The determinants of auditory distraction during reading: An eye-movement investigation

by

Martin Rachev Vasilev

Submitted in Partial Fulfilment of the Requirements for the
Degree of Doctor of Philosophy

Bournemouth University

September, 2018

Copyright statement

The determinants of auditory distraction during reading: An eye-movement investigation

by

Martin Rachev Vasilev

Abstract

Everyday reading rarely occurs in complete silence. Whether reading at the office, on the way to work, or in a cafeteria, people are often exposed to background sounds such as speech, noise or music that may distract them from their task. While a lot of research has focused on how background sounds affect readers' comprehension, less is known about their influence on the ongoing reading process. The present research investigated the effect of continuous and discrete background sounds on eye-movements during reading in an attempt to find out what makes such sounds distracting and how they affect online reading behaviour. The present investigation started with a meta-analysis of previous findings, which revealed that background speech, noise, and music all have a modest but reliably detrimental effect on reading comprehension. The first two experiments showed that intelligible speech disrupts eye-movements during reading mostly due to its semantic properties, which interfere with extracting the meaning of the text. This disruption was found to occur after the initial lexical processing of words and it resulted in more regressions and more re-reading fixations. However, participant's immediate comprehension of the text remained unaffected. Two further studies suggested that the increase in re-reading behaviour occurs in an attempt to maintain comprehension of the text under such distracting conditions because intelligible speech disrupted comprehension accuracy once participants could not selectively re-read the text. The final experiment showed that discrete deviant sounds also disrupt eye-movements during reading and lead to longer fixation durations when the sound is first heard. However, unlike intelligible speech, this type of distraction was likely due to saccadic inhibition of the oculomotor system. Taken together, the present results demonstrate that eye-movements during reading can reveal subtle auditory distraction effects that may not be detected in measures of comprehension accuracy and that they can give important theoretical insights into their cognitive and oculomotor origin. The findings are discussed in terms of theories of auditory distraction and computational models of eye-movement control during reading.

List of Contents

List of Figures

List of Tables

List of Abbreviations

ADHD: Attention-deficit/hyper-activity disorder

ANOVA: Analysis of variance

BF: Bayes factor

CI-DPA: Confidence interval divergence point analysis

d: Cohen's d (Effect size measure)

ERP: Event-related potentials

ESS: Effective sample size (in Markov Chain Monte Carlo sampling)

FFD: First fixation duration

g: Hedges' g (Effect size measure)

GD: Gaze duration

(G)LMM: (Generalised) Linear Mixed Models

MCMC: Markov Chain Monte Carlo

MMN: Mismatch negativity

RON: Reorientation negativity

RSVP: Rapid serial visual presentation

SFD: Single fixation duration

SPL: Sound pressure level

TMS: Transcranial magnetic stimulation

TVT: Total viewing time

Acknowledgements

First, I would like to thank Bernhard Angele and Julie Kirkby for supervising the work presented in this Thesis. I would also like to thank Daniel Rowan, who helped me create the sound stimuli for my first experiment, and who offered a valuable "outsider" perspective on my research. I am also grateful to Simon Liversedge, who offered constructive criticism and valuable suggestions on some of my experiments. I am greatly indebted to Fabrice Parmentier for sharing his knowledge on deviance distraction with me and for offering many useful suggestions on my final experiment that considerably improved its quality. I am also grateful to Tim Slattery for his feedback during my Transfer Report and beyond. I wish to thank Bournemouth University sincerely for funding my research and for supporting me over the years. I would also like to thank all the participants who kindly took part in my research (including Michelle, Kate, Alex, Ali, Aglaya, and my fellow PhD students at BU for their help with some last-minute data collection). I am also grateful to Martin Tomkins for his technical support during my degree and his seemingly endless supply of hardware laying around. I would also like to thank all PhD students from the office for their friendship and for being part of my PhD journey (especially the "evening crew"- Emmy, Camilo, and Vicky- who offered words of encouragement when I needed them and made sure I never run out of chocolate). I would also like to thank all my friends back home and around the world for keeping in touch despite the distance and my often limited free time. Finally, I would like to thank my family- my father Racho, my mother Tsetsa, my brother Darin, my niece Lori, and my nephew Simeon- for their support and for putting up with my long absences from home.

Author's declaration

Chapter 2 and parts of Chapter 1 have been published as:

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*(5), 567-597. doi: https://doi.org/10.1177/1745691617747398

At the time of submission of this Thesis, Chapters 3-5 were in preparation for resubmission, and Chapter 6 was under a second round of review.

Chapter 3 has been presented as:

Vasilev, M. R., Liversedge, S. P., Rowan, D., Kirkby, J. A., & Angele, B. (2017). *Auditory distraction by meaningful background speech during reading*. Talk presented at the 20th European Conference on Eye Movements, Wuppertal, Germany.

Chapters 3-5 have been presented as:

Vasilev, M. R., Liversedge, S. P., Rowan, D., Kirkby, J. A., & Angele, B. (2017). *Reading is disrupted by intelligible background speech: Evidence from eye-tracking*. Poster presented at the 58th Meeting of the Psychonomic Society, Vancouver, Canada.

CHAPTER 1: INTRODUCTION AND REVIEW OF THE LITERATURE

Reading is an important everyday skill that is indispensable in modern society. Although reading performance is best in silence when no distracting stimuli are present, such ideal conditions are rarely typical for daily life. Rather, much of everyday reading occurs in the presence of external auditory stimulation, such as noise from nearby traffic, music playing in the background, or a colleague talking on the phone. The interest in how auditory stimuli affect human performance is almost as old as modern psychology itself (e.g., Cassel & Dallenbach, 1918; Morgan, 1917). From the widespread use of personal radios among students in the 1940s (Henderson, Crew, & Barlow, 1945; Miller, 1947) to the rise in popularity of the TV (Armstrong, Boiarsky, & Mares, 1991; Cool, Yarbrough, Patton, Runde, & Keith, 1994) and mobile devices (Kallinen, 2002), researchers and educators alike have been interested in whether background sounds can distract students from reading and other study-related tasks.

Over the past eight decades, many studies have examined how experimental exposure to speech, noise, and music may affect reading comprehension. Although some interesting patterns of results have emerged, the research literature has been undermined by a fair number of inconsistent findings and the limited understanding of what properties of irrelevant sounds make them distracting. Additionally, the question of how irrelevant sounds influence the ongoing reading process on a moment-to-moment basis has received surprisingly little attention. While a lot has been learned about how different cognitive and oculomotor processes influence eye-movement control during reading (Rayner, 1998, 2009), most of these studies have been conducted in a quiet environment and thus have not taken into account the potential influence of auditory distractors. As a result, there is a limited understanding of how different auditory distractors affect readers' eye-movement behaviour or how any disruption at the level of eye fixations may be related to cognitive, oculomotor or comprehension processes during reading.

The overarching goal of the present Thesis was to help bridge research on auditory distraction and eye-movement control during reading in an attempt to better understand how auditory distractors influence the ongoing reading process. Therefore, the first and foremost aim of this Thesis was to investigate the effect of irrelevant background sounds on eye-movements during reading. The second aim was to investigate what acoustical and linguistic properties of irrelevant sounds give rise to distraction in eye-movements during reading, and how this distraction may be related to ongoing cognitive, oculomotor, and comprehension processes. The final aim of this Thesis was to consolidate some of the conflicting findings from previous reading comprehension experiments by performing a meta-analysis of the available evidence.

This introductory chapter starts with an overview of previous research on auditory distraction by continuous and discrete irrelevant sounds. Then, a few theories are outlined that could potentially account for auditory distraction effects observed in reading tasks. Afterwards, an introduction to eye-movements during reading is presented, along with the few eye-tracking studies to date that have investigated auditory distraction during reading. Finally, the research questions of this Thesis are outlined and motivated.

1.1. The Effect of Background Noise, Speech, and Music on Reading Comprehension

1.1.1. Background Noise

Background noise can be defined as any unwanted sounds that are not related to the reading task. Strictly speaking, some degree of background noise is always present during reading; however, the intensity of the background noise can vary enormously depending on the environment. A number of epidemiological studies have investigated the relationship between chronic exposure to noise and reading, and have suggested that chronic exposure to traffic noise is associated with lower reading ability in children (see Table 1 for an overview of the main findings). Interestingly, however, only very few studies to date have examined the effect of acute experimental exposure to noise. In one early study, Johansson (1983) found that the reading comprehension and reading speed of 10-year-old children did not differ between quiet conditions and conditions of continuous or intermittent acoustical noise. More recently, Dockrell and Shield (2006) investigated the effect of typical classroom noise (which is quite different from acoustical white or pink noise) on reading comprehension in 8-year-old children. Participants completed the Suffolk Reading Scale in one of three conditions: silence, noise consisting of children’s babble, and the same babble combined with intermittent environmental noise. The results showed that children performed better in the quiet condition than in the babble noise condition. Surprisingly, however, reading performance was best when the babble and the environmental noise were combined. Using a similar type of sound stimuli, Ljung, Sörqvist, and Hygge (2009) found that road traffic noise impaired the reading speed of 12- and 13-year-old children, but not their reading comprehension. However, a condition of children’s babble intermixed with irrelevant speech affected neither measure.

Study	Type of noise	Age /Grade of children	Measure	Main results
Papanikolaou et al. (2014)	Road traffic	4th -5th grade (9-10 years)	A passage with three questions	Children from low-noise schools had significantly better comprehension than children from medium- and high-noise schools
Shield and Dockrell (2008)	Environ-mental noise	2nd grade† (7 years)	Key Stage 1 examination	There was a (mostly) significant, negative correlation between external noise and reading scores (range: r= -.13 to -.43)
Lukas and DuPree (1980)	Road traffic	3rd and 6th grade	Standardized reading achievement test	Increasing noise levels were associated with poorer reading achievement
Bronzaft and McCarthy (1975)	Train traffic	2nd, 4th, and 6th grade	The Metropolitan Achievement Reading Test	Children on the noisy side of the school performed significantly worse than children on the quiet side; this difference corresponded to a reading delay of three to four months
Cohen, Glass, and Singer (1973)	Road traffic (at home)	2nd, 3rd,4th, 5th grade	The Metropolitan Achievement Reading Test	1) Moderate correlation (r= -.26 to -.31) between noise exposure and reading scores 2) Auditory discrimination accounted for 12% of the variance in reading scores
Stansfeld et al. (2005)	Road traffic/ Aircraft noise	9-10 years of age	Nationally standardized reading tests	1) Aircraft noise was significantly associated with an impairment in reading comprehension 2) Road traffic noise did not have a significant effect on reading comprehension
Hygge et al. (2002)	Aircraft noise	8 to 12 years of age	Standardized reading test (German)	1) Closing of an airport diminished the difference on the difficult items between children who were exposed to noise and children who were not exposed 2) Opening of a new airport led to more errors on the difficult items in children newly exposed to noise
Haines, Stansfeld, Brentnall, et al. (2001)	Aircraft noise	4th grade (8-9 years of age)	Suffolk Reading Scale, Level 2	1) Exposure to noise was not significantly associated with reading scores on the whole scale 2) Children from high-noise schools had significantly lower scores on the 15 most difficult items
Haines, Stansfeld, Job, Berglund, and Head (2001)	Aircraft noise	4th and 5th grade (8-10 years)	Suffolk Reading Scale, Level 2	Children from high-noise schools had significantly poorer reading comprehension than children from low-noise schools (~ 6-month reading delay)
Evans and Maxwell (1997)	Aircraft noise	1st and 2nd grade	Woodcock Reading Mastery Test	1) Noise exposure was significantly associated with lower reading scores (r = -.58) 2) Speech perception was a partial mediator
Evans, Hygge, and Bullinger (1995)	Aircraft noise	3rd and 4th grade	Standardized reading test (German)	1) Children from noisy schools made more errors on the reading test 2) The same trend was observed for the most difficult section of the word recognition test
Green, Pasternack, and Shore (1982)	Aircraft noise	2nd through 6th grade	Standardized reading ability test	There were significantly more students reading below their grade level in noisier compared to quieter schools
Cohen, Evans, Krantz, and Stokols (1980)	Aircraft noise	3rd and 4th grade	Standardized school reading test	No evidence that aircraft noise affects reading skills

Table 1. The effect of chronic exposure to noise on reading in children: A summary of the main results.

†The other group of children did not have a reading examination.

Studies of exposure to noise in adults have resulted in similarly mixed findings, sometimes even when done with the same materials (e.g., Martin, Wogalter, & Forlano, 1988, Experiments 4 and 5). While most studies have failed to find an effect of acoustical or environmental noise on reading comprehension (Gawron, 1984; Jahncke, Hygge, Halin, Green, & Dimberg, 2011; Johansson, Holmqvist, Mossberg, & Lindgren, 2012; Veitch, 1990), others have found such an effect after examining the mediating role of personality characteristics, such as introversion and extraversion (Furnham, Gunter, & Peterson, 1994; Ylias & Heaven, 2003). In summary, studies investigating the effect of background noise on reading comprehension have yielded inconsistent results, although some of them suggest that exposure to noise may be detrimental.

1.1.2. Background Speech

A specific kind of noise that often occurs in daily life is background speech. Compared to environmental and acoustical noise, background speech has specific acoustic properties that make it salient to listeners. Additionally, if the background speech is intelligible, it also carries semantic information (completely unintelligible background speech might also occur, but it is not very frequently encountered unless one is in a foreign country and does not understand the language). Perhaps owing to its semantic content, background speech is often rated as more distracting and more annoying than acoustical noise (Haapakangas et al., 2011; Haka et al., 2009; Landström, Söderberg, Kjellberg, & Nordström, 2002). Consistent with this subjective perception, intelligible background speech has been found to disrupt reading comprehension in a number of experiments (Armstrong, Boiarsky, & Mares, 1991; Baker & Madell, 1965; Martin et al., 1988; Sörqvist, Halin, & Hygge, 2010; however, see Venetjoki, Kaarlela-Tuomaala, Keskinen, & Hongisto, 2006).

Additionally, there is some evidence to suggest that this disruption effect may be larger for adult participants who have a poorer ability to immediately suppress the irrelevant background speech (Sörqvist et al., 2010; Sörqvist, Ljungberg, & Ljung, 2010).

A specific reading task that has been investigated in more detail in connection with background speech is proofreading. Proofreading is an important part of many professions, especially those related to teaching and publishing. Proofreading is a more cognitively demanding task than reading alone because it also requires allocating attention to look for mistakes, in addition to reading the text. There are generally two types of mistakes that have been investigated in proofreading studies: contextual mistakes that require understanding the meaning of the text to detect (e.g., problems with pronoun agreement), and non-contextual (i.e., spelling) mistakes that require only processing of the current word to detect. Due to the semantic content of intelligible speech, it can be hypothesized that background speech would disrupt the detection of contextual errors more than the detection of non-contextual errors.

Some support for this prediction was found in an early study by Weinstein (1977) who reported that background speech consisting of a radio news report significantly impaired the detection of contextual, but not the detection of non-contextual errors. However, Jones, Miles, and Page (1990) found exactly the opposite effect in another study. The authors manipulated both the intelligibility of background speech (which was played either normally or in reverse) and the intensity of the sound (50 vs 70 dB (A)). They found that the intensity of the sound did not affect proofreading performance, but that normal (i.e., intelligible) speech reduced the number of non-contextual errors that were detected. Critically, however, the intelligibility of speech did not affect the detection of contextual errors (Jones et al., 1990). More recently, Venetjoki et al. (2006) found that background speech reduced the overall accuracy on a similar proofreading task compared to continuous noise. However, even though the task included both contextual and non-contextual errors, there was no significant effect of background speech on either error type in isolation. In a similar study, Landström et al. (2002) found that background speech did not affect proofreading performance for either contextual or non-contextual errors compared to a condition of broadband noise (i.e., noise consisting of a wide range of frequencies). The auditory stimuli were presented at a comparable sound intensity level to Venetjoki et al. (2006), although the speech consisted of random spoken statements. Finally, Smith-Jackson and Klein (2009) also found no effect of background speech (intermittent or continuous) on overall proofreading accuracy.

Interestingly, a few studies have also suggested that the detrimental effect of background speech on reading and proofreading performance can be diminished by making the task harder and thus increasing participants' engagement with it (Halin, 2016; Halin, Marsh, Haga, Holmgren, & Sörqvist, 2013; Halin, Marsh, Hellman, Hellström, & Sörqvist, 2014). In a few experiments, Halin et al. showed that performance on a reading/ proofreading task was disrupted by background speech only when the text was formatted in a familiar font, but not when it was formatted in an unfamiliar (i.e., harder to read) font. Similarly, performance was disrupted only when the text was printed normally, but not when it was visually degraded (i.e., harder to read). Therefore, these results suggest that increasing task engagement may decrease the detrimental effect of background speech on reading comprehension and proofreading accuracy (see Sörqvist & Marsh, 2015, for a discussion).

Most studies that were considered so far have investigated only the end product of reading and proofreading (i.e., comprehension accuracy, proofreading accuracy, or the overall time taken to read the text). However, these studies do not tell us how the reading process is influenced on a moment-to-moment basis. More recently, several eye-tracking studies have addressed this question by showing that the effect of intelligible background speech on reading can also be found at the level of fixation durations and fixations probabilities (Cauchard, Cane, & Weger, 2012; Hyönä & Ekholm, 2016; Yan, Meng, Liu, He, & Paterson, 2017). One key finding from these studies is that background speech leads to an increase in the number of re-reading fixations (discussed in greater detail below). While these studies have been successful in explaining when disruption by background speech occurs during the reading process, one puzzling aspect is that none of the eye-tracking experiments have replicated the disruption effect in comprehension accuracy found in behavioural studies. It is currently not known why this inconsistency exists, but this raises questions about how reliable the effect of background speech on reading comprehension is.

In summary, background speech has been found to disrupt reading comprehension and proofreading accuracy in a number of experiments. Additionally, the available evidence suggests that this disruption is due to processing the semantic information of the speech sound (e.g., Martin et al., 1988). These effects appear to be more reliable than the effect of non-speech noise on reading, which has not been consistently replicated. Nevertheless, several recent studies have found no effect of background speech on reading comprehension, which casts doubt on its robustness and generalizability.

1.1.3. Background Music

Unlike noise and speech, which are usually a nuisance, playing music in the background is often done deliberately as a personal choice or a habit. Interest in the potential effect of background music on reading started in the first half of the 20th century with the popularity of personal radios and record players, and their use by students. However, these early studies did not paint a clear picture of the relationship between background music and reading. While some of them found that music can negatively impact reading comprehension in children and university students (Fendrick, 1937; Fogelson, 1973; Henderson, Crew, & Barlow, 1945), others found that background music either does not affect reading at all (Freeburne & Fleischer, 1952; Miller, 1947; Mitchell, 1949) or that it actually improves reading performance (Hall, 1952). Indeed, this controversy has persisted until the present day and even two out of the three eye-tracking studies to address this question (Cauchard et al., 2012; R. Johansson et al., 2012; Zhang, Miller, Cleveland, & Cortina, 2018) have failed to find any effect of background music on fixation durations or fixation probabilities.

To examine what conditions may give rise to distraction, some studies have investigated whether the effect of background music on reading comprehension is modulated by personality traits (Avila, Furnham, & McClelland, 2011; Furnham & Allass, 1999; Furnham & Bradley, 1997; Furnham & Stephenson, 2007; Furnham & Strbac, 2002; Furnham, Trew, & Sneade, 1999; Kou, McClelland, & Furnham, 2018). Based on Eysenck's (1967) theory of personality, these studies have predicted that individuals high in extraversion will be distracted less by background music than individuals high in introversion due to the extraverts' higher cortical arousal threshold. However, the results from these studies have been mixed. While some of them have found such an interaction between personality traits and background music (Daoussis & Mc Kelvie, 1986; Furnham & Bradley, 1997; Furnham & Strbac, 2002), others have not (Avila et al., 2012; Furnham & Allass, 1999; Furnham et al., 1999; Furnham & Stephenson, 2007; Kou et al., 2018). A number of factors may have contributed to these inconsistencies, such as the way in which participants were classified as introverts and extraverts, or the small sample size in some of the studies.

Another factor that has been considered is the genre of the music (Kallinen, 2002; Miller & Schyb, 1989; Mulliken & Henk, 1985; Tucker & Bushman, 1991). However, as the popularity of music genres changes with time, it is arguably better to investigate what aspects of the music may cause distraction. One factor that may play a role is participants' preference for the music. For example, Etaugh and colleagues (Etaugh & Michals, 1975; Etaugh & Ptasnik, 1982) reported that preferred music decreased reading comprehension scores, but only for students who rarely study while listening to music. In contrast, Johansson et al. (2012) found that participants had lower comprehension accuracy when listening to non-preferred music compared to a quiet control condition, but there was no such effect when they listened to preferred music. Additionally, they did not replicate the previous finding that participants' studying habits modulated the results. Adding further to the confusion, Perham and Currie (2014) found that preferred and non-preferred lyrical music (i.e., music with sung lyrics) is equally disruptive to reading comprehension, although they did not report data on students' studying habits.

The influence of background music on reading may also be modulated by the acoustic properties of the music. Some factors that have been considered are its informational load (Kiger, 1989), loudness and tempo (Thompson, Schellenberg, & Letnic, 2012), familiarity of the music to participants (Hilliard & Tolin, 1979) and its capability to induce a startle response (Ravaja & Kallinen, 2004). These results are quite interesting in terms of understanding what types of music may cause distraction, although they would benefit from further replication and extensions. In summary, previous studies suggest that certain types of music may be distracting, but a negative effect of background music on reading performance has not been consistently observed.

To summarise the discussion so far, the available evidence suggests that experimental exposure to background noise, speech, and music may disrupt reading performance. The effect of background noise and music appears to be less consistent, with many studies reporting non-significant effects on reading comprehension. Although the effect of background speech on reading appears to be more reliable, several recent experiments have also failed to find an effect in reading comprehension and proofreading tasks. Therefore, considerable uncertainty exists with respect to the magnitude of these distraction effects and what aspects of background sounds may be responsible for them.

1.2. Auditory Distraction by Deviant Sounds

While the available research on auditory distraction during reading has focused on continuous background sounds such as speech, noise, or music, little is known about how discrete sounds that involve subtle auditory changes may affect reading performance. Interestingly, there is a large body of evidence indicating that the brain automatically responds to changes in auditory stimulation that reach a certain threshold- a finding known as the *mismatch negativity* (MMN) effect (Näätänen, 1995; Näätänen, Gaillard, & Mäntysalo, 1978; Näätänen, Paavilainen, Rinne, & Alho, 2007; Näätänen & Michie, 1979). The MMN usually occurs when participants are presented with a block consisting of the same repeated sound (known as the “standard”), which is then occasionally replaced by an acoustically deviant sound (known as the “deviant”; Näätänen, 1995). The MMN is measured as a negative component of the event-related potential (ERP) that occurs some 150-200 ms after the onset of acoustical deviance (Näätänen et al., 2007; Tiitinen, May, Reinikainen, & Näätänen, 1994). This component is thought to reflect a pre-attentive mechanism for detecting auditory changes in the human brain (Berti & Schröger, 2001).

In addition to the MMN, a positive P3a component is also typically observed in response to deviant sounds, which peaks at around 300 ms after the deviant sound's onset (Horváth, Winkler, & Bendixen, 2008). The P3a component is thought to reflect the involuntary shift of attention towards the deviant sound (Berti & Schröger, 2001; Escera, Alho, Schröger, & Winkler, 2000; Schröger & Wolff, 1998a, 1998b). Finally, the P3a component can also be followed by a reorientation negativity (RON) component that peaks around 400-600 ms after the deviant sound's onset. The RON is in turn thought to reflect the refocusing of attention back to the main task (Berti, 2008; Schröger, Giard, & Wolff, 2000; Schröger & Wolff, 1998a).

Deviant sounds not only elicit specific electrophysiological responses in the brain, but they also give rise to behavioural distraction in the main task (Berti & Schröger, 2001; Escera, Alho, Winkler, & Näätänen, 1998; Parmentier, 2014; Schröger, 1996; Schröger & Wolff, 1998b). Such behavioural distraction is usually shown by increased reaction times in response to deviant sounds in tasks where participants need to categorize target stimuli, such as judging the parity of numbers presented on the screen (e.g., Leiva, Parmentier, & Andrés, 2015; Parmentier, Elford, Escera, Andrés, & Miguel, 2008; Parmentier, Elsley, Andrés, & Barceló, 2011) or judging the duration of the task-irrelevant sound (e.g., Berti & Schröger, 2001; Schröger & Wolff, 1998a, 1998b). Interestingly, behavioural distraction by deviant sounds does not appear to be a mere by-product of the respective electrophysiological responses, as it can vary substantially without an associated variance in the ERP components, and vice versa (Parmentier, 2014). This suggests that not all acoustical changes that are detected by the brain obligatorily cause behavioural distraction in reaction times or response accuracy measures.

Evidence from behavioural studies has shown that deviant sounds are distracting not because of their acoustical novelty *per se*, but rather because they violate the cognitive system's predictions (Parmentier et al., 2011; Vachon, Hughes, & Jones, 2012; see also Bubic, von Cramon, Jacobsen, Schröger, & Schubotz, 2009). In fact, deviant sounds fail to elicit behavioural distraction if their occurrence is entirely predictable- for example, when it is signalled by the appearance of a visual cue before the presentation of the sound (Horváth, Sussman, Winkler, & Schröger, 2011; Sussman, Winkler, & Schröger, 2003). Interestingly, the mechanism underlying deviance distraction does not appear to be restricted only to the auditory modality. For example, Parmentier, Ljungberg, Elsley, and Lindkvist (2011) found that unexpected vibro-tactile stimuli also yield behavioural distraction in categorization tasks much in the same way that acoustical deviant stimuli do. This suggests that the neural mechanism for detecting sensory changes may operate cross-modally. However, despite the potential relevance of deviance distraction in everyday life, no studies to date have investigated the effect of deviant sounds on reading.

1.3. Theories of Auditory Distraction

Over the last several decades, a number of theoretical accounts have attempted to explain the influence of irrelevant background sounds on cognitive performance. While most of them have been developed in simpler laboratory tasks (e.g., serial recall of items), they nevertheless make useful predictions that can also be applied to more complex cognitive tasks such as reading. Much of the impetus for the development of some of the early theories of auditory distraction was the original report of the *irrelevant speech effect* in serial recall memory (Colle & Welsh, 1976). This effect refers to the now classical finding that serial recall memory for visually presented items (e.g., digits) is reduced when participants listen to irrelevant speech that they are told to ignore compared to a baseline condition of silence (Colle, 1980; Ellermeier & Zimmer, 1997; Hughes & Jones, 2001; Jones, Madden, & Miles, 1992; LeCompte, 1995; Miles, Jones, & Madden, 1991; Neely & LeCompte, 1999; Salamé & Baddeley, 1982).

One of the earliest theoretical accounts to try to explain this finding in terms of auditory distraction is the *phonological interference* hypothesis. This account is based on Baddeley and Hitch's (1974, 1994) model of working memory, in which the phonological loop acts as an acoustic store where memories are registered and rehearsed through a process of sub-vocalization. Salamé and Baddeley (1982, 1987, 1989) reported a series of experiments in which they showed that memory for visually presented digits is impaired by unattended speech, but not by unattended acoustical noise. Additionally, a distraction effect was observed even if the speech sound was in a language that participants could not understand (Salamé & Baddeley, 1987; see also Colle & Welsh, 1976). The authors argued that this occurs because speech sounds automatically gain access to the phonological loop and thus interfere with the encoding and rehearsal of visually presented items. Even though this hypothesis is derived from a memory task, Salamé and Baddeley (1989) argued that a similar disruption may also be observed in more complex cognitive tasks such as reading.

Martin et al. (1988) were first to systematically test the phonological interference hypothesis in a reading comprehension task. In a series of experiments, they found that the disruptive effect of unattended speech was due to the semantic properties (i.e., meaning) of the speech, rather than its phonological features. More specifically, the authors found that

English speech (intelligible to participants) was more distracting that Russian speech (unintelligible to participants). Similarly, a continuous speech stream of random words was found to disrupt comprehension more than a continuous speech stream of non-words. To account for these results, Martin et al. (1988) argued that, unlike serial recall tasks, reading comprehension requires understanding the meaning of the text. Therefore, the semantic properties of the irrelevant speech can interfere with building the semantic representations of the text that is being read. This prediction will be referred to as the *semantic interference* hypothesis.

The *changing-state* hypothesis (Hughes & Jones, 2001; Jones & Macken, 1993; Jones, Madden, & Miles, 1992) is another prediction that is also derived from serial recall tasks. According to this hypothesis, interference is caused by background sounds that exhibit considerable acoustic variation, but not by steady-state, aperiodic sounds that do not have such variation (Jones et al., 1992). For example, a sound consisting of different consonants (e.g., “B, F, P, S, N”) should cause more interference than a sound made up of the same consonant (e.g., “M, M, M, M, M”) because it exhibits more acoustic variation. The hypothesized mechanism through which interference occurs is that changing-state sounds contain information about the serial order of their constituent sound elements (Hughes & Jones, 2001). This information can then interfere with maintaining the serial order of items in a memory task.

Although reading is a more complex cognitive task, it also involves maintaining the order of words in the sentence, as well as their syntactic relations. For example, since models of parallel word processing such as SWIFT (Engbert, Nuthmann, Richter, & Kliegl, 2005) assume that readers can process multiple words at the same time, they also have to assume, at least implicitly, that readers are somehow able to maintain information about the order of these words in the current sentence. Additionally, some models of reading comprehension (e.g., Kintsch, 1998) assume that word meanings are combined to form propositions or “idea units” according to their syntactic relationships (Kintsch & Rawson, 2005). Forming these units must also involve establishing and keeping track of the order of words in the sentence, as well as their syntactic relationships.

A final account that is relevant in a reading task is the *duplex theory* of auditory distraction (Hughes, Vachon, & Jones, 2005, 2007; Hughes, 2014; Sörqvist, 2010). According to this theory, auditory distraction can occur from two different processes: *interference-by-process* and *attentional capture* (Hughes, 2014). Interference-by-process (Marsh, Hughes, & Jones, 2008, 2009; Marsh & Jones, 2010) occurs when the background sound interferes with a process that is important for the main task. For example, in a reading task, the semantic processing of meaningful speech would interfere with the task because reading also requires semantic processing to extract the meaning of the text. Alternatively, auditory distraction can also be caused by attentional capture (Hughes et al., 2005; Vachon, Hughes, & Jones, 2012) where attention is temporally directed away from the main task by a sound that deviates from a repeated sequence of sounds. For example, the sound “B” in the sequence “AAAAAA**B**A” would capture attention because another “A” is expected in the sequence (Hughes, 2014). This proposition is motivated by the involuntary switch of attention and the corresponding P3a ERP component evoked by deviant sounds that have been extensively documented in the previous literature (see above).

In a reading task, the interference-by-process part of the duplex theory makes the same prediction as the semantic interference hypothesis by Martin et al. (1988) discussed earlier. The difference between the two accounts is very subtle: In Marsh et al.'s (2008, 2009) account, distraction occurs because processing the meaning of the background speech depends on the same process used for extracting the meaning of the text that is being read. In contrast, Martin et al. (1988) assume that it is the semantic properties of the speech that cause the interference. These two very similar views are difficult to disentangle empirically, and since they make the same general prediction in the present research, we will consider them together as theories of semantic distraction. The second part of the duplex theory- attentional capture- is a very interesting concept that is yet to be tested in more complex and ecologically-valid tasks such as reading. However, as the principles underlying deviance distraction and the involuntary capture of attention are often assumed to be broadly applicable to different tasks, subject populations, and methodologies (see Näätänen et al., 2007), similar type of distraction should also be observed in a reading task.

1.4. Eye-movements During Reading: An Introduction

The advances in eye-tracking technology in the past several decades have made it possible to record eye-movements accurately and with a high temporal precision while participants are reading a text. In the present research, the Eyelink 1000 system was used to record participants' eye-movements (SR Research Ltd., Ontario, Canada). The typical eye-tracking setup that was used in the present research is illustrated in Figure 1a. The Eyelink 1000 consists of a high-speed camera and an infrared illumination module that emits 890 nm infrared light directed at the eye. The emitted infrared light is reflected by the eye and this reflection is then used by the high-speed camera to calculate the gaze position of the eye in real time. As illustrated in Figure 1b, the camera uses the centre of the pupil and the corneal reflection (first Purkinje image) to track the position of the eye on the screen. Because the distance between the corneal reflection and the centre of the pupil changes with pure eye rotation (provided that the head is kept relatively stable), video-based eye-trackers such as the Eyelink 1000 can record movements of the eye after an initial calibration is performed (Duchowski, 2007). The standard Eyelink 1000 system can perform monocular tracking of the eye at a sampling rate of 1000 Hz (i.e., a new sample of the eye's position is obtained every millisecond) and has an average accuracy of 0.25 - 0.5 º per visual angle.

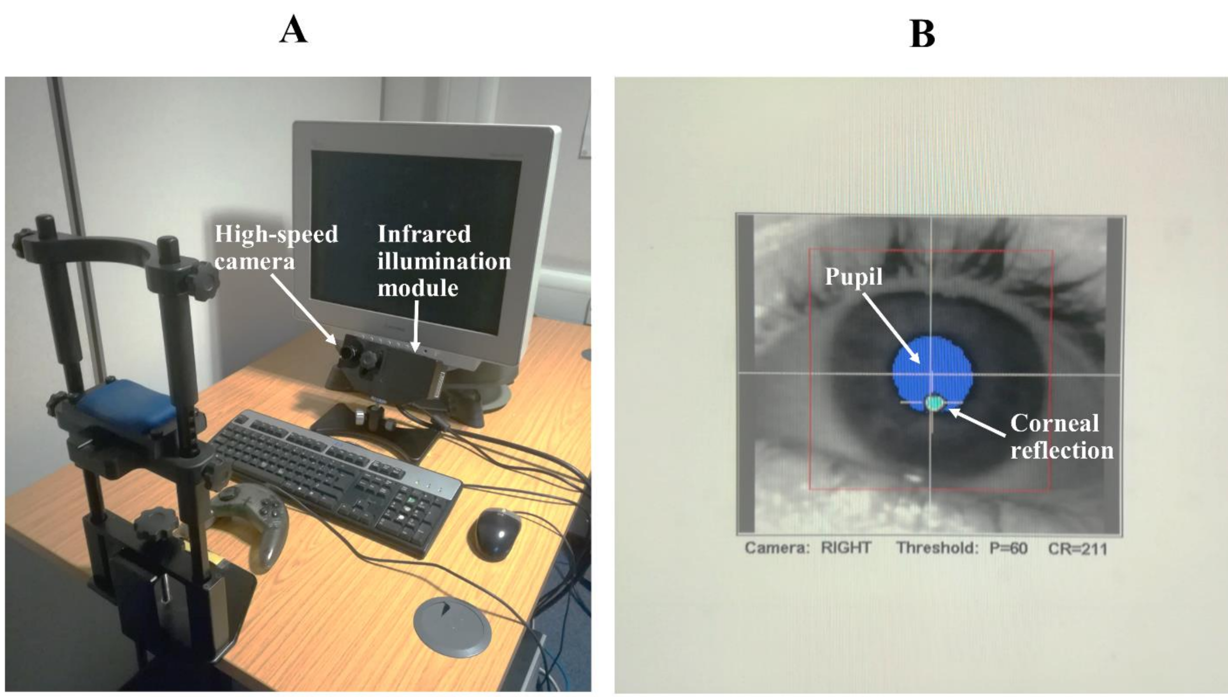

Figure 1. An illustration of the eye-tracking equipment used in the present research. Panel **A** shows the position of the eye-tracker and the rest of the equipment that was used in the experiments (with the exception of Chapter 6, where a Tower Mount set-up of the Eyelink 1000 was used). Panel **B** shows an image recorded by the high-speed camera while tracking the centre of the pupil and the corneal reflection of the right eye.

When reading a text, the eyes do not move smoothly along the page but rather alternate between quick jump-like movements (known as *saccades*) and short periods of relative stability (known as *fixations*). Because visual acuity decreases rapidly away from the point of fixation, readers need to fixate individual words in the text in order to bring them into the centre of vision for processing. Visual acuity is good in the *fovea*, or the central 2 º of visual angle, but it decreases steadily in the *parafovea* (2 to 5 º of visual angle) and is even poorer in the periphery (>5 º of visual angle; Rayner, 1998). While fixating words in foveal vision is crucial for their recognition, there is now a large body of evidence indicating that readers can also pre-process the upcoming word in parafoveal vision (Rayner, 1975, 1998, 2009; Schotter, Angele, & Rayner, 2012; Vasilev & Angele, 2017). When readers have a valid parafoveal preview of the upcoming word, fixation times are shorter once this word is subsequently fixated compared to when parafoveal preview is denied- a finding known as the *preview benefit* effect (Rayner, 1998). This benefit represents a processing advantage that is thought to give the cognitive system a head start in recognising the upcoming word in order to meet the neurological and oculomotor constraints imposed by saccadic eye-movements (Rayner & Reingold, 2015; Reichle & Reingold, 2013).

