech, nor the degree of semantic similarity between auditory distracters and visual memoranda as the Semantic-Interference-by-Content account asserts.

The Interference-by-Process account could be further extended to account for the disruption of reading comprehension by meaningful speech (Martin et al., 1988; Vasilev, Liversedge, et al., 2019). It is possible that lexical/semantic processing of background speech material impairs semantic integration processes (Marsh et al., 2009, 2014) that underpin reading comprehension. However, to date scant research has investigated how the relationship between the reading processes involved in the focal task and those applied automatically to background speech impacts reading performance. Indeed, there has been a preoccupation with investigating the nature of the distracter stimuli (e.g., acoustics or semantics; Ljung, Sörqvist, & Hygge, 2009) rather than the type of reading processes adopted to meet the demands of the focal task. Therefore, the present study represents the first attempt to manipulate the nature of the focal task in relation to a constant irrelevant speech stream in the context of text comprehension.

In the present study, we adopted two different task instructions to manipulate the nature of reading that participants engaged in as their focal task, whilst maintaining the same focal task stimuli (written sentences) and identical irrelevant speech stimuli across the different task conditions. In this way, we were able to determine how commonality (or differences) in the characteristics of the focal and to-be-ignored materials impact written language processing performance. In particular we manipulated the meaningfulness of irrelevant speech, contrasting normal speech in a participant's mother tongue with speech in a language foreign to the participants. For the focal tasks, written text was processed in two qualitatively different ways. In the first task, the semantic acceptability task, participants were required to simply read a Chinese sentence normally to form a judgement as to whether the sentence made sense. Since this task requires a judgement of meaning, we assumed that

participants would engage in semantic processing during text comprehension such that they formed a coherent representation of sentential meaning. In the second task, the non-character detection task, participants were required to read Chinese sentences and make a judgment as to whether any of the constituent characters in the sentence were non-characters. The non-characters were formed by transposing two radicals in a character, thereby rendering them illegal. Hence, the non-character detection task here may be regarded as being somewhat similar to a visual search task (cf. Maxfield, 1997). In such a task, participants likely process characters linguistically due to automatic activation, but they are not required to undertake more extensive semantic processing involving the integration of the individual meanings of words and the syntactic structure of the sentence to form a representation of sentential meaning. Instead, they need only process the characters within the sentences to identify one that is orthographically illegal.

During natural reading, very robust word frequency effects have been repeatedly demonstrated such that high frequency words typically receive fewer and shorter fixations than low frequency words. Importantly, however, under conditions of visual search in text, frequency effects on target words (i.e., search targets) are attenuated (e.g., Rayner & Fischer, 1996; Rayner & Raney, 1996; Wang, Sui, & White, 2019). As the word frequency effect is generally taken as a hallmark of normal word identification (Inhoff & Rayner, 1986; Paterson & Jordan, 2010; Sereno & Rayner, 2003), its absence for the visual search target suggests that the extent to which target words are lexically processed is reduced, and such a finding would be entirely consistent with the claim that during visual search readers process text for meaning to a lesser degree than they do when they are reading to form a semantic judgment.

The semantic acceptability task and non-character detection tasks are similar to tasks requiring detection of contextual errors, and non-contextual errors, in proofreading studies that have manipulated the presence or absence of concurrent background speech (Halin,

Marsh, Haga, Holmgren, & Sörqvist, 2014; Jones et al., 1990; Venetjoki, Kaarlela-Tuomaala, Keskinen, & Hongisto, 2006). However, in those studies, participants were requested to detect both contextual errors and non-contextual errors in one text simultaneously, likely yielding a dual-task-type paradigm. In the present study the type of error is manipulated at the task level to avoid any propensity for dual-tasking, or the prioritisation of detecting one type of error over another.

In sum, the current study explored the influence of the meaningfulness of background speech in two Chinese written text processing tasks, both of which involved linguistic processing, but only one of which required the formation of a semantic representation of sentential meaning. Also, we adopted eye-tracking methodology to examine the irrelevant speech effect in relation to moment-to-moment cognitive processes. A small number of previous eye-tracking studies have demonstrated that readers tend to make more fixations, more regressions and have longer rereading times when exposed to meaningful speech compared to silence (Cauchard et al., 2012; Hyönä & Ekholm, 2016; Vasilev, Liversedge, et al., 2019; Yan, Meng, Liu, He, & Paterson, 2018). In addition, Vasilev, Liversedge, et al. (2019) reported that meaningfulness of speech did not affect lexical frequency effects (see also Yan et al., 2018). In accordance with these studies, we expected increases in later eye-movement measures when the participants read for meaning in the presence of meaningful speech compared to silence.

A critical question concerned how changes in focal task processing relative to any irrelevant speech processing might modulate the disruption produced by irrelevant speech. Our primary hypothesis was based on the Interference-by-Process account. If the disruption produced by irrelevant speech occurs as a consequence of interference-by-process, then an interaction between task instruction and background sound should occur. That is, the meaningfulness of irrelevant speech should be more disruptive in the semantic acceptability

task than in the non-character detection task, since more aspects of processing in relation to the focal and irrelevant stimulus are shared in the former than the latter situation. Alternatively, however, finding that disruption by irrelevant speech occurs regardless of task instruction would be consistent with the Phonological-Interference-by-Content account, the Semantic-Interference-by-Content account, the Feature-Based Interference-by-Content account and the Attentional Diversion account, according to which changes in the reading process adopted to complete the focal task should not impact the magnitude of the disruption produced by irrelevant speech on the reading tasks.

Method

Participants

60 undergraduate students (mean age =22.5 years, *SD* =2.5; 52 females) recruited from Tianjin Normal University were randomly assigned to one of two between-participant groups: semantic acceptability vs. non-character detection instructions (i.e., 30 in each). A between-participants design was adopted to avoid any potential effects associated with performing one of the tasks transferring to and contaminating effects associated with the other task. All of the participants reported normal or corrected-to-normal vision, normal hearing and were native Chinese speakers. None of them could speak or understand Uyghur (the language used for the meaningless speech condition). Participants received some daily necessities for their participation. The study was approved by the research ethics committee at Tianjin Normal University.

Apparatus

Eye-movements were recorded using an EyeLink 1000 plus eye-tracker (SR Research, Canada). Sampling rate was 1000Hz. The reading materials were presented on a 24-inch screen with a resolution of 1024 × 768 pixels across a single line. At 70 cm viewing distance, each character subtended 1.4°. The participant's head was kept immobile by using a

head and chin rest. Forty-three participants had their right eye tracked and seventeen had their left eye tracked during binocular viewing.

Materials

Sounds. The background sounds were: Chinese speech; Uyghur speech; a control condition of silence. The speech material was a narrative taken from a Chinese-Uyghur bilingual book. Thus, the semantic content of the Chinese speech and the Uyghur speech was equivalent. They were recorded in the same adult female voice in Chinese and Uyghur (the speaker was a native Uyghur speaker as well as being an extremely proficient Chinese speaker. She had used Mandarin as her main language for communication since the age of 7 years and received a Secondary Certificate in Mandarin Proficiency, implying natural intonation and fluent expression in Mandarin). The speech sounds were sampled with 16-bit resolution, at a sampling rate of 44.1 KHz using Audacity 2.1.3 software. All the sounds were presented binaurally continually through headphones (Newmine MX660) for the full period during the experiment. The intensity of both types of speech was 58-70 dB(A). The ambient level for the silent condition was 45 dB(A). All the auditory stimuli were of sufficient duration (no less than 13 minutes and 57 seconds) to extend over the full period that the participants spent reading the text.

Sentences. Sentence stimuli consisted of 75 experimental sentences and 24 filler sentences. 20 participants who did not take part in the eye-tracking experiment were asked to rate the difficulty, naturalness and plausibility of each experimental sentence on a scale of 1 to 5 (1=very easy /very unnatural /very implausible; 5=very difficult /very natural /very plausible). Results showed that the experimental sentences were rated as quite easy (M=1.56, SD=0.21) , very natural (M=4.52, SD=0.27) and quite plausible (M=4.13, SD=0.32).

The experimental sentences were identical under the two tasks. The filler sentences were also very similar in the two tasks. In the non-character detection task, each filler

sentence included an orthographically illegal non-character in place of a legal character. The non-characters were formed by the reversal of their radicals (e.g., the non-character "果" was created from "课" in Figure 1). In the semantic acceptability task, the filler sentences were changed to make them implausible. This was achieved either by changing a small number of characters (on average 2.60 characters, *SD*=0.99 for 15 of the filler sentences), or reversing the order of the subject and object in the sentence (for the remaining 9 filler sentences). The characters that were changed were matched with the originals for syntactic class in order to maintain the grammatical structure of the sentence. These stimuli were also assessed for semantic acceptability in the prescreen procedure outlined earlier and they were rated as implausible (Plausibility: *M*=1.14, *SD*=0.12). Prior to the formal experiment, there were 6 practice sentences, two of which were filler items.

(Insert Figure 1 about here)

Design

A 2 × 3 mixed design was employed with task instruction (semantic acceptability vs. non-character detection) as a between-participants factor and background sound (Chinese speech vs. Uyghur speech vs. silence) as a within-participants factor.

The sentences were divided into three blocks, each consisting 25 experimental sentences and 8 filler sentences. The order of the three background sounds was counterbalanced across participants. Thus each block was presented under each sound condition an equal number of times across participants. The sentences in each block were presented randomly.

Procedure

Participants were tested individually and instructed to ignore background sounds. Before the experiment, the eye tracker was calibrated for each participant using a three-point dot matrix that was displayed across the line where each experimental or filler sentence

would appear. A validation procedure followed to ensure that each fixation was within 0.20° of each calibration point.

At the start of each trial, a fixation cue appeared on the left side of the display screen. When this was fixated, a sentence was displayed with the first character replacing the cue. Participants pressed one of two keys on a keyboard to make their semantic acceptability or non-character identification judgments, after which the sentence disappeared. The equipment was recalibrated whenever necessary. Participants wore headphones during the whole experiment which lasted approximately 25 min.

Analysis

Importantly, we only analysed data from the experimental sentences (data from the filler sentences were discarded). Thus, our analyses are based on participants reading natural sentences under the two task instructions. We undertook analyses of judgment accuracy, global (sentence level) and local (character level) eye-movements. Global eye-movement analyses provide a broad and general indication of the nature of overall eye-movement patterns during reading throughout the entirety of a trial, whereas local measures provide more detailed information focused on the time course of effects specifically associated with processing the individual constituent characters of the sentence. For the global analyses, we computed reading rate, fixation count, regression count, average fixation duration and average saccade length. For the local analyses, we computed first fixation duration (duration of the first fixation on a character), single fixation duration (the duration of a fixation when readers made only one first-pass fixation on a character), gaze duration (sum of all fixations on a character prior to a fixation on another character), regression path duration (the sum of all fixations from the first fixation on a character until a fixation to the right of that character and total fixation time (sum of all fixations on a character). Specifically, first fixation duration, single fixation duration and gaze duration include fixations during first pass, thus

they are usually taken to reflect early stages of linguistic processing in reading, while regression path duration and total fixation time are taken to reflect later stages of processing (Hyönä, Lorch, & Rinck, 2003; Liversedge, Paterson, & Pickering, 1998; Rayner, 1998)[2]. We computed these measures for all of the characters of the sentence other than the first and the last in the sentence.

Analyses were performed with linear-mixed effects models and run with the lme4 package (Bates, Maechler, & Dai, 2009), available in the R environment (R Development Core Team, 2008). Generalized linear mixed models (GLMM) were used to analyse accuracy. For each variable, a model was specified with participants and items as crossed random effects, with task instruction and background sound as fixed factors. Three successive difference contrasts were set up to analyse the independent effects of meaningfulness (Chinese speech vs. Uyghur speech), phonological and acoustical properties of speech (Uyghur speech vs. silence), and overall speech (Chinese speech vs. silence). Regression coefficient estimates (*b*), standard errors (*SE*), *t*-values (*z*-value for the accuracy) and *p*-values are reported. We first run a full random structure for participants and items. If the initial model failed to converge then the random structure was incrementally trimmed, beginning with the items level. Data except accuracy of judgment were log-transformed prior

[2] Note that these eye movement measures of reading are standard and most (if not all) are presented in studies such as this to allow readers to understand comprehensively the patterns of eye movements that occurred when reading was disrupted relative to when it was not. The reading time measures can be characterised as summing spatially contiguous fixations, temporally contiguous fixations or both (see Liversedge et al., 1998, for a more complete discussion of these issues and their importance in relation to the interpretation of eye movement data). Given the definitions of the measures, it is the case that fixations contributing to one measure may also contribute to other measures meaning that these measures are not independent (e.g., first and single fixation duration are subsumed within gaze duration, gaze duration is subsumed within total reading time, etc). Very importantly, however, any such dependency relation does not mean that an effect in one measure necessarily means that a similar effect will be observed in the other measures. This is certainly not the case. Indeed, only through consideration of whether effects occur in all measures, or instead in some of the measures but not others, is it possible to gain significant insight into the nature and time course of disruption to processing.

to analysis. Separate analyses were also performed for each task to tease apart the significant interactions.

Results

All the participants obtained high accuracy scores for their judgments (95%-100%; *M*=98.8%, *SD*=1.6%). The accuracy for non-character detection was slightly higher than that for judgments of semantic acceptability (99.4% vs. 98.1%; *b*=1.19, *SE*=0.37, *z*=3.20, *p*=0.001, *d*=0.12). No other significant effects were found. Clearly, the participants were able to perform the tasks adequately.

Data were excluded if (a) fixations were shorter than 80 ms or longer than 800 ms (7.44%), (b) the trial received less than 3 fixations (0.11%) or was not tracked successfully (0.36%), or (c) the value of the dependent measure was more than 3 standard deviations above the mean for each participant (0.92%).

Global analyses

Means and standard errors for reading rate, fixation count, regression count, average fixation duration and average saccade length broken down by task instruction and sound condition, are shown in Table 1. The results of the linear-mixed effects models, for each of these global measures, are shown in Table 2.

(Insert Table 1 & Table 2 about here)

Robust main effects of task instruction or background sound for all global measures were observed. First, participants made more fixations (*d*=-1.41), more regressions (*d*=-0.88), and had longer average fixation durations (*d*=-0.79) and shorter average saccade lengths (*d*=0.78), and thus had slower reading rates (*d*=1.71) in the semantic acceptability task than they did in the non-character detection task. Second, Chinese speech produced lower reading rates (Chinese speech vs. Uyghur speech, *d*=0.10; Chinese speech vs. silence, *d*=0.12) and more fixations (Chinese speech vs. Uyghur speech, *d*=-0.13; Chinese speech vs. silence, *d*=-

0.16) and regressions (Chinese speech vs. Uyghur speech, *d*=-0.15; Chinese speech vs. silence, *d*=-0.14) than Uyghur speech or silence. Whilst the Uyghur speech condition and the silence condition did not differ in these measures. For average fixation duration, participants made longer fixations when exposed to Chinese speech in comparison to silence (*d*=-0.08), but the difference (2 ms) between Chinese speech and Uyghur speech was not significant. No reliable difference emerged in average saccade length across the three background sound conditions.