The length of saccades during reading usually depends on the number of characters in the text that the eye travels rather than the actual distance in terms of degrees per visual angle (Morrison & Rayner, 1981; O’Regan, 1983). The average saccade during reading usually spans about 8 characters, although individual saccades can vary considerably and are often in the range of 2 to 18 characters (Rayner, 1978). While most saccades during reading are progressive and directed towards unexplored text, the eyes also occasionally go back to revisit previously-read words. These so-called *regressions* occur on 10-15 % of all saccades and are often related to text comprehension difficulties or incomplete processing of previous words in the sentence (Rayner, 1998).

When a word is fixated, the eyes most frequently land a little to the left of its centre- a finding known as the *preferred viewing location* effect (Rayner, 1979). This type of landing distribution may occur because readers aim for the centre of the word but undershoot this location due to saccadic range error (McConkie, Kerr, Reddix, & Zola, 1988). Readers also do not fixate every word in the text and around 15 % of all content words and 65 % of all function words are skipped (i.e., not fixated) during reading (Rayner, 2009). One factor that influences the fixation probability of words is their length. Longer words are less likely to be skipped and more likely to receive multiple fixations compared to shorter words (Kliegl, Grabner, Rolfs, & Engbert, 2004; Rayner & McConkie, 1976; Rayner, 1979). Therefore, unsurprisingly, longer words are also fixated longer due to the increase in re-fixation probability (Kliegl et al., 2004).

However, eye-movements during reading are influenced not only by low-level visual characteristics of the words, but also by their psycholinguistic properties. For example, readers can use the preceding sentence context to predict upcoming words in the text. This is demonstrated by the finding that more predictable words are more frequently skipped compared to less predictable words (Balota, Pollatsek, & Rayner, 1985; Rayner, Li, & Juhasz, 2005; Rayner, Slattery, Drieghe, & Liversedge, 2011; Rayner & Well, 1996). Additionally, in cases when more predictable words are fixated, they tend to receive shorter fixation durations compared to less predictable words (e.g., Kliegl, Nuthmann, & Engbert, 2006; Rayner & Well, 1996). Furthermore, fixation durations during reading are also sensitive to the lexical properties of words. More specifically, words with lower lexical frequency are fixated longer on average compared to words with higher lexical frequency (Inhoff & Rayner, 1986; Rayner & Duffy, 1986; Schilling, Rayner, & Chumbley, 1998). Similarly, fixation durations are also sensitive to the age of acquisition of words (Juhasz & Rayner, 2003, 2006), and this does not appear to be a mere consequence of cumulative lexical frequency effects (Juhasz, 2005; Juhasz & Rayner, 2006). Therefore, these findings clearly demonstrate that eye-movements during reading are sensitive to the ongoing cognitive processing of words in the text.

The detailed understanding of how cognitive and visual factors influence the reading process has led to the development of advanced computational models of eye-movement control during reading that can explain many of the empirical findings. The E-Z Reader model (Reichle, Pollatsek, Fisher, & Rayner, 1998; Reichle, Rayner, & Pollatsek, 2003; Reichle, Warren, & McConnell, 2009; Schotter, Reichle, & Rayner, 2014) is one such family of models which assumes that attention is allocated in a serial manner and that readers process the sentence one word at a time. In this model, word processing starts with a pre-attentive visual stage which reflects the time needed to propagate the visual features of the word from the retina to the brain (Reichle et al., 2009). This early stage is then followed by two lexical processing stages: familiarity check (L1) and lexical access (L2). The time needed to complete the two lexical stages is a function of the lexical frequency of the word and its predictability given the preceding sentence context.

In this model, completion of the familiarity check stage initiates the programming of the next saccade because word recognition (i.e., completion of the lexical access stage) is imminent and likely to happen soon (Reichle et al., 1998). Therefore, the E-Z Reader model assumes that eye-movements are under direct cognitive control because this initial stage of lexical processing triggers the programming of the next saccade (Rayner & Reingold, 2015). While the next saccade is being programmed, the model assumes that attention is shifted covertly to the upcoming word (see Posner, 1980, 2016), which allows readers to pre-process that word before it is directly fixated. The programming of the next saccade also happens in two stages: a labile stage (M1), which can be cancelled by other saccadic programs, and a non-labile stage (M2), which can no longer be cancelled. A more recent version of the E-Z Reader model (Reichle et al., 2009) has also implemented a post-lexical integration stage (I) that aims to provide a framework for modelling the effect of higher-level language processing on eye-movements during reading.

SWIFT is another computational model of eye-movement control during reading which assumes that attention is distributed in parallel to a few words at a time according to an attentional gradient (Engbert, Longtin, & Kliegl, 2002; Engbert et al., 2005; Risse, Hohenstein, Kliegl, & Engbert, 2014; Schad & Engbert, 2012). In this model, word processing starts as soon as words fall within a spatially distributed activation field that extends on either side of the current point of fixation (but that is assymetrically smaller to the left; Engbert et al., 2005). Similar to the E-Z Reader model, lexical processing is also carried out in two stages: first, word activation increases from zero until it reaches a maximum specified by the word's processing difficulty; and second, activation then decreases from this maximum until it reaches again zero (Risse et al., 2014). In SWIFT, the word's processing difficulty is estimated entirely from its lexical frequency. Word predictability also has an influence on the processing rate of words. However, as the predictability of words is independent from the visual input, its influence is decoupled from that of lexical frequency (Engbert et al., 2005).

Because a few words at a time can be processed in parallel in the SWIFT model, the target of the next saccade is determined through a competition of words with different lexical activations- the higher the activation of a word, the more likely it is that it will be selected as the next saccade target (Engbert et al., 2002, 2005). Additionally, the programming of the next saccade also happens in a labile and a non-labile stage. However, unlike the E-Z Reader model, saccadic programming is not directly determined by the cognitive processing of words but, instead, is initiated by a random saccadic timer. Nevertheless, this timer can be foveally inhibited by the processing difficulty of the currently-fixated word, which allows the model to account for effects of cognitive processing on eye-movements (Engbert et al., 2005). As such, SWIFT implements a mechanism of indirect cognitive control of eye-movements where the ongoing processing of words can modulate, but does not determine eye-movements.

The above-mentioned models have mostly implemented empirical effects at the level of individual words, such as lexical frequency and predictability, which are typically investigated in a single-sentence reading paradigm. However, there is also some research that has investigated how readers process higher-level language representations and larger pieces of text (see Rayner, Raney, & Pollatsek, 1995; Staub & Rayner, 2007). For example, when reading longer texts, participants spend less time during the initial, first-pass reading, but more time during the second-pass re-reading of the text compared to reading single unconnected sentences (Radach, Huestegge, & Reilly, 2008). This suggests that participants can adapt their reading strategy to the format of the text. Additionally, when participants are asked to read the same passage twice, they are about 9-14 % faster to do so the second time (Hyönä & Niemi, 1990; Rayner et al., 1995; see also Kaakinen & Hyönä, 2007). This speed-up is mostly due to participants' making fewer and shorter fixations during the second reading of the text (Raney & Rayner, 1995). Finally, when participants are asked to proofread the text stimuli, rather than the typical instruction to read them for comprehension, they make shorter saccades, have longer first-pass reading, are more likely to re-fixate words, and show stronger lexical frequency effects (Kaakinen & Hyönä, 2010; Schotter, Bicknell, Howard, Levy, & Rayner, 2014). All these findings suggest that the format and the instructions of the reading task can modulate the allocation of attention and participants' eye-movement behaviour.

1.5. Auditory Distraction during Reading: Evidence from Eye-movements

There are only a few studies to date that have investigated the effects of background speech and acoustical noise on eye-movements during reading. In the first study of this kind, Johansson, Holmqvist, Mossberg, and Lindgren (2012) recorded participants' eye-movements while they read texts in four auditory sound conditions: preferred music, non-preferred music, background sounds recorded from a café, and silence. The authors found that none of the sound conditions influenced fixation durations or fixation probabilities during reading compared to the baseline of silence.

In a similar study, Cauchard, Cane, and Weger (2012) investigated the effect of instrumental music and background speech consisting of a radio programme on eye-movements during paragraph reading. In addition to the sound manipulation, the study also had an interruption condition in which the text disappeared for 60s on half of the trials when participants fixated a target word in the paragraph. During the interruption time, an unrelated audio news story was played. Cauchard et al. found that instrumental music generally had no effect on participants' eye-movements during reading. In contrast, intelligible background speech led to longer gaze durations, longer reading and re-reading times, and more fixations compared to silence. However, because participants' reading was interrupted by the unrelated audio news story, this may have influenced their reading behaviour in this study. For example, the interruption by the unrelated audio story may have acted as a secondary source of distraction that could have amplified the disruption in response to the (primary) unrelated speech in the experiment. Therefore, the auditory disruption effects observed in this study may have been confounded by this interruption manipulation.

More recently, Hyönä and Ekholm (2016) reported a series of experiments that investigated how background speech affects reading of syntactically complex sentences. In Experiment 1, they found that listening to intelligible speech (Finnish) did not result in significantly longer fixation durations compared to either speech in an unfamiliar language (Italian) or silence. In this sense, the authors did not find evidence for the phonological disruption hypothesis. In the remaining three experiments, Hyönä and Ekholm found that scrambled Finnish speech is more disruptive than both silence and normal, non-scrambled speech. The scrambled Finnish speech was created by randomizing the order of words in the text and reading them aloud with an intonation that resembles that of coherent speech. Interestingly, the disruption by background speech in their experiments was most consistently shown by an increase in sentence re-reading time (Hyönä & Ekholm, 2016).

Hyönä and Ekholm (2016) also found that scrambled speech created from the to-be-read text was not more distracting than scrambled speech created from an unrelated text. Additionally, scrambled speech from an unrelated text that was semantically, but not syntactically, anomalous was not more distracting than scrambled speech that was both semantically and syntactically anomalous. These results point to two conclusions. First, they suggest that scrambled speech is disruptive not because of similarity in semantic content between the speech and the text, but rather because both sources of information are calling on the same semantic processes for analysing meaning (Hyönä & Ekholm, 2016). Second, they also suggest that the syntactic anomaly of scrambled speech does not *per se* make it more distracting to readers.

In a similar study, Yan, Meng, Liu, He, and Paterson (2017) investigated distraction effects by background speech in readers of Mandarin Chinese. Participants read single sentences with a target word lexical frequency manipulation in three background sound conditions: intelligible (i.e., Mandarin) speech, meaningless speech (the same speech scrambled in 60 ms segments), and silence. The scrambling method used in this study did not leave the individual words intact as in Hyönä and Ekholm's (2016) experiments, but it preserved the general acoustic variation that is present in normal speech. Yan et al. found that intelligible speech resulted in longer reading times, more fixations, and more regressions compared to both meaningless speech and silence. Additionally, the otherwise ubiquitous lexical frequency effect was eliminated in the two speech conditions, but only for the first fixation duration on the target word. This suggests that background speech may have a very early influence on the language processing system by delaying access to the lexical representation of words.

Finally, Zhang et al. (2018) investigated the effect of background music on eye-movements while reading academic-level passages. In Experiment 1, participants brought their own music to the lab, while in Experiment 2 all participants listened to the same song segments from popular music pieces that were selected by the experimenter. The music segments in Experiment 2 consisted of only the first verse and the first chorus of the song, with each segment separated by a 1-2 s silence gap from the next segment. The results showed that the first-pass reading of the text, as measured with gaze durations, was not significantly affected by the music condition compared to the silence baseline. However, background music had a significant effect on the word length by lexical frequency interaction, which the authors took as evidence that the irrelevant music may have impaired the sub-lexical processing of low-frequency words. Additionally, consistent with previous studies on the effect of intelligible speech on eye-movements (Cauchard et al., 2012; Hyönä & Ekholm, 2016; Yan et al., 2017), background music also led to longer total viewing time, text re-reading time, and greater regression probability. This suggests that background music may also lead to a similar increase in re-reading behaviour as intelligible background speech. Finally, Zhang et al.'s (2018) Experiment 2 suggested that the disruption in re-reading measures is exacerbated for a short period of time after a new song has started playing.

In summary, the available eye-tracking studies of auditory distraction during reading have provided the first important evidence regarding how task-irrelevant sounds may affect the reading process. First, they suggest that intelligible speech disrupts the ongoing reading process- this disruption may start as early as the lexical processing of individual words (Yan et al., 2017), but it also appears to continue in the post-lexical stages when it leads to an increase in re-reading fixations (Cauchard et al., 2012; Hyönä & Ekholm, 2016; Yan et al., 2017). Second, background noise from a café does not appear to affect fixation durations or fixation probabilities during reading (R. Johansson et al., 2012), which points to the fact that not all irrelevant sounds obligatorily cause distraction. Finally, there is mixed evidence regarding whether background music may also disrupt eye-movements during reading. While earlier studies have generally failed to find evidence for such disruption (Cauchard et al., 2012; R. Johansson et al., 2012), Zhang et al.'s (2018) recent data raise the possibility that music may have a similar disruption effect to that of intelligible speech.

1.6. Scope and Research Questions of this Thesis

The available literature has yielded many important results about when task-irrelevant sounds can disrupt reading efficiency. However, due to the mixed findings from reading comprehension experiments, it is not known whether background speech, noise, and music can all reliably disrupt participants' comprehension. Additionally, as most previous studies have only investigated the final product of the reading process (i.e., comprehension), less is known about how background sounds influence the ongoing reading process. While a small number of eye-tracking studies have provided the first clues as to how irrelevant sounds may disrupt eye-movements during reading, it is currently not well understood how these sounds impact the oculomotor or cognitive processing systems. Furthermore, the mixed findings from previous experiments have made it challenging to assess what linguistic or acoustical properties of irrelevant sounds are responsible for the observed distraction. It is also not immediately clear how the disruption in eye-movement measures may be related to comprehension difficulties when reading the text. Finally, previous studies have only examined continuous irrelevant sounds that are usually played for the whole duration of the trial. Therefore, it is currently not known whether discrete sounds that involve subtle auditory changes can also influence the ongoing reading process.

In summary, the overarching goals of this Thesis were as follows:1) to establish whether and to what extent backgrounds sounds can disrupt reading comprehension based on the available evidence, and to find out what the nature of this disruption is; 2) to test whether intelligible speech disrupts the reading process by interfering with the lexical access of words, and whether this interference is phonological or semantic in nature; 3) to examine how intelligible speech affects reading comprehension and online integration processes; 4) to test whether the increase in re-reading behaviour in response to intelligible speech is related to maintaining an accurate comprehension of the text; and 5) to investigate whether discrete deviant sounds that violate participants' expectations can influence eye-movements during reading.

In Chapter 2, the first research question was addressed by making a statistical synthesis of previous findings from reading comprehension experiments in order to find out whether, and to what extent, auditory stimuli can interfere with reading comprehension. As comprehension is the end product of the reading process, a better understanding of how auditory stimuli affect this variable is the necessary first step in building theoretical frameworks that can explain auditory distraction during reading and in understanding how eye-movements may be disrupted by task-irrelevant sounds. To do this, a Bayesian meta-analysis approach was adopted that makes it possible to quantify the degree of belief, given the data, that background sounds can disrupt reading. Second, Bayesian meta-regression models were used to test the predictions derived from existing theories on auditory distraction and to estimate how likely it is that they can explain the existing data.

The second research question of whether intelligible speech disrupts the initial lexical processing of words is addressed in Chapter 3. This is an important theoretical question as lexical identification plays a major role in computational models of eye-movement control during reading such as E-Z Reader (Reichle et al., 1998) and SWIFT (Engbert et al., 2005). If intelligible speech affects the lexical identification of words, this would suggest that it has a very early influence on the language processing system. Additionally, the second aim of this study was to investigate whether the phonological or semantic information of intelligible speech may disrupt lexical identification and sentence reading more broadly. To answer these questions, participants' eye-movements were recorded while they read single sentences with a target word lexical frequency manipulation.

The third research question of how intelligible speech affects comprehension and online integration processes is addressed in Chapter 4. More specifically, this study attempted to find out whether participants' immediate reading comprehension is disrupted only when they have to answer more difficult questions and whether intelligible speech makes it harder to integrate text information across multiple sentences. In this experiment, participants' eye-movements were recorded while they read short paragraphs. Additionally, participants answered either easy yes/no comprehension questions or more difficult multiple-choice questions. As longer passages are processed differently compared to single unrelated sentences (Radach et al., 2008), this experiment also made it possible to replicate and extend the main auditory distraction effects from Chapter 3 by using more ecologically-valid reading stimuli.

The fourth research question of this Thesis was whether the increase in regressions and re-reading fixations in response to intelligible speech is due to an attempt to maintain an accurate comprehension of the text. In fact, while previous studies have suggested that intelligible speech leads to an increase in re-reading behaviour (Cauchard et al., 2012; Hyönä & Ekholm, 2016; Yan et al., 2017), it is currently not well understood why this occurs. Chapter 5 tested one possible explanation for this: namely, that the increase in re-reading behaviour is due to a transient comprehension difficulty caused by the irrelevant speech that participants try to overcome and still comprehend the text. This research question was tested in two experiments (one behavioural and one eye-tracking) in which participants were prevented from re-reading previous words in the text.

The final research question that will be examined in this Thesis is how the ongoing reading process is influenced by discrete deviant sounds. While distraction by deviant sounds has been well-documented at both the behavioural (Dalton & Hughes, 2014; Hughes, 2014; Parmentier, 2014) and electrophysiological level (Berti, 2008; Escera et al., 2000; Näätänen et al., 2007), it is currently not known whether eye-movements during reading are sensitive to discrete deviant sounds. This is an important theoretical question as most of the studies on deviance distraction have been conducted using simple categorization tasks, and it is not well understood if deviance distraction can also be observed in more complex everyday tasks such as reading. In Chapter 6, this question is addressed by developing a new experimental paradigm in which a sequence of five short sounds is played once participants fixate five target words in the sentence.

CHAPTER 2: A BAYESIAN META-ANALYSIS OF PREVIOUS FINDINGS

The review of the literature in Chapter 1 showed that background noise, speech, and music may be detrimental to reading comprehension, but that considerable uncertainty exists as to the reliability and the magnitude of such distraction effects. This uncertainty makes it difficult to draw firm conclusions about the experimental effects and their real-world significance. Are background sounds reliably disruptive to reading, and is this disruption large enough to be of any practical significance? Additionally, after 80 years of research on the topic, what theoretical conclusions can be made about the types of background sounds that are disruptive to reading?

The present study addressed these questions by performing a Bayesian random-effects meta-analysis of studies investigating experimental exposure to noise, speech, or music in the background. Both studies with adults and children were considered. Bayesian inference is especially suited to answer these questions because it enables us to directly quantify the uncertainty of the estimate of auditory distraction effects, given the available evidence. This in turn makes it possible to derive the probability, given the data, that background noise, speech, and music can distract readers from their task. Bayesian meta-analytical models have traditionally been used in biology and medicine (e.g., Sutton & Abrams, 2001; Sutton, Abrams, Jones, Sheldon, & Song, 2000), but more recently have also been introduced to psychology and linguistics (Jäger, Engelmann, & Vasishth, 2017; Marsman et al., 2017; Vasishth, 2015; Vasishth, Chen, Li, & Guo, 2013; see also Kruschke & Liddell, 2017). As such, they have been successfully used to address contentious research questions, such as the processing of relative clauses in Chinese (Vasishth et al., 2013), and the extent to which readers can pre-process words in parafoveal vision (Vasilev & Angele, 2017).

There are two available (non-Bayesian) meta-analyses to date that have addressed how background noise and music affect a wide range of behavioural and cognitive tasks (Kämpfe, Sedlmeier, & Renkewitz, 2011; Szalma & Hancock, 2011). While the results from these meta-analyses are quite interesting, their more general focus on all types of cognitive tasks does not make it possible to make firm conclusions about reading in particular. Interestingly, Kämpfe et al. reported a separate analysis of reading-only studies and estimated the general effect of music to be r= -0.11 (d= -0.22). However, this estimate was based on only eight studies and thus does not include most of the currently available data. Therefore, one of the contributions of the present meta-analysis was to estimate the general effect of background noise, speech, and music on reading, and to calculate the probability, given all the available evidence, that these auditory stimuli are detrimental to reading performance.

The second and more important goal of the present analysis was to investigate what aspects of background sounds give rise to distraction. Although it can be informative to estimate the overall size of the effects, as previous meta-analyses have done, this does not tell us what it is about these sounds that makes them distracting. As it was discussed previously, there are a few theories that make specific predictions about what type of auditory stimuli should be distracting. Therefore, the second aim of the study was to test the predictions of these theories using Bayesian meta-regression models (Welton, Sutton, & Cooper, 2012). As some of the theories outlined above were not originally developed in reading comprehension tasks, it is important to keep in mind that the present study is not a strict test of these theories. Rather, it aims to find out whether they can accommodate the existing evidence in reading tasks, and if not, to pave the way for the development of future theories.

2.1. Predictions

All of the predictions in the present analyses are summarised in Figure 2. The phonological interference hypothesis (Salamé & Baddeley, 1982) makes the unique prediction that all types of speech sounds should be equally distracting because they all gain access to the phonological store. Therefore, both intelligible speech (i.e., in participants' native language) and unintelligible speech (i.e., in a foreign language) should be equally distracting. Additionally, the phonological interference hypothesis is not capable of explaining distraction by non-speech background noise and non-lyrical music because neither sound gains access to the phonological store.

The semantic interference (Martin et al., 1988) and interference-by-process (Marsh et al., 2008) accounts both make the prediction that only intelligible speech that can be processed semantically by participants would cause distraction. Therefore, intelligible speech should be more distracting than unintelligible speech. Additionally, they also predict that: 1) lyrical music should be more distracting than non-lyrical music because the former contains lyrics that are intelligible to participants; and 2) intelligible speech should be more distracting than lyrical music because, on average, continuous speech has more semantic content than lyrical music[1]. However, since lyrical music that is intelligible to participants contains not only semantic, but also phonological information, it is not possible to rule out any involvement of phonology in this effect.

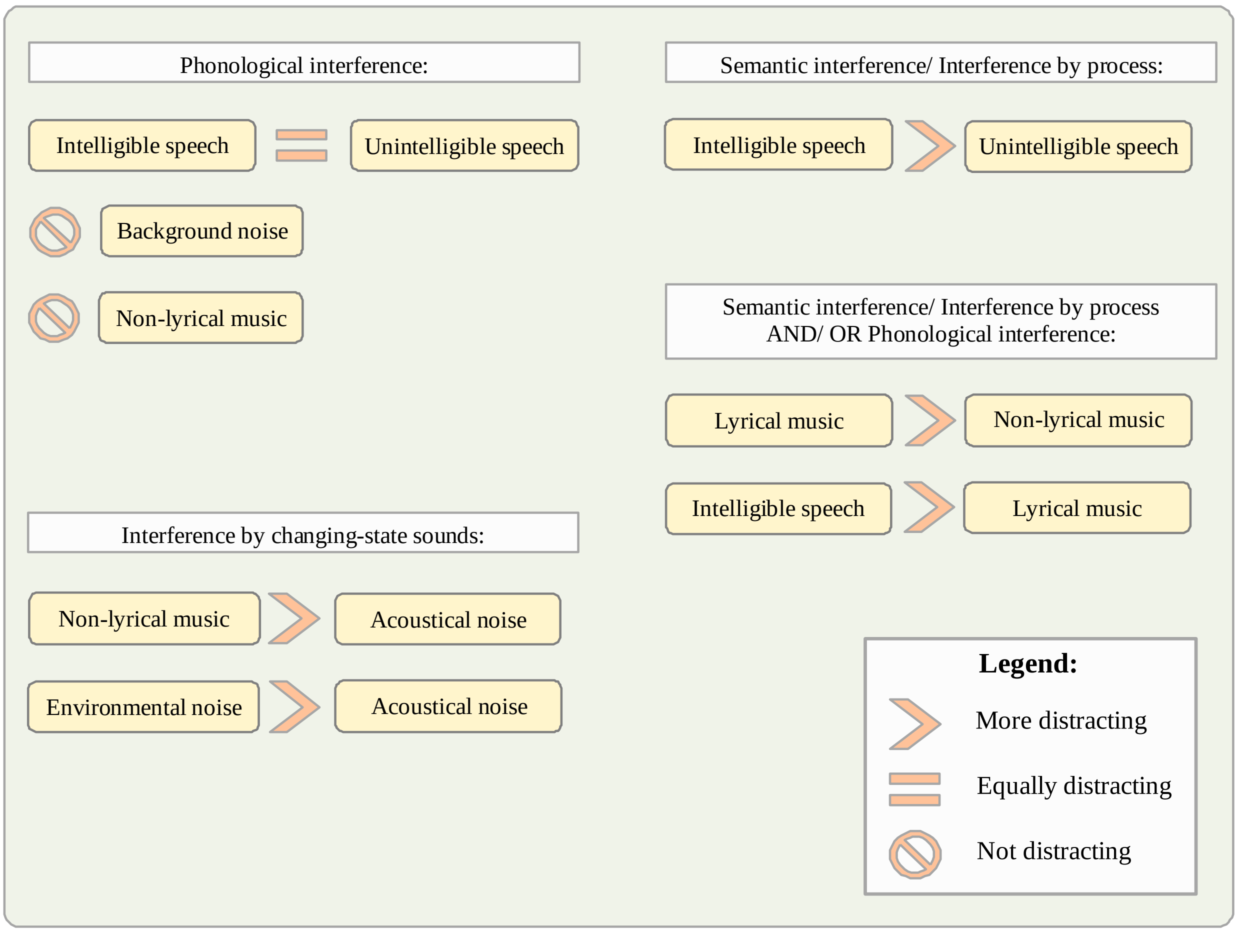

Figure 2. A schematic summary of the predictions in Chapter 2 derived from theories on auditory distraction.

[1] It should be noted that the amount of semantic content may differ depending on the type of music. Nevertheless, the lyrical music examined in this analysis also contained instrumental sections that didn't have lyrics. This was determined by manually examining the music that was played in the original studies. Therefore, even though lyrics were present in the music, this wasn't the case for the whole duration of the song.

Finally, the changing-state hypothesis (Jones, Madden, & Miles, 1992) predicts that sounds exhibiting considerable acoustic variation should be more distracting than steady-state sounds that do not exhibit such variation. This leads to two predictions. First, non-lyrical music should be more distracting than acoustical noise (e.g. white or pink noise). This is because the former exhibits more acoustic variation than the latter. Non-lyrical music is the strongest test of this prediction because it avoids any potential confounds from spoken language that would be present in lyrical music. Second, more complex environmental noise (e.g. traffic noise or office noise containing phones ringing, indistinct chatter, etc.) should again be more distracting that steady-state acoustical noise because it also exhibits more acoustic variation.

2.2. Method

The goal of a meta-analysis is to pool together evidence from multiple studies in order to estimate some parameter of interest (e.g., the true difference in comprehension accuracy between reading in silence and reading with music in the background). A Bayesian meta-analysis differs from the classical (frequentist) meta-analysis in the sense that it uses Bayesian inference to estimate the parameter and the uncertainty surrounding this estimate. Before performing the analysis, the researcher needs to express their prior belief about the parameter in terms of a probability distribution. This is known as the *prior probability distribution* and it reflects the researcher's belief about the parameter prior to observing the data. After the data are collected, a *likelihood function* is constructed, which essentially tells us how probable the data are for different values of the parameter (Lynch, 2007). The result of Bayesian inference is a *posterior probability distribution*, which is the researcher's updated belief about the parameter *given* the observed data.

The posterior probability distribution is derived from Bayes' theorem, which states that the posterior distribution is proportional to the product of the prior probability distribution and the likelihood (i.e., Posterior $\propto$ Prior x Likelihood; see Lynch, 2007 for more details). In the meta-analysis, the observed means are the empirical effect sizes (that is, the differences between conditions) reported in the original studies. In contrast, the posterior mean of the effect sizes is simply the mean of the posterior probability distribution that is derived from the Bayesian meta-analysis. Therefore, the posterior mean reflects our updated belief about the size of the effect (i.e., the difference) in light of the observed data.

One important part of any meta-analysis is to assess the data for publication and other reporting biases. One common way to do this is to use what is known as a *funnel plot* (Egger, Smith, Schneider, & Minder, 1997; Sterne et al., 2011). This is a scatter plot of all the effect sizes included in the meta-analysis against some measure of their precision, such as the standard error or the inverse of the standard error. More precise studies (i.e., the ones with smaller standard error) will appear more narrowly at the top of the plot, while less precise studies (i.e., the ones with larger standard error) will scatter more widely at the bottom. When there is no bias or heterogeneity between studies, the scatter of the plot will resemble a symmetrical inverted funnel (Sterne et al., 2011). *Funnel plot asymmetry* can occur if studies are missing from one side of the plot, thus creating an asymmetrical funnel shape. For example, this can happen if publication or other reporting biases are preventing the dissemination of studies with negative findings (however, reporting biases are not the only possible source of asymmetry, and other factors need to be explored as well; see Sterne et al., 2011).

2.2.1. Literature Search

The search of the literature was conducted by following the PRISMA guidelines (Moher et al., 2009). A flowchart of the process is presented in Figure 3. Google Scholar, Scopus, the Web of Science, and ProQuest Dissertations were searched with the following keywords: “background noise AND reading”, “background speech AND reading”, and “background music AND reading”. The search for each of the three background sounds was done separately. The literature search covered articles published before the 25th of June, 2017. Additionally, the reference lists of all screened articles, as well as those of previous literature reviews and meta-analyses on similar topics (Beaman, 2005; Clark & Sörqvist, 2012; Dalton & Behm, 2007; Kämpfe et al., 2011; Klatte, Bergström, & Lachmann, 2013; Shield & Dockrell, 2003; Szalma & Hancock, 2011), were also examined.

When searching the literature, it is important to consider relevant studies that have been conducted but have never been published in a peer-reviewed journal or an edited book (i.e., the so-called file-drawer problem; Rosenthal, 1979). This issue was addressed through some of the databases that were searched. ProQuest Dissertations contains more than 2 million doctoral and masters’ dissertations (Lefebvre, Manheimer, & Glanville, 2008), which often contain unpublished research. Additionally, Google Scholar indexes a wide range of unpublished sources, such as conference proceedings, dissertations, reports, and pre-prints. Furthermore, author searches were carried out for researchers who have done work on this topic in the last two decades. These searches included researcher networking websites such as ResearchGate.net and Academia.edu that also contain unpublished research (e.g., conference presentations or unpublished manuscripts). In the present meta-analysis, unpublished studies accounted for 17 % of all screened records, thus showing that the search strategy was effective in locating them (unpublished studies typically make up 8-10% of all sources in systematic reviews and meta-analyses; Clarke & Clarke, 2000; Lefebvre et al., 2008). The unpublished studies came from different sources, such as dissertations, conference proceedings, reports, and unpublished manuscripts.

The identified articles were evaluated against the inclusion criteria presented in Appendix A. In short, the studies had to experimentally manipulate background noise, speech or music in a reading or a proofreading task, have a sound methodological design, and include reading in silence as a baseline condition. The inclusion criteria were developed prior to the meta-analysis with the help of a smaller, qualitative review of the literature. Epidemiological studies of chronic exposure to traffic noise in children were not included because they answer a qualitatively different question and are often confounded by other variables, such as social deprivation (Haines, Stansfeld, Head, & Job, 2002). Overall, 44 % of the experiments whose eligibility was assessed were included in the meta-analysis. Although the inclusion rate may appear to be low, it was necessary to ensure that only studies that are similar enough to be analysed together are included. Information about the included studies and their effect sizes is presented in Appendix B.

2.2.2. Dependent Measures

The main dependent variable was reading comprehension accuracy, which was available for 54 of the studies (83.1 %). Therefore, most of the reported analyses are based on reading comprehension accuracy. Moreover, effect sizes for reading speed were available for 13 studies (20 %), and these were analysed separately. Finally, experiments reporting proofreading accuracy (N=7; 10.7 %) were also analysed for completeness, but this was again done separately from the analysis on reading comprehension accuracy.

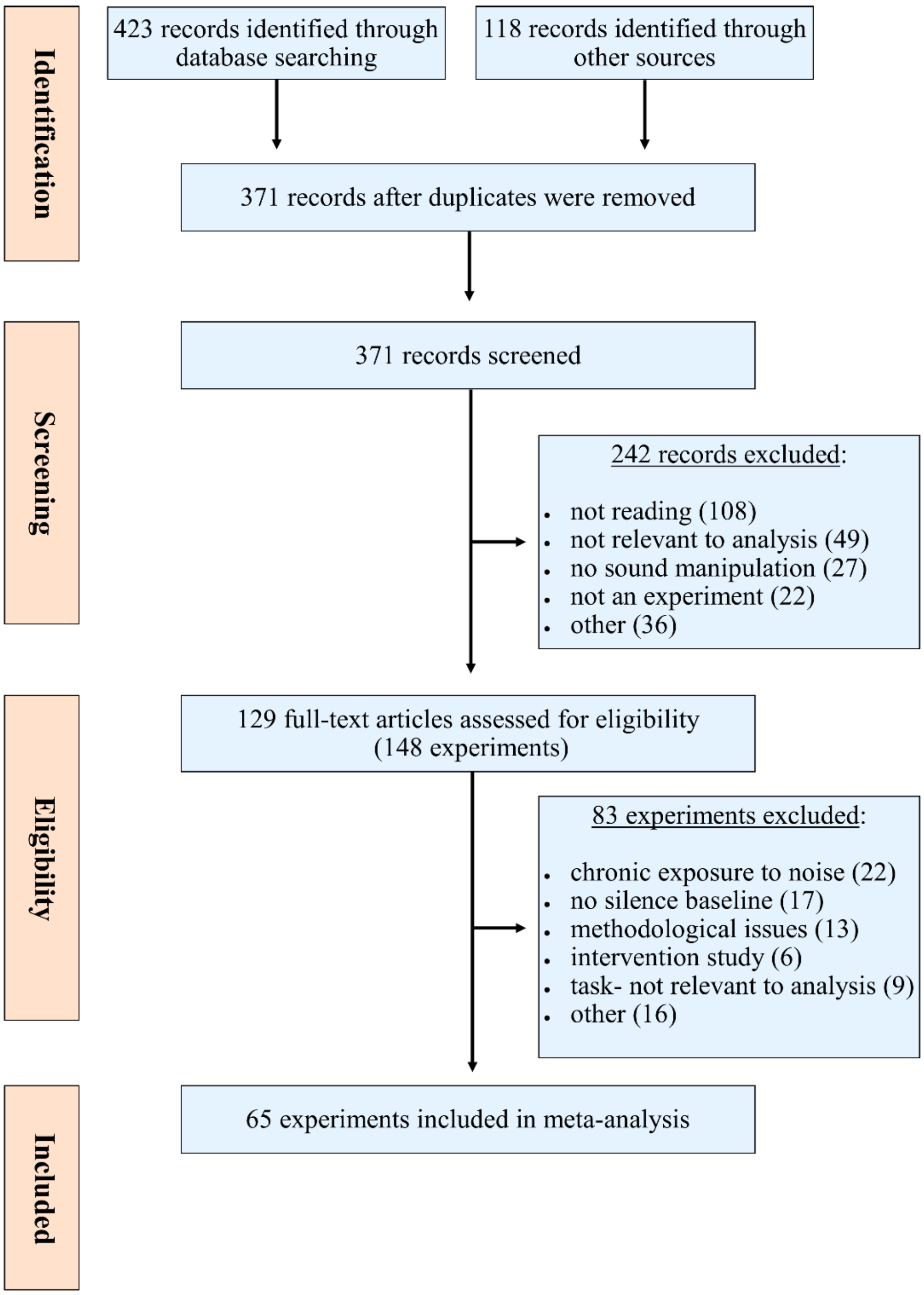

Figure 3. A flowchart illustrating the stages of the literature search process in Chapter 2.

For the meta-regression analyses, additional information about the type of sound manipulation was also extracted (e.g., whether the noise was environmental or acoustical, whether the music was lyrical or non-lyrical). If a study contained a background music manipulation, the songs were manually examined by the author in order to determine whether they were lyrical or non-lyrical. Only studies that could be unambiguously classified as either lyrical or non-lyrical were added to this meta-regression analysis.