The two-way interactions between task instruction and sound condition (Chinese speech vs. Uyghur speech, or Chinese speech vs. silence), in which we were most interested, were significant in reading rate, fixation count and regression count. Two sets of separate analyses were conducted, one for each of the two task instructions. For the semantic acceptability task, separate analyses showed no more interference in the Uyghur speech condition relative to the silence condition. However, there were significant differences between Chinese speech and Uyghur speech (reading rate: *b*=0.13, *SE*=0.02, *t*=5.44, *p*<0.001, *d*=0.30; fixation count: *b*=-0.10, *SE*=0.02, *t*=-4.74, *p*<0.001, *d*=-0.26; regression count: *b*=-0.16, *SE*=0.03, *t*=-4.79, *p*<0.001, *d*=-0.26), and Chinese speech and silence (reading rate: *b*=0.15, *SE*=0.04, *t*=4.06, *p*<0.001, *d*=0.35; fixation count: *b*=-0.12, *SE*=0.03, *t*=-4.18, *p*<0.001, *d*=-0.30; regression count: *b*=-0.16, *SE*=0.05, *t*=-3.11, *p*=0.004, *d*=-0.24). It means that, Chinese speech markedly disrupted semantic acceptability judgments. And while the global measures of average fixation duration and average saccade length did not produce robust effects, there was a numerical trend in the same direction. Namely, average fixation duration was longer and average saccade length was smaller in the Chinese speech condition than that in other two sound conditions. For the non-character detection task, there were no effects of speech on any of the global measures (*ps*≥ 0.39).

From these global measures, we can summarize that the participants who read for meaning had lower accuracy and reading rate in comparison to the participants who were asked to detect a non-character in a sentence. This is consistent with the previous studies examining the influence of task difficulty (or processing depth) on eye-movements in reading (Rayner, 1978, 1998, 2009; Rayner, Chace, Slattery, & Ashby, 2006; Reichle, Vanyukov, Laurent, & Warren, 2008). Importantly, there were reliable interactions between task instruction and background sound, with meaningful speech impairing task performance when participants were engaged in semantic comprehension, but not when engaged in non-character detection. This indicates that the nature of focal task processing determines, to some extent at least, the magnitude of the disruptive influence of background speech on processing. Next, we will consider local processing measures to examine the effects of task instruction and background speech on character processing.

Local analyses

Table 3 shows means and standard errors for local eye-movement measures. The results of the linear-mixed effects models are shown in Table 4.

(Insert Table 3 & Table 4 about here)

Again, there were robust main effects of task instruction. These were significant in all the local measures: first fixation duration (d=-0.70), single fixation duration (d=-0.70), gaze duration (d=-0.73), regression path duration (d=-0.75) and total fixation duration (d=-1.19). All of these measures of reading time were longer in the semantic acceptability task than in the non-character detection task.

Another interesting point was that the effect of speech did not emerge in the first fixation duration and single fixation duration results, but did appear in gaze duration and in later measures. That is, Chinese speech caused longer gaze durations (Chinese speech vs. Uyghur speech, d=-0.09; Chinese speech vs. silence, d=-0.08), regression path durations

(Chinese speech vs. Uyghur speech, d=-0.15; Chinese speech vs. silence, d=-0.16) and total fixation durations (Chinese speech vs. Uyghur speech, d=-0.15; Chinese speech vs. silence, d=-0.18) than Uyghur speech or silence (though the contrast 'Chinese speech vs. Uyghur speech' in gaze duration was only marginal). Clearly, these effects demonstrate that speech did not impact the earliest stages of character recognition, but did affect the likelihood that readers made a refixation on a character, and the time they spent re-reading the preceding portion of the sentence as well as the character itself. Again, no differences between Uyghur speech and silence were found.

Importantly, the interactions between task instruction and sound condition (Chinese speech vs. Uyghur speech, or Chinese speech vs. silence) were only significant for regression path duration and total fixation duration. Separate analyses revealed no difference across the three sound conditions in the non-character detection task ($ps \geq 0.51$). However, Chinese speech increased regression path duration and total fixation durations compared with reading in the presence of Uyghur speech (regression path duration: b=-0.10, SE=0.02, t=-5.33, p<0.001, d=-0.29; total fixation duration: b=-0.07, SE=0.02, t=-4.64, p<0.001, d=-0.31) or reading in silence (regression path duration: b=-0.12, SE=0.03, t=-4.31, p<0.001, d=-0.33; total fixation duration: b=-0.08, SE=0.02, t=-3.97, p<0.001, d=-0.34) for the semantic acceptability task.

To summarize the findings from the local analyses at the level of the individual character, fixation times were longer when participants were required to judge plausibility than when they were required to identify a non-character during sentence reading. More interestingly, meaningful speech exclusively disrupted processing in the semantic acceptability task, and this disruption was restricted to relatively late measures (regression path duration and total fixation duration). In contrast, meaningful speech did not cause impairment of non-character detection. Additionally, meaningless speech had little influence

on either task. Overall, these results, alongside the results from the global analyses, demonstrate a very important point. Specifically, the extent to which a speech distracter exerts an influence over processing is not determined solely by the nature of the speech distractor itself. Instead, it appears that the disruptive influence of a distracter is determined by the nature of the speech comprising that distracter in relation to the nature of processing required for the focal task.

Discussion

This study investigated the role of processing in the focal task in relation to auditory distraction effects in on-line text processing. We investigated this question in an eye-movement experiment in which participants were required to either engage in reading for meaning, or alternatively read a sentence in order to detect a non-character in the presence of concurrent background sounds. We anticipated that when participants read text for meaning, an increased depth of processing would occur that would involve semantic evaluation, whereas when readers searched for a non-character, we expected participants' processing to be shallower with much less by way of interpretation of sentential meaning. We also manipulated the nature of the distracting speech such that it was either meaningful (Chinese for our Chinese participants), or meaningless (Uyghur, a language that was not known to our participants, but which clearly contained phonological speech properties that were readily discernible to our participants).

Our results showed important interactive effects. These effects were such that while both meaningful speech and meaningless speech had no significant influence on processing in the non-character detection task, meaningful speech (Chinese), but not meaningless speech (Uyghur) did produce substantial disruption in the semantic acceptability task. To be clear, disruption by speech on text processing only occurred when the speech was meaningful and the focal task required semantic comprehension. Our primary conclusion from this basic

pattern of effects is that the extent to which auditory distraction effects occur in a reading situation depends both on the nature of the distraction sounds and the precise task that the participant is engaged in. To us, these results are in line with the Interference-by-Process account of distraction effects (Marsh et al., 2008; 2009) and run contra to that expected by Interference-by-Content accounts.

Recall that the Phonological-Interference-by-Content account predicted that any speech-like sound with phonological content could cause impairment, while the Semantic-Interference-by-Content account predicted an unconditional, detrimental effect from meaningful speech as it would interfere with semantic representation of the written text regardless of the characteristics of the reading task. Further, the Feature-Based Interference-by-Content accounts predicted a graded disruptive effect whereby the greatest disruption would emerge from meaningful background speech with a reduced level of disruption arising from meaningless background speech compared with silence. The absent effects of Uyghur speech on both the semantic acceptability and non-character detection tasks reported here cast doubt on the Phonological-Interference-by-Content account and the Feature-Based Interference-by-Content account. Moreover, that disruption from meaningful background speech only occurred for the semantic acceptability task further supports the Interference-by-Process account over the Semantic-Interference-by-Content account that assumes unconditional disruption by meaningful speech regardless of whether the focal task requires extensive semantic processing.

The pattern of findings reported here is also at odds with any account that assumes the disruption produced by background speech is a function of the extent to which those distracters capture attention away from the focal task (Bell et al., 2012; Cowan, 1995, 1999). Generally, this Attentional Diversion account assumes that auditory changes within background sound cause an orienting response away (or attentional capture) from the focal

task. An extension of this account to deal with the effects of meaningfulness observed on reading reported here is to assume that semantic features within the sound produce attentional capture. Therefore, in relation to the current study, the Attentional Diversion view proposes that a disruptive effect of changing sound should be observed for both semantic acceptability and non-character detection tasks, which is inconsistent with the pattern of data we observed. On this account it is unclear why the semantic properties of irrelevant sound should be endowed with greater attentional capturing power for the semantic acceptability task—whereby an effect of background speech meaningfulness was observed, relative to the non-character detection task, where an effect of background speech meaningfulness did not occur. Furthermore, the Attentional Diversion account has also met with difficulty in the domain of short-term memory wherein only tasks that require serial processing are prone to disruption via acoustic changes within background speech (e.g., Beaman & Jones, 1997; Hughes, Vachon, & Jones, 2007).

Previous findings suggest that at least some semantic analysis of background speech occurs regardless of whether semantic processing is a determinant of focal task performance. For example, Röer et al. (2017) report that participants produce a greater number of category-exemplars in the context of a verbal fluency task, if those exemplars were presented earlier as to-be-ignored distracters during a serial recall task. Further, Vachon, Marsh, and Labonté (2019) demonstrate that the presentation of a category-change within a to-be-ignored sequence (e.g., onion, carrot, sprout, lettuce, parsnip, lorry, cabbage, swede) disrupts performance of a concurrent visually-based focal task. Given Röer et al.'s and Vachon et al.'s findings demonstrating that some semantic analysis of background speech occurs regardless of whether semantic processing is a determinant of focal task performance, and given the automaticity of semantic processing for written words (MacLeod, 1991), there is no reason to

suppose that the semantic features of background speech that determine distraction should be more potent with respect to attentional capture in an arguably more difficult reading task—semantic acceptability—over an arguably easier reading task—non-character detection (for similar conclusions, see Marsh et al., 2009). On the contrary, Halin, Marsh, and Sörqvist (2015) found that participants were able to answer more questions concerning a background speech story when it had been presented during performance of an easier task (n-1-back) than a more difficult task (n-2-back), suggesting that peripheral processing of the meaning of background speech is reduced under high levels of central load. If the difficulty of a task can be gauged by performance levels on that task, then it would be expected that semantic processing of background speech, whether the result of capture or not, should be greater for the easier non-character detection task than for the more difficult semantic acceptability task, which is the opposite of what we observed. This pattern of findings undermines any account that assumes that attentional capture by the semantic properties of background speech occurs to the extent that the task is difficult.

Furthermore, a raft of studies has shown that increasing task-difficulty reduces distraction of task performance by background sound. Indeed, in a number of studies task-difficulty has been manipulated by making the to-be-attended material a little harder to read (Ball, Threadgold, Solowiej, & Marsh, 2018; Halin, Marsh, Hellman, Hellström, & Sörqvist, 2014; Halin, Marsh, Haga, et al., 2014; Hughes, Hurlstone, Marsh, Vachon, & Jones, 2013; Marsh et al., 2018). For example, in the context of a serial short-term memory task, Hughes et al. (2013) increased task difficulty (specifically inducing greater encoding load) by overlaying the to-be-remembered items with Gaussian visual noise. The manipulation eliminated the disruptive impact that an auditory deviant (a change in voice within the background speech sequence) has on serial recall performance but had no impact on the changing-state effect (the disruption produced by a sequence of changing- relative to steady-

state tokens). The explanation for this was that high encoding load increases task-engagement and may potentiate a blocking mechanism such that attentional capture is resisted. The changing-state effect was unaffected by task difficulty because it is not attributable to attentional capture but rather a clash between a preattentive process applied to the sound and the deliberate serial rehearsal process (the two tasks requiring similar order-based processes). So here, then, making the task difficult does not necessarily modulate all forms of auditory distraction especially in circumstances wherein similar processes, be they serial order based or semantic, are applied to to-be-attended and to-be-ignored material.

In contrast to the Interference-by-Content and Attentional Diversion accounts, the Interference-by-Process account provides a clear and straightforward explanation for the qualitatively different auditory distraction effects we observed in the two present tasks. In the semantic acceptability task emphasis was placed on semantic comprehension, and semantic analysis also appeared to occur relatively automatically for the meaningful auditory distractor stimuli, resulting in processing interference. In contrast, participants processed the meaning of the sentences to a much lesser degree in the non-character detection task (Rayner & Fischer, 1996; Rayner & Raney, 1996; Wang et al., 2019), and consequently the automatic analysis of speech meaning had little influence on the detection of nonwords (i.e., there was far less shared processing interference).

One point that is worth dealing with in relation to the present results concerns the possibility that the two tasks, processing sentences for semantic acceptability compared with processing sentences to detect a non-character may be fundamentally different. To be clear, it might be argued that the non-character detection task is an entirely visually mediated search task that does not engage linguistic processing (as distinct from the semantic acceptability task that is entirely linguistically mediated). If this were true, then processing in the two tasks might be considered independent and this might explain the differential results. In our view,

this suggestion is very unlikely. As mentioned earlier, semantic processing of words is known to be automatic and reflexive (MacLeod, 1991), and based on Halin et al., (2015), there is no evidence to suggest that semantic aspects of background speech should have increased distraction effects in a semantic acceptability task relative to a non-character detection task.

Despite this, the fact remains that in our experiment participants spent longer reading sentences for meaning than they did when making judgments as to the presence of a non-character. At some level, this is entirely to be expected since reading for meaning requires that participants carefully integrate the individual words into a message level representation, which itself requires frequent re-reading of the text to ensure that the representation of sentential meaning is well-formed and coherent. By contrast, when readers make a judgment as to the presence or absence of a non-character in the sentence, then they are simply required to process each character of the sentence, ensure that all conform with orthographic rules and potentially identify them where possible. The degree to which text is re-inspected is far reduced in such a situation. To focus on the global processing time differences that exist between tasks and conclude on that basis that the two tasks require processing that is categorically different in nature, in our view, is erroneous.

Beyond this assertion, we might also test this suggestion more directly. If the semantic acceptability task requires reading, then a very straightforward hypothesis is that there should be a strong relationship between reading time measures and one of the most basic linguistic characteristics of the words in the sentences, that is, their lexical frequency (as has been demonstrated very robustly by over 40 years of eye movement research). In contrast, if the non-character detection task draws solely upon visual processes underlying search, and does not engage linguistic processing, then no such relationship between fixation durations and lexical frequency of words should exist. The alternative possibility (that we favour) is that linguistic processing is engaged in both tasks and to this extent they are

fundamentally very similar, and therefore, any relationship between the linguistic characteristics of the words in the sentences and eye movement behaviour should exist to a comparable level in both the semantic acceptability and non-character detection tasks. Clearly, these two positions are distinct and discriminable. We, therefore, undertook a comprehensive set of analyses of our eye movement data to explore this. We first segmented our sentences into word units and defined each individual word in each sentence (excluding the first and last in the sentence as is standard in eye movement experiments investigating reading) as a region of interest. We based our criterion for a word on the SUBTLEX-CH database (Cai & Brysbaert, 2010). In this way, we obtained 497 words in total from 75 experimental sentences. Five of these words had extremely high frequencies that represented outliers (>3SD from the mean) in our sample, and therefore, these were removed. Basic properties of the remaining 492 words are summarized in Table 5. Next, for each word we computed five local eye movement measures: first fixation duration, single fixation duration, gaze duration, regression path duration and total fixation duration. All of these measures are known to provide very robust index of lexical frequency effects in reading. Next, we correlated each of these reading time measures with the lexical frequency of each of the words. The frequency data were log-transformed prior to correlation analysis. As can be seen from Table 6, our results were very clear. We obtained very robust and quite comparable correlations for each of our 5 measures with lexical frequency and this occurred for both tasks. Furthermore, the effects we observed were, if anything, stronger for the non-character detection task than for the semantic acceptability task. Based on these results, we feel it is very clear that both our tasks engaged linguistic processing, drawing upon comparable underlying linguistic processes. We find these comprehensive and very robust analyses much more compelling in relation to the question of the comparability of our two tasks than simple comparison of overall mean reading/judgment times.

Another main goal of the present study was to use eye-tracking methods to investigate which specific linguistic processes associated with reading are affected by the concurrent speech. Globally speaking, there were substantive distraction influences for meaningful speech on the eye-movement record when the task required readers to form a semantic evaluation of the written sentence. However, when readers simply scanned for a non-character, global disruption of eye-movement behaviour by meaningful speech was minimal. Specifically, when the irrelevant sounds were meaningless, there was little evidence of disruption regardless of whether readers evaluated the sentence for meaning or processed a sentence to detect a non-character. These effects were very robust, occurring in all the global eye-movement measures apart from average fixation duration and average saccade length (though these measures did show consistent numerical differences).