2.2.3. Effect Size Calculation

Standardized effect sizes of the mean difference (*g*) and their variances were calculated from the reported descriptive statistics. This was done by first calculating Cohen's *d* for the respective design of the study and then applying Hedges' *g* (Hedges & Olkin, 1985) correction for small sample bias. The effect sizes were calculated with formulas 12.11-12.22 from Borenstein (2009). In all effect sizes, silence was the control condition. Therefore, the effects represent the standardized mean difference between reading in the experimental sound condition and the control condition of reading in silence. If descriptive statistics were unavailable or incomplete, the effect sizes were calculated by digitalizing graphs (Rohatgi, 2015) or converted/ approximated from the reported test statistics by using existing formulas (Borenstein, 2009; Lajeunesse, 2013)[2]. In the analysis of reading comprehension accuracy and proofreading accuracy, studies were coded so that negative effect sizes indicate lower comprehension/ proofreading accuracy in the experimental sound condition. Similarly, in the analysis of reading speed, negative effect sizes also indicate slower reading speed in the

[2] Four studies did not contain any information that made it possible to calculate the effect sizes. As all of the studies were more than 25 years old, it was not possible to obtain the data from the authors. Therefore, these studies were discarded (they did not count towards the number of included studies). We explored the implications of this through statistical simulations and found no evidence that failing to include these studies biased the results (see Appendix C).

experimental sound condition compared to silence. One effect size was excluded as an outlier (see Figure A1 in Appendix C).

Because 55.5% of the studies used a within-subject design, it was necessary to estimate the population correlation (ρ) between the control and experimental conditions (Borenstein, 2009; Szalma & Hancock, 2011). Eight statistically-independent estimates were obtained from experiments for which the raw data were available, as well as from one study (Miller, 1947) that reported the required statistics. These represented a wide range of experimental sound types and included both reading comprehension and reading speed measures. We followed Szalma and Hancock's (2011) approach to meta-analyse the obtained correlations and to obtain a weighted estimate of ρ. The resulting weighted value of 0.74 was used for calculating the effect sizes for all within-subject design studies.

Effect sizes from within- and between-subject studies are calculated with different standard deviation metrics and are thus not necessarily comparable (S. B. Morris & DeShon, 2002). Consistent with previous work (Kämpfe et al., 2011; Szalma & Hancock, 2011), the effect sizes from within-subject studies were transformed to make them comparable to the effect sizes of between-subject studies. This was done using Formula 11 from Morris and DeShon (2002). Additionally, because some studies yielded more than one effect size, care was taken to avoid statistical non-independence in the analyses (see Noble, Lagisz, O'dea, & Nakagawa, 2017 for a recent overview). If a study contributed multiple effect sizes per analysis, these were averaged together to include only one effect size for that study (Lipsey & Wilson, 2001)[3].

[3] One exception was the meta-regression model comparing lyrical vs. non-lyrical music. We show in Appendix C that the way the effect sizes were chosen did not influence the conclusions from this analysis.

2.2.4. Publication Bias

In the present meta-analysis, 12.3% of all included studies were from the so-called grey literature (i.e. they were not formally published in a peer-reviewed journal or in an edited book at the time of analysis). To assess the data for publication and other related biases, we performed a number of visual and statistical tests using the "meta" (Schwarzer, 2007) and "metafor" (Viechtbauer, 2010) R packages. The visualization of the results for reading comprehension is presented in Figure 4 (see Appendix C for reading speed). The funnel plots (Panels **a** and **b**) indicated that there was some heterogeneity in the data, but there was no clear evidence of asymmetry that could indicate publication bias. This was confirmed by a funnel plot test of asymmetry based on a weighted linear regression of the effect estimates on their standard errors (Sterne et al., 2011), which revealed no statistically significant evidence for asymmetry for either reading comprehension ($t(52) = -0.42$, $p= 0.67$) or reading speed ($t(11)= 0.08$, $p= 0.93$; proofreading accuracy was not considered here because funnel plot tests of asymmetry are not recommended when there are fewer than 10 studies; Sterne et al., 2011). Additionally, meta-regression analyses (Figure 4e-f) indicated that the size of auditory distraction effects was not predicted by the impact factor of the journal or the year of publication. In summary, there was no evidence to suggest that publication bias may have influenced the conclusions from the meta-analysis.

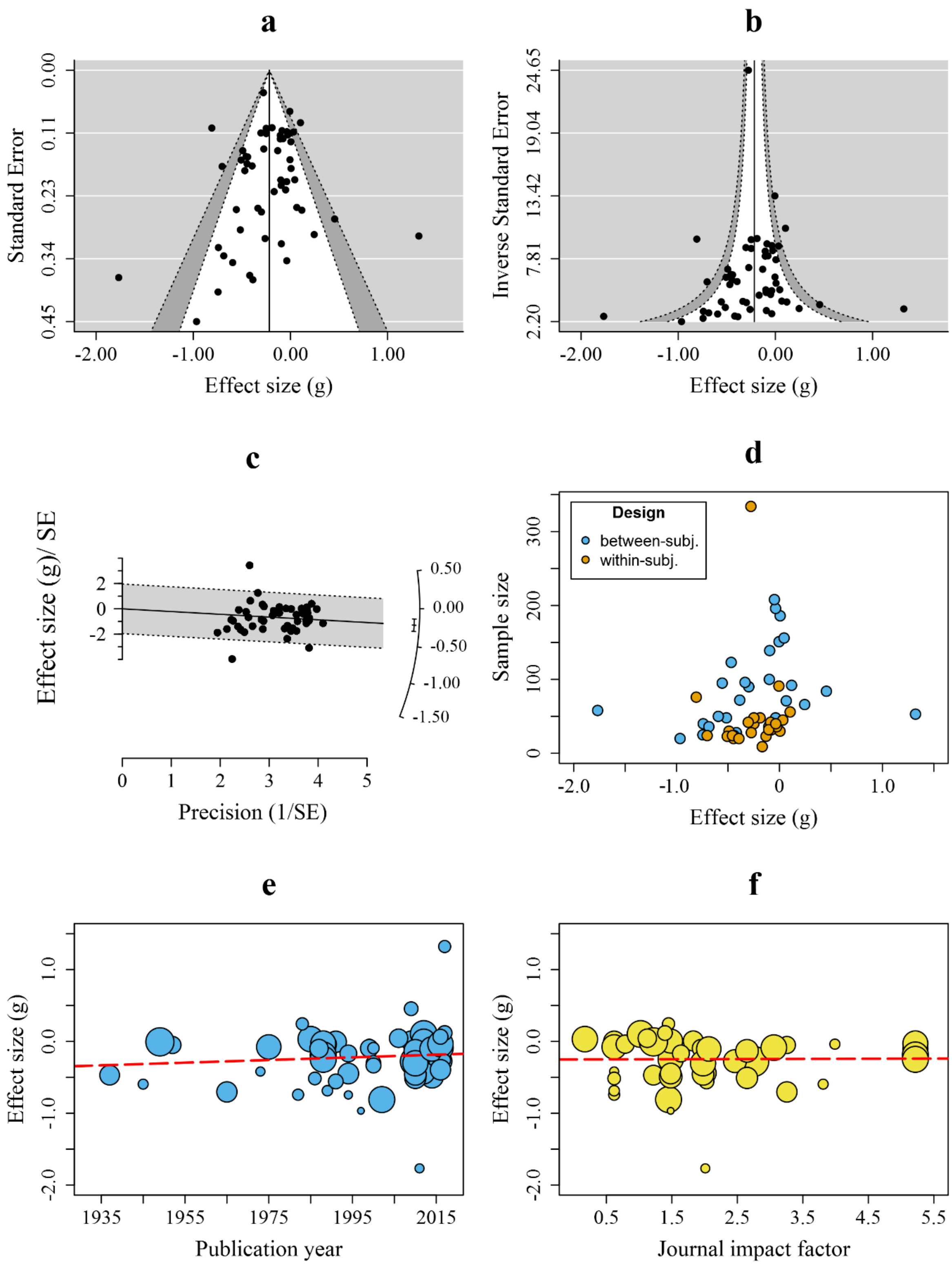

Figure 4. Visual assessment of publication and other related biases for reading comprehension accuracy in Chapter 2 (presentation format adapted from Nakagawa, Noble, Senior, & Lagisz, 2017, Figure 6). **a**: Funnel plot of effect sizes against their standard error.

White and dark grey bounds indicate the 95% and 99% pseudo-confidence intervals, respectively. Studies with smaller standard error should appear at the top, while studies with larger standard error should scatter at the bottom of the plot. **b**: The same funnel plot of effect sizes against the inverse of their standard error. In this funnel plot, more precise studies (i.e. the ones with smaller standard error) should appear at the top, while less precise studies (i.e., the ones with larger standard error) should scatter at the bottom of the plot. **c**: Radial (Galbraith) plot of the *z*-statistic of each study (y axis) against the inverse of the standard error (x axis). Shading shows *z*-value bounds of ± 2. The vertical scatter of effect sizes shows how much heterogeneity there is in the data and the shading shows the approximate 95 % confidence interval where, on average, 95% of the studies are expected to lie (Anzures-Cabrera & Higgins, 2010). **d**: Plot of effect sizes against their sample size, broken down by study design type. **e** and **f**: Meta-regression models examining whether the size of effects is predicted by publication year (**e**) or impact factor of the journal where the study was published (**f**). Both models show that this was generally not the case. Red dotted line shows the meta-regression slope.

2.2.5. Data Analysis

2.2.5.1. Meta-analysis.

The common choice in meta-analysis is between a fixed-effect and a random-effects model. A fixed-effect model assumes that all effect sizes that are combined together are estimating the same true underlying effect, which we will call θ. Therefore, the effect size of the *i*-th study, T_i, is assumed to come from a normal distribution with some mean θ and variance σ_i^2:

$$T_i \sim Normal(\theta, \sigma_i^2) \quad i = 1, 2, 3, \dots, n \tag{1}$$

In this model, any variability in the estimate is due to sampling error alone. On the other hand, a random-effects model relaxes this assumption by explicitly allowing for variability in the true effect size between studies (Welton et al., 2012). In this case, the observed effect size of the *i*-th study T_i is assumed to be generated by a unique underlying true effect for that *i*-th study, denoted here by θ_i. This unique underlying effect θ_i is in turn assumed to come from a normal distribution with some (unknown) mean θ and between-study variance τ^2:

$$T_i \sim Normal(\theta_i, \sigma_i^2) \quad i = 1, 2, 3, \dots, n \qquad (2)$$

$$\theta_i \sim Normal(\theta, \tau^2)$$

Therefore, the true effect sizes of individual studies in a random-effects meta-analysis can be informally thought of as random samples from a normal distribution of effect sizes (Welton et al., 2012).

In the present meta-analysis, a random-effects model was chosen *a priori* because some between-study heterogeneity was expected due to differences in design, sound intensity levels, participants, reading materials, and so forth. A random-effects model can naturally account for such sources of variability between studies and is often the model of choice in studies on language processing (e.g. Jäger et al., 2017; Vasishth et al., 2013; Vasilev & Angele, 2017). The full Bayesian model was defined as follows (Jäger et al., 2017; Schmid & Mengersen, 2013):

$$T_i \mid \theta_i, s_i^2 \sim Normal(\theta_i, s_i^2) \quad i = 1, 2, 3, \dots, n \qquad (3)$$

$$\theta_i \mid \theta, \tau^2 \sim Normal(\theta, \tau^2),$$

$$\theta \sim Uniform\ (-10, 10),$$

$$\tau \sim Uniform(0, 10)$$

where: T_i is the observed effect size (in Hedges' g) in the *i*-th study

θ_i is the true auditory distraction effect in the *i*-th study

s_i^2 is the true sampling variance of the *i*-th study, estimated from the within-study variance of the sampling distribution of study *i*

θ is the unknown true auditory distraction effect estimated by the model

τ^2 is the unknown between-study variance

In this model, precision was defined as the inverse of the within-study variance of the sampling distribution. The last two lines in Equation 3 indicate the prior probability distributions used for θ and τ. In the present analysis, we used Uniform priors that assign equal probability to any value on these intervals. As these are vague priors, they have very little to no influence on the results. This was confirmed by doing a sensitivity analysis of the main results with alternative priors: *Normal* $(0, 10^4)$ for θ and *Normal* $(0, 10^4)$ *I(0,)* for τ (normal distribution truncated at 0). The sensitivity analysis indicated that the choice of priors did not influence the results (see Appendix C).

2.2.5.2. Meta-regression.

Although random-effects meta-analysis can account for heterogeneity between studies, it does not tell us what causes this heterogeneity in the first place (Welton et al., 2012). However, it is possible to use meta-regression models to investigate how different study characteristics (e.g. whether the background music was lyrical or non-lyrical) are associated with the observed effect sizes. Meta-regression models are similar to the ordinary least-squares regression, but with the crucial difference that the estimate is adjusted by the precision of the studies (i.e., the inverse of the within-study variance of the sampling distribution; Welton et al., 2012). The model from Equation 3 was extended by adding a regression coefficient β for the underlying effect of the covariate (the added parameters are formatted in bold; Jäger et al., 2017; Welton et al., 2012):

$$T_i \mid \theta_i, \boldsymbol{\beta}, s_i^2 \sim Normal(\theta_i + \boldsymbol{\beta x_i}, s_i^2) \quad i = 1, 2, 3, \dots, n \qquad (4)$$

$$\theta_i \mid \theta, \tau^2 \sim Normal(\theta, \tau^2),$$

$$\theta \sim Uniform\ (-10, 10),$$

$$\tau \sim Uniform(0, 10)$$

$$\boldsymbol{\beta \sim Uniform\ (-10, 10)}$$

where: β is the regression coefficient for the underlying effect of the covariate x_i.

θ_i is the true auditory distraction effect in the *i*-th study, adjusted for the covariate effect x_i

θ is the unknown true auditory distraction effect, also adjusted for the covariate effect x_i

All remaining parameters have the same interpretation as in Equation 3.

The contrasts used for the covariate x_i are presented in Table 2. These contrasts were used to test the predictions outlined in the introduction.

2.2.5.3. Posterior sampling. The posterior probability distribution was sampled with JAGS (Plummer, 2003) using the R software, v. 3.31 (R Core Team, 2016). Five Markov Chain Monte Carlo (MCMC) chains were run with 100 000 iterations each. Checks were made to ensure that the starting values of the MCMC chains did not influence the results. The first 3000 iterations were discarded as burn-in. A thinning interval of 5 was used for the MCMC chains (i.e., every fifth sample was retained) to reduce the influence of auto-correlation. The summary of the posterior distribution was based on 20 000 samples per chain (excluding the burn-in period). Convergence was assessed with visual inspection of the trace plots and with Gelman and Rubin's (1992) convergence diagnostic. The diagnostics suggested that convergence had been achieved in all models.

Comparison	Covariate levels		Contrast coding	
	Level 1	Level 2	Level 1	Level 2
Non-lyrical vs lyrical music	non-lyrical	lyrical	-1	1
Lyrical music vs intelligible speech	music	speech	-1	1
Unintelligible vs intelligible speech	unintelligible	intelligible	-1	1
Acoustical vs environmental noise	acoustical	environmental	-1	1
Acoustical noise vs instrumental music	noise	music	-1	1
Child vs adult participants	child	adult	-1	1

Table 2. Type of meta-regression comparisons and the contrast coding of covariates.

The effective sample size (ESS) of the MCMC chains was calculated for every parameter and contrast of interest. The ESS is the size of the MCMC chain after adjusting it for auto-correlation (Kass, Carlin, Gelman, Neal, & Carlin, 1998; Kruschke, 2015). All of the present analyses had an ESS greater than 10 000, as recommended by Kruschke (2015). This was necessary for achieving a stable estimation of the credible interval limits, because this estimation depends on sparse regions of the posterior probability distribution that are sampled less often by the MCMC chain (Kruschke, 2015).

The results are presented as the estimate of the effect sizes of interest and their corresponding 95 % credible intervals. Unlike the classical confidence intervals, credible intervals have the intuitive interpretation that they contain the true auditory distraction effect with 95% probability because the values within this interval make up 95% of the posterior probability distribution (see Morey, Hoekstra, Rouder, Lee, & Wagenmakers, 2016). All probabilities reported in the paper are the posterior probability, given the data, that auditory distraction effects exist. A more detailed summary of Bayesian methods and their interpretation is beyond the scope of this paper. However, Nicenboim and Vasishth (2016) provide an accessible overview.

2.3. Results

2.3.1. Meta-analysis

The results from the meta-analysis are presented in Table 3. Additionally, forest plots are presented in Figure 5 for the main measure of comprehension accuracy. To interpret the magnitude of the effects, we will consider J. Cohen's (1988) guidelines of 0.20 for small effects, 0.50 for medium effects, and 0.80 for large effects. Overall, there was a small negative effect for reading comprehension (g= -0.21), which indicates that background sounds generally impaired comprehension accuracy. Consistent with the review of the literature, background speech had a stronger negative impact on reading comprehension (g= -0.26) compared to both background noise and music (g= -0.17 and -0.19, respectively). Nevertheless, the effect for all three sound types was fairly small in size.

Reading speed and proofreading accuracy were also impaired by background sounds. However, the effect sizes for these two measures were very small and the 95% credible intervals all included 0 as a plausible value for the effect (note that this does not allow us to conclude that there is no true effect, just that it is possible that the true effect size is 0). Interestingly, however, the probability that these effects are negative was very high in all analyses (more than 90%). This means that, although the size of the effects was small, there was a very high probability that background speech, noise, and music are detrimental to reading comprehension, reading speed, and proofreading accuracy.

Type of analysis	N	Mean ES (g)	95% CrI	p(ES<0 | Data)	τ^2	ESS
Reading comprehension						
All sounds	54	-0.21	[-0.30, -0.13]	> 0.99	0.06	91803
Noise	12	-0.17	[-0.33, 0.002]	0.97	0.06	92499
Speech	20	-0.26	[-0.36, -0.17]	> 0.99	0.02	47662
Music	36	-0.19	[-0.34, -0.05]	> 0.99	0.13	93678
Reading speed						
All sounds	13	-0.06	[-0.15, 0.02]	0.92	0.01	20915
Speech	6	-0.08	[-0.20, 0.03]	0.92	0.01	28612
Proofreading accuracy						
Speech and Noise	7	-0.14	[-0.42, 0.04]	0.94	0.04	40097
Speech [a]	6	-0.09	[-0.30, 0.07]	0.90	0.02	41296

Table 3. Posterior effect size estimates of auditory distraction effects and 95% credible intervals from the meta-analysis. N: number of studies in the analysis. p(ES<0 | Data): probability that background sounds are detrimental to reading, given the data (i.e., probability that the effect size is smaller than 0). CrI: credible interval. τ^2: estimated between-study variance. ESS: effective sample size of the MCMC chains for the main parameter of interest (θ).

[a] intelligible speech only

Although one can use Bayes factors to perform hypothesis testing (e.g., Rouder & Morey, 2011; Rouder, Morey, & Province, 2013), the emphasis in the present meta-analysis was on estimating the magnitude of auditory distraction effects. The findings from this meta-analysis suggest that there are almost certainly non-null effects, even if their magnitude is small. Therefore, even if a Bayes factor were to favour a null hypothesis relative to some alternative hypothesis, the prior probability of the null hypothesis being exactly true would be negligible in this case. Because of this, the posterior probability of the null hypothesis would remain small.

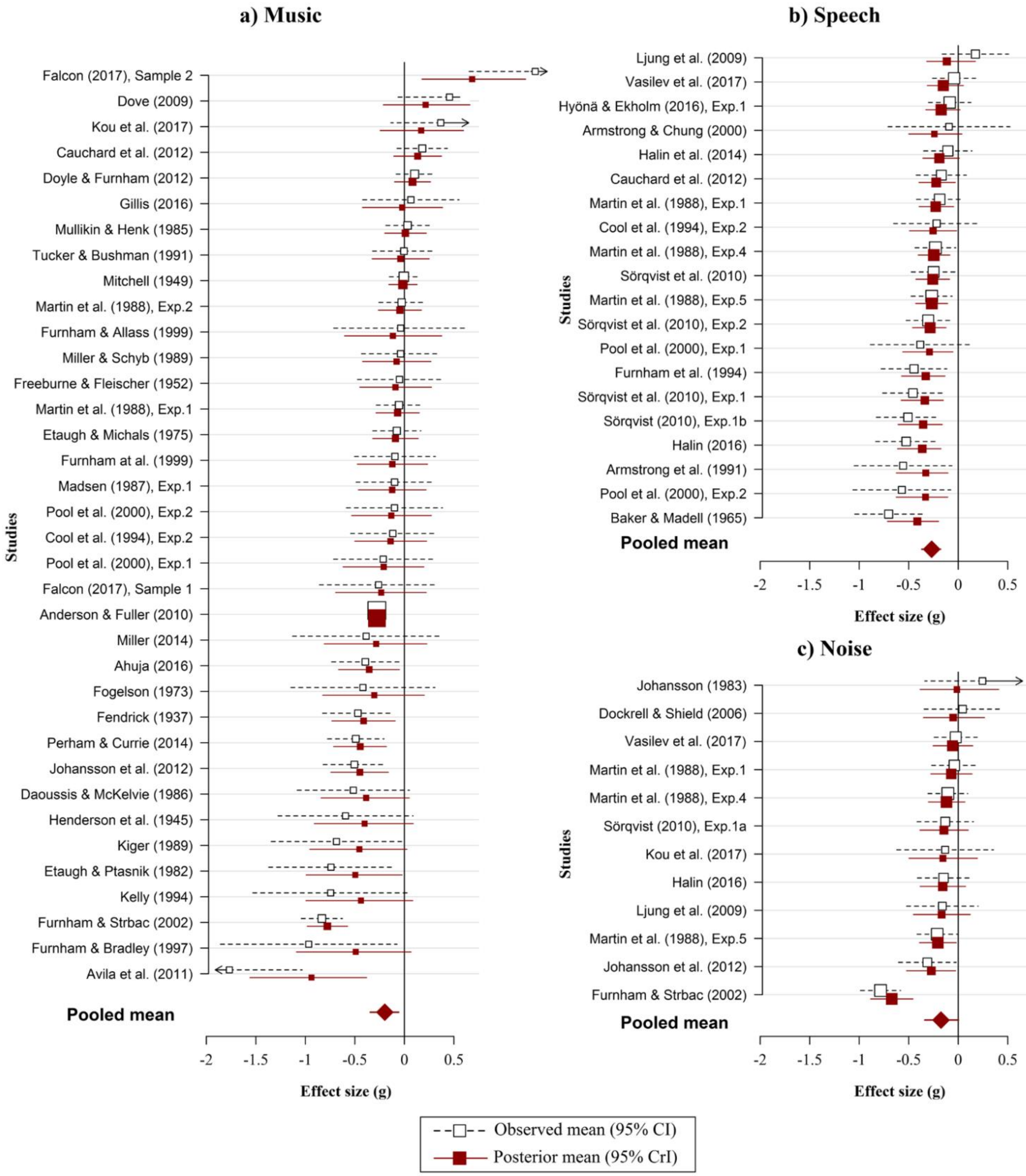

Figure 5. Forest plot for the main effect of background music (**a**), speech (**b**), and noise (**c**) on reading comprehension in Chapter 2. Plotted are the observed (i.e. empirical) effect sizes with their 95% confidence intervals, and the posterior effect size estimates from the meta-analysis model with their corresponding 95% credible intervals. The size of squares is proportional to the weight of each study (i.e., the inverse of the within-study variance of the

sampling distribution). The pooled estimate from the meta-analysis is shown by the dark red diamond at the bottom of each panel (with 95% credible intervals).

Because both studies with adult and child participants were included in the analyses above, we carried out meta-regression models to test whether the effect sizes differed between adults and children. Only reading comprehension was considered in these analyses, as there were too few child studies to reliably estimate differences in reading speed, and all proofreading studies were done with adults. The results are presented in Table 4. They show the estimated mean difference between studies with children compared to studies with adults, after adjusting for their precision in the analysis. Overall, the difference between adults and children was very close to 0, thus showing that background sounds were equally detrimental to reading comprehension for both children and adults. One exception was that background noise impaired reading comprehension in children slightly more than it did in adults, but the mean difference was still quite small (g= 0.05). Additionally, the effect was not highly reliable as there was only 73% probability of a true mean difference. Taken together, these results suggest that effect sizes for reading comprehension did not generally differ between adults and children. For this reason, child and adult studies were analysed together in all remaining analyses.

2.3.2. Meta-regression

The results from the meta-regression models testing the theoretical predictions outlined in the introduction are presented in Figure 6 and Figure 7. Recall that the models yield a regression slope, which shows the estimated mean difference between the two groups, after adjusting for the precision of individual studies. Consistent with the semantic, but not with the phonological interference hypothesis, there was 99% probability that intelligible speech was more distracting than unintelligible speech (mean difference: g= -0.12). Additionally, in line with both the semantic and phonological interference hypotheses, there was 95% probability that lyrical music was more distracting than non-lyrical music (mean difference: g= -0.19). Interestingly, however, intelligible speech and lyrical music did not differ between one another, and the estimated probability of a true difference was only 54% (with 50% being no difference, since the posterior probability density would lie evenly to the left and right side of 0). This last result is surprising because, arguably, most people perceive lyrical music as subjectively less distracting than intelligible speech. For example, it can be speculated that students may be more likely to choose to study while listening to lyrical music in the background than they are to study while listening to an audio book. However, the present results suggest that both lyrical music and intelligible speech are equally distracting.

Analysis	Number of studies		Mean	95 % CrI	$p(ES_{CH} >$	ESS
	children	adults	diff. (g)		$ES_A \mid Data)$	
Reading comprehension						
All sounds	18	36	-0.01	[-0.10, 0.08]	0.43	30623
Noise	5	7	0.05	[-0.13, 0.22]	0.73	29974
Speech	5	15	0.00	[-0.12, 0.12]	0.51	30263
Music	13	23	0.02	[-0.12, 0.17]	0.64	18498

Table 4. Mean difference in the effect size between child and adult studies: Meta-regression results. Mean diff: Posterior estimate of the mean difference (in Hedges' g) between adult and child participants. CrI: credible interval. $p(ES_{CH} > ES_A)$: probability that the effect size for child participants is bigger than the effect size for adult participants, given the data. ESS: effective sample size of the MCMC chains for the main parameter of interest (β).

Consistent with the changing-state hypothesis, there was 90% probability that environmental noise was more distracting than acoustical noise (mean difference: g= -0.10). However, there was only 55% probability of a difference between non-lyrical music and acoustical noise, thus suggesting that the two background sound types did not generally differ. As Figure 6b shows, the size of both effects, as estimated by a random-effects meta-analysis, was very close to 0. This result is contrary to the predicted difference from the changing-state hypothesis.

2.4. Discussion

The present study investigated the magnitude of auditory distraction effects during reading and how compatible these effects are with existing theories of distraction. We will first consider the overall size of the effects and then discuss their theoretical implications. The main findings from the meta-analysis can be summarized as follows. First, background speech, noise, and music all had a negative effect (indicating distraction) on reading comprehension accuracy. The magnitude of the effects was small, but highly reliable, meaning that there was very high probability that these sounds are detrimental to reading comprehension given the available evidence. Second, auditory distraction effects measured with reading comprehension did not generally differ between adults and children. Finally, background speech, noise, and music had a very small, negative effect on reading speed, and background speech and noise also had a small, negative effect on proofreading accuracy. Although both effects proved to be smaller than the ones observed in reading comprehension, there was still high probability that they were negative (>90%).

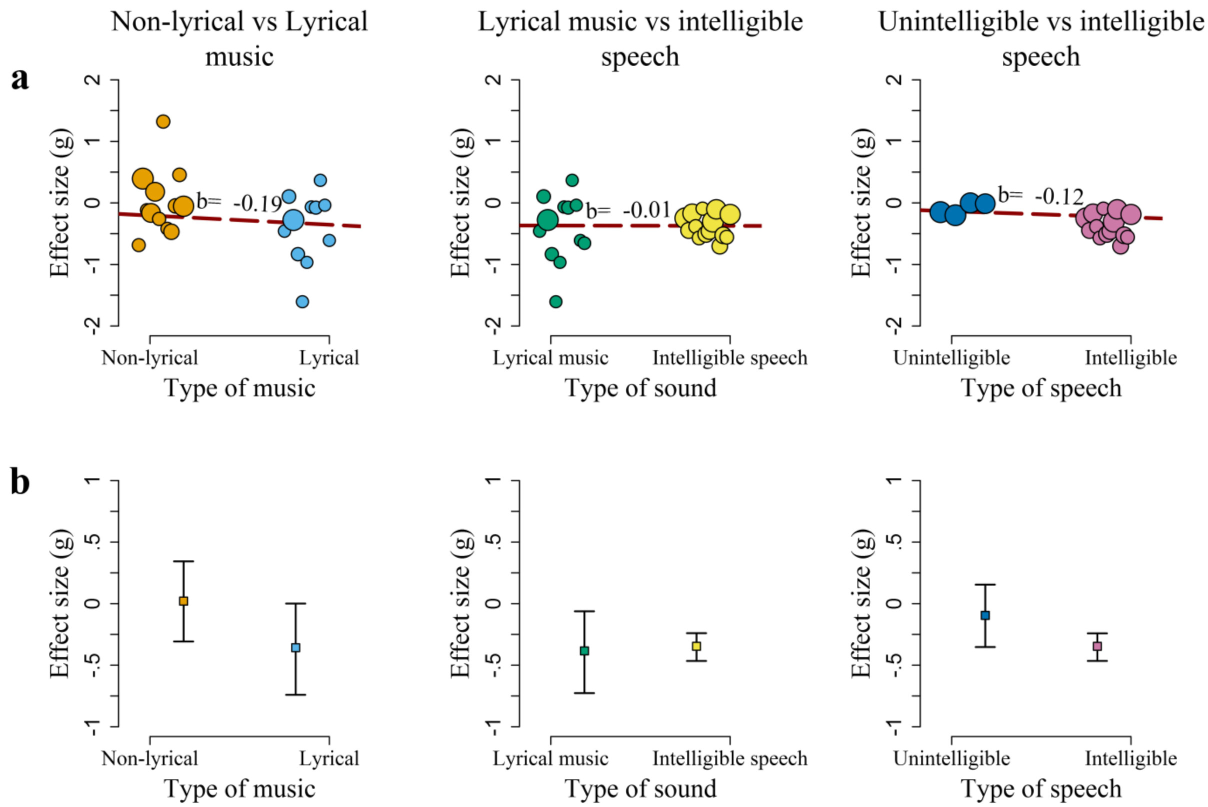

Figure 6. Results of the meta-regression models testing the predictions of the semantic and phonological interference hypotheses in Chapter 2. Panel **a** shows the regression slope and the observed effect size of the studies included in the analysis. The slope indicates the mean difference estimated by the meta-regression model (in terms of Hedges's *g*) between the two groups. The size of circles is proportional to the weight of individual studies (inverse of the within-study variance of the sampling distribution). Panel **b** shows the posterior effect size for each group, as estimated by a random-effects meta-analysis of the simple effect. Error bars show the 95% credible intervals. Effective sample size of the MCMC chains for β (panel **a,** from left to right): 11455, 24381, 54689. Effective sample size of the MCMC chains for θ (panel **b**, from left to right): 98478, 95721, 97382, 32748, 15048, 34152.

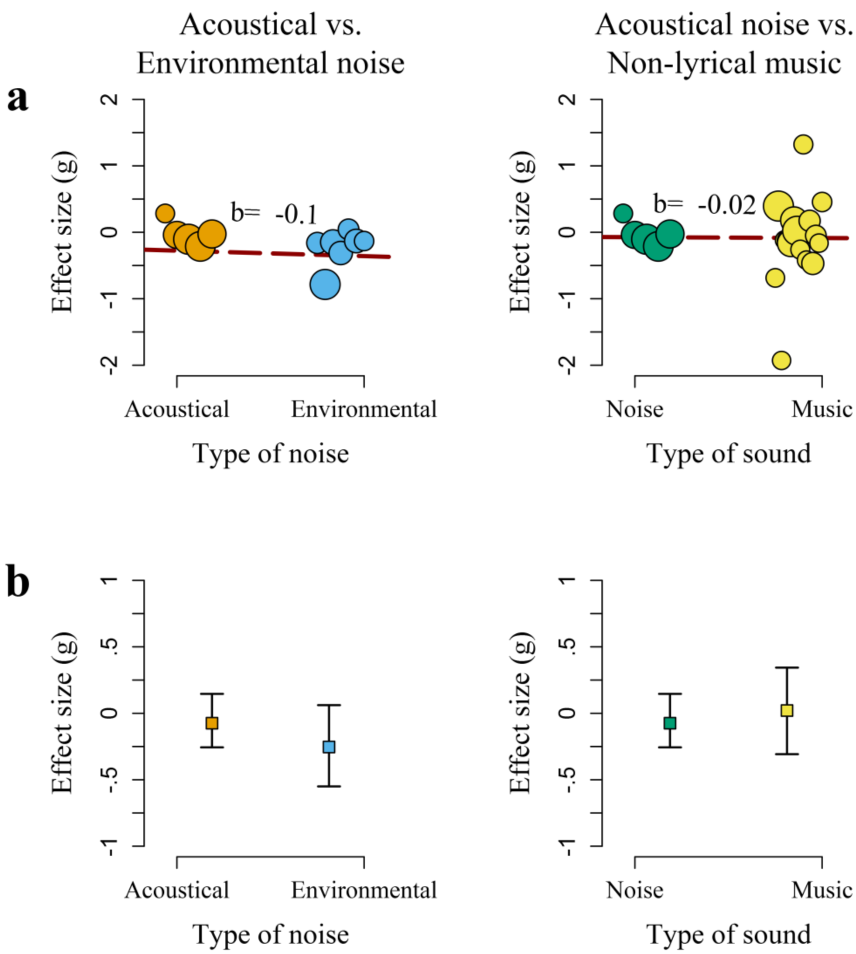

Figure 7. Results of the meta-regression models testing the predictions of the changing-state hypothesis in Chapter 2. Panel **a** shows the regression slope and the observed effect size of the studies included in the analysis. The slope indicates the mean difference estimated by the meta-regression model (in terms of Hedges's *g*) between the two groups. The size of circles is proportional to the weight of individual studies (inverse of the within-study variance of the sampling distribution). Panel **b** shows the posterior effect size for each group, as estimated by a random-effects meta-analysis of the simple effect. Error bars show the 95% credible intervals. Effective sample size of the MCMC chains for β (panel **a,** from left to right): 36063, 13062. Effective sample size of the MCMC chains for θ (panel **b**, from left to right): 31904, 89200, 31904, 98478.

The present results provide the first comprehensive analysis of auditory distraction effects in a reading task. As the review of the literature in Chapter 1 showed, interest in this topic has a very long history that precedes the cognitive revolution, and indeed, most of the work on auditory distraction in other cognitive tasks. Traditionally, much of the interest in auditory distraction in reading tasks has been due to its practical implications for reading outside the psychological laboratory, such as studying for an exam, reading in the classroom, or any kind of work that involves reading in a busy office. However, the inconclusive and sometimes contradictory evidence has made it difficult to arrive at clear conclusions until now. The present results advance our understanding of this topic by showing that external auditory input almost always comes at a cost for reading efficiency. Even though the observed cost was modest, especially for measures such as reading speed and proofreading accuracy, there was still relatively high probability that it reflects a true effect in the population. Therefore, the present study resolves some of the controversy highlighted in the introduction by showing that general auditory distraction effects by background noise, speech, and music almost certainly exist, but that their magnitude is small.

Given that there was very high probability that background speech, noise, and music are detrimental to reading comprehension, why have some of the previous findings been so inconsistent? One possibility is that some of the original studies may not have had sufficient statistical power to detect the underlying effects. Figure 8 shows the relationship between sample size and statistical power for a range of effect sizes, including the ones observed in the present meta-analysis (see Wallisch, 2015). This is for illustrative purposes only, as statistical power is influenced not only by sample size and the magnitude of the true effect, but also by other factors, such as the reliability of the measure, missing data, sampling control and so on (Hansen & Collins, 1994). Nevertheless, as Figure 8 clearly shows, statistical power is related to sample size and generally a larger number of participants are required to achieve sufficient statistical power of detecting some of the auditory distraction effects observed in the present study. This suggests that, although most of the observed effects are negative in sign, statistical significance may not always be achieved if the underlying effect is small and the experiment is underpowered.