Turning next to the local eye movement measures, here we found that the early measures (first fixation duration, single fixation duration, and gaze duration) were exclusively influenced by the task and did not show robust interactive effects of task and the meaningfulness of speech. In contrast, later measures of regression path duration and total fixation duration did show interactive effects. Clearly, the interactive effects stemmed from refixation behaviour. That is to say, the disruptive effect of speech caused readers to make additional fixations on words and make regressions to re-read and effectively form a coherent interpretation of the sentence. Our results are consistent with other studies investigating auditory distraction on on-line reading (Cauchard et al., 2012; Hyönä & Ekholm, 2016; Vasilev, Liversedge, et al., 2019), which reported disruption was mostly due to an increase in rereading time. Furthermore, these results fit neatly with the work by Vasilev, Liversedge, et al. (2019) who have shown that lexical frequency effects during sentence reading are not modulated by auditory distraction (see also Vasilev, 2019; Vasilev, Parmentier, Angele, & Kirkby, 2019). Frequency effects are a hallmark of efficient lexical identification. The fact

that they are not influenced by meaningful speech, just like the present findings, suggests that lower level linguistic processes such as word identification are not impacted negatively by meaningful or meaningless auditory stimuli. To recap briefly, it appears that shallow aspects of perceptual and linguistic processing are barely influenced by the presence of meaningful speech, whereas the same speech sounds disrupt higher order processing associated with the formation of a representation of sentential meaning. Recall that according to the Interference-by-Process account, the disruption by meaningful speech that appeared in the semantic acceptability task occurred due to shared processing interference in semantic processing associated with the construction of a representation of sentence meaning. As semantic interpretation represents a relatively late stage of linguistic processing in reading, it seems reasonable that significant distraction effects occurred mainly on the later eye-movement measures which are reflective of such higher order processing (Booth & Weger, 2013; Hyönä et al., 2003; Reichle, Warren, & McConnell, 2009).

It should be noted that the present findings are inconsistent with those of Jones et al. (1990), which showed that meaningful speech reduced detection rate for non-contextual errors (i.e., misspellings or omission errors), but exerted no effects on detection of contextual errors (i.e., grammatical errors or inappropriate words). Notably, different from our study, Jones et al. required participants to detect both non-contextual errors and contextual errors in one text simultaneously, which made the task somewhat like a dual-task paradigm. Jones et al. suggested that under irrelevant speech conditions, the readers' limited attentional capacity was primarily directed towards meaning and away from the physical features of the text. Thus attentional resources allocated to the detection of nonwords were reduced, while detection of contextual errors was little affected. However, other studies (Halin, Marsh, Haga, et al., 2014; Venetjoki et al., 2006; Weinstein, 1974, 1977) failed to replicate the results of Jones et al. (1990). For example, recently, Halin, Marsh, Haga, et al. (2014) found

background speech even improved detection of spelling errors of function words, but its presence made it more difficult for participants to detect contextual errors of function words: a result that is in line with the findings of the present study.

One of the most striking aspects of the present findings concerns the pervasive influence of task instruction on our dependent measures. Interestingly, whether participants were instructed to evaluate sentences for meaning, or simply scan a sentence for a non-character, affected response accuracy (accurate non-character detection rates were higher than were the rates of accurate semantic judgments), affected global reading measures (all of which together showed that readers made more and longer fixations, shorter saccades and took longer overall to read sentences for meaning than was the case when sentences were scanned for non-characters), and local measures of character processing (with local analyses of reading time being longer in the semantic acceptability task than the non-character detection task)[3]. These results are consistent with previous studies exploring eye-movement control in reading and target-word search (e.g., Rayner & Raney, 1996). All of the data in

[3] It is interesting that in the present results, the non-character detection task produced significantly shorter fixations than the semantic acceptability task. This pattern of effects is in contrast to the results of Kaakinen and Hyona (2010), and Schotter, Bicknell, Howard, Levy, and Rayner (2014) who found longer fixation durations when their participants engaged in proofreading compared to reading for comprehension. We can offer two possible explanations for this difference. First, the present study used a between-participants design, whereas both the earlier studies used a within-subjects design in which the comprehension task was performed first, followed by the proofreading task. It is possible that after performing a reading task participants may find it difficult to disengage such processing, and therefore, the longer fixations during the proofreading could reflect a dual task situation. Also, since targets in the proofreading task were words with internal letter transpositions, these may have orthographically activated the lexical entries for the corresponding words to such a degree that target identification was challenging. By contrast, radical transpositions in Chinese characters, as used here in the non-character detection task, involved changes that were, arguably, more substantive. Also, these transpositions frequently produced characters with radical positional frequency violations, perhaps making target detection easier. Of course, these suggestions are speculative. Nonetheless, it remains the case that both these possibilities may have contributed to the differences between the present results and those of Kaakinen and Hyona (2010), and Schotter et al. (2014). We are grateful to an anonymous reviewer for bringing this point to our attention.

these studies support the suggestion that reading and evaluating sentences for meaning is more effortful and requires deeper cognitive processing than does processing sentences in order to determine whether a non-character is present (Perkins & Brutten, 1992; Reichle et al., 2008). Of further note is that in both tasks we exclusively evaluated data from those trials in which the sentence was meaningful (in the semantic acceptability task), or did not contain a non-character (in the non-character detection task), confirming that none of the effects we report are attributable to the presence of any oddity in the sentence.

To sum up, the study reported here is one of very few (if any) that has investigated the role of focal task processes in the susceptibility to auditory distraction during online linguistic processing. We found that the auditory distraction on tasks involving text processing is determined jointly by both the properties of speech and the nature of the focal task. And this detrimental effect primarily occurs in relation to higher order (post-lexical) linguistic processing. Further work should explore the theoretical reach of the Interference-by-Process account. For example, it is possible that the disruption observed to writing by background meaningful speech (Sörqvist et al., 2012) could be underpinned by disruption to reading processes as observed in the current study, since writing tasks often require re-reading of produced text (Torrance, Johansson, Johansson, & Wengelin, 2016).

Acknowledgments

This research was supported by the Project of Key Research Institute of Humanities and Social Sciences in Universities of Ministry of Education of China (Grant No. 15JJD190003), and the TNU Funding Programme for Novice Ph.D. Students (Grant No. 2017BSXR004). The data files, R scripts, and reading stimuli used in the present research are openly available at: https://osf.io/dh8se/

References

Anderson, J. R. (1983). *The architecture of cognition*. Cambridge, MA: Harvard University Press.

Baddeley, A. D. (1979). Working memory and reading. In P. A. Kolers, M. E. Wrolstad, & H. Bouma (Eds.), *Processing of visible language*, Vol. I, pp. 355-370. New York: Plenum Press.

Baddeley, A. D. (1986). *Working Memory*. Oxford: Oxford University Press.

Ball, L. J., Threadgold, E., Solowiej, A., & Marsh, J. E. (2018). Can Intrinsic and Extrinsic Metacognitive Cues Shield Against Distraction in Problem Solving? *Journal of Cognition, 1*, 15. doi.org/10.5334/joc.9

Banbury, S., & Berry, D. C. (1998). The disruption of speech and office-related tasks by speech and office noise. *British Journal of Psychology*, *89*, 499-517. doi.org/10.1111/j.2044-8295.1998.tb02699.x

Bates, D., Maechler, M., & Dai, B. (2009). lme4: linear mixed-effects models using S4 classes. 2009. R Package, version, 999375–31.

Beaman, C. P. (2004). The irrelevant sound phenomenon revisited: What role for working memory capacity? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*, 1106– 1118. doi.org/10.1037/0278-7393.30.5.1106

Beaman, C. P., & Jones, D. M. (1997). Role of serial order in the irrelevant speech effect: Tests of the changing-state hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23*, 459-471. doi.org/10.1037/0278-7393.23.2.459

Bell, R., Röer, J. P., Dentale, S., & Buchner, A. (2012). Habituation of the irrelevant sound effect: Evidence for an attentional theory of short-term memory disruption. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*, 1542-1557. doi.org/10.1037/a0028459

Booth, R. W., & Weger, U. W. (2013). The function of regressions in reading: Backward eye movements allow rereading. *Memory & Cognition, 41*, 82-97. doi.org/10.3758/s13421-012-0244-y

Boyle, R., & Coltheart, V. (1996). Effects of irrelevant sounds on phonological coding in reading comprehension and short-term memory. *Quarterly Journal of Experimental Psychology, 49A*, 398-416. doi.org/10.1080/027249896392702

Cai, Q., & Brysbaert, M. (2010). SUBTLEX-CH: Chinese word and character frequencies based on film subtitles. *PLOS ONE, 5*, e10729. doi.org/10.1371/journal.pone.0010729

Cauchard, F., Cane, J. E., & Weger, U. W. (2012). Influence of background speech and music in interrupted reading: An eye-tracking study. *Applied Cognitive Psychology*, *26*, 381-390. doi.org/10.1002/acp.1837

Colle, H. A., & Welsh, A. (1976). Acoustic masking in primary memory. *Journal of Verbal Learning and Verbal Behavior, 15*, 17-31. doi.org/10.1016/S0022-5371(76)90003-7

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review, 108*, 204–256. doi.org/10.1037/0033-295X.108.1.204

Cowan, N. (1995). *Attention and memory: An integrated framework.* New York: Oxford University Press.

Cowan, N. (1999). An embedded-processes model of working memory. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 62–101). New York, NY: Cambridge University Press.

Gathercole, S. E., & Baddeley, A. D. (1993). *Working memory and language*. Hove, U.K.: Erlbaum.

Haarmann, H. J., Davelaar, E. J., & Usher, M. (2003). Individual differences in semantic short-term memory capacity and reading comprehension. *Journal of Memory and Language, 48*, 320-345. doi.org/10.1016/S0749-596X(02)00506-5

Halin, N., Marsh, J. E., Haga, A., Holmgren, M., & Sörqvist, P. (2014). Effects of speech on proofreading: Can task-engagement manipulations shield against distraction? *Journal of Experimental Psychology: Applied*, *20*, 69-80. doi.org/10.1037/xap0000002

Halin, N., Marsh, J. E., Hellman, A., Hellström, I. & Sörqvist, P. (2014). A shield against distraction. *Journal of Applied Research in Memory and Cognition, 3*, 31-36. doi.org/10.1016/j.jarmac.2014.01.003

Halin, N., Marsh, J. E., & Sörqvist, P. (2015). Central load reduces peripheral processing: Evidence from incidental memory of background speech. *Scandinavian Journal of Psychology, 56*, 607–612. doi.org/10.1111/sjop.12246

Hughes, R. W., Hurlstone, M. J., Marsh, J. E., Vachon, F., & Jones, D. M. (2013). Cognitive control of auditory distraction: Impact of task difficulty, foreknowledge, and working memory capacity support duplex-mechanism account. *Journal of Experimental Psychology: Human Perception and Performance, 39*, 539–553. doi.org/10.1037/a0029064

Hughes, R. W., Vachon, F., & Jones, D. M. (2007). Disruption of short-term memory by changing and deviant sounds: Support for a duplex-mechanism account of auditory distraction. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*, 1050–1061. doi.org/10.1037/0278-7393.33.6.1050

Hyönä, J., & Ekholm, M. (2016). Background speech effects on sentence processing during reading: An eye movement study. *PLOS ONE, 11*, e0152133. doi.org/10.1371/journal.pone.0152133

Hyönä, J., Lorch Jr, R. F., & Rinck, M. (2003). Eye movement measure to study global text processing. In J. Hyönä, R. Radach, & H. Deubel (Eds.), *The mind's eye: Cognitive and applied aspects of eye movement research* (pp. 313–334). Amsterdam: Elsevier Science.

Inhoff, A. W., & Rayner, K. (1986). Parafoveal word processing during eye fixations in reading: Effects of word frequency. *Perception & Psychophysics, 40*, 431-439. doi.org/10.3758/BF03208203

Johansson, R., Holmqvist, K., Mossberg, F., & Lindgren, M. (2012). Eye movements and reading comprehension while listening to preferred and non-preferred study music. *Psychology of Music, 40*, 339-356. doi.org/10.1177/0305735610387777

Jones, D. M. (1995). The fate of the unattended stimulus: Irrelevant speech and cognition. *Applied Cognitive Psychology*, *9*, 23-38. doi.org/10.1002/acp.2350090704

Jones, D. M., & Macken, W. J. (1993). Irrelevant tones produce an irrelevant speech effect: Implications for phonological coding in working memory. *Journal of Experimental Psychology: Learning, Memory and Cognition, 19*, 369–381. doi.org/10.1037/0278-7393.19.2.369

Jones, D., Madden, C., & Miles, C. (1992). Privileged access by irrelevant speech to short-term memory: The role of changing state. *Quarterly Journal of Experimental Psychology, 44*, 645-669. doi.org/10.1080/14640749208401304

Jones, D. M., Marsh, J. E., & Hughes, R. W. (2012). Retrieval from memory: Vulnerable or inviolable? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*, 905–922. doi.org/10.1037/a0026781

Jones, D. M., Miles, C., & Page, C. (1990). Disruption of proofreading by irrelevant speech: Effects of attention, arousal or memory? *Applied Cognitive Psychology*, *4*, 89-108. doi.org/10.1002/acp.2350040203

Jones, D. M., & Tremblay, S. (2000). Interference in memory by process or content? A reply to Neath (2000). *Psychonomic Bulletin & Review*, *7*, 550-558. doi.org/10.3758/BF03214370

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, *99*, 122-149. doi.org/10.1037/0033-295X.99.1.122

Kaakinen, J. K., & Hyönä, J. (2010). Task effects on eye movements during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*, 1561-1566. doi.org/10.1037/a0020693

Liversedge, S. P., Paterson, K. B., & Pickering, M. (1998). Eye movements and measures of reading times. In G. Underwood (Ed.), *Eye guidance in reading and scene perception* (1st ed.). Amseterdam: Elsevier.

Ljung, R., Sörqvist, P., & Hygge, S. (2009). Effects of road traffic noise and irrelevant speech on children's reading and mathematical performance. *Noise and Health*, *11*, 194-198. doi.org/10.4103/1463-1741.56212

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin, 109*, 163-203. doi.org/10.1037/0033-2909.109.2.163

Marsh, J. E., Hughes, R. W., & Jones, D. M. (2008). Auditory distraction in semantic memory: A process-based approach. *Journal of Memory and Language*, *58*, 682-700. doi.org/10.1016/j.jml.2007.05.002

Marsh, J. E., Hughes, R.W., & Jones, D. M. (2009). Interference by process, not content, determines semantic auditory distraction. *Cognition*, *110*, 23-38. doi.org/10.1016/j.cognition.2008.08.003

Marsh, J. E., Ljung, R., Jahncke, H., MacCutcheon, D., Pausch, F., & Ball, L. J., et al. (2018). Why are background telephone conversations distracting? *Journal of Experimental Psychology: Applied, 24*, 222-235. doi.org/10.1037/xap0000170

Marsh, J. E., Perham, N., Sörqvist, P., & Jones, D. M. (2014). Boundaries of semantic distraction: Dominance and lexicality act at retrieval. *Memory & Cognition, 42*, 1285-1301. doi.org/10.3758/s13421-014-0438-6

Martin, R. C., Wolgalter, M. S., & Forlano, J. G. (1988). Reading comprehension in the presence of unattended speech and music. *Journal of Memory and Language*, *27*, 382-398. doi.org/10.1016/0749-596X(88)90063-0

Maxfield, L. (1997). Attention and semantic priming: A review of prime task effects. *Consciousness and Cognition*, *6*, 204-218. doi.org/10.1006/ccog.1997.0311

Meng, Z., Lan, Z., Yan, G., Marsh, J. E., & Liversedge, S. P. (2020). *Task demands modulate the effects of speech on text processing* [Data set and code book]. OSF. doi.org/10.17605/OSF.IO/DH8SE