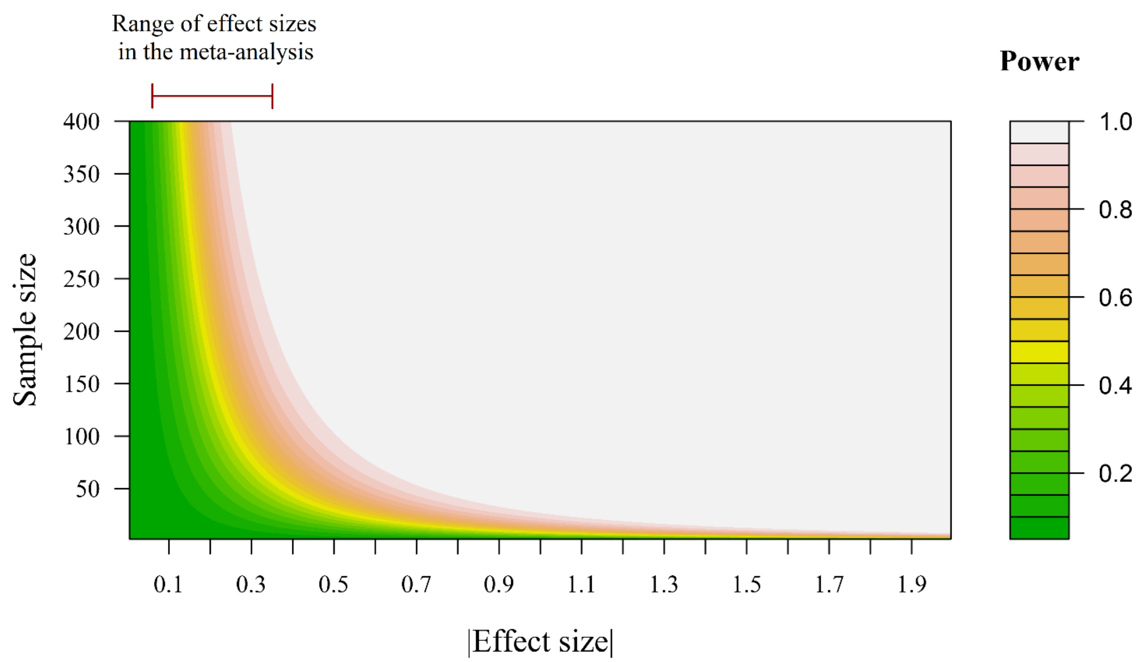

Figure 8. An illustration of the sample sizes needed to achieve different levels of statistical power for a range of realistic effect sizes. Dark red interval at the top shows the range of effect sizes observed in the present meta-analysis. Desirable levels of statistical power are depicted by warm colours. Statistical power was calculated with the “pwr” R package (Champely, 2012) and is based on an independent-samples *t*-test with equal groups, and an α level of 0.05 (two-tailed).

2.4.1. Implications for Theories of Auditory Distraction

The second goal of the present study was to investigate what properties of background sounds make them distracting and to test what theoretical frameworks can explain the results. This is an important question as not all studies have explicitly considered the theoretical implications of their work, with some researchers taking a more applied approach of simply testing whether certain types of sounds are distracting to readers or not.

More broadly, the present analyses provide a glimpse into how well readers can maintain focus on the main task (reading) while listening to a competing stream of auditory input that they try to ignore. The meta-regression results provided a few key insights into the nature of auditory distraction effects, as measured with reading comprehension accuracy.

First, lyrical music was found to be more distracting than non-lyrical music, but equally as distracting as intelligible speech. Second, intelligible speech was in turn more distracting than unintelligible speech. Finally, environmental noise was more distracting than acoustical noise, but there was no reliable difference between non-lyrical music and acoustical noise. These results provide strong support for the notion that the presence of language in background sounds is the strongest contributor to auditory distraction. Indeed, the two largest distraction effects were found for lyrical music (g= -0.35) and intelligible speech (g= -0.34). This last finding is consistent with both the semantic interference (Martin et al., 1988) and interference-by-process (Marsh et al., 2008) accounts, which predict that either the semantic content of speech/ sung lyrics or the actual process of trying to extract their meaning can distract readers from their main task. Nevertheless, these two accounts don't have a mechanism that can explain distraction by non-speech background noise.

The present findings are generally not consistent with the phonological interference account for two reasons. First, it predicts that all speech sounds should be equally distracting because they all would gain access to the phonological store; however, intelligible speech was reliably more distracting than unintelligible speech. Additionally, background noise, which would not gain access to the phonological store, was also found to cause distraction. Finally, the results are only partially consistent with the changing-state account (Jones et al., 1992), which predicts that sounds with greater acoustic variation would cause greater distraction. This is because environmental noise was more distracting than acoustical noise (consistent with the theory), but non-lyrical music was not more distracting than acoustical noise (not consistent with the theory). In both cases, environmental noise and non-lyrical music exhibit more acoustic variation that acoustical (e.g., white or pink) noise.

What type of theoretical framework could account for the present results? Clearly, none of the theories considered so far can account for all the findings. While some theories were successful in accounting for some of the effects, the present results suggest that new theoretical models are needed that can explain all the evidence. This is not necessarily a limitation of existing theoretical accounts because, as noted previously, not all of them were originally designed to account for distraction effects in a reading task. Additionally, these theories offer very useful mechanisms through which auditory distraction can occur. In this sense, it is more useful to consider a hypothetical model that can explain the data from reading tasks by taking into account the contribution of these theories.

One such framework could be a two-component model in which noise and speech influence reading through separate processes. In the first component, background noise would cause a small decrement in comprehension. The present data cannot fully explain why this disruption by noise occurs and more research is needed to understand this mechanism. There was some evidence that noise exhibiting greater acoustic variation is associated with greater distraction (see Jones et al., 1992), but other potential mechanisms need to be explored as well. The second component would cause greater decrements in comprehension from intelligible speech (see Marsh et al., 2008; Martin et al., 1988). Recent evidence suggests that the cognitive process of trying to analyse the meaning of the speech may be enough to cause distraction (Hyönä & Ekholm, 2016). Whether the semantic content and semantic representation of the speech sound are processed and cause additional distraction is an open question that needs to be explored in future research. This second component would also account for the effect of background music. This is because the present results suggest that distraction by background music is effectively reduced to distraction from the sung lyrics, since music without lyrics was not found to be distracting (see Figure 6b).

The predictions of this model could be further tested through future experimental work. For example, previous research has mostly focused on measuring differences in reading comprehension, while only few studies to date have used reading speed as a dependent variable. This in turn did not make it possible to evaluate the model based on this measure. However, the two-component model would make the same prediction for reading speed: background noise should lead to a modest decrease in reading speed, while intelligible background speech should lead to a greater decrease in reading speed due to interference from semantically processing the speech. Measuring eye-movements during reading could also provide a more detailed view of auditory distraction because eye fixations are sensitive to the ongoing cognitive processing of the text (see Rayner, 1998). For example, no studies to date have examined how acoustical or environmental noise may affect fixation durations or fixation probabilities during reading. If the assumption of the first component of the model is correct, there should be an increase in either fixation durations or the number of fixations when readers are exposed to noise in the background.

A stronger test of semantic interference by intelligible speech (i.e., the second component of the model) would be to study two participant populations with the same speech sounds. For example, monolingual speakers of French should be distracted by French speech (intelligible), but not by the same speech, translated into and spoken in a foreign language, such as German (unintelligible). Conversely, monolingual speakers of German should be distracted by the same German speech (intelligible), but not by the French speech (unintelligible). If the magnitude of auditory distraction by intelligible speech is the same in the two populations, this would provide strong evidence for semantic interference by background speech. Additionally, lyrical music has only rarely been used to study distraction due to semantic interference. For example, the proposed model predicts that a lyrical song in the participants' native language would cause distraction because the lyrics are intelligible, while the same song in a foreign language would not cause distraction because the lyrics are unintelligible (see Chew, Yu, Chua, & Gan, 2016). Likewise, the model predicts that an instrumental version of the same song would also not cause any distraction. Another promising avenue would be to investigate distraction by intelligible speech and lyrical music in second language learners in order to determine the role of language proficiency in semantic interference. This could be done by having participants read a text in their native language while listening to background speech in their second language. The second component of the model predicts that distraction will increase as a function of language proficiency because more proficient speakers of the second language would be better at semantically processing the background speech.

2.4.2. Practical Implications

The present results also have some practical implications for settings where readers are exposed to distracting background sounds. For example, there is evidence that listening to music when studying or working is a commonplace. In one survey, university students reported listening to music 62% of the time when studying or doing homework (David, Kim, Brickman, Ran, & Curtis, 2015). Additionally, Calderwood, Ackerman, and Conklin (2014) found that 59% of university students played music in the background when they were asked to study as they normally do. There is also some evidence that listening to music at work is common, with 80% of employees reporting that they listen to music during working hours (Haake, 2006). In this sense, there are many situations in daily life in which people can choose to listen to music while doing reading-related tasks. The present results have direct implications for reading in educational and work settings because they suggest that listening to lyrical music should be avoided when reading a text for comprehension. This is because lyrical music contains intelligible language in the form of sung lyrics, and this type of music was found to be disruptive to reading comprehension. Instead, readers can avoid this disruption by listening to non-lyrical (i.e., instrumental) music because it does not contain any intelligible language.

In the two-component model outlined above, intelligible lyrical music and intelligible speech are assumed to be equally distracting. In fact, intelligible background speech is often present in many work settings, particularly in open-plan offices and other shared areas that have poor acoustic privacy (e.g., Haapakangas, Hongisto, Eerola, & Kuusisto, 2017; Haapakangas, Hongisto, Hyönä, Kokko, & Keränen, 2014; Schlittmeier & Liebl, 2015). The present results suggest that intelligible speech is likely to impair performance on office tasks that require reading for comprehension, proofreading or processing the meaning of written information. Because of this, limiting the amount of intelligible speech in open-plan offices is likely to improve reading performance among office workers. In cases where this is difficult to achieve for practical reasons, acoustically masking the background speech (e.g. with natural sounds) might be helpful as this will decrease its intelligibility and therefore its negative impact (Haapakangas et al., 2011; Jahncke, Björkeholm, Marsh, Odelius, &

Sörqvist, 2016; see also Hongisto, 2005). Furthermore, the present results and the proposed model also suggest that readers exposed to background noise will likely incur a modest cost in terms of reduced comprehension. This suggests that external environmental noise should be limited in settings where reading is common, such as in schools or in libraries. Finally, the practical implications of the present findings would apply equally to both adults and children because the two groups did not generally differ in terms of auditory distraction during reading.

2.4.3. Limitations

While meta-regression is a very useful tool for testing how auditory distraction differs between background sounds or age groups, the present results are only observational in nature (Thompson & Higgins, 2002). Therefore, direct evidence from laboratory experiments and direct comparisons between the different factors are required to verify these results. Nevertheless, we anticipate that our findings, which are based on all the available evidence, will prove to be very useful in guiding future experimental research and advancing our theoretical understanding of how auditory distraction during reading occurs.

Additionally, some of the meta-regression analyses were based on a small number of studies. However, this is not necessarily a limitation in the Bayesian approach that we have adopted here because the results simply reflect our best understanding of auditory distraction effects given the currently available data. Once more data is available, the present results can be easily updated via Bayes' theorem, which will lead to an even more precise estimate of the effects.

2.4.4. Future Directions

The present study grouped background sounds into broad categories, such as noise, speech, or music. However, real-word sounds that readers are routinely exposed to do not always belong to only one of these categories. Rather, different sounds may be present at the same time, such as music playing from a TV, background speech from a nearby conversation, and environmental noise from nearby traffic. Currently, there is a limited understanding of how different types of sounds may interact to increase or decrease distraction. For example, there is some evidence that acoustical noise, when intermixed with background speech, can reduce the negative impact of the speech sound by reducing its intelligibility (Haapakangas et al., 2011; Hongisto, 2005; Venetjoki et al., 2006). Therefore, more research is needed to investigate sounds that are more complex and thus more realistic of auditory distraction in the real world. Additionally, previous research has not investigated the behavioral aspects of auditory distraction: for example, whether participants' motivation and goals can influence how distracted they are by different background sounds during reading.

Another question that deserves more attention is how auditory distraction may differ between age groups. Studies with adults and children have usually been done in isolation, which makes it challenging to assess how these groups differ under the same experimental conditions. The present meta-regression analyses are arguably the only possible way of addressing this question with the currently available data. However, experiments directly comparing adults and children are needed to make firm conclusions. Traditionally, a lot of research has focused on large-scale epidemiological studies of chronic exposure to noise in schools such as the RANCH (Stansfeld et al., 2005) and West London studies (Haines, Stansfeld, Job, Berglund, & Head, 2001; Haines et al., 2001). Because of this, surprisingly little is known about the effect of experimental exposure to noise on reading in children. Eye-movement recordings may be particularly helpful in studying this topic as they can reveal subtle auditory distraction effects that may not appear in behavioural measures such as comprehension accuracy (Cauchard et al., 2012; Hyönä & Ekholm, 2016; Yan et al., 2017). Longitudinal studies of reading development have already made successful use of eye-tracking to study processes such as the development of the perceptual span (Sperlich, Meixner, & Laubrock, 2016), and this method also holds promise in understanding how children's susceptibility to distraction may change during the school years and beyond.

Eye-tracking technology and ERP recordings are useful methods because they can provide rich data about the time course of auditory distraction effects during reading. We anticipate that this type of evidence will be crucial for gaining a better understanding of when and how these effects occur, and what their theoretical nature is. The field of eye-movements during silent reading has already seen the successful development of advanced computational models such as the E-Z Reader (Reichle, Pollatsek, Fisher, & Rayner, 1998) and SWIFT (Engbert et al., 2005), which can simulate many empirical findings. Similarly, a more precise quantification of the time course of auditory distraction effects can move the field forward by making it possible to build computational models that can simulate these processes and to generate new predictions.

2.4.5. Conclusion

Auditory distraction during reading has been a topic of interest for the last 80 years and, as the surge of recent publications shows, it is likely to continue to be an active area of research in the future. The present study was the first attempt to make a comprehensive statistical synthesis of auditory distraction effects in a reading task. The results showed that background noise, speech and music are almost always distracting, even if the distraction effects are often small in size. Sounds that contain intelligible language (i.e., speech or lyrical music) were particularly distracting, most likely due to their semantic properties that interfere with processing the written text. The present findings also have some practical implications. For example, they suggest that listening to instrumental music while reading does not affect the comprehension of the text, whereas listening to lyrical music does. Additionally, readers exposed to background noise would likely incur a cost in terms of reduced comprehension, even if this cost is very small. Finally, the recent interest in measuring eye-movements during reading in the presence of background auditory input heralds the emergence of a new sub-field that may give an even more precise understanding of how and when auditory distraction occurs. The rest of this Thesis will focus on gaining a better understanding of how background speech, noise, and discrete auditory distractors affect eye-movements during reading.

CHAPTER 3: THE EFFECT OF INTELLIGIBLE SPEECH ON LEXICAL PROCESSING DURING SENTENCE READING

Previous studies investigating the effect of background speech and noise on reading comprehension have painted a mixed picture (see Chapter 1 for an overview). For example, while some of them have found that intelligible speech is detrimental to reading and proofreading performance (Jones, Miles, & Page, 1990; Martin et al., 1988; Sörqvist, Halin, & Hygge, 2010), others have failed to find such an effect (Haka et al., 2009; Landström, Söderberg, Kjellberg, & Nordström, 2002; Ljung, Sörqvist, & Hygge, 2009; Venetjoki, Kaarlela-Tuomaala, Keskinen, & Hongisto, 2006). Similarly, studies on the effect of acoustical noise on reading in adults have also resulted in mixed findings. Some of them have found no evidence that acoustical noise is detrimental to reading comprehension (Gawron, 1984; Jahncke, Hygge, Halin, Green, & Dimberg, 2011; Veitch, 1990), while others have found that it can be detrimental to some people depending on their personality characteristics (Furnham, Gunter, & Peterson, 1994; Ylias & Heaven, 2003). Therefore, the evidence from behavioural studies is inconclusive, but it suggests that at least some sounds may be disruptive to reading.

One limitation of behavioural studies is that they have focused only on the end product of reading (i.e., comprehension). However, recording participants' eye-movements makes it possible to investigate how the reading process unfolds in time and to uncover subtle auditory distraction effects that may not be apparent in comprehension measures. A better understanding of the time course of these effects is also crucial for developing theoretical frameworks that can explain how auditory stimuli interfere with the reading process. While theories of semantic and phonological distraction make very specific predictions about the types of speech sounds that should disrupt reading, these predictions are mostly descriptive in nature and they do not tell us which aspects of the reading process are affected. Therefore, eye-tracking evidence has the potential to advance our theoretical understanding of auditory distraction by making it possible to formulate more precise and quantitative predictions in a reading task.

Previous research has indicated that background speech has a direct influence on eye-movements during reading (Cauchard et al., 2012; Hyönä & Ekholm, 2016; Yan et al., 2017). However, it is currently not well understood whether background speech influences the early stages of word processing or if its effect is constrained only to the later processes of sentence integration. Additionally, it is not clear what properties of the speech sound give rise to the distraction in eye-movements during reading. In the present experiment, we sought to answer two research questions: 1) Is the disruption by intelligible speech in eye-movements semantic or phonological in nature (or a combination of the two)? and 2) Does intelligible speech influence the lexical processing of words?

The few available eye-tracking studies to date have provided the first clues as to how intelligible speech may disrupt reading. With the exception of Hyönä and Ekholm's (2016) Experiment 1, all previous studies seem to suggest that intelligible speech leads to an increase in re-reading fixations. However, it is not immediately clear what properties of background speech give rise to the disruption. For example, it is not known whether the disruption is due only to the semantic properties of speech, or if phonology also plays a role.

The statistical synthesis of previous findings from behavioural experiments in Chapter 2 did not find evidence for phonological distraction in measures of reading comprehension, but this result may not necessarily extend to measures of eye-movements during reading. Additionally, the comparison between the phonological and semantic distraction accounts from Chapter 2 was based on a small number of studies. Therefore, more evidence from eye-movements during reading is required to investigate the role of phonology in distraction by intelligible speech.

While the manipulation in Hyönä and Ekholm's (2016) Experiment 1 could make the theoretical distinction between phonological and semantic distraction, the authors reported no disruption by either intelligible or unintelligible speech. Therefore, their results did not provide support for either the phonological or semantic disruption hypothesis. One possible explanation for this finding is that the foreign (i.e., unintelligible) speech material used in their study was taken from a language course, while the native speech was an excerpt from a novel. Therefore, the lack of a statistically significant difference may have occurred because the two speech sounds potentially differed in properties such as intonation, content, and rate of speech. The present research made a more stringent test of the semantic and phonological disruption theories by using intelligible and unintelligible speech that are more closely matched on these variables, and by including an acoustical noise condition that contains no phonological information but that has an amplitude spectrum similar to that of speech.

Additionally, there is conflicting evidence about which stages of the reading process are influenced by intelligible speech. For example, Hyönä and Ekholm (2016) reported that the effect of scrambled speech was mostly evident in re-reading fixations, while Yan et al. (2017) observed the same effect for intelligible speech. These findings suggest that the effect of background speech is mostly evident in second-pass reading measures. However, Cauchard et al. (2012) also reported an effect on gaze durations, and Yan et al. found that intelligible speech eliminated the frequency effect for first fixation durations. The last two findings seem to suggest that the early stages of word processing may also be affected. If the initial processing of words is disrupted, this may occur because the semantic properties of speech interfere with accessing the lexical information of words. This is an important theoretical question that has not been addressed in an alphabetical language before.

The first goal of the present experiment was to investigate whether the phonological or semantic properties of speech (or a combination of the two) is responsible for the disruption by intelligible speech in eye-movements during reading. We used a paradigm in which participants read single sentences that were presented concurrently with the sounds. Importantly, participants heard the sound stimuli only for the duration that they were actually reading, thus reducing potential habituation effects (Banbury & Berry, 1997). Additionally, the speech stimuli were carefully matched and consisted of single declarative sentences that were unrelated to each other. This was similar to the reading stimuli, which also consisted of unrelated declarative sentences. Furthermore, only naturally-occurring speech was used (i.e., without any scrambling) and this speech was spoken at a consistent rate throughout the whole experiment. Finally, because participants' comprehension was assessed immediately after reading a sentence, it was possible to test whether background sounds have an immediate effect on reading comprehension. This is an important question because most behavioural studies to date have had a delay between reading the text and the subsequent comprehension assessment (e.g., due to other tasks intervening in between;

Martin et al., 1988) and any observed differences may not be due to deficits in immediate text comprehension (see Sörqvist et al., 2010).

The present study had four background sound conditions that made it possible to differentiate between the phonological and semantic distraction accounts: Gaussian noise filtered to have an amplitude spectrum similar to that of long-term average speech (referred to as 'speech-spectrum noise'), Mandarin speech, English speech, and silence (the control condition). According to the phonological distraction account (Salamé & Baddeley, 1982, 1987) irrelevant speech should disrupt the ongoing reading process regardless of whether it is intelligible or unintelligible because it automatically gains access the phonological store of working memory capacity. However, speech-spectrum noise would not be expected to cause such disruption because it would not gain access to the phonological store. Therefore, if the disruption is phonological in nature, we would expect English speech to be more distracting than speech-spectrum noise, but equally as distracting as Mandarin speech. On the other hand, if the disruption is semantic in nature (Marsh et al., 2008, 2009; Martin et al., 1988), we would expect English speech to be more distracting than Mandarin speech because participants can understand the former language but not the latter.

It should be noted that Mandarin phonology differs from English phonology in a number of ways, such as the use of distinct tones, the smaller number of syllables, the lack of polysyllabic words, and the high number of homophones (Duanmu, 2006). Nevertheless, the phonological distraction account (Salamé & Baddeley, 1982, 1987) does not predict that the disruption occurs due to phonetic similarity between the irrelevant speech sound and the text that participants are reading. Rather, it occurs because the irrelevant speech gains access to the phonological store of working memory where it interferes with the storing and rehearsal of the phonological information of the text that participants are reading. Therefore, the actual language of the irrelevant speech is often not thought to be of critical importance. In fact, phonological distraction in behavioural experiments has been observed with a range of different languages, including Arabic (Baddeley & Salamé, 1986; Salamé & Baddeley, 1987), German (Colle & Welsh, 1976), Russian (Klatte, Lee, & Hellbruck, 2002), and Japanese (Ellermeier & Zimmer, 1997), to name a few.

One possibility is that the disruption by intelligible speech is not either entirely semantic or entirely phonological in nature, but rather a combination of the two. To test for this possibility, we will distinguish between two versions of the phonological distraction account. In the strong version, any distraction effects are attributed to phonology alone. As a result, English speech should be more distracting than Noise but equally as distracting as Mandarin speech. In the weaker version of the theory, phonology is responsible for some, but not all of the distraction effects. Therefore, the weaker version of the theory predicts that Mandarin should be more distracting than speech-spectrum noise (indicating some contribution of phonology), but less distracting than English speech (indicating that the rest of the disruption effect can be attributed to semantic interference).

The second goal of the present experiment was to test whether intelligible speech interferes with the lexical processing of words. Yan et al.'s (2017) study suggests that intelligible speech may disrupt lexical processing in readers of Mandarin, but, interestingly, this effect was found only in first fixation durations. This suggests that the disruption of lexical access by intelligible speech is limited only to the very first fixation on words. In the present experiment, we tested whether lexical processing is affected in readers of English by manipulating the lexical frequency of a target word in each sentence. Previous research has shown that lower frequency words are fixated longer than higher frequency words (Inhoff & Rayner, 1986; Rayner, 2009). Therefore, as the frequency effect reflects the difficulty inherent in the lexical access of words, the present study tested whether intelligible background speech interferes with lexical access. For example, in any model of word identification where word representations accrue activation constantly (e.g., Morton, 1969; Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001), we might expect that English speech makes it harder to accumulate activation in order to identify a word compared to the other sound conditions. In this case, we should find a stronger word frequency effect in this condition compared to the other background input conditions because low frequency words require more activation for lexical access than high frequency words. In this sense, we would expect the disruption effect of intelligible English speech to be greater for low frequency words than for high frequency words.

3.1. Summary of Predictions

The following predictions were tested in the present experiment:

H1: If the disruption by intelligible speech is entirely phonological in nature, English speech should be more distracting than Silence and Noise, but equally as distracting as Mandarin speech (strong form of phonological interference).

H1.2: If the disruption by intelligible speech is only partially phonological in nature, Mandarin speech should be more distracting than Noise (weaker form of phonological interference).

H2: If the disruption by intelligible speech is entirely semantic in nature, English speech should be more distracting than Silence, Noise, and Mandarin; additionally, prediction H1.2 above should not be supported by the data (strong form of semantic interference).

H2.1: If the disruption by intelligible speech is a combination of semantic and phonological interference, English speech should be more distracting than Silence, Noise, and Mandarin; additionally, prediction H1.2 above should also be supported by the data (combination of phonological and semantic interference).

H3: If intelligible speech interferes with the lexical access of words, there should be greater disruption by English speech for low frequency compared to high frequency words.

Based on the available evidence (e.g., Hyönä & Ekholm, 2016; Yan et al., 2017), we expected to find support for predictions H2 and H3 above.

3.2. Method

3.2.1. Participants

Forty university students (28 female) participated for course credit or a payment of £8. Their mean age was 22.4 years (*SD*= 5.2 years; range: 18-40 years). All participants were native speakers of British English, reported normal or corrected-to-normal vision, normal hearing, and no prior diagnosis of reading disorders. Participants were naïve as to the purpose of the experiment. None of them had any knowledge of Mandarin Chinese. Ethical approval for the study was obtained from the Bournemouth University Research Ethics Committee (protocol No. 11663).

The statistical power of study design was 0.831 for an average effect size of d= 0.47 based on the method described in Westfall (2015). The expected value of d= 0.47 was determined by calculating the effect size for all disruption effects by background speech reported in Hyönä and Ekholm (2016) and then taking their average. As the current power exceeds the recommended value of 0.80 (J. Cohen, 1988), the present experiment was sufficiently powered to detect auditory disruption effects by background speech.

3.2.2. Materials

3.2.2.1. Sentence stimuli. The reading material consisted of 128 English sentence frames (see Figure 9b for an example and Appendix D for a complete list). Their average length was 13.2 words. Each sentence frame had a target word position which could contain either a low-frequency or a high-frequency word (picked using the SUBTLEX-UK database; Van Heuven, Mandera, Keuleers, & Brysbaert, 2014). The target word was never one of the first or last three words in the sentence frame. The target words were an equal number of adjectives and nouns. High and low frequency target words were matched on word length, bigram frequency, and neighbourhood size using the N-watch software (Davis, 2005). This information is presented in Table 5. Additionally, cloze-task predictability norms (Taylor, 1953) were obtained from 21 undergraduate students who did not participate in the eye-tracking study. High and low frequency target words did not differ significantly in their predictability given the preceding sentence frame, $t(127) = 0.97$, $p = 0.33$.

	High-frequency words				Low-frequency words			
	Mean	SD	Min	Max	Mean	SD	Min	Max
Word length (in letters)	5.6	1.1	3	7	5.6	1.1	3	7
Lexical frequency[1]	160	146	46	779	3	2	0.06	10
Bigram token frequency	1282	925	129	5173	1279	994	83	7050
Neighbourhood size	2.8	3.8	0	22	2.8	3.6	0	20
Predictability	0.01	0.04	0	0.29	0.01	0.03	0	0.24

Table 5. Descriptive statistics for the target words in Chapter 3.

[1]in counts per million.

3.2.2.2. Auditory stimuli. The sound stimuli consisted of three types of sound: speech-spectrum noise, English speech, and Taiwanese Mandarin speech. The English speech was taken from the BKB corpus (Bench, Kowal, & Bamford, 1979). The corpus consists of short sentences spoken in British English that last for about 1-2 seconds (e.g. “The house had nine rooms.”). Thirty-two sound files were created by concatenating seven speech sentences and removing the silence gaps. Each speech sentence appeared only once in the sound files. In half of the speech sound files, the speaker was female; in the other half, the speaker was male. The speech-spectrum noise was created by filtering Gaussian noise by the average amplitude spectrum of the English BKB sentences in male voice.

Thirty-two Mandarin sound files were created in the same way as the English ones. The speech sentences were taken from Kuo (2006), who translated 240 sentences from the BKB (Bench et al., 1979) and IHR (MacLeod & Summerfield, 1990) corpora. Therefore, the Mandarin speech sentences were intended for the same audience and had the same sentence structure as the English ones. The average speech rate in the experiment was matched between the English speech (*M*= 3.16 words per second) and the Mandarin speech (*M*= 3.08 words per second) condition, $t(62)= 1.10$, $p= 0.28$.

The four sound conditions (Silence, Noise, Mandarin, and English) were presented in blocks of 32 sentences. The sentences within each block appeared in random order. The order of the blocks and the assignment of sound conditions to the sentences were counterbalanced with a full Latin square design. The frequency of the target word was also counterbalanced.

3.2.3. Apparatus

An Eyelink 1000 was used to record participants' eye-movements. Viewing was binocular, but only the right eye was recorded. The sampling frequency was 1000 Hz. Participants rested their head on a chin-and-forehead rest. The sound stimuli were administered binaurally through noise-cancelling headphones (Bose QuietComfort 25) at a sound pressure level (SPL) of 59-61 dB(A). The SPL was measured with a RadioShack digital meter (model 33-2055) over a 2-minute interval. The amplitude resolution of the sounds was 32 bits. The sampling frequency was 22 kHz for the English speech and speech-spectrum noise, and 44 kHz for the Mandarin speech. The sounds were played on an Intel HD Audio integrated sound card.

The experiment was run using the EyeTrack 0.7.10h software (Stracuzzi, 2004) on a PC with Microsoft Windows XP. The stimuli were presented on a 20-inch Mitsubishi Diamond Pro 2070 monitor with a screen resolution of 1024 x 768 and a refresh rate of 150 Hz. The sentences were displayed in Courier New 14pt. font and appeared as black text over white background on a single line in the middle of the screen. The width of each letter was

11 pixels. Participants sat 60 cm away from the monitor and at this distance each letter subtended approximately .40º of visual angle.

3.2.4. Procedure

Participants were instructed to focus on what they were reading and to ignore any sounds they may hear. Participants wore the headphones throughout the whole experiment. A three-point calibration of the eye-tracker was performed at the beginning of the experiment and it was then repeated as required. The calibration error was kept at < .30º of visual angle. All beeps during calibration and drift check were turned off. The experiment started with six practice trials, followed by the experimental trials. The trial presentation is illustrated in Figure 9. The experiment lasted for 30-40 minutes.

All trials began with a drift check, after which a black square appeared with a 50-pixel offset from the left edge of the screen. Once participants fixated the square, the sentence was presented, with the first letter of the first word at the centre of the square. The onset of the background sound was simultaneous with the onset of the sentence. Participants used a button on a gamepad controller to terminate the trial once they finished reading the sentence. However, there was a trial timeout that corresponded to the length of the speech sound that was playing. In other words, if a participant did not terminate the trial by pressing a button, the trial ended automatically when the speech sound finished playing. For the English and Mandarin sound conditions, the timeout corresponded to the length of the individual speech files (between 9.2- 12.6 s). The same timeouts were randomly assigned to the sentences in the silence and noise conditions. There was a yes/no comprehension question after 34% of trials. For example, in the sentence "The house was immediately recognisable by its green fence and big windows.", the question was: "Did the house have small windows? Yes/ No".

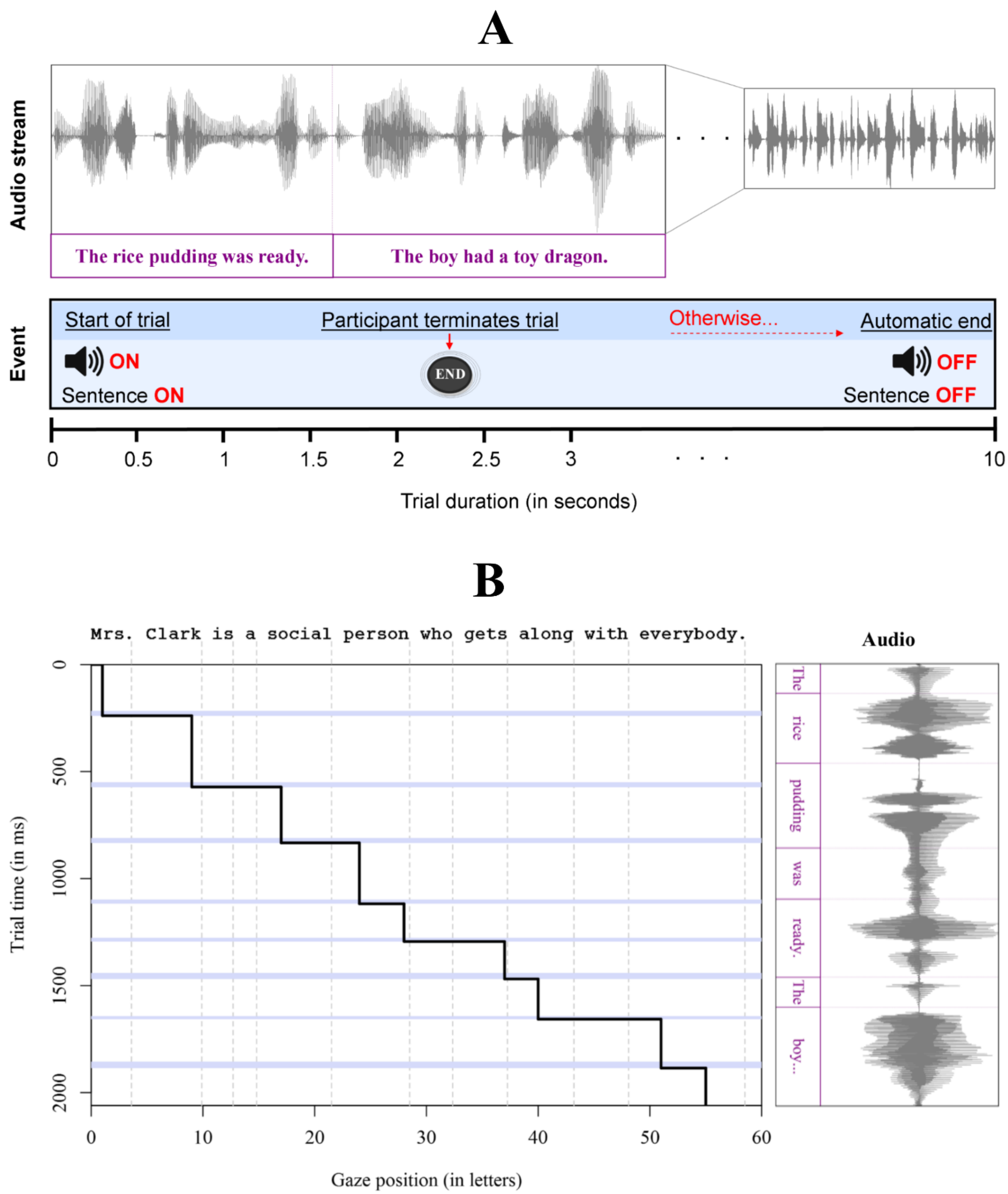

Figure 9. An illustration of the stimuli presentation in Chapter 3. Panel **A** shows the events during the trial and the speech sound that was playing. The sentence and the speech sound

were simultaneously presented at the start of the trial. Trials were normally terminated by the participant by pressing the button. If the participant did not press the button, the trial was automatically terminated when the sound stopped playing. Panel **B** shows the timeline (including gaze position and auditory input) of a sample trial that was terminated by the participant. Horizontal blue lines show the saccades and the right-hand side shows the audio that was playing while they were reading. Vertical dotted lines indicate the word boundaries. In the sample sentence, the target word ("social") is high frequency; in the low frequency condition it was replaced by the word "chatty".

3.2.5. Data Analysis

Several measures of global reading were analysed in the present study: total sentence reading time (the sum of all fixations on the sentence), probability of regression, number of first-pass and second-pass fixations, saccade length, and saccade landing position. In addition to this, the three standard local fixation duration measures were computed for the target word: 1) first fixation duration (FFD; the duration of the first fixation on the word); 2) gaze duration (GD; the sum of all fixations on the word before moving to another word); and 3) total viewing time (TVT; the sum of all fixations on the word, including second-pass reading). FFD and GD are often considered to be first-pass measures of reading because they capture the initial fixations on words during the first, progressive reading of the sentence (Clifton Jr., Staub, & Rayner, 2007). Once readers move to the right of a word or make a regression to previous words, the first-pass reading is considered to have been completed (Rayner, Sereno, Morris, Schmauder, & Clifton, 1989). On the other hand, TVT is often considered to be a measure of second-pass reading as it includes all fixations made during first-pass reading and all fixations made during regressions back to the word (Rayner et al., 1989). Finally, comprehension accuracy was also analysed between the sound conditions.

When analysing data from psycholinguistic experiments, it is important to take into account variability not only across subjects, but also across items. This is because the results may be specific to the language materials used in the study and may not necessarily generalise to language more broadly (H. H. Clark, 1973; Coleman, 1964). Traditionally, this has been addressed by performing separate Analyses Of Variance (ANOVAs) that average observations across subjects (F1) and items (F2). More recently, (Generalised) Linear Mixed Models ((G)LMMs) have been used as an alternative method because they allow participants and items to be specified as crossed random effects (Baayen, Davidson, & Bates, 2008). The advantage of this method is that it makes it possible to account for variability across subjects and items within the same model. Additionally, LMMs are more robust when analysing missing data compared to ANOVAs (Baayen et al., 2008), which is a common scenario when working with eye-movement data.