Murphy, D. R., Bailey, H., Pearson, M., & Albert, G. (2018). The irrelevant speech effect among younger and older adults: The influence of background noises on reading comprehension. *Experimental Aging Research, 44*, 162-178. doi.org/10.1080/0361073X.2017.1422561

Neath, I. (2000). Modeling the effects of irrelevant speech on memory. *Psychonomic Bulletin and Review, 7*, 403-423. doi.org/10.3758/bf03214356

Neely, C. B., & Lecompte, D. C. (1999). The importance of semantic similarity to the irrelevant speech effect. *Memory & Cognition, 27*, 37-44. doi.org/10.3758/BF03201211

Oberauer, K., & Lange, E. B. (2008). Interference in working memory: Distinguishing similarity-based confusion, feature overwriting, and feature migration. *Journal of Memory and Language, 58*, 730–745. doi.org/10.1016/j.jml.2007.09.006

Page, M. P. A., & Norris, D. G. (2003). The irrelevant sound effect: What needs modelling, and a tentative model. *Quarterly Journal of Experimental Psychology Section A*, *56*, 1289-1300. doi.org/10.1080/02724980343000233

Paterson, K. B., & Jordan, T. R. (2010). Effects of increased letter spacing on word identification and eye guidance during reading. *Memory & Cognition, 38*, 502-512. doi.org/10.3758/MC.38.4.502

Perkins, K., & Brutten, S. R. (1992). The effect of processing depth on ESL reading comprehension. *Journal of Research in Reading, 15*, 67-81. doi.org/10.1111/j.1467-9817.1992.tb00023.x

Rayner, K. (1978). Eye movements in reading and information processing. *Psychological Bulletin, 85*, 618-660. doi.org/10.1037/0033-2909.85.3.618

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*, 372-422. doi.org/10.1037/0033-2909.124.3.372

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology*, *62*, 1457-1506. doi.org/10.1080/17470210902816461

Rayner, K., Chace, K. H., Slattery, T. J., & Ashby, J. (2006). Eye movements as reflections of comprehension processes in reading. *Scientific Studies of Reading, 10*, 241-255. doi.org/10.1207/s1532799xssr1003_3

Rayner, K., & Fischer, M. H. (1996). Mindless reading revisited: Eye movements during reading and scanning are different. *Perception & Psychophysics, 58*, 734-747. doi.org/10.3758/BF03213106

Rayner, K., & Raney, G. E. (1996). Eye movement control in reading and visual search: Effects of word frequency. *Psychonomic Bulletin & Review, 3*, 245-248. doi.org/10.3758/BF03212426

Reichle, E. D., Vanyukov, P. M., Laurent, P. A., & Warren, T. (2008). Serial or parallel? Using depth-of-processing to examine attention allocation during reading. *Vision Research, 48*, 1831-1836. doi.org/10.1016/j.visres.2008.05.007

Reichle, E. D., Warren, T., & McConnell, K. (2009). Using EZ Reader to model the effects of higher level language processing on eye movements during reading. *Psychonomic Bulletin & Review, 16*, 1-21. doi.org/10.3758/PBR.16.1.1

Röer, J. P., Körner, U., Buchner, A., & Bell, R. (2017). Semantic priming by irrelevant speech. *Psychonomic Bulletin & Review, 24,* 1205-1210. doi.org/10.3758/s13423-016-1186-3

Salamé, P., & Baddeley, A. D. (1982). Disruption of short-term memory by unattended speech: Implications for the structure of working memory. *Journal of Verbal Learning and Verbal Behavior*, *21*, 150-164. doi.org/10.1016/S0022-5371(82)90521-7

Schotter, E. R., Bicknell, K., Howard, I., Levy, R., & Rayner, K. (2014). Task effects reveal cognitive flexibility responding to frequency and predictability: Evidence from eye movements in reading and proofreading. *Cognition, 131*, 1-27. doi.org/10.1016/j.cognition.2013.11.018

Sereno, S. C., & Rayner, K. (2003). Measuring word recognition in reading: Eye movements and event-related potentials. *Trends in Cognitive Sciences, 7*, 489-493. doi.org/10.1016/j.tics.2003.09.010

Sörqvist, P. (2010). Effects of aircraft noise and speech on prose memory: What role for working memory capacity? *Journal of Environmental Psychology, 30*, 112-118. doi.org/10.1016/j.jenvp.2009.11.004

Sörqvist, P., Nöstl, A., & Halin, N. (2012). Disruption of writing processes by the semanticity of background speech. *Scandinavian Journal of Psychology, 53*, 97-102. doi.org/10.1111/j.1467-9450.2011.00936.x

Torrance, M., Johansson, R., Johansson, V., & Wengelin, Å. (2016). Reading during the composition of multi-sentence texts: An eye-movement study. *Psychological Research*, *80*, 729-743. doi.org/10.1007/s00426-015-0683-8

Vachon, F., Marsh, J. E., & Labonté, K. (2019). The automaticity of semantic processing revisited: Auditory distraction by acategorical deviation. *Journal of Experimental Psychology: General.* Advance online publication. doi.org/10.1037/xge0000714

Vasilev, M. R. (2019). The determinants of auditory distraction during reading: An eye-movement investigation (Doctoral dissertation, Bournemouth University).

Vasilev, M. R., Kirkby, J. A., & Angele, B. (2018). Auditory distraction during reading: A Bayesian meta-analysis of a continuing controversy. *Perspectives on Psychological Science,13*, 567-597. doi.org/10.1177/1745691617747398

Vasilev, M. R., Liversedge, S. P., Rowan, D., Kirkby, J. A., & Angele, B. (2019). Reading is disrupted by intelligible background speech: Evidence from eye-tracking. *Journal of Experimental Psychology: Human Perception and Performance.* Advance online publication. doi.org/10.1037/xhp0000680

Vasilev, M. R., Parmentier, F. B., Angele, B., & Kirkby, J. A. (2019). Distraction by deviant sounds during reading: An eye-movement study. *Quarterly Journal of Experimental Psychology, 72*, 1863-1875. doi.org/10.1177/1747021818820816

Venetjoki, N., Kaarlela-Tuomaala, A., Keskinen, E., & Hongisto, V. (2006). The effect of speech and speech intelligibility on task performance. *Ergonomics, 49*, 1068-1091. doi.org/10.1080/00140130600679142

Wang, X., Sui, X., & White, S. J. (2019). Searching for a word in Chinese text: Insights from eye movement behaviour. *Journal of Cognitive Psychology, 31*, 145-156. doi.org/10.1080/20445911.2019.1585435

Weinstein, N. D. (1974). Effect of noise on intellectual performance. *Journal of Applied Psychology, 59*, 548-554. doi.org/10.1037/h0037338

Weinstein, N. D. (1977). Noise and intellectual performance: A confirmation and extension. *Journal of Applied Psychology, 62*, 104-107. doi.org/10.1037/0021-9010.62.1.104

Yan, G. L., Meng, Z., Liu, N. N., He, L. Y., & Paterson, K.B. (2018). Effects of Irrelevant Background Speech on Eye Movements during Reading. *Quarterly Journal of Experimental Psychology, 71*, 1270-1275. doi.org/10.1080/17470218.2017.1339718

Table 1

Mean (standard error) for each global eye-movement measure, broken down by task instruction and sound condition.

		Reading rate (char/min)	**Fixation count**	**Regression count**	**Average fixation duration (ms)**	**Average saccade length (char)**
Semantic acceptability	Chinese speech	361 (6)	14.72 (0.21)	4.24 (0.09)	218 (1)	3.43 (0.04)
	Uyghur speech	409 (6)	13.32 (0.20)	3.66 (0.08)	215 (1)	3.48 (0.04)
	Silence	421 (7)	13.07 (0.19)	3.68 (0.08)	214 (1)	3.44 (0.04)
Non-character detection	Chinese speech	812 (11)	7.42 (0.11)	2.21 (0.06)	194 (1)	4.44 (0.05)
	Uyghur speech	829 (12)	7.40 (0.13)	2.17 (0.06)	193 (1)	4.48 (0.06)
	Silence	825 (11)	7.38 (0.12)	2.19 (0.05)	193 (1)	4.38 (0.05)

Table 2

Output from the liner-mixed effects models, for each of the five global measures. Significant p-values are marked in bold.