The data were analysed with (G)LMMs by using the “lme4” package v.1.1-12 (Bates, Machler, Bolker, & Walker, 2014) in R 3.3.0 (R Core Team, 2016). *P*-values for LMM models were calculated with the lmerTest package v.2.0-33 (Kuznetsova, Brockhoff, & Christensen, 2017). Fixation durations were log-transformed in all analyses. Low and high frequency target words were coded as 0.5 and -0.5, respectively. Treatment contrasts with English speech as the baseline were used for the effect of background sound. Additionally, to test whether phonology may account for some, but not all disruption effects, a separate comparison between Mandarin and Noise was done.

The results were adjusted for multiple comparisons using the Holm-Bonferroni (Holm, 1979) correction in order to avoid an increase in Type 1 error probability due to the additional comparison between Mandarin and Noise. Background sound was entered as a fixed effect in the models; frequency was also a fixed effect in the target word analyses. Random intercepts, as well as random slopes for the sound condition were specified for subjects and items (Baayen, Davidson, & Bates, 2008; this corresponds to the “maximum” model for the main variable used for inferences, see Barr, Levy, Scheepers, & Tily, 2013)[4]. Results were considered statistically significant if the adjusted *p*-values were ≤ 0.05. Effect sizes in Cohen’s *d* (J. Cohen, 1988) are reported as a measure of the magnitude of the effects.

3.3. Results

The average trial duration was 3.8 s (*SD*= 1.74 s). There were 0.5% of trials where timeout was reached before participants pressed the end button and these were excluded from the data. Furthermore, 5.2% of the fixation duration data were excluded because of blinks. Additionally, trials in which FFD was above 800 ms, GD was above 2000 ms, or TVT was above 3000 ms were removed as outliers from all analyses (0.1% of data). The number of outliers excluded per condition did not differ significantly (χ^2 (2) = 0.4, p= 0.82). If fixation duration was an outlier in any of the three measures, the whole trial was removed from the analysis. Fixations shorter than 80 ms that occurred within one letter space of another fixation were combined with that fixation.

3.3.1. Comprehension Accuracy

The mean comprehension accuracy in the experiment is presented in Figure 10. There were no significant differences in comprehension accuracy across the sound conditions (all *p*s ≥ 0.20). Auditory speech sounds did not appear to affect comprehension accuracy which remained high across all conditions.

[4] The following random slopes for background sound were removed due to convergence failure: random slope for items for saccade length, GD, and TVT; random slope for both participants and items for regression probability and number of first-pass fixations.

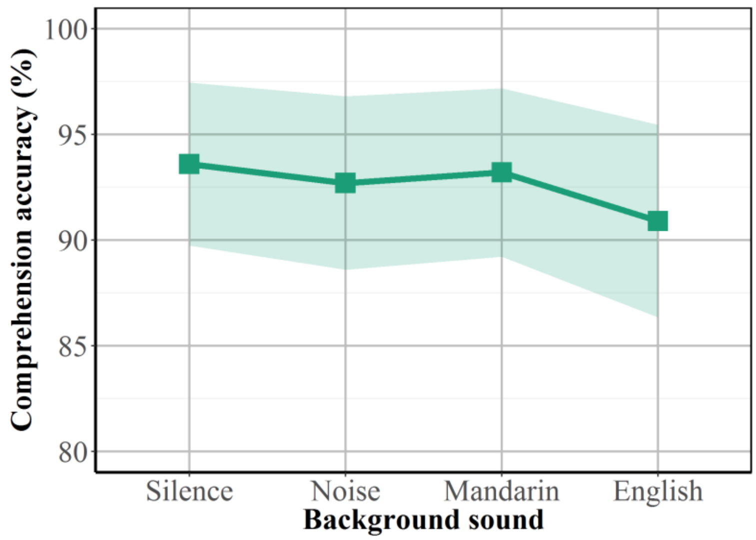

Figure 10. Mean descriptive statistics for reading comprehension accuracy in the background sound conditions in Chapter 3. Shading indicates the standard error.

3.3.2. Global Reading

Descriptive statistics of global reading on the whole sentence are presented in Table 6. The total sentence reading time was significantly longer in English speech compared to Silence (b= -0.07, SE= 0.03, t= -2.52, p= 0.03, d= -0.23), Noise (b= -0.12, SE= 0.03, t= -4.61, p<0.001, d= -0.27) and Mandarin speech (b= -0.06, SE= 0.02, t= -2.61, p= 0.02, d= -0.14). The remaining analyses indicated that this was due to more second-pass fixations in English speech compared to all other sound conditions (Silence: b= -0.24, SE= 0.07, z= -3.36, p= 0.001, d= -0.14; Noise: b= -0.41, SE= 0.08, z= -5.37, p< 0.001, d= -0.18; Mandarin: b= -0.22, SE= 0.05, z= -3.99, p< 0.001, d= -0.10). As is evident from the descriptive statistics in Table 6, there was no difference in the number of first-pass fixations (all ps ≥ 0.80). English speech also resulted in a significantly greater regression probability compared to all other sound conditions (Silence: b= -0.09, SE= 0.02, z= -3.52, p< 0.001, d= -0.03; Noise: b= -0.14, SE= 0.03, z= -5.46, p< 0.001, d= -0.04; Mandarin: b= -0.08, SE= 0.02, z= -

3.31, p= 0.002, d= -0.05). There were no significant differences in saccade length (all ps ≥ 0.35) or word landing position (all ps ≥ 0.13) across the sound conditions.

Sound condition	Total sentence reading time (in ms)	Word landing position (in letters)	Saccade length (in letters)	Regression probability	Number of fixations (per word)		
					1st-pass	2nd-pass	Total
Silence	3040 (1244)	2.81 (2.14)	8.86 (8.11)	.23 (.42)	1.03 (.57)	.48 (.77)	1.51 (.84)
Noise	2960 (1354)	2.86 (2.15)	8.72 (7.69)	.22 (.41)	1.04 (.56)	.44 (.74)	1.48 (.82)
Mandarin	3150 (1426)	2.85 (2.16)	8.91 (8.38)	.23 (.42)	1.04 (.59)	.51 (.82)	1.55 (.92)
English	3370 (1616)	2.86 (2.16)	8.73 (8.15)	.24 (.43)	1.03 (.61)	.62 (.93)	1.65 (1.02)

Table 6. Mean of global reading measures per background sound condition in Chapter 3 (SDs in parentheses).

The planned comparison between Mandarin and Noise indicated that participants made significantly more second-pass fixations in Mandarin speech compared to Noise (b= -0.20, SE= 0.08, z= -2.43, p= 0.02, d= 0.09). However, as Table 6 shows, this effect was in part driven by the slightly better reading performance under Noise compared to Silence. No other differences between Noise and Mandarin speech were significant (all ps> 0.052). In summary, the results supported most strongly hypothesis H2, which stated that the disruption by intelligible speech is only semantic in nature. Hypothesis H2.1, which stated that the disruption has both a semantic and a phonological component, received only weak support because evidence for contribution of phonology was found in only one measure (number of second-pass fixations).

3.3.3. Target Word Analysis

Fixation durations on the target word are shown in Figure 11 and the results of the LMMs are shown in Table 7. There were robust frequency effects on the target word. However, contrary to hypothesis H3, the contrasts between English speech and the

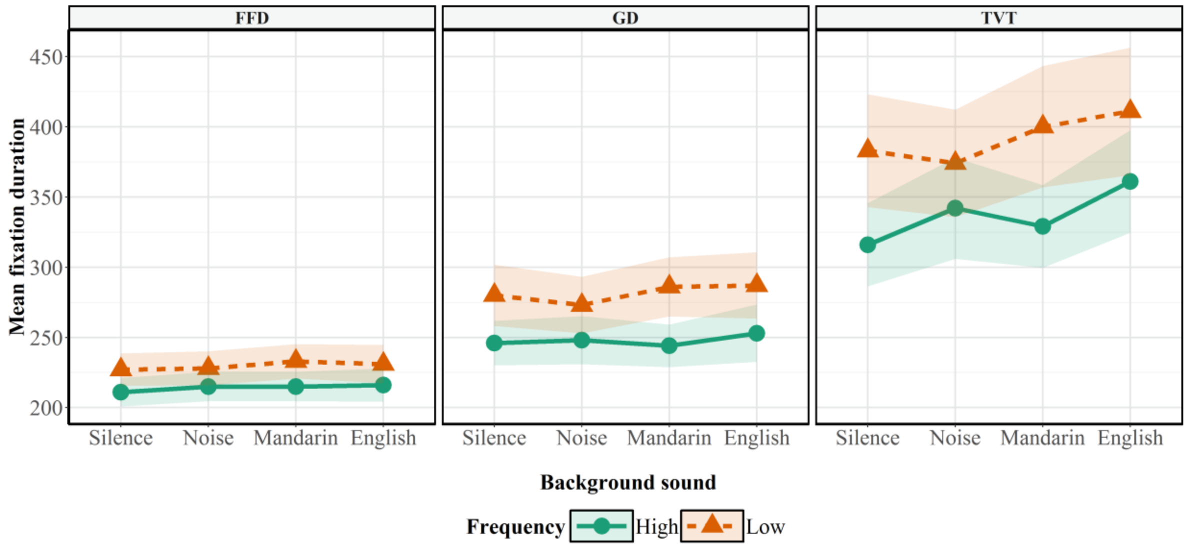

Figure 11. Mean fixation durations on the target word for the different background sound conditions in Chapter 3, broken down by target word frequency. FFD: first fixation duration. GD: gaze duration. TVT: total viewing time. Shading indicates the standard error.

remaining sound conditions failed to interact with target word lexical frequency[5]. Consistent with the results from global reading measures, the effect of English speech was not found on first-pass measures, but only on TVT, which includes re-fixations during second-pass reading. This is because English speech led to a greater number of re-reading fixations. English speech resulted in longer TVT compared to Silence (d= -0.15) and Noise (d= -0.12). The difference between English and Mandarin for TVT (d= -0.09) did not reach significance on the target word, but it was significant in the analysis of all words in the sentence (see Appendix E). No differences between Mandarin and Noise were significant (all $ps \geq 0.16$).

[5] In order to test the possibility that the target word analysis did not have sufficient statistical power to detect an interaction effect, frequency norms were obtained for all words in the sentence. The frequencies were then entered into a model that included all the fixations for all words in the sentence. The results (presented in Appendix E) were consistent with the target word analyses and showed no significant interactions with lexical frequency.

Therefore, the fixation duration analysis also supported hypothesis H2 that the disruption by intelligible speech is only semantic in nature.

Fixed effect	FFD				GD[1]				TVT[1]			
	b	SE	t	p	b	SE	t	p	b	SE	t	p
Intercept	5.35	.02	265.9	**<.001**	5.49	.03	167.2	**<.001**	5.78	.05	119.02	**<.001**
Freq	.05	.02	2.90	**<.01**	.11	.02	4.80	**<.001**	.12	.03	4.33	**<.001**
Eng vs Slc	-.02	.02	-.93	.36	-.01	.02	-.52	.61	-.08	.03	-2.90	**.01**
Eng vs Noise	<-.01	.01	-.08	.93	-.02	.02	-.72	.47	-.07	.03	-2.64	**.02**
Eng vs Mnd	.02	.02	1.06	.36	.01	.02	.38	.70	-.04	.02	-1.46	.30
Freq: Eng vs Slc	.02	.03	.86	.72	.01	.03	.32	.75	.05	.04	1.28	.40
Freq: Eng vs Noise	<-.01	.03	-.03	.98	-.02	.03	-.66	.51	-.02	.04	-.52	.60
Freq: Eng vs Mnd	.02	.03	.87	.72	.03	.03	.79	.43	.02	.04	.53	.60

Table 7. LMMs analysis of fixation duration measures on the target word in Chapter 3. Freq: lexical frequency. Eng: English. Slc: Silence. Mnd: Mandarin. FFD: first fixation duration. GD: gaze duration. TVT: total viewing time. Statistically significant *p*-values are formatted in bold.

[1] Background sound was removed as a slope for items due to convergence failure.

3.3.4. Post-hoc Analysis

Because many of the effects in the present analyses were due, at least in part, to an increase in second-pass fixations, additional exploratory analyses were conducted to investigate where re-reading fixations occurred in the sentence. In this analysis, we compared the number and distance of re-reading fixations that were made after the start of a regression until participants made a progressive fixation (i.e., until they fixated a new word in the sentence that they had not already fixated). To determine the location of re-reading fixations, we calculated their distance (in words) in relation to the most recently fixated word in the sentence before the regression (see Figure 12 for an illustration of the method). If English speech interferes with the integration of recently-read words into the sentence context (i.e., “local” disruption), we would expect re-reading fixations to occur in close proximity to the source of the difficulty, that is, the most recently fixated word in that sentence. In contrast, if this disruption is due to a failure to maintain the representation of the previous part of the sentence in working memory, we would expect that fixations will be more distant from the most recently fixated word, presumably in order to re-activate the previous sentence context.

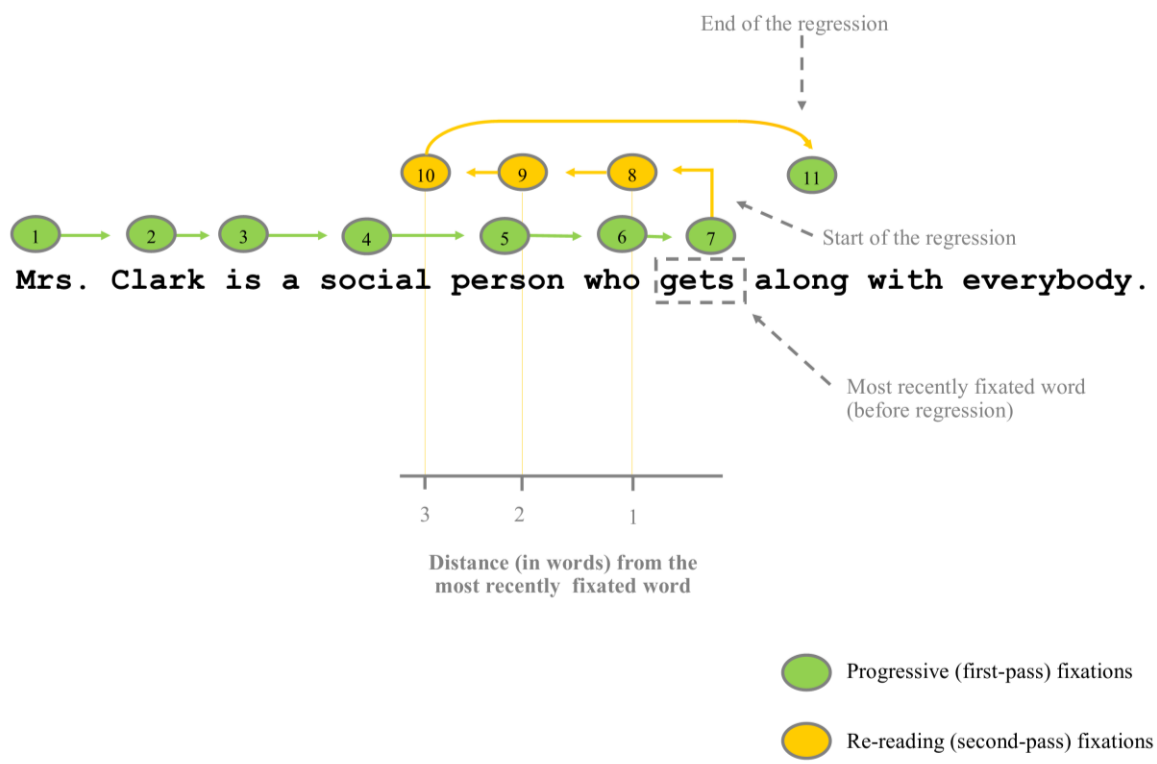

Figure 12. An illustration of the method for calculating the distance of re-reading fixations in Chapter 3 (see Figure 13 below). The distance of each re-reading fixation was determined by how many words participants went back from the most recently fixated word (“gets”) before making a progressive fixation (№ 11). Fixation numbers show the order in which the fixations occurred in the trial.

The results from the analysis are plotted in Figure 13. The number of re-reading fixations decreased with increasing distance from the most recently fixated word in the sentence (*b*= -0.03, *SE*= 0.005, *z*= -6.68, *p*< 0.001). Critically, however, English interacted significantly with distance (*b*= 0.01, *SE*= 0.005, *z*= 2.12, *p*= 0.03), thus showing that the mean difference between English and Silence became smaller with increasing distance. This trend is apparent in Figure 13 where a clear increase in the number of re-reading fixations can be seen only when the distance was five words or less. Therefore, re-reading fixations were mostly constrained to words that were close to the most recently fixated word in the sentence.

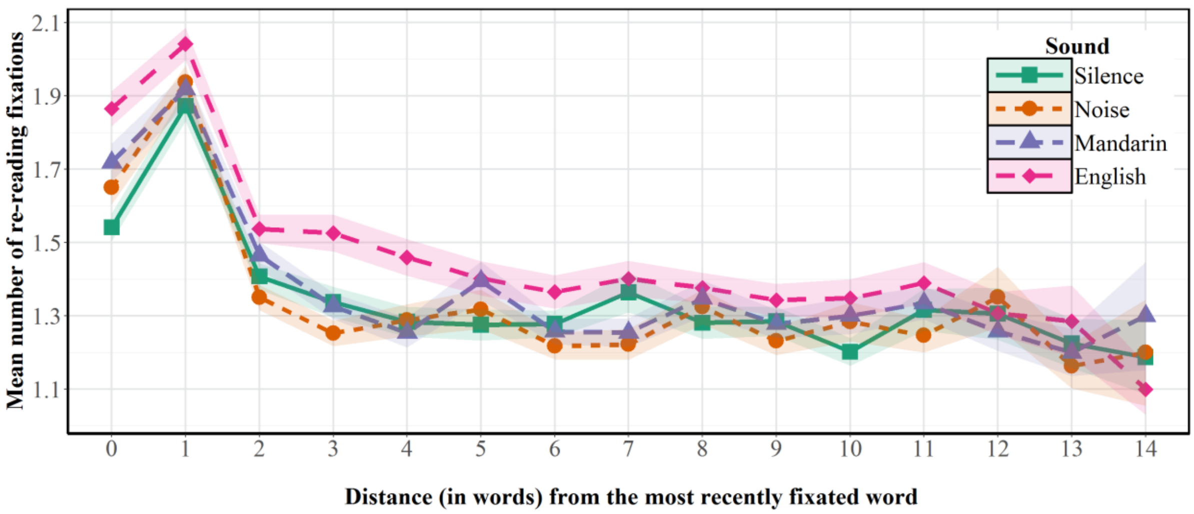

Figure 13. Position of re-reading fixations for the different sound conditions in Chapter 3 as a function of distance from the most recently fixated word. Shading indicates the standard error.

3.4. Discussion

The present experiment investigated auditory distraction effects by intelligible background speech on reading single sentences. There were two main questions of the study: (1) Is the disruption by intelligible speech semantic or phonological in nature (or a combination of the two)? And (2) does intelligible speech affect the lexical processing of words? In terms of the first question, English speech increased the overall sentence reading time compared to silence. This was found to be mostly caused by making more regressions and more second-pass fixations when re-reading words. The present experiment provided support for the theoretical prediction that this disruption effect is semantic in nature (Marsh et al., 2008, 2009; Martin et al., 1988). This was because English speech resulted in longer sentence reading times compared to Mandarin speech, and this arose due to readers making more regressions and more re-reading fixations. Therefore, the present results support the semantic disruption account and are in line with Yan et al.'s (2017) and Hyönä and Ekholm's (2016) Experiments 2-4.

Because English speech was consistently more disruptive than Mandarin speech, this provides evidence against the strong form of phonological disruption view (hypothesis H1), which predicted that any speech sound (intelligible or not) would cause interference because it gains access to the phonological store of working memory capacity (Salamé & Baddeley, 1982). Nevertheless, there was limited support for the view that phonology may account for some, but not all, of the disruption effects (hypothesis H1.2). This was because Mandarin speech led to more second-pass fixations compared to Noise. However, this effect warrants further replication as it was found in only one measure and it was partially driven by the fact that participants made fewer second-pass fixations in Noise compared to Silence. This is especially true because a facilitation effect of acoustical noise has generally not been reported in previous studies (e.g. C. R. Johansson, 1983; R. Johansson et al., 2012; Landström et al., 2002; Martin et al., 1988). Overall, the present results are largely consistent with Hyönä and Ekholm's (2016) Experiment 1, in the sense that the authors did not find any evidence to support the phonological disruption account. Therefore, taken together with Hyönä and Ekholm's (2016) findings, the present results suggest that phonology plays little if any role in auditory distraction by intelligible speech. In this sense, while we acknowledge that there was a hint in the data for a contribution of phonology, the present results are most readily explained by hypothesis H2, which predicted that the disruption effect is only semantic in nature.

The results from global reading measures agree with those of Yan et al. (2017), who also reported longer sentence reading times, more fixations and greater regression probability with intelligible speech in the background. However, the present experiment provided greater insight by showing that the increase in fixations was entirely due to more second-pass (i.e., re-reading) fixations. Additionally, the present results advance our theoretical understanding of disruption by intelligible speech by showing that these effects are due to the semantic content of the speech. Therefore, one of the novel contributions of the present experiment was to show that semantic disruption is observed in eye-movement measures when comparing naturally-occurring speech sounds: English speech, which could be processed semantically by participants, led to greater disruption in second-pass reading measures compared to Mandarin speech, which could not be processed semantically.

The second aim of the present study was to investigate whether lexical processing is affected by intelligible speech. Most importantly, contrary to hypothesis H3, the results indicated that intelligible speech did not make the lexical access of low frequency words more difficult. Indeed, robust frequency effects were observed in all background sound conditions. On the surface, this result may appear to be contrary to Yan et al.'s (2017) finding that intelligible speech eliminated the frequency effect in first fixation duration for Mandarin readers. However, Yan et al. also observed the same effect for meaningless (i.e., scrambled) speech. This in turn argues against disruption to lexical access due to semantic interference because the two speech conditions did not differ between one another.

Therefore, both Yan et al.'s study and the present experiment provide converging evidence that the semantic properties of speech do not affect lexical access of words during normal reading. This suggests that the semantic disruption by intelligible speech occurs in the post-lexical stages of the reading process.

Interestingly, the above-mentioned finding by Yan et al. (2017) still suggests that some property of background speech (other than its meaning) can interfere with lexical identification during the first fixation on words. At present, it is not clear what properties of irrelevant speech were responsible for the disruption in their experiment. Therefore, more research is needed to understand what causes this effect in readers of Mandarin Chinese. Using a lexical decision paradigm might be particularly helpful in studying this topic due to the greater control over the timing of the stimuli. Additionally, testing monolingual Mandarin Chinese and monolingual English speakers with the same speech stimuli in the two languages could help dissociate any cross-language differences.

Experiment 1 also showed that the initial reading of words was not influenced by English speech. This was due to the fact that measures of first-pass reading (FFD and GD) did not show any disruption effects by English speech. This suggests that intelligible speech did not cause an overall slowing down of the initial word processing and it also did not lead to inefficient word identification because the effect was not modulated by word frequency. However, robust disruption effects were found in measures of second-pass reading. This suggests that intelligible speech disrupted reading on a more global level, as participants made more re-reading fixations and more regressions compared to unintelligible (Mandarin) speech.

Evidence from behavioural studies has also shown that intelligible speech can disrupt performance on other tasks, such as free recall, that require the use of semantic processing (Marsh et al., 2008, 2009; Marsh, Perham, Sörqvist, & Jones, 2014; Marsh, Sörqvist, Hodgetts, Beaman, & Jones, 2015; see Marsh & Jones, 2010 for a review). One task that is more similar to reading and also requires the retrieval of concepts from semantic memory is verbal fluency (e.g., retrieving examples of the semantic category “animals”). Consistent with the interference-by-process account, Jones, Marsh, and Hughes (2012) showed that verbal, but not phonemic, fluency is disrupted by intelligible speech. The former task relies on semantic processing, while the latter does not. Interestingly, the present experiment suggests that, unlike verbal fluency, reading is not disrupted at the stage of retrieving word concepts from semantic memory. Rather, this disruption occurs later when participants need to combine the meaning of individual words to comprehend the sentence that they are reading.

The lack of disruption in retrieving word concepts provides support for the interference-by-process account (Marsh et al., 2008, 2009), which stipulates that the nature of the main task determines when intelligible speech is distracting. In the context of verbal fluency, the task is to retrieve word concepts from semantic memory according to a certain rule. In contrast, reading imposes different task demands because retrieving the concepts of individual words is not enough for comprehension- readers also need to combine these concepts to form the meaning of the sentence. Therefore, the present results also hold implications for understanding the effect of intelligible speech on cognition more broadly by showing that the cognitive process that is disrupted by the speech sound depends on the demands of the main task.

The post-hoc analysis of re-reading fixations provided important insight into the nature of the disruption to processing that intelligible speech caused. Even though this analysis was not pre-planned and should only be considered as exploratory, the results suggest that English speech made it more difficult to integrate recently-read words into the sentence context. This was because the increase in re-reading fixations occurred in close proximity to the initial, first-pass fixations on words, presumably, those words that were the source of processing difficulty (i.e., the origin of the regression). Sentence comprehension is assumed to involve the retrieval of concepts from memory that are used to inform and construct the meaning of the sentence in relation to broader general world knowledge. Also, such knowledge is used to generate expectations and understand new concepts, as well as to disambiguate sentential ambiguities (Griffiths, Steyvers, & Tenenbaum, 2007). However, because auditory English speech and written English sentences both convey semantic meaning, it seems likely that the processing difficulty we observed derives from disruption to semantic processes associated with the construction of a representation of sentential meaning.

It seems likely that there are two possible causative accounts for such disruption: it may arise due to competition, or even conflict (i.e., inconsistency) between the two representations of meaning (one deriving from the auditory speech and the other from text reading); alternatively, the processing cost may derive from the cognitive burden associated with processing two, rather than one, sources of sentential meaning. Hyönä and Ekholm (2016) tested the first alternative by presenting scrambled speech that consisted either of the text that participants were reading or of an unrelated text. They found that the two scrambled speech conditions did not differ between each other, which led them to suggest that the observed semantic interference is not due to competing semantic representations between the text and the speech sound. The second interpretation would be consistent with both Hyönä and Ekholm's (2016) results and the interference-by-process account (Marsh et al., 2008, 2009), which predicts that the disruption occurs because both the speech and the written text rely on the same process for analysing meaning.

A further interesting finding from the present experiment was that none of the background sounds impaired participants' comprehension of the sentences. This suggests that, while the efficiency with which readers were able to construct a representation of sentential meaning was reduced, readers were still able to attain an understanding of the sentence that they were reading. This is consistent with previous eye-tracking studies (Cauchard et al., 2015; Hyönä & Ekholm, 2016; Yan et al., 2017), but not with other behavioural studies (e.g. Martin et al., 1988; Sörqvist et al., 2010). Given that there was evidence for semantic disruption in the eye-movement measures, why have none of the eye-tracking studies so far found effects in comprehension accuracy? Indeed, because extracting the semantic content of the sentence is crucial for comprehension, it might be argued that a semantic disruption effect should also be found in comprehension accuracy measures.

One possible way to explain this apparent inconsistency is that the comprehension questions in previous eye-tracking studies may have been quite easy to answer, whereas those from behavioural studies may have been more taxing. Indeed, almost all eye-tracking studies investigating reading share something in common: comprehension assessment is usually carried out through the presentation of questions requiring a binary (e.g., "yes/no") answer, and the average comprehension accuracy is almost always 80% or better. In this sense, it is possible that no difference in comprehension accuracy was found because the questions were not as challenging as those used in behavioural studies. If this is the case, then comprehension accuracy should be disrupted when questions are more difficult and probe a deeper level of text comprehension. An alternative explanation is that the immediate comprehension of short texts is not disrupted by intelligible speech, regardless of the difficulty of questions. If this is the case, then the disruption observed in eye-movement measures must be due to a transient difficulty in processing the meaning of the sentence, which readers can overcome and still achieve approximately the same level of comprehension. These two different possibilities will be tested in Chapters 4 and 5, respectively.

In summary, the present experiment examined what properties of intelligible speech (phonology, semantics, or a combination of the two) give rise to distraction in eye-movements during reading and whether intelligible speech disrupts the early stages of reading by interfering with the lexical access of words. The results provided strong support for the hypothesis that the disruption in eye-movements is semantic in nature, and there was only limited evidence to suggest that there is any contribution of phonology to this effect. Additionally, there was no evidence that intelligible speech interferes with the lexical access of words, which suggests that the disruption occurs after the individual words in the text have been lexically identified. Furthermore, intelligible speech disrupted only second-pass measures of reading, which points to the fact that readers had difficulties integrating the meaning of the sentence that they had just read. Finally, despite the evidence for semantic disruption in second-pass reading measures, comprehension accuracy was not affected. The next two chapters will explore two different potential explanations for the lack of disruption in comprehension accuracy by intelligible speech.

CHAPTER 4: THE EFFECT OF INTELLIGIBLE SPEECH ON COMPREHENSION AND ONLINE INTEGRATION PROCESSES

In Chapter 3, intelligible background speech resulted in an increase in the second-pass reading of single sentences that was characterised by making more regressions and more re-reading fixations on previous words. This increase in re-reading behaviour was found to be caused by semantic interference from the irrelevant speech sound that likely created a temporary difficulty in processing the meaning of the sentence. However, despite the reliable disruption that was observed in eye-movement measures, there was no associated decrease in comprehension accuracy in that experiment. This last result is consistent with previous eye-movement studies that have also failed to find any disruption of comprehension accuracy by intelligible speech (Cauchard et al., 2012; Hyönä & Ekholm, 2016; Yan et al., 2017), but is contrary to some behavioural studies that have found evidence of such disruption (e.g., Baker & Madell, 1965; Halin, 2016; Martin et al., 1988; Sörqvist, Halin, et al., 2010). If the effect of intelligible speech on eye-movements arises from semantic interference, it is not immediately clear why comprehension is not affected given that semantic processing is necessary to achieve an accurate comprehension of the sentence. One possible way to explain the lack of an effect in comprehension accuracy in Chapter 3, as well as in previous eye-movement studies, is that the comprehension questions may have been too easy to answer, which in turn may have prevented the detection of the effect.

In much of the literature on eye-movements during reading, comprehension assessment has often been used as a means of ensuring that participants are reading the text stimuli for comprehension rather than as a dependent measure that is of critical theoretical importance. As a result, many studies, particularly those using a single-sentence reading paradigm, have assessed comprehension with simple two-choice questions (e.g., “Yes/No”) that appear on a subset of the sentence stimuli (typically, after 25-50% of all sentences). However, this type of assessment may not be best suited to detect semantic disruption effects in comprehension accuracy for two reasons. First, the assessment is somewhat infrequent because it typically does not occur after every trial and this can lead to reduced statistical power due to the smaller number of observations per sound condition. Second, because the questions are not demanding and can often be answered just by recognising words or phrases from the sentence (see Wotschack & Kliegl, 2013), they may not always be sensitive enough to distinguish whether participants actually understood the meaning of the sentence. This could be problematic when studying semantic interference by intelligible speech as comprehension may remain unaffected if the questions do not assess a deeper level of text understanding.

Therefore, the first aim of the present experiment was to test whether the lack of disruption in comprehension accuracy in Chapter 3 and in previous eye-tracking studies may be due to the fact that the comprehension questions were too easy to answer. In the present study, short paragraphs were used instead of single sentences because they offer a more ecologically-valid reading task and allow for greater opportunity to construct comprehension questions that are more demanding of readers. In the present experiment, a question difficulty manipulation was added in which participants either answered easy questions that were comparable in their difficulty to those used in Chapter 3 or more difficult questions that required a deeper level of text understanding.

The second aim of the present experiment was to test whether intelligible speech disrupts the integration of information across multiple sentences. The results from Chapter 3 suggested that intelligible speech made it harder to integrate the meaning of individual words in order to construct the meaning of the currently-read sentence because participants made more regressions and more re-reading fixations. However, it is not immediately clear whether this difficulty extends beyond the level of the currently-read sentence. In other words, is the disruption by intelligible speech limited only to the individual sentences that make up the text, or is there additional disruption due to integrating information across multiple sentences?

While there have been previous attempts to answer this question, the evidence so far has been inconclusive. For example, Cauchard et al. (2012) examined whether intelligible speech leads to longer re-reading times when participants regress back to a sentence that has already been read once and exited to the right (i.e., sentence look-back time; see Hyönä, Lorch, & Rinck, 2003). They found that intelligible speech led to significantly longer sentence look-back times compared to the silence baseline. Additionally, the authors reported that the effect in sentence look-back time accounted for 27% of the overall increase in reading time in their experiment (Cauchard et al., 2012). Therefore, this suggests that a non-trivial amount of the increase in re-reading fixations by intelligible speech may be due to a difficulty in integrating meaning across multiple sentences. Interestingly, Hyönä and Ekholm (2016) also analysed sentence look-back times in their experiments. However, they found a difference in look-back times in only one out of four experiments: more specifically, scrambled intelligible speech led to longer look-back times compared to the silence condition in their Experiment 3. While it is not immediately clear why this effect was not found in the remaining three experiments, this discrepancy shows that more evidence is required to understand how intelligible speech may affect the integration of meaning across multiple sentences.

4.1. Question difficulty manipulation

Although most studies of eye-movements during reading have not explicitly considered how different types of comprehension assessment may influence reading patterns, two recent experiments have manipulated the difficulty of questions in order to answer this question (Weiss, Kretzschmar, Schlesewsky, Bornkessel-Schlesewsky, & Staub, 2017; Wotschack & Kliegl, 2013). Wotschack and Kliegl (2013) were first to manipulate the difficulty and frequency of comprehension questions in a single-sentence reading paradigm. In their experiment, participants were assigned to two question difficulty groups in a between-subjects design[6]. In the easy-question group, participants' comprehension was assessed with a three-choice question that could typically be answered by visual word recognition alone. In the difficult-question condition, participants' comprehension was also assessed with a three-choice question, but the answers had less verbatim overlap with the sentence, thus making it harder to find the correct answer (Wotschack & Kliegl, 2013). In addition to this, the comprehension questions appeared after 27% of the sentences in the easy-question condition, but after 100% of the sentences in the difficult-question condition. Therefore, the comprehension questions in the difficult-question condition were both more

[6] The study also had two groups of participants (young adults and older adults). Each participant group was divided into the two question difficulty groups, thereby yielding a 2 (age: young vs older adults) x 2 (question difficulty: easy vs difficult) between-subjects design. In the present discussion, we will only consider the results from the young adult group since this corresponds to the participant population used in this Thesis and in previous studies on auditory distraction. The results from the older adult participant group largely agreed with the ones of the younger adult group, with the exception that the question difficulty manipulation affected first-pass reading more strongly in older adults than in young adults.

frequent and harder to answer. Wotschack and Kliegl (2013) found that participants in the difficult-question condition made fewer first-pass single fixations, but, at the same time, they also made more regressions and had longer total viewing time compared to the easy-question group. These results show that task demands (operationalized as the difficulty and frequency of questions in this study) can have an influence on eye-movements during reading.

In a similar study, Weiss et al. (2017) also divided participants into two question difficulty groups. The reading stimuli in this study consisted of three distinct types of sentences: 1) sentences that contained semantic reversal anomalies; 2) relative clause sentences; and 3) garden path sentences. In this study, the difficult questions required participants to make the correct thematic assignment of the noun phrases in the sentence, whereas the easy questions did not require that (Weiss et al., 2017). However, unlike Wotschack and Kliegl's (2013) study, the frequency of questions did not differ between the two difficulty conditions. Weiss et al. (2017) found that participants in the difficult-question group had longer go-past times and were more likely to regress back to previous parts of the sentence when they came close to the sentence’s end. These findings are consistent with Wotschack and Kliegl's (2013) study, which also showed that difficult questions led to an increase in second-pass reading. However, unlike Wotschack and Kliegl's (2013) study, Weiss et al. (2017) did not find any influence of question difficulty on first-pass reading measures.

4.2. Present Study

Similar to the two studies above, the present experiment also had two question difficulty conditions: 1) an easy condition in which the questions could typically be answered by recognising words and phrases from the text; and 2) a difficult condition in which the answers to the questions were paraphrased and participants needed to understand the meaning of the whole paragraph to find the correct answer. In the present study, the frequency of comprehension questions was kept the same in the two difficulty conditions. The easy condition was comparable to the comprehension questions from Chapter 3. The difficult condition required understanding the meaning of the main topics in the paragraph and making inferences based on that meaning. If English speech affects only participants' understanding of the meaning of the text but not their ability to answer questions based on recognising words and phrases from the text, there should be an interaction between English speech and question difficulty, with greater disruption in comprehension accuracy on the difficult compared to the easy questions.

While the question difficulty manipulation was modelled after Wotschack and Kliegl's (2013) study, it should be mentioned that the two question difficulty conditions in the present study corresponded to different levels of text comprehension. For example, in the model proposed by Kintsch (1998), text comprehension occurs at different levels, such as the processing of individual words (linguistic level), the forming of propositional and syntactic relations between words (microstructure level), and the recognition of global topics of meaning and their interrelationships in the text (macrostructure level; Kintsch & Rawson, 2005). The easy question condition corresponded to text comprehension at the linguistic level since it was largely related to processing the meaning of individual words or short phrases. On the other hand, the difficult question condition roughly corresponded to the macrostructure level because it required understanding the main topics in the paragraphs and making inferences based on this information. However, it should be noted that, because the present paragraphs were relatively short, their macrostructure was not very complex. As a result, the macrostructure of the text was mostly assessed with questions that required integrating the meaning of the individual sentences that made up the paragraph.

The same four background sound conditions were used as in Chapter 3: Silence, speech-spectrum noise, Mandarin speech and English speech. Based on the findings from Chapter 3, we expected to observe more re-reading fixations and more regressions when the text was read in the auditory context of English speech compared to both Mandarin speech and silence. Additionally, we expected that English speech would lead to more regressions to previously-read sentences and to longer sentence look-back times. This was because we expected that English speech would disrupt the integration of the currently-read sentence into the context of previously-read sentences, thus prompting participants to re-visit previous sentences more often.

4.3. Predictions

The same predictions of the phonological disruption (Salamé & Baddeley, 1982, 1987) and semantic disruption theories (Marsh et al., 2008, 2009; Martin et al., 1988) from Chapter 3 were again tested in the present experiment:

H1: If the disruption by intelligible speech is entirely phonological in nature, English speech should be more distracting than Silence and Noise, but equally as distracting as Mandarin speech (strong form of phonological interference).

H1.2: If the disruption by intelligible speech is only partially phonological in nature, Mandarin speech should be more distracting than Noise (weaker form of phonological interference).

H2: If the disruption by intelligible speech is entirely semantic in nature, English speech should be more distracting than Silence, Noise, and Mandarin; additionally, prediction H1.2 above should not be supported by the data (strong form of semantic interference).

H2.1: If the disruption by intelligible speech is a combination of semantic and phonological interference, English speech should be more distracting than Silence, Noise, and Mandarin speech; additionally, prediction H1.2 above should also be supported by the data (combination of phonological and semantic interference).

Consistent with the results from Chapter 3, we expected that hypothesis H2 would be most strongly supported by the data. Additionally, based on the question difficulty manipulation, the following prediction was made:

H3: English speech should disrupt comprehension accuracy only when participants are answering difficult, but not easy, comprehension questions.

4.4. Method

4.4.1. Participants

Forty-eight Bournemouth University students (33 female) participated for course credit or a payment of £10. Their mean age was 19.8 years (*SD*= 1.7 years; range: 18 - 27 years). None of them had participated in the experiment from Chapter 3. Participants were native speakers of British English, reported normal or corrected-to-normal vision, normal hearing, no prior diagnosis of reading disorders, and no prior knowledge of Mandarin Chinese. Participants were naïve as to the purpose of the experiment. Ethical approval for the experiment was obtained from the Bournemouth University Research Ethics Committee (protocol No. 14005). The statistical power of the experiment was 0.859 based on the same average effect size used for the power calculation in Chapter 3 (d= 0.47). This indicates that the experiment was sufficiently powered.

4.4.2. Materials and Design

The reading materials consisted of 24 paragraphs[7]. Each paragraph was four sentences long and had an average length of 89.7 words (SD= 6.2 words; range: 77 to 103 words). The topic of the paragraphs was usually a short description of a person, a place or an event. Real names and specific details were avoided to prevent participants from using their prior knowledge to answer the questions. An example paragraph is provided below:

> Many tourists visiting the land-locked country were not aware of the pristine lake that was situated near its eastern border. Because it was surrounded by a forest and there were no major roads going there, the lake was mostly known only by the locals. However, with its crystal-clear waters and unforgettable scenery, the unspoiled lake was a dream place to relax. According to one local legend, the lake's water had rejuvenating powers and many people from the region would go there in the summer for a swim.

Each paragraph contained two yes/no questions that could be answered by visual word recognition alone ("easy" condition), and two multiple-choice questions with four answers that required understanding the meaning of the paragraph to answer ("difficult" condition). An example of the easy questions is "Did the lake have unforgettable scenery? Yes/No". An example of the difficult questions is:

---

[7] The full set of stimuli is available from the author upon request.

What can be said about the water in the lake?

1) It was murky and shallow

2) It was believed to alleviate stress and chronic medical conditions

3) It was believed to make you feel younger and more energetic

4) It was thought to be suitable for drinking

The answers to multiple-choice questions were paraphrased to prevent participants from recognizing words or phrases from the paragraph in order to find the correct answer (Wotschack & Kliegl, 2011). In easy question condition, one question was based on the first two sentences of the paragraph, and the other question was based on the last two sentences. In the difficult question condition, both questions required understanding the main topics of meaning in the paragraph and making inferences based on that meaning.

Ten undergraduate students who did not take part in the eye-tracking experiment participated in a pilot study in which they read the paragraphs, answered the comprehension questions, and rated the difficulty of questions on a scale from 1 (easy) to 5 (difficult). Each of the comprehension questions appeared on a separate screen and participants could not go back to re-read the text to help them answer the questions. The two question difficulty conditions were presented in separate blocks that were counterbalanced across participants. Because the easy questions had only two answers and the difficult questions had four answers, participants' comprehension was analysed as accuracy above chance level. This controlled for the difference in chance level performance between the easy (50%) and difficult (25%) questions. Comprehension accuracy was significantly better on the easy (*M*= 43.7.8 %; *SD*= 16.6%) compared to the difficult questions (*M*= 31.2 %; *SD*= 34.6%), *t*= -

0.06., *SE*= 0.02, *t*= -3.72, *p*< 0.001. This shows that participants understood the paragraphs sufficiently well in both question difficulty conditions. Additionally, questions in the difficult condition (*M*= 2.70; *SD*= 1.33) were rated as significantly more difficult than questions in the easy condition (*M*= 1.61; *SD*= 1.02), *b*= 1.06, *SE*= 0.09, *t*= 10.98, *p* < 0.001. Finally, participants spent more time reading the paragraphs in the difficult questions' block (*M*= 34.8 s; *SD*= 14.48 s) compared to the easy questions' block (*M*= 30.9 s; *SD*= 10.13 s), *b*= 3.48, *SE*= 1.43, *t*= 2.43, *p*= 0.01.

The speech stimuli were taken from the same two corpora used in Chapter 3 (Bench et al., 1979; Kuo, 2006). Six English and six Mandarin sound files were created by concatenating 40 unique speech sentences; each speech file lasted for at least 60 s[8]. Silence gaps were removed to create a continuous stream of speech. Half of the files contained speech that was spoken by a female actor and the remaining half contained speech spoken by a male actor. The English and Mandarin conditions were matched on average rate of speech (English speech: 3.09 words per second; Mandarin speech: 3.08 words per second). The same speech-spectrum noise as in Chapter 3 was used.

The two question difficulty conditions were presented in two separate blocks. Within each question difficulty block, the different sound conditions were also blocked. The assignment of paragraphs to conditions and the order of experimental blocks were counterbalanced with a full Latin square design. At the start of each question difficulty

[8] Half of the Mandarin speech sounds were looped for the last 2s because the sentences were not long enough to create 60 s of unique speech. The looped speech was reached on only one trial and the seven fixations that occurred during that time were removed from further analysis.

block, there were two practice paragraphs (read in silence) that were used to introduce participants to the different type of comprehension questions.

4.4.3. Apparatus

Participants' eye-movements were recorded with an Eyelink 1000 at a sampling frequency of 1000 Hz. Viewing was binocular, but only the right eye was recorded. Participants rested their head on a chin-and-forehead rest. Similar to Chapter 3, the sound stimuli had an amplitude resolution of 32 bits and a sampling frequency of 22 kHz (for the English speech and speech-spectrum noise), and 44 kHz for the Mandarin speech. The sounds were played binaurally at 59-61 dB(A) SPL via Bose QuietComfort 25 noise-cancelling headphones. The sounds were played on an Intel HD Audio integrated sound card.

The experiment was programmed using the EyeTrack 0.7.10h software (Stracuzzi, 2004) and was run on a PC with a Microsoft Windows XP operating system. The paragraphs were presented on a 20-inch Mitsubishi Diamond Pro 2070 monitor with a screen resolution of 1024 x 768 and a refresh rate of 150 Hz. The text was formatted in a Courier New 14pt. font and appeared as black text over white background on the screen. The width of each letter was 11 pixels. Participants sat 60 cm away from the monitor and at this distance each letter subtended approximately .40º of visual angle. The paragraphs appeared with a 50-pixel offset on the x axis and 150-pixel offset on the y axis of the screen. The text was double-spaced and aligned to the left. Line breaks occurred at the empty space between words, but with the condition that there should be at least 50 pixels to the right of the last letter on the line. All paragraphs fitted on a single screen. Participants pressed buttons on a gamepad controller to terminate the trial and to answer the comprehension questions.

4.4.4. Procedure

Participants were calibrated on a 9-point calibration grid. The calibration accuracy was monitored with a drift check before each trial and participants were recalibrated whenever that was necessary. The average calibration error was kept at $\leq 0.4^{\circ}$. Each trial started with a black gaze box that appeared at 50 pixels on the x-axis and 150 pixels on the y-axis of the screen. Once participants fixated the box, the paragraph appeared on the screen, with the first letter of the first sentence presented in the middle of where the box was. The onset of the background sound was simultaneous with the appearance of the paragraph on the screen. Each question difficulty block started with the two practice paragraphs. Participants were not informed about the difficulty of the questions prior to the experiment and were simply told that some of them will require a yes/no answer, while others will require a multiple-choice answer. The paragraphs and each of the comprehension questions appeared for a maximum of 60 s on the screen. This duration was determined to be sufficient based on the pilot results. The experiment lasted for about 40-50 minutes.

4.4.5. Data Analysis

A few measures of global reading were analysed: number of first- and second-pass fixations, intra-sentence, inter-sentence regression probability, saccade length, and saccade landing position. In the present experiment, we use the term “intra-sentence” regression to denote the probability of making a regression within the currently-read sentence. This is the traditional measure of regression probability that was reported in Chapter 3 and in most of the existing literature. In contrast, “inter-sentence” regression refers to cases where participants regress to a previously-read sentence. This distinction was introduced to test whether background speech disrupts only the integration of text information within sentences or also integration between sentences. Additionally, sentence re-reading time and sentence look-back time were also analysed. Sentence re-reading time was defined as the sum of all re-reading fixations within the currently-read sentence before the eyes moved on to the next sentence (Liversedge, Paterson, & Pickering, 1998). Sentence look-back time was defined as the sum of all re-reading fixations in a sentence when participants regress back from a subsequent sentence (Hyönä et al., 2003). Furthermore, the three local measures of word reading were also analysed: FFD, GD, and TVT. In the analyses of local reading measures, all words in all sentences were included. Finally, comprehension was analysed as accuracy above chance level between the different sound and question difficulty conditions.

The data were analysed with (G)LMMs by using the "lme4" package v.1.1-12 (Bates et al., 2014) in the R statistical software v.3.3.1 (R Core Team, 2016). *P*-values for the LMM models were computed with the lmerTest package v.2.0-33 (Kuznetsova et al., 2017). Background sound and question difficulty were entered as fixed effects in the models. Random intercepts, as well as random slopes for background sound and question difficulty were specified for both participants and items (Baayen, Davidson, & Bates, 2008; Barr, Levy, Scheepers, & Tily, 2013)[9]. Treatment contrasts were used for the background sound condition (with English speech as the baseline). Sum contrasts were used for the question difficulty condition (-1: easy; 1: difficult). Fixation durations were log-transformed in all analyses. Similar to Chapter 3, an additional comparison between Mandarin speech and Noise was also done to test the weaker version of the phonological disruption account. The

[9] Due to convergence failure, the following random slopes were removed: background sound was removed as a random slope for items for saccade length, number of first fixations, gaze duration, and sentence re-reading time; question difficulty was removed as a random slope for items for inter-regression probability and saccade landing position.

results were again adjusted for multiple comparisons with the Holm-Bonferroni (Holm, 1979) procedure to avoid an increase in Type 1 error probability. The results were considered to be statistically significant if the adjusted p values were ≤ 0.05. Effect sizes in Cohen's d (J. Cohen, 1988) are also reported.

4.5. Results

4.5.1. Comprehension Accuracy

The results for comprehension accuracy are presented in Figure 14. There was a main effect of question difficulty ($b_1 = 0.33$, $SE = 0.03$, $t = 9.89$, $p < 0.001$; $b_2 = 0.33$, $SE = 0.03$, $t = 9.49$, $p < 0.001$; $d = -0.41$), indicating that comprehension was significantly lower on the difficult compared to the easy questions. However, there was no significant difference in comprehension accuracy between English and Silence, English and Noise, or Mandarin and Noise (all ps > 0.12). The difference between English and Mandarin was significant by subjects ($b_1 = 0.06$, $SE = 0.02$, $t = 2.51$, $p = 0.03$), but not by items ($b_2 = 0.06$, $SE = 0.03$, $t = 2.07$, $p = 0.10$). Therefore, there were generally no significant differences in comprehension accuracy between the sound conditions and the hint of an effect in the comparison between English and Mandarin was driven by the slightly better accuracy in Mandarin compared to Silence. There were also no significant interactions between background sound and question difficulty for any of the comparisons (all ps ≥ 0.61). In this sense, there was no support for hypothesis H3, which stated that English speech would disrupt comprehension accuracy only for the difficult, but not for the easy questions.

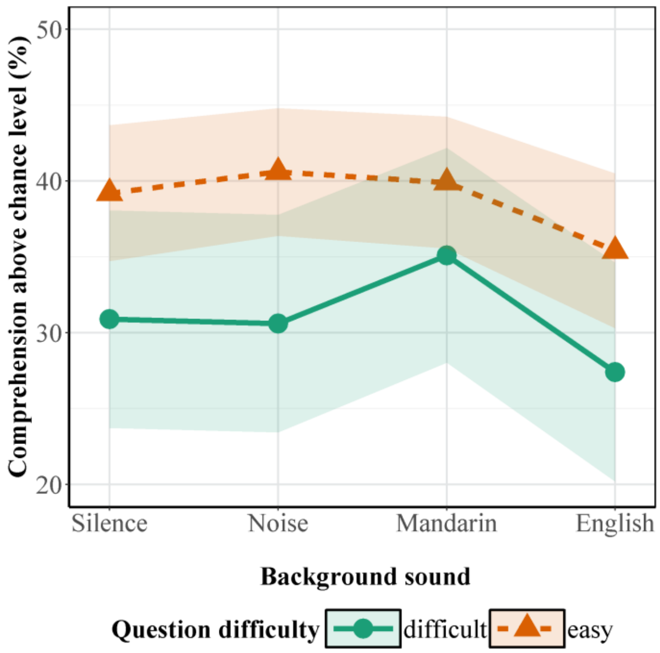

Figure 14. Mean comprehension accuracy in Chapter 4, broken down by question difficulty condition. Shading indicates the standard error.

Although there was no significant difference in comprehension accuracy between English and Silence, it is not immediately obvious why the lack of effect occurred. It is important to determine whether there is no true difference in comprehension accuracy when the text is read in silence and under conditions of English speech (i.e., the null hypothesis is true), or alternatively, whether such a difference does exist (i.e., the alternative hypothesis is true), but the present experiment was not sufficiently powered to detect it. Bayes factors were used to discriminate between these two possibilities (see Dienes, 2014, 2016). Bayes factor regression analyses (Rouder & Morey, 2012) were carried out with the "BayesFactor" R package[10] (Morey, Rouder, & Jamil, 2015). This test yields a Bayes factor, which is the

[10] A prior width of r= $\sqrt{2}/2$ was used in the analyses. We show in Appendix F that the choice of prior did not influence the conclusions from these analyses.

posterior odds of the null and the alternative hypothesis, given the data. Bayes factors greater than 1 favour the alternative hypothesis, whereas Bayes factors smaller than 1 favour the null hypothesis.

The comparison between English speech and Silence in comprehension accuracy showed substantial evidence in support of the null hypothesis of no difference (subjects: BF= 0.18; items: BF= 0.21; see Jeffreys, 1961; Wetzels et al., 2011)[11]. Additionally, the analysis also favoured the null hypothesis of no interaction between question difficulty and the contrast between English and Silence (subjects: BF= 0.15; items: BF= 0.21). The remaining contrasts between English and Mandarin, English and Noise, and Mandarin and Noise also favoured the null hypothesis of no difference and no interaction with question difficulty (range of BFs: 0.12 - 0.44). Therefore, the Bayes factor analysis suggested that there was no true mean difference in the contrast between English and Mandarin that was significant by subjects in the LMM analysis above. In summary, the BF analyses provided direct evidence that there is no difference in comprehension accuracy between English speech and Silence. They also confirmed the LMM results by showing that the effect of English speech on comprehension is not modulated by the difficulty of the questions.

4.5.2. Pre-processing of Eye-tracking Data

Fixation durations were manually pre-processed with the EyeDoctor software (Stracuzzi & Kinsey, 2009) to align the vertical position of fixations (whenever necessary), and to remove blinks from the data (5.81 % of all fixations). Fixations shorter than 80 ms that occurred within one letter of another fixation were combined with that fixation. Any

[11] The same analysis of question accuracy on the data from Experiment 1 yielded a BF of 0.16, thus confirming the same conclusion.

remaining fixations shorter than 80 ms were excluded (1.4 % of the data). Additionally, fixations greater than 1000 ms were excluded as outliers (0.11 % of the data). Furthermore, for the analyses of word time reading measures, FFD longer than 1000 ms, GD longer than 2200 ms, or TVT longer than 3000 ms were discarded as outliers (0.1 % of the data). Although cut-offs of 800 ms for FFD and 2000 ms for GD are typically used in single-line reading studies (e.g. Risse & Kliegl, 2014; Schotter, Lee, Reiderman, & Rayner, 2015), using them resulted in a highly disproportionate number of outliers excluded per sound condition ($\chi 2$ (3) = 14.548, p= 0.002). Increasing the cut-offs by 200 ms ensured there were no significant differences in the number of outliers excluded per condition (χ^2 (3) = 4.09, p= 0.27), while still removing the longest fixation durations that may not reflect normal reading[12]. This was justified by the fact that participants were reading paragraphs which naturally contained longer compound words that are not typically used in single-line reading studies such as the one from Chapter 3.

4.5.3. Global Reading Measures

The descriptive statistics for global reading measures are presented in Table 8 and Table 9. The results from the (G)LMMs are presented in Table 10 for all dependent measures, with the exception of saccade landing position, for which the results are reported in the text. English speech resulted in significantly longer paragraph reading time (d= -0.47), greater intra-sentence regression probability (d= -0.14), and more second-pass fixations (d= -0.15) compared to Silence. The difference between English and Noise was significant for paragraph reading time (d= -0.37), saccade length (d= 0.02), intra-sentence regression

[12] A re-analysis of the data with the outlier cut-offs from Chapter 3 did not change the main results or the conclusions from the analysis.

probability (d= -0.13), and number of second-pass fixations (d= -0.14). The contrast between English and Mandarin was significant for paragraph reading time (d= -0.36), saccade length (d= 0.02), intra-sentence regression probability (d= -0.09), number of first-pass fixations (d= -0.05) and number of second-pass fixations (d= -0.11). There were no differences in saccade landing position for any of the experimental conditions (all ps ≥ 0.07).

Sound condition	Question difficulty	Paragraph reading time (in s)	Intra-sentence regression probability	Inter-sentence regression probability	Number of fixations (per word)		
					1st-pass	2nd-pass	Total
Silence	difficult	25.9 (8.70)	.25 (.43)	.11 (.31)	.81 (.83)	.28 (.61)	1.09 (1.02)
Silence	easy	24.3 (8.00)	.25 (.43)	.08 (.27)	.78 (.79)	.26 (.61)	1.05 (0.99)
Noise	difficult	27.0 (9.79)	.27 (.44)	.11 (.32)	.83 (.85)	.30 (.66)	1.13 (1.08)
Noise	easy	24.6 (9.59)	.24 (.43)	.09 (.28)	.79 (.81)	.26 (.59)	1.04 (1.01)
Mandarin	difficult	26.9 (8.85)	.27 (.45)	.12 (.32)	.81 (.86)	.31 (.68)	1.12 (1.11)
Mandarin	easy	25.0 (9.50)	.27 (.44)	.08 (.27)	.77 (.79)	.28 (.65)	1.05 (1.01)
English	difficult	30.5 (11.54)	.32 (.47)	.15 (.36)	.85 (.99)	.40 (.84)	1.25 (1.33)
English	easy	28.8 (10.54)	.31 (.46)	.12 (.33)	.82 (.88)	.36 (.80)	1.18 (1.23)

Table 8. Mean descriptive statistics of global reading measures in Chapter 4 (SDs in parenthesis).

The comparison between Mandarin and Noise revealed a significant difference only for intra-sentence regression probability (b= 0.11, SE= 0.05, z= -2.29, p= 0.022, d= -0.03). There were no significant differences for any other measures (all ps ≥ 0.07). Therefore, similar to Chapter 3, the results supported most strongly hypothesis H2, which stated that disruption effects by intelligible speech are only semantic in nature. There was very limited evidence in support of hypothesis H2.1, which stated that the disruption by intelligible speech has both a semantic and a phonological component. However, similar to Chapter 3, this support was found only in one measure (intra-sentence regression probability), and even this measure was not the same as the one from Chapter 3 (number of second-pass fixations).

Sound	Question difficulty	Saccade length	Landing position
Silence	difficult	8.47 (5.63)	2.90 (2.29)
Silence	easy	8.47 (5.48)	2.88 (2.28)
Noise	difficult	8.50 (5.74)	2.90 (2.30)
Noise	easy	8.50 (5.37)	2.87 (2.28)
Mandarin	difficult	8.42 (5.70)	2.94 (2.33)
Mandarin	easy	8.52 (5.72)	2.83 (2.24)
English	difficult	8.30 (5.71)	2.93 (2.31)
English	easy	8.47 (5.63)	2.85 (2.25)

Table 9. Mean saccade length and saccade landing position in Chapter 4 (in letters).

The results also showed a significant main effect of question difficulty for two of the dependent measures. Participants made more inter-sentence regressions (d= 0.10) and more second-pass fixations (d= 0.05) when answering difficult compared to easy questions. These results show that the block of paragraphs with difficult questions prompted participants to adopt a more careful reading strategy, in which they made more re-reading fixations, and regressed more often to previous sentences. Additionally, the contrast between English speech and Noise interacted significantly with question difficulty for inter-sentence regression probability and number of second-pass fixations. For both measures, the interaction was due to the fact that the difference between English speech and Noise was smaller in the difficult compared to the easy question condition.

Effect	Paragraph reading time				Saccade length				Intra-sentence regression probability			
	b	SE	t	p	b	SE	t	p	b	SE	z	p
Intercept	3.32	.05	63.9	**<.001**	8.52	.21	40.4	**<.001**	-.86	.06	-13.6	**<.001**
Eng vs Slc	-.18	.04	-4.64	**<.001**	.19	.10	1.89	.13	-.32	.06	-5.61	**<.001**
Eng vs Noise	-.14	.02	-5.77	**<.001**	.23	.09	2.69	**.02**	-.32	.06	-5.79	**<.001**
Eng vs Mnd	-.13	.02	-5.43	**<.001**	.20	.07	2.79	**.02**	-.22	.05	-4.58	**<.001**
Diff	.03	.02	1.55	.12	-.06	.05	-1.41	.32	.04	.02	1.66	.20
Diff: Eng vs Slc	-.02	.02	-.90	.74	.05	.05	.96	.67	-.02	.02	-0.78	.43
Diff: Eng vs Noise	.02	.02	1.08	.56	.06	.05	1.23	.44	.04	.02	1.92	.09
Diff: Eng vs Mnd	.02	.02	.89	.75	.01	.05	.16	.87	<-.01	.02	-.16	.87
Effect	Inter-sentence regression probability				Number of 1st-pass fixations				Number of 2nd-pass fixations			
	b	SE	z	p	b	SE	z	p	b	SE	z	p
Intercept	-2.84	.25	-11.5	**<.001**	-.21	.04	-5.65	**<.001**	-1.07	.07	-15.4	**<.001**
Eng vs Slc	-.25	.24	-1.04	.59	-.03	.02	-1.78	.15	-.35	.05	-7.11	**<.001**
Eng vs Noise	-.27	.19	-1.43	.31	-.03	.02	-1.65	.20	-.35	.06	-6.42	**<.001**
Eng vs Mnd	-.14	.18	-.76	.90	-.05	.02	-2.33	**.04**	-.27	.04	-6.16	**<.001**
Diff	.18	.06	3.00	**.01**	.01	.01	1.52	.13	.06	.02	2.39	**.02**
Diff: Eng vs Slc	-.01	.03	-.19	.85	.01	.01	.57	.91	-.01	.02	-0.69	.49
Diff: Eng vs Noise	.11	.03	3.46	**.001**	.01	.01	.69	.91	.04	.02	2.61	**.02**
Diff: Eng vs Mnd	.01	.03	.35	.72	.01	.01	1.51	.26	.01	.02	.63	.53

Table 10. Results from (G)LMMs on global measures of reading in Chapter 4. Eng: English. Slc: Silence. Mnd: Mandarin. Diff: question difficulty. Statistically significant *p*-values are formatted in bold.

One question of particular interest in the present experiment was how intelligible speech affects the integration of information across sentences. To determine this, we compared the disruption in sentence re-reading time and sentence look-back time. If the disruption is limited only to the currently-read sentence, there should an increase in sentence re-reading time, but not in sentence look-back time. On the other hand, if intelligible speech affects sentence integration processes, such a disruption should also be observed in look-back time. The descriptive statistics are plotted in Figure 15. English speech resulted in significantly longer sentence re-reading time compared to Silence (*b*= -0.38, *SE*= 0.04, *t*= -9.14, *p*< 0.001, *d*= -0.43), Noise (*b*= -0.34, *SE*= 0.05, *t*= -7.62, *p*< 0.001, *d*= -0.36), and Mandarin (*b*= -0.26, *SE*= 0.04, *t*= -7.10, *p*< 0.001, *d*= -0.30). However, the difference between Mandarin and Noise was not significant (*b*= -0.08, *SE*= 0.05, *t*= -1.61, *p*= 0.12, *d*= 0.06), thus providing support for hypothesis H2 that the disruption is only semantic in nature. There were no differences in look-back time for any of the background sound comparisons (all *p*s ≥ 0.16). This suggests that the increase in re-reading behaviour was mostly constrained to the currently-read sentence as the difference in look-back time did not reach statistical significance. In other words, English speech disrupted only the processing of the current sentence and did not lead to longer re-reading times of previous sentences.

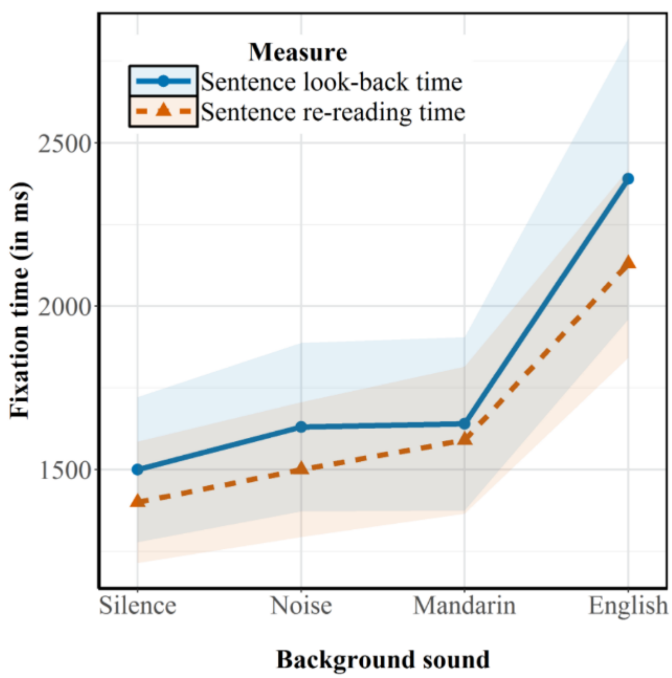

Figure 15. Mean sentence re-reading time and sentence look-back time in Chapter 4. Shading indicates the standard error.

4.5.4. Word-level Reading Measures

The descriptive statistics for local fixation duration measures are shown in Figure 16 and the results from LMMs are presented in Table 11. English speech resulted in significantly longer FFD (d= -0.05), GD (d= -0.08), and TVT (d= -0.21) compared to silence. Additionally, the difference between English and Noise was significant for TVT (d= -0.17), and the difference between English and Mandarin was also significant for both GD (d= -0.05) and TVT (d= -0.16). The difference between English and Noise, and English and Mandarin for FFD did not reach statistical significance, but was still in the expected direction. Therefore, the disruption effects in TVT from Chapter 3 were replicated; additionally, there were also some effects in first-pass reading measures (FFD and GD). Consistent with Chapter 3, there were no differences between Mandarin and Noise in word-level reading measures (all ps ≥ 0.56). In summary, the analysis of local word-level reading measures supported hypothesis H2, which stated that the disruption effect by intelligible speech is only semantic in nature. Contrary to hypotheses, H1.2 and H2.1, there was no evidence for a contribution of phonology.

Furthermore, there was a significant effect of question difficulty for TVT (d= 0.08), which indicated that TVT was longer when participants were answering difficult compared to easy questions. Finally, question difficulty interacted significantly with the comparison between English and Noise for FFD. This was because FFD was longer in English speech compared to Noise when the questions were easy (d= 0.05), but not when they were difficult. There were no other significant interactions between question difficulty and background sound (all ps ≥ 0.1).

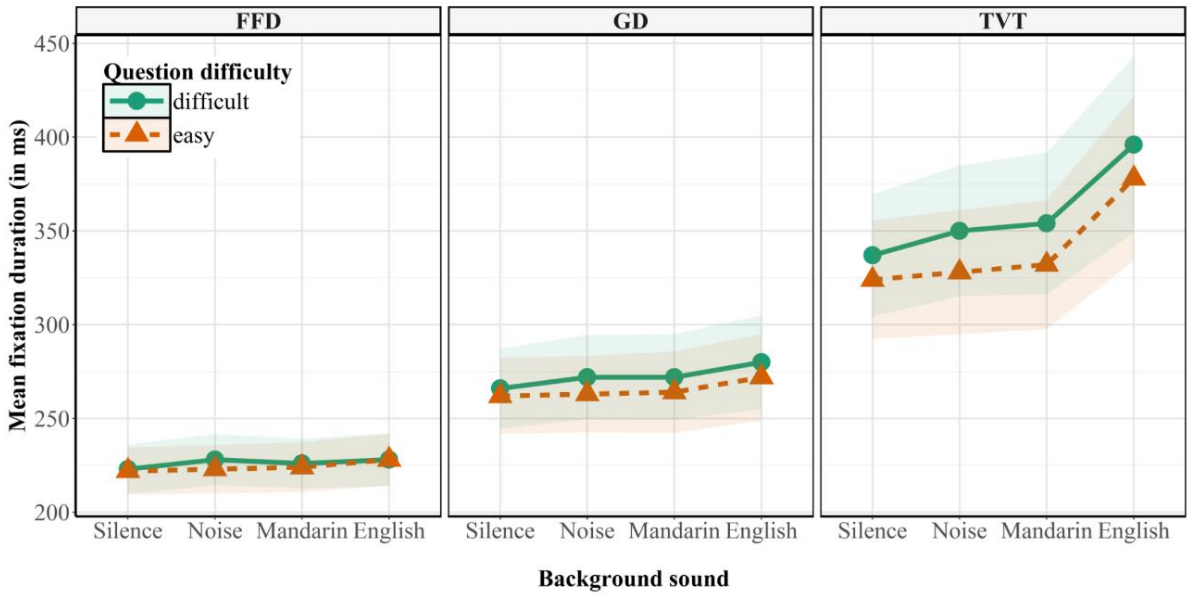

Figure 16. Mean descriptive statistics for local word-level reading measures in Chapter 4, broken down by question difficulty condition. FFD: first fixation duration. GD: gaze duration. TVT: total viewing time. Shading indicates the standard error.

Effect	FFD				GD				TVT			
	b	SE	t	p	b	SE	t	p	b	SE	t	p
Intercept	5.35	.02	340.3	**<.001**	5.49	.02	273.5	**<.001**	5.73	.03	197.1	**<.001**
Eng vs Slc	-.02	.01	-2.33	**.05**	-.03	.01	-3.54	**.002**	-.10	.02	-5.31	**<.001**
Eng vs Noise	-.01	.01	-1.24	.45	-.02	.01	-2.24	.06	-.09	.02	-5.48	**<.001**
Eng vs Mnd	-.01	.01	-1.97	.11	-.02	.01	-3.01	**.01**	-.08	.02	-5.21	**<.001**
Diff	<-.01	<.01	-.76	.87	.01	<.01	1.17	.24	.02	.01	2.97	**.004**
Diff: Eng vs Slc	.01	<.01	1.28	.39	<-.01	.01	-.08	1	<.01	<.01	<.01	1
Diff: Eng vs Noise	.01	<.01	2.73	**.01**	.01	.01	1.32	.37	.01	.01	1.95	.10
Diff: Eng vs Mnd	.01	<.01	1.39	.33	<.01	.01	.89	.75	.01	.01	1.61	.21

Table 11. Results from LMMs on local word-level measures of reading in Chapter 4. Eng: English. Slc: Silence. Mnd: Mandarin. Diff: question difficulty. FFD: first fixation duration. GD: Gaze duration. TVT: Total viewing time. Statistically significant *p*-values are formatted in bold.

4.6. Discussion

The present experiment investigated the effect of intelligible background speech on comprehension accuracy and online integration processes during paragraph reading. The eye-movement measures replicated the disruption effects of intelligible speech found in measures of second-pass reading in Chapter 3. In fact, the amount of disruption was greater than what was observed in the single-sentence reading paradigm used in Chapter 3. This was because, on average, the size of the effects in Cohen's *d* was 76 % greater in the comparison between English speech and Silence and 84% greater in the comparison between English speech and Mandarin speech. Additionally, unlike Chapter 3, there was evidence that intelligible speech also disrupted first-pass reading. More specifically, gaze durations were longer in English speech compared to both Mandarin speech and Silence, and first fixation durations were also longer in English speech compared to Silence (but not compared to Mandarin). Participants also made more first-pass fixations in English speech compared to Mandarin (but not compared to Silence). In this sense, the disruption in paragraph reading was greater than the disruption in sentence reading (Chapter 3) because the magnitude of the effects in second-pass measures was greater and there was at least some evidence that first-pass reading measures were also affected. Because reading connected sentences requires the construction of a discourse model of the text (see Gernsbacher & Foertsch, 2000; O'Brien & Cook, 2015), the greater magnitude of the disruption in paragraph reading may be due to a difficulty in constructing a coherent discourse of the paragraph (Kehler, 2004; Wolf & Gibson, 2005) when readers are listening to intelligible speech in the background.

While the text stimuli were longer in the present experiment and participants may have had more opportunity to go back and re-read the text, the probability of making a regression within the current sentence was comparable in the two experiments (23% in Chapter 3 vs. 25% in the present experiment in the silence condition). Additionally, the probability of making a regression to previous sentences (9.5% in the silence condition) was more than twice as low, thus suggesting that such regressions were not as common as regressions within the currently-read sentence. Therefore, the stronger effects in measures of second-pass reading are not likely to be explained by the text stimuli being longer. In the present experiment, participants also made 22.1% fewer first-pass fixations and 40.2% fewer second-pass fixations compared to Chapter 3. However, at the same time, fixation durations increased by 5.7 % for FFD and by 9.7 % for TVT across all conditions. This suggests that, compared to Chapter 3, participants made fewer but longer fixations in both first-pass and second-pass reading.