	Reading rate				**Fixation count**				**Regression count**				**Average fixation duration**				**Average saccade length**			
	b	*SE*	*t*	*p*	*b*	*SE*	*t*	*p*	*b*	*SE*	*t*	*p*	*b*	*SE*	*t*	*p*	*b*	*SE*	*t*	*p*
Intercept(grand mean)	6.26	0.04	153.84	**<0.001**	2.23	0.04	58.93	**<0.001**	0.86	0.04	19.73	**<0.001**	5.31	0.01	446.49	**<0.001**	1.65	0.03	48.134	**<0.001**
task instruction (semantic acceptability vs. non-character detection)	0.74	0.08	9.35	**<0.001**	-0.62	0.07	-8.60	**<0.001**	-0.60	0.09	-7.01	**<0.001**	-0.11	0.02	-4.62	**<0.001**	0.25	0.07	3.71	**<0.001**
Sound (Chinese vs. Uyghur)	0.07	0.02	3.27	**0.002**	-0.06	0.02	-3.19	**0.002**	-0.10	0.03	-3.39	**0.001**	-0.01	0.01	-1.59	0.12	0.005	0.01	0.41	0.69
Sound (Chinese vs. silence)	0.08	0.02	3.50	**<0.001**	-0.07	0.02	-3.57	**<0.001**	-0.08	0.03	-2.64	**0.01**	-0.01	0.01	-2.10	**0.04**	-0.01	0.02	-0.79	0.44
Sound (Uyghur vs. silence)	0.01	0.02	0.45	0.66	-0.01	0.02	-0.52	0.61	0.02	0.03	0.60	0.55	-0.00	0.01	-0.58	0.56	-0.02	0.01	-1.21	0.23
task instruction × Sound (Chinese vs. Uyghur)	-0.12	0.04	-2.76	**0.008**	0.09	0.04	2.37	**0.02**	0.11	0.06	1.92.	**0.06**	0.01	0.01	1.04	0.30	-0.01	0.02	-0.54	0.59
task instruction × Sound (Chinese vs. silence)	-0.14	0.04	-3.14	**0.003**	0.11	0.04	2.89	**0.005**	0.15	0.06	2.35	**0.02**	0.02	0.01	1.41	0.16	-0.01	0.03	-0.42	0.68
task instruction × Sound (Uyghur vs. silence)	-0.02	0.05	-0.48	0.63	0.02	0.04	0.64	0.52	0.03	0.06	0.62	0.54	0.01	0.01	0.43	0.67	0.00	0.03	0.02	0.99

Table 3

Mean (standard error) for each local eye-movement measure, broken down by task instruction and sound condition.

		First fixation duration (ms)	**Single fixation duration (ms)**	**Gaze duration (ms)**	**Regression path duration (ms)**	**Total fixation duration (ms)**
Semantic acceptability	Chinese speech	221 (1)	221 (1)	227 (2)	395 (5)	287 (3)
	Uyghur speech	218 (1)	218 (1)	222 (1)	356 (5)	266 (2)
	Silence	217 (1)	217 (1)	222 (1)	350 (5)	264 (2)
Non-character detection	Chinese speech	195 (1)	195 (1)	197 (1)	275 (4)	207 (1)
	Uyghur speech	195 (1)	195 (1)	196 (1)	274 (4)	208 (1)
	Silence	195 (1)	195 (1)	197 (1)	278 (4)	207 (2)

Table 4

Output from the liner-mixed effects models, for each of the five local measures. Significant p-values are marked in bold.

	First fixation duration				**Single fixation duration**				**Gaze duration**				**Regression path duration**				**Total fixation duration**			
	b	*SE*	*t*	*p*	*b*	*SE*	*t*	*p*	*b*	*SE*	*t*	*p*	*b*	*SE*	*t*	*p*	*b*	*SE*	*t*	*p*
Intercept(grand mean)	5.32	0.01	382.25	**<0.001**	5.32	0.01	378.58	**<0.001**	5.33	0.01	371.54	**<0.001**	5.70	0.03	220.21	**<0.001**	5.45	0.02	326.60	**<0.001**
task instruction (semantic acceptability vs. non-character detection)	-0.12	0.03	-4.21	**<0.001**	-0.12	0.03	-4.17	**<0.001**	-0.13	0.03	-4.46	**<0.001**	-0.29	0.05	-5.67	**<0.001**	-0.26	0.03	-8.04	**<0.001**
Sound (Chinese vs. Uyghur)	-0.01	0.01	-1.05	0.30	-0.01	0.01	-0.86	0.40	-0.01	0.01	-1.86.	**0.07**	-0.06	0.01	-4.05	**<0.001**	-0.03	0.01	-3.61	**<0.001**
Sound (Chinese vs. silence)	-0.01	0.01	-1.56	0.13	-0.01	0.01	-1.33	0.19	-0.02	0.01	-2.13	**0.04**	-0.06	0.02	-3.71	**<0.001**	-0.04	0.01	-3.89	**<0.001**
Sound (Uyghur vs. silence)	-0.00	0.01	-0.58	0.57	-0.00	0.01	-0.50	0.62	-0.00	0.01	-0.31	0.76	-0.01	0.01	-0.41	0.68	-0.01	0.01	-0.81	0.42
task instruction × Sound (Chinese vs. Uyghur)	0.01	0.01	1.04	0.31	0.01	0.01	1.02	0.31	0.02	0.01	1.34	0.19	0.08	0.03	2.99	**0.004**	0.07	0.02	3.87	**<0.001**
task instruction × Sound (Chinese vs. silence)	0.02	0.01	1.32	0.19	0.02	0.01	1.38	0.17	0.02	0.01	1.41	0.16	0.11	0.03	3.27	**0.002**	0.08	0.02	3.42	**0.001**
task instruction × Sound (Uyghur vs. silence)	0.00	0.01	0.35	0.73	0.01	0.01	0.46	0.65	0.00	0.01	0.18	0.86	0.03	0.03	0.95	0.35	0.00	0.02	0.16	0.87

Table 5

Basic properties of the 492 words in the experimental sentences.

	MIN	***MAX***	***M***	***SD***
Word frequency (counts-per-million)	0.03	12811.05	535.78	1510.98
Number of strokes	1	36	13.54	5.71
Word length (character)	1	4	1.79	0.58

Table 6

Pearson coefficients between word frequency and each local eye-movement measure under two tasks.

		First fixation duration	Single fixation duration	Gaze duration	Regression path duration	Total fixation duration
Semantic acceptability	Word frequency	-.164**	-.137**	-.409**	-.181**	-.475**
Non-character detection	Word frequency	-.304**	-.302**	-.448**	-.308**	-.453**

Note: Word frequency was log-transformed.

**p<.01

Task instruction	Sentences type	Example sentences
Semantic acceptability	Experimental	我和同桌准备这周末去图书大厦购买习题册。 (My deskmate and I are going to the Book Shop to buy exercise books this weekend.)
	Filler	妹妹今天在**体育**课上学习了函数和方程的概念。 (My sister learned the concepts of function and equation in **physical education** class today.)
Non-character detection	Experimental	我和同桌准备这周末去图书大厦购买习题册。 (My deskmate and I are going to the Book Shop to buy exercise books this weekend.)
	Filler	妹妹今天在**数学**果上学习了函数和方程的概念。 (My sister learned the concepts of function and equation in **mathematics** clssa today.)

Figure 1. Example sentences used in the two tasks. The experimental sentences were identical across the two tasks. The different characters in the filler sentences are marked in bold or with underlining (no boldface or underlining of these characters was presented in the formal experiment). In the semantic acceptability task, the filler sentences were implausible. In the non-character detection task, each filler sentence was plausible, but included an orthographically illegal non-character in place of a legal character. For example, the non-character "果 (corresponding to *clssa* in translation)" was created from the legal character "课 (*class*)".