Similar to Chapter 3, the results provided strong evidence in support of hypothesis H2 that the disruption by intelligible speech is only semantic in nature (Marsh et al., 2008, 2009; Martin et al., 1988). This was because English speech resulted in greater disruption compared to the other sound conditions in measures of both second-pass reading and first-pass reading (gaze durations). Therefore, because English speech resulted in a greater disruption compared to Mandarin speech, there was again no support for the strong form of the phonological disruption account (H1; Salamé & Baddeley, 1982, 1987) stating that any disruption is only phonological in nature. However, there was limited support for the weaker version of the phonological disruption account (H1.2) because Mandarin speech resulted in a greater intra-sentence regression probability compared to Noise. This suggests that there may be very limited contribution of phonology to the disruption effects by intelligible speech (which would be consistent with hypothesis H2.1), but this was found in only one measure and the same effect was not observed in Chapter 3 in that same measure. Therefore, the present data can be best explained by hypothesis H2, which stated that the disruption by intelligible speech is only semantic in nature. We will revisit what role, if any, phonology may play in distraction by intelligible speech in the General Discussion (Chapter 7), but for now we note that there was very limited evidence in support of a contribution by phonology.

One of the contributions of the present experiment was that it investigated how information is integrated across multiple sentences. Generally speaking, there was no evidence to suggest that the integration of information across sentences is disrupted by intelligible speech because participants did not make more regressions to previous sentences when listening to English speech in the background compared to Silence or Mandarin speech. Additionally, the time that they spent re-reading the sentence during such regressions (i.e., look-back time) also did not differ significantly between the sound conditions. This is largely consistent with Hyönä and Ekholm’s (2016) findings, because the authors also reported no effects in look-back times in three out of their four experiments (the only significant difference in their research was between silence and scrambled speech in Experiment 3). Furthermore, there was no difference between (non-scrambled) intelligible speech and silence in Hyönä and Ekholm’s (2016) Experiments 1 and 3, which is also in agreement with the present results. Interestingly, Cauchard et al.'s (2012) finding that intelligible speech led to longer sentence look-back times is contrary to both the present findings and Hyönä and Ekholm’s (2016) results. Therefore, further research is required to determine the boundary conditions under which such an effect may be observed. We speculate that this discrepancy could potentially be due to differences in the speech stimuli or the text that participants were reading. These are potential mediating factors that have not been thoroughly investigated so far in studies on auditory distraction by intelligible speech.

At any rate, the present study suggests that the increase in re-reading behaviour in response to intelligible speech is constrained only to the currently-read sentence and does not also extend to previously-read sentences. Therefore, the observed disruption in second-pass reading in the present research is likely not related to a difficulty in integrating text meaning across multiple sentences. Rather, it likely reflects a transient difficulty in integrating the meaning of individual words within the current sentence in order to form the meaning of that sentence.

While the difference was not significant, it is also worth noting that English speech resulted in a numerically greater look-back time compared to Silence and this difference was similar in its numerical magnitude to the disruption effect in sentence re-reading time. An examination of the participant means indicated that there was a considerable between-subject variability. Because of this, future studies should investigate whether individual differences may modulate the effect of intelligible speech on sentence look-back time. For example, the time that participants spend re-reading previous sentences could be related to their ability to suppress the irrelevant background speech (see Sörqvist, Halin, et al., 2010; Sörqvist, Ljungberg, & Ljung, 2010). In any case, the present results still suggest that the type of text that participants are reading has an effect on the magnitude of the disruption effects because paragraph reading resulted in stronger disruption compared to single-sentence reading. While the stronger effects in measures of second-pass reading may not be due to a difficulty in integrating text meaning across sentences, they likely arise from the need to construct a coherent discourse representation of the paragraph. Therefore, the discourse representation of the paragraph and the richer text context may amplify this transient disruption that occurs when processing the meaning of the currently-read sentence.

Even though intelligible speech resulted in a considerable disruption of eye-movements, comprehension accuracy remained unaffected in both question difficulty conditions. This suggests that participants could maintain a similar level of text comprehension with English speech in the background, even when the questions probed a deeper level of text understanding. This points to the fact that the disruption observed in eye-movement measures in the English speech condition reflects participants' attempt to successfully attain comprehension in the distracting reading conditions. The results from eye-movement measures provide converging evidence to the same effect. The experimental block with difficult comprehension questions led to a change in eye-movement behaviour, which was characterised by more regressions to previous sentences and longer word re-reading times. However, the disruption effect by English speech did not interact with question difficulty, thus suggesting that the amount of disruption did not depend on the task demands imposed by the question difficulty manipulation. In this sense, there was no evidence that the disruption effect in eye-movement measures increased in the block with difficult questions. Rather, participants were able to adapt to the different task demands, and the magnitude of the disruption was proportional to these demands.

The effect of question difficulty on eye-movements further suggests that participants can make strategic decisions about the nature of the reading task and that they can adjust their reading behaviour accordingly. This finding is in line with the results by Wotschack and Kliegl (2013) and Weiss et al. (2017), who also found that answering more difficult comprehension questions led to an increase in re-reading behaviour. The increase in the number of fixations that participants made and the probability of making a regression to previous sentences in the condition with difficult questions could be due to an attempt to engage in more effective discourse processing in order to develop a richer representation of the meaning of the text. This may occur in response to the expectation that participants will be asked more difficult and more detailed comprehension questions. Similar evidence of such “meta” control over eye-movements has also been found in response to the type of text that participants are reading. For example, participants make more regressions and have longer fixation durations when reading scientific texts compared to reading newspaper articles or light fiction (Rayner, Pollatsek, Ashby, & Clifton, 2012; Rayner et al., 1995).

While there was robust disruption by intelligible speech in eye-movement measures, comprehension accuracy in the present experiment remained unaffected. This suggests that intelligible speech does not degrade the meaning of the text that has been read, at least in the short term and when reading single sentences or short paragraphs. Even though a number of behavioural experiments have reported a disruption in comprehension accuracy (e.g., Baker & Madell, 1965; Halin, 2016; Martin et al., 1988; Sörqvist, Halin, et al., 2010), the present research is not necessarily inconsistent with such studies because it only shows that the immediate comprehension of short sentences and paragraphs is not affected by intelligible speech when participants can re-read previous words and sentences. This difference in the results is not likely to be explained by the greater difficulty of comprehension questions in previous studies because the average accuracy in the studies cited above was 34.1% above chance level (range: 21.2- 43.3%). The average accuracy above chance level on the difficult question in the present research was 31%. Therefore, the difficult questions were, on average, slightly more challenging than the questions used in previous reading comprehension studies.

There are a few possible reasons why a disruption in comprehension may have been observed in previous research. For example, it is possible that intelligible speech may disrupt the transfer of text meaning to long-term memory. In fact, many of the behavioural experiments have had a delay between the reading task and the comprehension assessment, often even with other tasks in between (e.g., Boman, 2004; Knez & Hygge, 2002; Martin et al., 1988). Additionally, the present research used text stimuli that were relatively short and easy to understand. Therefore, it may be the case that intelligible speech disrupts the comprehension of longer and more complex texts that require making inferences between different paragraphs or larger topics of meaning.

Furthermore, the speech stimuli were also relatively simple and they may not have been very engaging to our participants. Therefore, it may be more difficult to maintain comprehension of the text when the intelligible speech is more engaging. This could be because engaging speech makes it harder to selectively attend to the text and filter out the irrelevant speech sound. There is some evidence to suggest that the content of the speech may influence the amount of distraction. For example, hearing only one side of a telephone conversation is more distracting than hearing both sides, presumably because the former type of speech is less predictable than the latter (Emberson, Lupyan, Goldstein, & Spivey, 2010; Marsh et al., 2018). In a similar fashion, engaging speech may be more likely to attract attention away from the main task and thus lead to a greater disruption in comprehension accuracy. These are all avenues that need to be explored by future research.

Because the difficult questions received a difficulty rating of 2.7 on a 5-point scale in the pilot study, it could be argued that the lack of interaction between question difficulty and background sound in comprehension accuracy may be due to the fact that the difficult questions were still not challenging enough to detect such an effect. However, the fact that the block with difficult questions prompted participants to read the paragraphs more carefully clearly suggests that the difficult questions were more challenging than the easy ones. Additionally, the difficulty rating in the pilot study was subjective in nature and thus may not perfectly correlate with participants' performance on the questions (i.e., one can judge the questions to be easy and still answer them incorrectly)[13]. Furthermore, as mentioned above, the difficult questions resulted in slightly lower accuracy above chance level compared to previous behavioural studies. Therefore, even though the difficult questions in the present study were still fairly challenging, future studies may wish to utilise even more difficult questions to make a more rigorous test of the hypothesis that intelligible speech disrupts comprehension only when the questions are difficult to answer. However, it should be kept in mind that if the questions are so difficult that they lead to accuracy that is close to chance-level performance, they will have a poor psychometric sensitivity to detect any potential auditory distraction effects.

In summary, the present study investigated how intelligible speech affects comprehension processes and the integration of information across multiple sentences. The results replicated the disruption effects of intelligible speech on eye-movements from Chapter 3 and were most readily explained by the theoretical view that the observed

[13] The point-biserial correlation between accuracy and difficulty rating in the pilot data was r= – 0.48 overall (r= -0.36 on the difficult and r= -0.34 on the easy questions). This supports the view that the difficulty rating is only moderately related to participant's performance on the comprehension assessment.

disruption is only semantic in nature. Additionally, intelligible speech did not affect the comprehension of short paragraphs even when the questions were more difficult to answer and required a deeper level of text understanding. This suggests that participants could maintain the immediate comprehension of the paragraphs when faced by distracting intelligible speech. Interestingly, the increase in re-reading behaviour occurred only in the currently-read sentence and not in previously-read sentences, thus suggesting that the difficulty of integrating text meaning occurred only at the sentence level, and did not extend to the paragraph level. At the same time, the magnitude of the disruption effects in measures of second-pass reading was greater compared to the single-sentence reading study from Chapter 3. This suggest that the disruption by intelligible speech is greater in paragraph reading compared to single-sentence reading, which could potentially be due to the increase in context and the need to construct a discourse model of the text when reading paragraphs. The next Chapter will examine more closely the link between regressions and comprehension accuracy when listening to intelligible speech in order to better understand how participants are able to maintain an accurate comprehension of the text under such distracting conditions.

CHAPTER 5: THE ROLE OF REGRESSIONS IN DISTRACTION BY INTELLIGIBLE SPEECH

The results from Chapters 3 and 4 indicated that intelligible speech disrupts the ongoing reading process by prompting participants to make more regressions and more re-reading fixations on previously-read words. Additionally, there was strong evidence indicating that this disruption was mostly due to the semantic properties of the irrelevant speech that interfere with processing the meaning of the written text. This finding is consistent with the results from previous eye-tracking studies (Cauchard et al., 2012; Hyönä & Ekholm, 2016; Yan et al., 2017), which have also demonstrated that irrelevant speech leads to an increase in re-reading behaviour. However, while these studies have been successful in showing how the ongoing reading process is disrupted by intelligible speech, little is known about why this increase in re-reading behaviour occurs in the first place. Furthermore, as participants' comprehension was not affected by the speech sound in any of these studies, it is not clear how participants can still maintain an accurate comprehension of the text. Therefore, it is theoretically important to better understand the link between regressive eye-movements and reading comprehension when listening to intelligible speech in the background.

One possible explanation for the lack of disruption in comprehension may be that participants tend to re-read the text more frequently when listening to intelligible speech in order to compensate for the experienced distraction. This explanation assumes that participants are actively trying to comprehend the passages at the same level as when they are reading them in silence. However, the intelligible speech that is playing in the background occasionally leads to a transient interference in processing the meaning of the written text that disrupts participants' task performance. In order to compensate for the experienced disruption and still comprehend the passage, participants initiate a regression to resolve the processing difficulty before continuing to read the unexplored text. Therefore, under this explanation, the observed increase in regressions and re-reading fixations is simply a manifestation of the transient interference in processing the meaning of the written text that is caused by listening to intelligible speech in the background. We will refer to this potential explanation as the *distraction re-reading* hypothesis.

Although this hypothesis has not been formally tested before, there are a few sources of evidence that make it plausible. First, in Chapter 4, there were robust disruption effects by intelligible speech in measures of second pass-reading, but there was no associated disruption of comprehension accuracy. Given that participants in Chapter 4 used, on average, only about a half of their allocated time to read the paragraphs, they would have had more than enough time to selectively re-read the text in order to overcome the disruption and still achieve an accurate comprehension of the paragraphs. Additionally, the post-hoc analysis from Chapter 3 indicated that re-reading fixations in the intelligible speech condition occurred in close proximity to the most recently fixated word in the sentence- presumably, the word where the processing difficulty was first encountered. This suggests that the disruption by intelligible speech was transient in nature because re-reading fixations occurred close to where the progressive reading of the text was interrupted by the speech sound.

Previous evidence from eye-movements during reading further lends plausibility to this hypothesis. First, it is known that regressive eye-movements play an important role in resolving temporary sentence ambiguities (e.g., Frazier & Rayner, 1982; Meseguer, Carreiras, & Clifton, 2002; Staub, 2007). For example, Frazier and Rayner (1982) found that when readers encounter temporary syntactic ambiguities while reading garden-path sentences, they use regressive eye-movements to selectively re-analyse parts of the sentence that can help them resolve the ambiguity. This finding led the authors to suggest that readers do not typically backtrack to the beginning of the text when they encounter sentential ambiguities, but that they use the information they have acquired to selectively re-read parts of the text that can help them recover from an incorrect previous interpretation of the sentence. In a similar fashion, readers may also be selectively re-reading the words whose meaning was corrupted or whose meaning could not be integrated within the sentence context due to interference from the irrelevant speech sound.

Additionally, Rayner, Chace, Slattery, and Ashby (2006, Experiment 2) investigated how anaphor-antecedent inconsistencies influence eye-movements during reading. They found that participants made more regressive eye-movements when there was an inconsistency between the anaphor and its antecedent in the text (Rayner et al., 2006). This again suggests that readers use regressions to resolve online comprehension difficulties when processing the meaning of the text. Studies investigating discourse processes during reading have also linked regressive eye-movements to higher-level representations of meaning. For example, Hyönä (1995) found that participants made more regressive fixations when a sentence introduced a new subtopic in the text. This was thought to reflect integration processes that give readers more time to wrap up the meaning of the sentence, possibly because they are not yet ready to encounter new information before consolidating what has been read so far. Similarly, Blanchard and Iran-Nejad (1987) found that a surprising ending to a story led to more re-reading fixations compared to a non-surprising one. The authors argued that this may occur because, when readers encounter the surprising ending, the cognitive system is put "on hold" until all comprehension processes have been completed and new information can be acquired again. In summary, the studies above suggest that regressions play an active role in online text comprehension and are reflective of the immediate difficulties in processing the meaning of the text.

More direct evidence regarding the role of regressions in comprehension comes from two recent studies that have manipulated what participants see during a regression. In the first study, Booth and Weger (2013, Experiment 3) presented short statements for reading (e.g., "Andy is a good driver but his cousin David is not."). If participants made a regression while reading the sentence, one target word was changed using Rayner's (1975) gaze-contingent boundary paradigm. Critically, however, this change altered the meaning of the sentence (e.g., "Andy is a good dancer but his cousin David is not."). Afterwards, participants' comprehension of the sentence was probed with a question to determine which version of the sentence they had understood (the pre-change, the post-change one, or a baseline option that matched neither version). The authors found that, when participants fixated the changed word during a regression, they were more likely to select the post-change version of the sentence compared to when it was not fixated during a regression. This suggests that readers use regressions to further process the meaning of words and that this additional processing can influence their subsequent understanding of the sentence.

In another study, Schotter, Tran, and Rayner (2014) investigated the link between regressions and reading comprehension by eliminating participants' ability to do any additional visual processing of the text during regressions. In their study, Schotter et al. used a new manipulation (the so-called *trailing mask* paradigm) to experimentally prevent participants from obtaining any useful information from words if they are re-fixated during a regression. In this paradigm, each word is permanently masked by a string of 'x's once participants make a saccade to the right of it, thus making it impossible to re-read the word during a regression. Schotter et al. found that when participants were reading in the trailing mask condition, their comprehension was negatively affected compared to when they were reading in the normal (i.e., unmasked) text condition. This finding further suggests that regressions are important for maintaining an accurate comprehension of the sentences.

In the present research, we tested the distraction re-reading hypothesis, which stated that the increase in regressions and re-reading fixations is crucial for maintaining the immediate text comprehension in the face of distraction by intelligible speech. This was done by presenting the same paragraphs from Chapter 4, but this time in a way that prevented participants from selectively re-reading previous parts of the text. If regressive eye-movements and re-reading fixations are important for maintaining comprehension of the text when listening to distracting intelligible speech, we would expect to see a disruption in comprehension accuracy when participants are no longer able to re-read the text. Additionally, if regressions mostly support deeper levels of text comprehension, we would expect to see greater disruption in comprehension on the difficult compared to the easy questions. This is because answering the difficult questions required a deeper level of text understanding since the correct answers were paraphrased and could not be found by simply recognizing words or phrases from the text.

In Chapter 5, these predictions were tested with two different paradigms that prevented re-reading of the previous text. In Experiment 1, re-reading was prevented by presenting the paragraphs one word at a time using rapid serial visual presentation (RSVP; Forster, 1970). This was done to obtain preliminary evidence about the feasibility of the distraction re-reading hypothesis and to compare the comprehension accuracy results to the ones from Chapter 4. In Experiment 2, this hypothesis was tested more directly by comparing a condition in which participants could re-read the text to a condition in which they could not re-read the text within the same experiment. In Experiment 2, re-reading behaviour was rendered useless by masking the previous text with 'x's using Schotter et al.'s (2014) trailing mask paradigm.

5.1. Experiment 1

In Experiment 1, participants were prevented from making eye-movements to previous or upcoming words in the text, which effectively eliminated their control over what parts of the text they see. This was achieved by presenting the same paragraphs from Chapter 4 in RSVP mode (RSVP; Forster, 1970). In the RSVP paradigm, words are presented one by one at a constant rate at the centre of the screen. Because the RSVP stream is automatic and participants cannot go back to previous screens, no re-reading of the text is possible. We predicted that, when second-pass reading is eliminated, comprehension accuracy would be lower in the English speech condition compared to both the silence and Mandarin speech conditions. This was because the most robust disruption effects in both Chapter 3 and Chapter 4 were observed in measures of second-pass reading.

In Experiment 1, a few steps were taken to make the reading conditions in RSVP mode similar to the ones in Chapter 4. First, reading speed was determined from the average reading speed in Chapter 4 on the same materials (240 words per minute in the silence condition). Second, reading speed was further slowed down to 200 words per minute to account for the fact that participants could not pre-process the upcoming word in parafoveal vision (see Schotter et al., 2014, p.1225). To determine the amount of adjustment, we used pooled estimates of parafoveal pre-processing from the boundary paradigm (Rayner, 1975) that were calculated in a recent meta-analysis (Vasilev & Angele, 2017)[14]. Finally, a 500 ms blank screen was inserted at the end of each sentence to allow for sentence wrap-up effects and processing the "buffer" of words in the RSVP stream to occur (Just, Carpenter, & Woolley, 1982; Masson, 1983). Masson (1983) showed that this small addition to the RSVP procedure significantly improves comprehension accuracy on short passages similar to the ones used in this study.

Based on the distraction re-reading hypothesis, we predicted that English speech would significantly disrupt comprehension compared to both the Mandarin speech and silence conditions. No such disruption was predicted for the Mandarin and speech-spectrum noise conditions. Additionally, we also predicted that the amount of disruption in comprehension accuracy by intelligible speech would be significantly greater for the difficult compared to the easy comprehension questions.

[14] With 240 words per minute, the fixation time per word is 60/240= 0.25 s. Vasilev and Angele (2017, Table 2) estimated that valid preview of the upcoming word translates into 47.2 ms shorter TVT when that word is subsequently fixated (0.0472 s). By adding this number to the fixation time per word in Experiment 1, we obtain 0.25+ 0.0472= 0.2972 s adjusted fixation time per word. This in turn translates into an adjusted reading rate of 60/ 0.2972= 201.8843 words per minute (rounded down to 200).

5.1.1. Method

5.1.1.1. Participants.

Forty-eight Bournemouth University students participated for course credit (43 female). Their mean age was 19.7 years (*SD*= 3.5 years; range: 18-42 years). None of them had participated in the previous two experiments. Participants were native speakers of British English who reported normal or corrected-to-normal vision, normal hearing, no prior diagnosis of reading disorders, and no prior knowledge of Mandarin Chinese. The study had the same statistical power as the experiment from Chapter 4 and was therefore sufficiently powered. The study was approved by the Bournemouth University Research Ethics Committee (protocol No. 14005).

5.1.1.2. Materials.

The same reading and auditory stimuli from Chapter 4 were used. The conditions were blocked in the same way as in Chapter 4.

5.1.1.3. Apparatus.

The experiment was programmed in PsychoPy (Peirce, 2007) and was run on a Hewlett-Packard EliteDesk 800 G1 SFF PC with 8GB RAM and a Windows 7 operating system. The paragraphs were presented on a 24" BENQ XL2411 monitor with a screen resolution of 1920 x 1080 pixels and a refresh rate of 60 Hz. The words in the paragraph were formatted in a Courier New pt.16 font and appeared as black text over white background at the centre of the screen. Participants were seated approximately 85 centimetres from the monitor. At this distance, the stimuli subtended approximately the same degree per visual angle as in Chapter 4. The sound stimuli were presented binaurally through

Bose QuietComfort 25 noise-cancelling headphones at approximately 60 dB(A). The sounds were played on an Intel HD Max integrated sound card.

5.1.1.4. Procedure.

Before the reading task, participants read three practice sentences to get used to the RSVP mode of presentation. Similar to Chapter 4, each of the question difficulty blocks started with two practice paragraphs in order to introduce participants to the different comprehension difficulty conditions. Before each trial, participants pressed a button on the keyboard to start the RSVP presentation. The presentation of the paragraphs is illustrated in Figure 17.

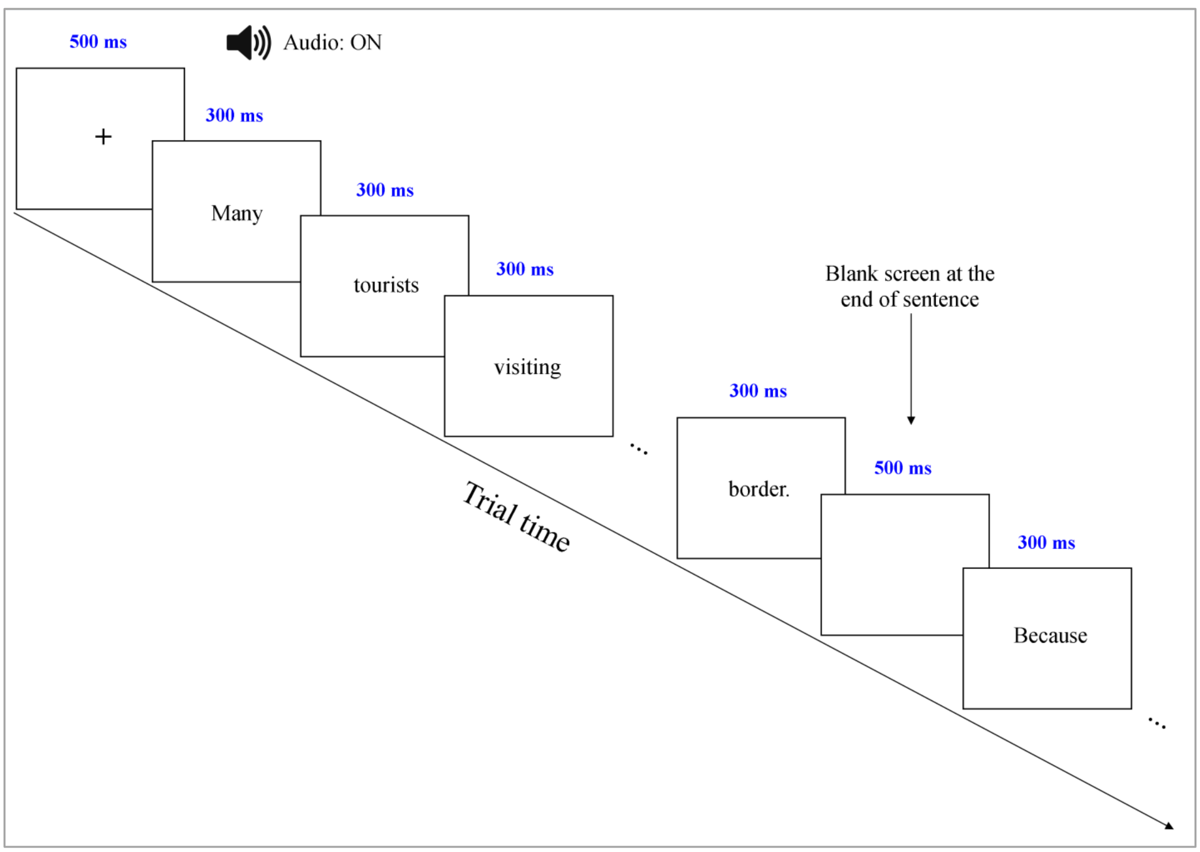

Figure 17. An illustration of the RSVP presentation in Chapter 5, Experiment 1. Each trial started with a fixation cross. The paragraph was then presented one word at the time, with

each word staying on the screen for 300 ms. A 500 ms blank screen was inserted at the end of each sentence. Background sounds were played simultaneously with the appearance of the first word on the screen.

Each trial started with a fixation cross at the centre of the screen, which remained there for 500 ms before it was replaced by the first word in the paragraph. The sound stimuli were played as soon as the first word appeared on the screen. The paragraphs were presented at a speed of 200 words per minute. Each word appeared at the centre of the screen and stayed there for 300 ms before it was replaced by the next word. Capitalization and punctuation of the paragraphs were preserved. There was a 500-ms blank interval at the end of each sentence. Participants were instructed to look at the centre of the screen and to read all words until the whole paragraph has been presented. Each paragraph was followed by the same comprehension questions used in Chapter 4. Participants pressed a button on the keyboard to select the correct answer. The experiment lasted for 30-35 minutes.

5.1.1.5. Data analysis.

Comprehension accuracy was the only dependent variable and it was analysed with LMMs by using the “lme4” package v.1.1-12 (Bates et al., 2014) in the R statistical software v.3.3.1 (R Core Team, 2016). Comprehension accuracy was analysed as accuracy above chance level due to the different chance levels in the two question difficulty conditions (50% for the easy questions and 25% for the difficult questions). Two separate models are reported for participants (b_1) and items (b_2) because analysing the data in terms of comprehension accuracy above chance level requires calculating the mean accuracy for each condition and then subtracting the chance level performance from it. Random intercepts and random slopes for background sound and question difficulty were added for both participants and items.

The same contrast coding as in Chapter 4 was used (question difficulty: -1= easy, 1= difficult; background sound: treatment contrast coding with English speech as the baseline). *P*-values were calculated with the lmerTest package v.2.0-33 (Kuznetsova et al., 2017). Similar to Chapters 3-4, *p*-values were adjusted with the Holm-Bonferroni (Holm, 1979) correction due to the additionally contrast between Mandarin and Noise. The results were considered as statistically significant if the adjusted *p*-values were ≤ 0.05.

5.1.2. Results

The mean comprehension accuracy is presented in Figure 18. There was a statistically significant main effect of question difficulty (b_1= -0.05, *SE*= 0.02, *t*= -3.06, p= 0.004; b_2= -0.05, *SE*= 0.02, *t*= -2.44, p= 0.03; *d*= -0.21), thus showing that accuracy was better on the easy compared to the difficult questions. Additionally, comprehension accuracy was significantly lower in the English speech condition compared to all other sound conditions (Silence: b_1= 0.10, *SE*= 0.02, *t*= 4.56, $p< 0.001$; b_2= 0.11, *SE*= 0.03, *t*= 3.74, $p< 0.001$; *d*= 0.22; Noise: b_1= 0.08, *SE*= 0.02, *t*= 3.46, *p*= 0.001; b_2= 0.08, *SE*= 0.03, *t*= 2.87, *p*= 0.009; *d*= 0.17; Mandarin: b_1= 0.08, *SE*= 0.02, *t*= 3.34, *p*= 0.002; b_2= 0.08, *SE*= 0.03, *t*= 2.71, *p*= 0.017; *d*= 0.17). However, there was no significant difference between Mandarin and Noise (*p*s ≥ 0.94). There were also no significant interactions between background sound and question difficulty (all *p*s≥ 0.99), thus showing that the amount of disruption did not differ as a function of question difficulty. This indicates that comprehension was equally impaired by English speech in the two question difficulty conditions.

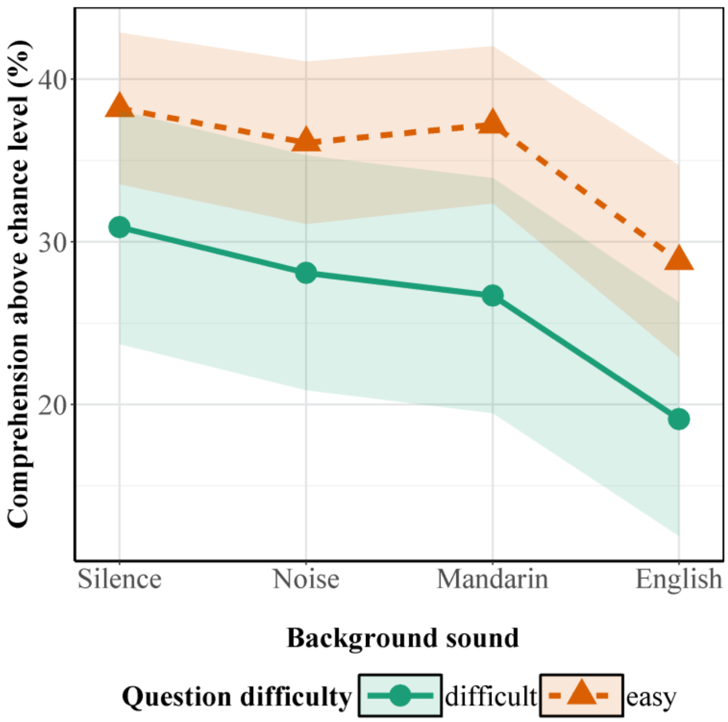

Figure 18. Mean comprehension accuracy above chance level for the different sound conditions in Chapter 5, Experiment 1 as a function of question difficulty. Shading indicates the standard error.

Bayes factor regression analysis (Morey et al., 2015; Rouder & Morey, 2012) of the difference in comprehension accuracy between English speech and Silence indicated decisive evidence in support of the alternative hypothesis of a true difference (subjects: BF= 153.6; items: BF= 19.4). Additionally, the comparisons between English and Noise (subjects: BF= 4.7; items: BF= 2.6) and English and Mandarin (subjects: BF= 4.5; items: BF= 2.7) also favoured the alternative hypothesis of a true mean difference. However, consistent with the LMM analysis, the contrast between Mandarin and Noise favoured the null hypothesis of no difference (subjects: BF= 0.11; items: BF= 0.15). Furthermore, the interaction between question difficulty and the sound conditions also favoured the null hypothesis of no difference (range of BFs: 0.16- 0.24). Therefore, the Bayesian analyses confirmed the finding that English speech disrupted comprehension more than the remaining sound conditions, but that the magnitude of this disruption was not modulated by the difficulty of questions.

5.1.3. Discussion

Experiment 1 showed that preventing participants from re-reading words significantly impaired their comprehension accuracy when listening to intelligible speech in the background. While the overall mean difference was modest (10.5%), the Bayesian analyses now favoured the alternative hypothesis of a true difference between the two conditions. Although it can be argued that the RVSP mode of presentation may have made the reading task harder in Experiment 1, participants' comprehension accuracy in the silence condition was very similar to the one observed in Chapter 4 (this includes both the main experiment and the pilot data). This suggests that participants could still comprehend the paragraphs at approximately the same level as in the experiment in Chapter 4 in the absence of English speech.

The results from Experiment 1 suggest that comprehension is negatively affected by intelligible speech when participants cannot go back to selectively re-read the text. The increase in regressions in the previous two experiments (Chapters 3-4) suggested that re-reading behaviour is necessary for developing and maintaining an accurate representation of the text. Once such behaviour was prevented in Experiment 1, this presumably limited participants' ability to maintain this representation under conditions of intelligible background speech. In this sense, the findings from Experiment 1 are in principle consistent with the notion that regressive eye-movements are necessary for maintaining comprehension of the text when listening to intelligible speech.

Experiment 1 therefore provided some preliminary evidence in support of the distraction re-reading hypothesis, which stated that regressions and re-reading fixations are important for maintaining an accurate comprehension of the text when listening to intelligible speech in the background. Nevertheless, this experiment has a few limitations that do not make it possible to conclusively rule out other alternative explanations for the observed pattern of results. One such limitation is that, unlike the study in Chapter 4, reading was no longer self-paced because participants were artificially constrained to fixating each word for 300 ms. As a result, Experiment 1 cannot unambiguously demonstrate that the disruption in comprehension accuracy by intelligible speech is due only to preventing participants from re-reading the text. It is also possible that this may have occurred because reading was no longer self-paced and participants had no control over how long they fixated each individual word. Additionally, the words in the text were presented in a spatially different way as they appeared one-by-one in the middle of the screen. This differs from Chapter 4 where the words were presented as normal text over multiple lines. Finally, the conclusions from Experiment 1 are also limited by the fact that any comparisons to the results from Chapter 4 are by necessity based on a set of two independent samples. Therefore, the strongest test of this hypothesis would be to compare a condition where re-reading of previous text is possible to a condition where re-reading of previous text is not possible within the same experiment.

Experiment 2 was designed to address the limitations above and to make a more rigorous test of the distraction re-reading hypothesis. In this study, the trailing mask paradigm by Schotter et al. (2014) was used to prevent participants from re-reading previous words in the text. In this paradigm, each word in the text is permanently masked by a string of ‘x’s immediately after participants make a saccade to the right of that word. While participants can still make regressions to previous words in the trailing mask condition, such regression do not yield any useful information because the text has already been masked by the time these words are re-fixated. This effectively eliminates participants’ ability to selectively re-read previous parts of the text in order to resolve comprehension difficulties that arise due to interference from the irrelevant speech.

5.2. Experiment 2

The experiment had a 2 x 2 x 2 within-subject design with the following factors: background sound (English speech vs silence), reading condition (normal text vs trailing mask text), comprehension question difficulty (easy vs difficult). To preserve statistical power and because the critical comparison for the present hypothesis is between silence and English speech, the Mandarin and speech-spectrum noise conditions from Experiment 1 were removed. Similar to Experiment 1, we expected that English speech will disrupt comprehension compared to the silence condition, but only in the trailing mask condition when participants cannot re-read the text. Additionally, we also expected to replicate the disruption effects by English speech from Chapter 4 in measures of second-pass reading. This was again hypothesized to occur only in the normal, but not in trailing mask condition because participants cannot re-read the text after it has been masked.

5.2.1. Method

5.2.1.1. Participants.

Forty-eight Bournemouth University students participated for course credit or a payment of £10 (29 female). Their mean age was 20.6 years (*SD*= 2.4 years; range: 18-32 years). One more participant was tested, but their data were excluded due to tracking problems. All participants were native speakers of British English who reported normal or corrected-to-normal vision, normal hearing and no prior diagnosis of reading disorders. None of them had participated in any of the previous experiments. All participants were naïve as to the purpose of the experiment. Ethical approval of the study was obtained from the Bournemouth University Research Ethics Committee (protocol No. 14005). The study had the same statistical power as the one in Chapter 4 and was therefore sufficiently powered.

5.2.1.2. Materials and design.

The reading materials consisted of the same paragraphs that were used in Chapter 4 and also in Experiment 1 of the present Chapter. The question difficulty manipulation was also the same as these two studies. The English speech was taken from the BKB (Bench et al., 1979) and IHR (MacLeod & Summerfield, 1990) corpora. Twelve 60 s speech files were created by concatenating between 40 to 42 unique speech sentences each and removing the silence gaps between sentences. Half of the sound files contained speech spoken by a male British English speaker and the remaining half contained speech spoken by a female British English speaker.

There were two reading conditions in the experiment: *normal text* (i.e., with no visual changes on the screen) and *trailing mask* text. In the trailing mask condition, each word in the text was permanently masked by a string of ‘x’s once participants made a saccade to the right of that word (see Figure 19b for an illustration). The gaze-contingent masking mechanism involved placing in invisible boundary (Rayner, 1975) at the first pixel after the end of each word. Once a boundary was crossed, the word immediately before the boundary was permanently masked with ‘x’s (see Schotter et al., 2014 for more details). The empty spaces between words were kept in the masked text, which helped preserve its general outline. This type of masking was identical to the one used by Schotter et al. (2014).

Because Experiment 2 used paragraphs instead of single sentences, it was necessary to extend Schotter et al.'s (2014) trailing mask manipulation for use in a multiple-line reading paradigm. This was needed as the error in tracking the vertical position of the eye can cause incorrect triggering of the display changes in the experiment. Pilot testing indicated that the least obtrusive way to implement this was to add a gaze-contingent check (a small square) at the end of each line that participants had to fixate to indicate they had finished reading the current line. At the start of each trail, only the first line was visible. Once the gaze-contingent check at the end of the first line was triggered, the square immediately disappeared and the next line was automatically revealed [15]. This procedure was then repeated until the whole paragraph had been presented (see Figure 19a for an illustration of the method). To avoid delays associated with having to fixate exactly within the square, the line check was triggered immediately after participants' gaze moved to right of the last word on the line (i.e., the square and the space around it simply acted as a catchment area).

[15] To ensure that the trailing mask is accurately triggered on the next line, the display changes started when participants made a rightwards (i.e., progressive) saccade to a new word. This was necessary as the return sweep saccade from the end of the previous line to the beginning of the next line can sometimes undershoot the line start, which may be followed by a corrective saccade to the left (Andriessen & de Voogd, 1973; Hofmeister, Heller, & Radach, 1999; Rayner, 1998). Such undershoot fixations are generally not thought to be related to text processing (Abrams & Zuber, 1972) and are much shorter than the average fixation during reading. In Experiment 2, participants made a corrective saccade to the left that landed on a previous word on 41.1% of all line crosses. The average duration of the undershoot fixation was 110 ms (SD= 59 ms). The advantage of allowing readers to make a return sweep to the next line was that it kept the reading process more natural. This approach was preferred because a pilot study in which participants had to fixate a gaze box at the start of each new line was found to be too disruptive to the reading process due to the delays in triggering the gaze boxes.

The text stimuli were presented in this way in all trials in order to keep the reading conditions constant throughout the experiment. Similar to Experiment 1 and Chapter 4, the background sound and question difficulty conditions were presented in separate blocks. The order of items within each block was randomised. Similar to Schotter et al.'s (2014) experiment, the normal text and trailing mask text trials were intermixed within blocks, but participants received a cue before the start of each trial that told them what type of text they will be reading. In the present study, a black gaze box at the start of each trial indicated that participants will be reading normal text, whereas a blue gaze box indicated that they will be reading the trailing mask text. All blocks and conditions were counter-balanced with a full Latin square design across participants.

5.2.1.3. Apparatus.

Participants' eye-movements were recorded with an Eyelink 1000 at a sampling frequency of 1000 Hz. The resolution noise was < 0.01° and the velocity noise was < 0.5° on average. Viewing was binocular, but only the right eye was recorded. The head was stabilized with a chin-and-forehead rest to reduce artefacts related to head movements. The experiment was programmed in Matlab 2014a (MathWorks, 2014) by using the Psychophysics toolbox (Brainard, 1997; Pelli, 1997) and Eyelink libraries (Cornelissen, Peters, & Palmer, 2002).

The experiment was run on a PC with a Microsoft Windows XP operating system. The paragraphs were presented on a 20-inch Mitsubishi Diamond Pro 2070 monitor with a screen resolution of 1024 x 768 and a refresh rate of 150 Hz. The text was displayed with the same dimensions and spatial layout as in Chapter 4. The only exception was that some of the lines were made shorter to make enough space for the fixation check at the end of each line.

The fixation check was a 16 x 16-pixel black square that was situated 3 letter spaces (33 pixels) to the right of the last word on the line. All paragraphs fitted on a single screen. The display changes in the experiment were completed on average within 9.12 ms of the eye moving to the right of each individual word (SD= 1.98 ms).

5.2.1.4. Procedure.

Participants were tested individually in a session that lasted for about 40-45 minutes. Participants were instructed to ignore the background speech and to focus on reading the paragraphs. They were also instructed that the paragraph will be revealed line by line and that they will need to fixate a small square at the end of each line to reveal the next line. Furthermore, participants were informed that the words in some paragraphs will be masked by ‘x’s after they have read them, but that they should try to read the text as normally as possible. They were also told that the colour of the gaze box before each paragraph will indicate what type of text they will be reading.

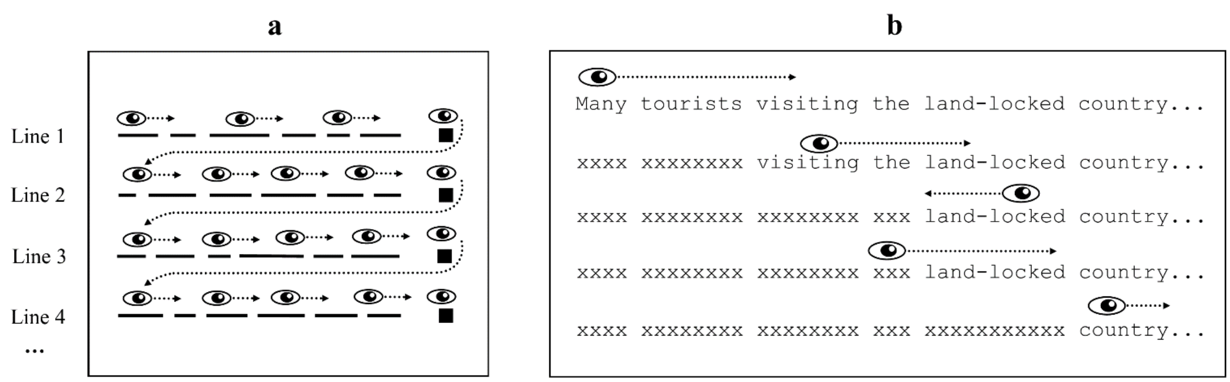

Figure 19. An illustration of the text stimuli presentation in Chapter 5, Experiment 2. Panel **a** shows a schematic representation of the line-by-line text presentation (with horizontal lines representing the text). At the start of each trial, only the first line was visible. Participants then revealed each new line of text by fixating a small black square at the end of each line until the whole paragraph was revealed. Panel **b** shows an example of the trailing mask

reading condition. Words were permanently masked by a string of ‘x’s once the eye moved to the right of each word.

Before the start of the experiment, participants were calibrated on a 9-point calibration grid. The calibration was then monitored with a drift check before the start of each trial. The calibration error was kept at ≤ 0.4°. The beeps during calibration and drift check were turned off. Each question difficulty block started with two practice trials. One practice trial was displayed in the normal text condition, while the other one was displayed in the trailing mask condition. All trials started with a gaze box (black in the normal text condition and blue in the trailing mask condition) that was centered at the location of the first letter on the first line. Once the gaze box was fixated for 100 ms, it disappeared and the text was immediately displayed on the screen. The onset of the sound in the English speech condition was simultaneous with the appearance of the text stimuli. Participants clicked the left button of the mouse to indicate that they had finished reading the paragraph and also to choose the correct answer to the comprehension questions. Termination of the trial was possible only after all lines of the text had been revealed. Similar to Experiment 1 and Chapter 4, there was a 60 s trial timeout for both the paragraphs and the comprehension questions.

5.2.1.5. Data analysis.

The experiment had a 2 (background sound: English speech vs silence) x 2 (reading condition: trailing mask text vs normal text) x 2(comprehension question difficulty: easy vs difficult) within-subject design. The same measures of global reading from Chapter 4 were analysed: paragraph reading time, number of first- and second-pass fixations, intra-sentence, inter-sentence regression probability, saccade length, and saccade landing position.

Additionally, FFD, GD, and TVT were calculated for each word in the paragraph and were analysed as local word-level reading measures. The data were analysed with (G)LMMs by using the lme4 package v. 1.1-12 (Bates et al., 2014) in R v. 3.30 (R Core Team, 2016). Fixation durations were log-transformed in all analyses. Sum contrast coding was used for all three independent variables: background sound (Silence: -1; English: 1), reading condition (trailing mask: -1; normal text: 1), comprehension question difficulty (easy: -1; difficult: 1). Participants and items were added as random intercepts in all analyses (Baayen et al., 2008). Background sound, reading condition, and question difficulty were added as random slopes for participants and items in all analyses (Barr et al., 2013). Similar to Experiment 1, the only exception to this rule was that comprehension accuracy was analysed with two separate models for participants (b_1) and items (b_2). For consistency purposes, *p*-values are reported for all analyses (calculated with the lmerTest package v.2.0-33; Kuznetsova et al., 2017). The results were considered statistically significant if the *p*-values were ≤ 0.05.

5.2.2. Results

Similar to Chapter 4, fixation were manually pre-processed with the EyeDoctor software (Stracuzzi & Kinsey, 2009) to re-align their vertical position if necessary and to remove blinks from the data (4.47 % of all fixations). Fixations shorter than 80 ms that occurred within one character of another fixation were combined with that fixation. All other fixations smaller than 80 ms were excluded from the data (2.4 %). Additionally, any fixations longer than 1000 ms were excluded as outliers (0.29 %). In the analysis of word-level reading measures, words with FFD longer than 1000 ms, GD longer than 2200 ms, and TVT longer than 3000 ms were excluded from the data (0.22 % of observations). There were no significant differences in the number of observations excluded per condition (all *p*s $\geq$ 0.11).

5.2.2.1. Comprehension accuracy.

The descriptive statistics for comprehension accuracy in Experiment 2 are presented in Figure 20. The LMM analysis indicated a main effect of question difficulty (b_1= -0.06, *SE*= 0.01, *t*= -5.62, *p*< 0.001; b_2= -0.06, *SE*= 0.01, *t*= -4.97, *p*< 0.001; *d*= -1.67), which was due to comprehension being significantly lower on the difficult compared to the easy questions. Additionally, there was a main effect of background sound (b_1= 0.02, *SE*= 0.01, *t*= 2.24, *p*= 0.03; b_2= 0.02, *SE*= 0.01, *t*= 2.08, *p*= 0.04; *d*= 0.33), which shows that accuracy was significantly lower in the English speech compared to the silence condition. Furthermore, the main effect of reading condition was also significant (b_1= 0.03, *SE*= 0.01, *t*= 3.57, *p*= 0.001; b_2= 0.03, *SE*= 0.01, *t*= 3.15, *p*= 0.002; *d*= 0.49), which was due to comprehension being lower in the trailing mask compared to the normal reading condition. In line with the distraction re-reading hypothesis, there was a significant interaction between background sound and reading condition (b_1= -0.03, *SE*= 0.01, *t*= -3.67, *p*< 0.001; b_2= -0.03, *SE*= 0.01, *t*= -3.16, *p*= 0.002). This was due to accuracy being lower in English speech compared to Silence, but only in the trailing mask (*d*= -0.65) and not in the normal reading condition (*d*= 0.12).

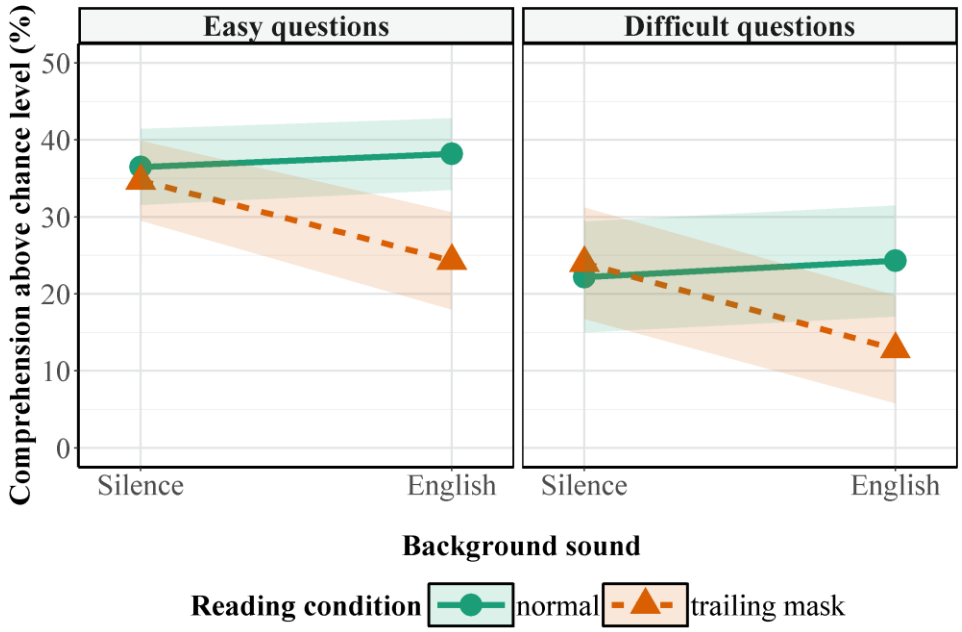

Figure 20. Mean comprehension accuracy above chance level in Experiment 2, Chapter 5. Shading indicates the standard error.

Bayes factor regression analyses (Morey et al., 2015; Rouder & Morey, 2012) also supported the alternative hypothesis that there is an interaction between background sound and reading condition (subjects: BF= 11.43; items: BF= 9.40). There was strong evidence in support of the alternative hypothesis that comprehension accuracy is disrupted by English speech in the trailing mask condition (subjects: BF= 27.81; items: BF= 13.70). Conversely, the null hypothesis of no difference in comprehension accuracy between English speech and Silence was supported for the normal reading condition (subjects: BF= 0.13; items: BF= 0.18). Consistent with the LMM analysis, the null hypothesis of no interaction between background sound and question difficulty was supported (subjects: BF= 0.11; items: BF= 0.14). Sensitivity analyses using a range of realistic priors indicated that the results were not influenced by the chosen prior distribution (r= $\sqrt{2}/2$; see Appendix G). In summary, both the Bayes factor and LMM analyses were in line with the distraction re-reading hypothesis, which predicted that comprehension accuracy would be disrupted by English speech only when participants cannot selectively re-read the previous text. The magnitude of the disruption in comprehension accuracy was not modulated by whether participants were answering easy or difficult questions.

5.2.2.2. Global reading measures.

The descriptive statistics for global reading measures are presented in Table 12 and Table 13, and the results from (G)LMMs are presented in Table 14 and Table 15. English speech resulted in significantly longer paragraph reading time (d= 0.24), greater intra-sentence regression probability (d= 0.21), and more second-pass fixations (d= 0.04) compared to Silence. Additionally, saccades landed further away from the beginning of the word when participants were answering difficult compared to easy questions (d= 0.03). Furthermore, the trailing mask condition resulted in significantly shorter paragraph reading time (d= 0.50), smaller intra-sentence (d= 0.19) and inter-sentence (d= 0.16) regression probability, fewer first-pass (d= 0.11) and second-pass (d= 0.18) fixations, and saccades that landed further away from the beginning of the word (d= 0.03) compared to the normal reading condition.

There was a statistically significant interaction between background sound and question difficulty for inter-sentence regression probability. This was due to participants making fewer inter-sentence regressions in English speech compared to Silence, but only when they were answering difficult comprehension questions (d= -0.02). Additionally, background sound interacted significantly with reading condition for paragraph reading time, intra-sentence regression probability, number of second-pass fixations, and saccade length. The interaction was due to participants taking longer to read the paragraphs (d= 0.27), making more intra-sentence regressions (*d*= 0.09), more second-pass fixations (*d*= 0.08), and having shorter saccade length (*d*= -0.03) in the English speech compared to the Silence condition, but only when reading was normal and not when the text had a trailing mask. Therefore, the interactions replicate the results from Chapter 4 by showing that English speech disrupts these measures under normal reading conditions (i.e., without any visual masking).

Sound condition	Reading condition	Paragraph reading time (in s)	Intra-sentence regression probability	Inter-sentence regression probability	Number of fixations (per word)		
					1st-pass	2nd-pass	Total
		Easy questions					
Silence	normal	28 (8.3)	.25 (.43)	.08 (.27)	.80 (.8)	.27 (.66)	1.07 (1.02)
Silence	mask	25.3 (6.5)	.20 (.40)	.05 (.21)	.73 (.75)	.18 (.59)	.91 (.94)
English	normal	29.8 (9.9)	.30 (.46)	.08 (.28)	.80 (.82)	.34 (.86)	1.14 (1.19)
English	mask	25.5 (6.9)	.19 (.39)	.05 (.22)	.73 (.83)	.17 (.76)	.90 (1.17)
		Difficult questions					
Silence	normal	28.4 (7.7)	.26 (.44)	.10 (.30)	.81 (.82)	.29 (.81)	1.10 (1.14)
Silence	mask	25.1 (6)	.20 (.40)	.05 (.22)	.72 (.78)	.18 (.69)	.90 (1.06)
English	normal	31.5 (9.2)	.30 (.46)	.09 (.29)	.82 (.83)	.35 (.78)	1.17 (1.16)
English	mask	25.8 (8.1)	.20 (.40)	.04 (.21)	.71 (.76)	.18 (.61)	.89 (.96)

Table 12. Mean descriptive statistics of global reading measures in Chapter 5, Experiment 2 (SDs in parenthesis).

Sound condition	Reading conditio n	Question difficulty	Saccade length (in characters)	Landing position (in characters)
Silence	normal	easy	9.08 (6.94)	2.66 (2.37)
Silence	mask	easy	8.88 (6.32)	2.72 (2.39)
English	normal	easy	8.96 (6.60)	2.69 (2.33)
English	mask	easy	8.78 (6.45)	2.74 (2.38)
Silence	normal	difficult	9.16 (6.48)	2.71 (2.34)
Silence	mask	difficult	8.82 (6.21)	2.85 (2.42)
English	normal	difficult	8.88 (6.53)	2.75 (2.38)
English	mask	difficult	8.90 (6.58)	2.80 (2.38)

Table 13. Mean saccade length and landing position in Chapter 5, Experiment 2 (SDs in parenthesis).

Finally, question difficulty also interacted significantly with reading condition for paragraph reading time, inter-sentence regression probability, and number of first-pass fixations. This was due to longer paragraph reading times (d= 0.12), greater inter-sentence regression probability (d= 0.05), and more first-pass fixations (d= 0.02) when participants were answering difficult, as opposed to easy questions, but only when reading was normal and not when the text was presented with a trailing mask. This also replicates the question difficulty effects from Chapter 4 by showing that answering difficult comprehension questions leads to a change in reading behaviour that is characterized by more fixations and more regressions to previous sentences.

Effect	Paragraph reading time				Saccade length				Intra-sentence regression probability			
	b	SE	t	p	b	SE	t	p	b	SE	z	p
Intercept	10.2	.03	309.7	**<.001**	9.09	.19	47.9	**<.001**	-1.37	.09	-15.9	**<.001**
Sound	.02	.006	3.40	**.002**	-.03	.05	-.71	.47	.05	.02	2.54	**.01**
Diff	.01	.008	1.39	0.17	.002	.05	.03	.97	.03	.02	1.53	.12
RC	.07	.01	6.10	**<.001**	.05	.07	.78	.43	.24	.03	8.47	**<.001**
Sound: Diff	.006	.004	1.44	.15	.006	.02	.28	.77	.01	.008	1.52	.12
Sound: RC	.02	.004	4.16	**<.001**	-.05	.02	-2.43	**.01**	.06	.008	7.51	**<.001**
Diff: RC	.01	.004	2.47	**.01**	-.01	.02	-.48	.63	.001	.008	.14	.88
Sound: Diff: RC	.004	.004	.92	.36	-.04	.02	-1.79	.07	-.01	.008	-1.57	.11

Effect	Inter-sentence regression probability[1]				Number of 1^{st}-pass fixations				Number of 2^{nd}-pass fixations			
	b	SE	z	p	b	SE	z	p	b	SE	z	p
Intercept	-2.88	.08	-38.5	**<.001**	-.28	.02	-12.0	**<.001**	-1.66	.09	-19.0	**<.001**
Sound	-.05	.04	-1.15	.25	-.001	.005	-.15	.88	.06	.02	2.56	**.01**
Diff	.05	.04	1.29	.19	.001	.006	.26	.79	.03	.02	1.58	.11
RC	.23	.05	4.82	**<.001**	.06	.009	6.03	**<.001**	.29	.03	9.76	**<.001**
Sound: Diff	-.05	.01	-3.45	**.001**	<.001	.004	-.09	.92	.005	.007	.68	.49
Sound: RC	-.01	.01	-.68	.49	.002	.004	.44	.65	.06	.007	8.72	**<.001**
Diff: RC	.04	.01	2.87	**.004**	.01	.004	2.82	**.005**	.01	.007	1.53	.12
Sound: Diff: RC	-.01	.01	-.92	.35	.005	.004	1.30	.19	-.007	.007	-1.12	.26

Table 14. Results from (G)LMMs for global reading measures in Chapter 5, Experiment 2. Sound: background sound. Sound: background sound. Diff: question difficulty. RC: reading condition. Statistically significant *p*-values are formatted in bold.

[1] Reading condition was removed as a random slope for items due to convergence failure.

Effect	Saccade landing position			
	b	SE	t	p
Intercept	2.74	.05	54.2	**<.001**
Sound	.006	.01	.60	.54
Diff	.04	.02	2.56	**.01**
RC	-.04	.01	-3.11	**.004**
Sound: Diff	-.004	.007	-.49	.62
Sound: RC	.008	.007	1.14	.25
Diff: RC	-.01	.007	-1.47	.14
Sound: Diff: RC	.01	.007	1.31	.19

Table 15. Results from LMMs for saccade landing position in Chapter 5, Experiment 2. Sound: background sound. Sound: background sound. Diff: question difficulty. RC: reading condition. Statistically significant *p*-values are formatted in bold.

5.2.2.3. Word-level reading measures.

The descriptive statistics for word-level reading measures are displayed in Figure 21 and the LMM results are shown in Table 16. Consistent with the findings from Chapter 4, English speech resulted in significantly longer fixation duration for all three local reading measures compared to Silence (FFD: d= 0.04; GD: d= 0.04; TVT: d= 0.06). The reading condition also affected fixation durations: the trailing mask resulted in significantly longer FFD (d= -0.05) and GD (d= -0.08) compared to the normal reading condition. This suggests that reading the text with a trailing mask prolonged the first-pass fixation time on words. Conversely, the trailing mask condition resulted in significantly *shorter* TVT (d= 0.08) compared to the normal reading condition. This last effect was in the opposite direction because participants made fewer second-pass fixations in the trailing mask condition (which count towards TVT), presumably because the masked text did not provide any useful information and participants developed the strategy of avoiding it.

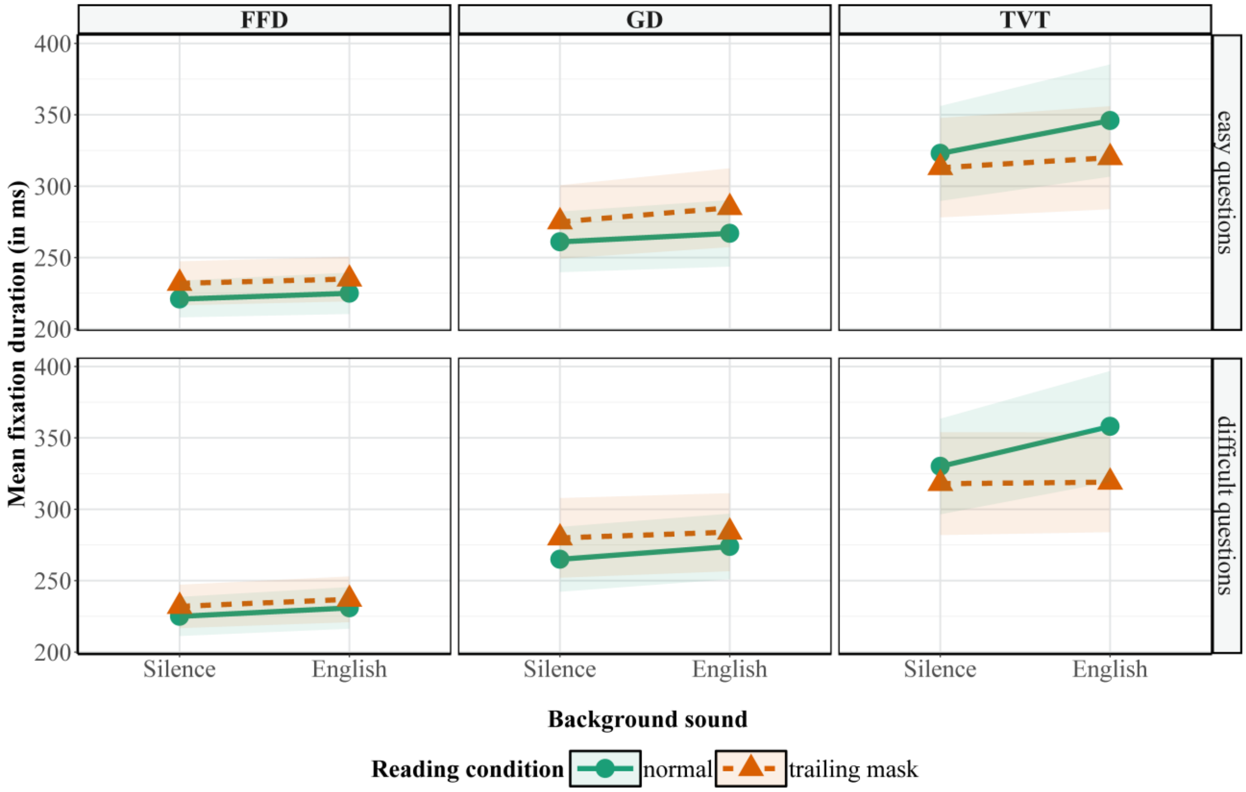

Figure 21. Mean descriptive statistics of local word-level reading measures in Chapter 5, Experiment 2. FFD: first fixation duration. GD: gaze duration. TVT: total viewing time. Shading indicates the standard error.

Background sound interacted significantly with reading condition for TVT, but not for FFD or GD. This was due to TVT being longer in the English speech condition compared to Silence (*d*= 0.10), but only when reading type was normal and not in the trailing mask condition. This replicates the result from Chapter 4 where TVT was also disrupted by English speech under normal reading conditions. Additionally, there was a significant two-way interaction between question difficulty and reading condition for both FFD and TVT. This also replicates the results from Chapter 4 by showing that FFD (*d*= 0.05) and TVT (*d*= 0.04) were longer when participants were answering difficult compared to easy questions, but only in the normal reading condition (which was equivalent to the reading mode in

Chapter 4). Finally, there was a significant three-way interaction for GD between background sound, question difficulty and reading condition. This was due to GD being longer in English speech compared to silence for all conditions, except in the trailing mask condition when the comprehension questions were difficult.

Effect	FFD				GD				TVT			
	b	SE	t	p	b	SE	t	p	b	SE	t	p
Intercept	5.36	0.01	369.7	**<.001**	5.48	.02	324.2	**<.001**	5.61	.02	270.8	**<.001**
Sound	.006	.003	2.27	**.03**	.009	.003	2.68	**.01**	.02	.004	3.62	**.001**
Diff	.005	.003	1.98	.054	.006	.003	1.76	.08	.009	.005	1.97	.055
RC	- .01	.004	-3.86	**<.001**	- .01	.004	-3.38	**.002**	.03	.007	4.36	**<.001**
Sound: Diff	.001	.002	.68	.49	<-.001	.002	-.006	.99	.001	.002	.34	.73
Sound: RC	.001	.002	.75	.45	.001	.002	.49	.61	.01	.002	5.22	**<.001**
Diff: RC	.003	.002	2.27	**.02**	.003	.002	1.40	.16	.005	.002	2.55	**.01**
Sound: Diff: RC	.001	.002	.47	.63	.004	.002	1.97	**.049**	.003	.002	1.55	.12

Table 16. Results from LMMs for local word-level reading measures in Experiment 2, Chapter 5. Sound: background sound. Diff: question difficulty. RC: reading condition. Statistically significant *p*-values are formatted in bold.

5.2.3. Discussion

Experiment 2 tested the distraction re-reading hypothesis, which predicted that intelligible speech will have a negative effect on the immediate comprehension of short paragraphs only when participants cannot go back to selectively re-read the text. The results provided support for this hypothesis because comprehension accuracy was significantly disrupted when re-reading of previous words was prevented in the trailing mask condition, but no such disruption was observed in the normal reading condition. At the same time, English speech resulted in a significant disruption of second-pass measures during normal reading, thus replicating the results from Chapter 4. Therefore, the present results provide direct empirical evidence that the increase in re-reading behaviour when listening to intelligible speech is related to maintaining an accurate comprehension of the paragraphs. As there were no significant interactions with the question difficulty condition, it appears that the disruption in comprehension occurs regardless of whether participants are answering easy or difficult questions. This is consistent with the results from Experiment 1 in this Chapter. The lack of an effect in inter-sentence regression probability also replicates the finding from Chapter 4 that intelligible speech does not seem to affect the integration of meaning across sentences. Finally, Experiment 2 also replicated the question difficulty effect on eye-movement measures from Chapter 4, which showed that participants made more fixations, more regressions to previous sentences and had longer TVT when answering difficult compared to easy questions.

While Experiment 2 replicated the main findings from Chapter 4, there may be a few apparent inconsistencies regarding the measures in which the effects were found. Before considering them, it is important to note that a direct replication of the intelligible speech and question difficulty effects from Chapter 4 can be shown in this experiment by a significant two-way interaction between each of the two factors and reading condition. This is because only the conditions with normal text presentation (and not the trailing mask one) corresponded to the reading conditions from Chapter 4. On the other hand, a main effect of background sound or question difficulty shows that the respective effect was observed in both the normal and the trailing mask condition. This would still be consistent with the findings from Chapter 4, but it would suggest that the effect is not limited only to normal reading.

The effect of intelligible speech in Experiment 2 was observed in the same dependent variables as in Chapter 4, apart from saccade length, which did not differ between the English and silence condition in Chapter 4. Nevertheless the difference in Chapter 4 was still in the expected direction and English speech also differed significantly from both Mandarin speech and Noise in that experiment. Additionally, while there was no interaction between background sound and reading condition for FFD and GD in the present experiment, the main effect of background sound was significant for both variables. This is still consistent with the results from Chapter 4 because it suggests that first-pass fixation durations generally increased in the English speech condition regardless of whether the text was normal or had a trailing mask. This is not surprising because the trailing mask manipulation had no effect on the first-pass fixations of words. Therefore, first-pass fixation durations generally increased in the presence of intelligible speech regardless of the reading condition. Finally, the only inconsistent finding with respect to question difficulty was that this effect was not found in the number of second-pass fixations. However, while not significant, the mean difference was still in the expected direction.

In summary, Experiment 2 found evidence that regressions and re-reading fixations allow readers to maintain the immediate comprehension of short paragraphs when listening to intelligible speech in the background. This suggests that readers use regressive eye-movements to resolve temporary comprehension difficulties that arise from semantic interference due to the irrelevant speech sound (see Marsh et al., 2008, 2009). While the present results demonstrate the link between regressive saccades and immediate text comprehension when reading under distracting conditions, they do not exclude the possibility that comprehension may still be negatively affected even if selective re-reading of the text is possible. Clearly, there is nothing that prevents readers from making regressions to previous words and sentences in everyday life situations. Additionally, re-reading has also been possible in previous studies that have shown disruption in comprehension accuracy by intelligible speech (e.g., Baker & Madell, 1965; Martin et al., 1988; Sörqvist, Halin, et al., 2010). This is not necessarily inconsistent with the present results, because they only show that readers can maintain the immediate comprehension of short paragraphs that are fairly easy to understand for skilled readers. For example, it is possible that the strategy of selectively re-reading the previous text may not be enough to compensate for semantic disruption when readers are processing longer and more complex texts (e.g., university level textbooks). Furthermore, background speech may also disrupt only long-term text comprehension, which could explain why comprehension has been disrupted in the above studies since they have generally used longer texts with greater delay between reading and the comprehension assessment. These are all possibilities that need to be explored in future research.

5.3. General Discussion

The present research investigated the role of regressive eye-movements on reading comprehension when listening to distracting intelligible speech in the background. This was the first attempt to directly examine how the disruption observed in eye-movement measures is related to participants' comprehension of the text. This is an important theoretical question as not all studies have found such disruption in measures of comprehension (see Chapters 1 and 3), but the disruption has been consistently observed in measures of second-pass reading (Chapters 3-4; Cauchard et al., 2012; Hyönä & Ekholm, 2016, Experiments 2-4; Yan et al., 2017). The present experiments found evidence that regressions support comprehension when listening to distracting irrelevant speech and that they allow readers to overcome the experienced distraction and still maintain an accurate comprehension of the text. Therefore, this is one potential explanation of why intelligible speech consistently disrupts measures of second-pass reading, but this does not necessarily translate into a decrease in comprehension accuracy. If readers can compensate for the additional processing difficulty by making more regressive saccades and more re-reading fixations, such a disruption in comprehension may not occur.

Traditionally, the role of regressive eye-movements during reading has remained elusive because regressions are not typically under the experimenter's control. Rayner (1998) recognised this problem in his classical review of the literature by noting that our understanding of regressions is limited by the fact that they cannot be induced experimentally in any easy way. Because of this, most of the evidence showing a link between regressions and online comprehension processes comes from studies that have manipulated certain properties of the text and then investigated how this affects regressive eye-movements (e.g., Blanchard & Iran-Nejad, 1987; Frazier & Rayner, 1982; Hyönä, 1995; Meseguer et al., 2002; Rayner et al., 2006). Recent studies have adopted a more direct approach by controlling what participants see when they make a regression and then studying how this affects comprehension processes (e.g., Booth & Weger, 2013; Inhoff & Weger, 2005; Schotter et al., 2014).

The present research builds upon these findings by showing that readers make use of regressions to overcome semantic distraction when their attentional resources are overtaxed by the irrelevant speech sound that causes a temporary difficulty in processing the meaning of the text (see Marsh et al., 2008, 2009). Therefore, the present findings are instrumental in demonstrating that regressions are triggered not only by comprehension difficulties caused by properties of the text itself (e.g., Frazier & Rayner, 1982; Meseguer et al., 2002; Rayner et al., 2006), but also by external auditory stimulation that is not relevant to the text in any way. In this sense, one function of regressions is likely to temporarily stop progressive reading behaviour and to re-direct the eyes to any immediate problems with the semantic processing of the text before readers can continue exploring the unread text. Because of this, the present research provides further evidence that regressions are a key component of the reading process and are necessary for achieving an accurate comprehension of the text (Schotter et al., 2014).

Of course, maintaining an accurate comprehension of the text is not the only purpose of regressions. For example, it has been argued that the majority of regressions during reading do not occur because of problems with comprehension, but rather due to oculomotor error or word identification problems (Vitu, 2005; Vitu & McConkie, 2000). Such regressive saccades are very short and are often assumed to land on the previous word in the sentence (Bicknell & Levy, 2011; Inhoff, Greenberg, Solomon, & Wang, 2009; Vitu, 2005). While this may be the norm for silent reading, the results from Chapters 3-4 clearly demonstrate that the frequency of regressive saccades due to comprehension difficulties can increase in different auditory environments. Intelligible speech in particular appears to put strain on online comprehension processes because it interferes with the semantic processing of the text and leads to an increase in the frequency of regressive saccades.

It is not likely that the increase in regressions and re-reading fixations by intelligible speech is due to word identification problems or oculomotor error for two reasons. First, Chapter 3 showed that intelligible speech did not disrupt the lexical processing of words. As word frequency and predictability are the two key variables that are theorised to influence word recognition times (e.g., Engbert et al., 2005; Reichle et al., 1998), this shows that intelligible speech disrupts the reading process only after word identification has been completed[16]. Additionally, intelligible speech did not affect saccade landing profiles in any of the three eye-tracking experiments, which also argues against an explanation based on systematic oculomotor error.

Traditionally, regressions have not played an important role in computational models of eye-movement control because such models have mostly been concerned with progressive reading behaviour (Vitu, 2005). For example, early versions of the E-Z Reader model (Pollatsek, Reichle, & Rayner, 2006; Reichle et al., 1998, 2003) have assumed that regressions to previous words occur when higher-level linguistic or comprehension processes are disrupted. Therefore, by modelling eye-movements only when such processes are running smoothly (Reichle et al., 2003), these models have avoided the problem of having to simulate regressions to previous words. Nevertheless, a more recent version of the E-Z Reader model (Reichle et al., 2009) has made a step in this direction by adding a post-lexical integration stage that can be used to simulate the effects of higher-level language processing. In this framework, the integration stage reflects the time needed to integrate the currently fixated word into the higher-level linguistic representation of the text. If this integration fails, the model can initiate a regression to the previous word in the sentence with a certain probability.

[16] In Chapter 3, word predictability was controlled for when examining the effect of intelligible speech on lexical processing of words.

In the SWIFT model (Engbert et al., 2002, 2005), regressive saccades to previous words occur due to incomplete word recognition and are a natural consequence of the target selection mechanism of the model. Because the probability of selecting a word as the next saccadic target is a function of its lexical activation, any word with a non-zero activation can be selected as a target for a regressive saccade (Engbert et al., 2002). As noted by Vitu (2005), the line of text in this model can be thought of as a saliency map where the saliency of each word is a function of its lexical activation. Therefore, the eyes will be drawn towards the word with the highest saliency at a given point in time. Interestingly, in their model, Engbert et al. (2002) distinguished between "local" regressions that are executed to words within the attentional window and "global" regressions that can occur at any point in time if a word has not been fully processed prior to its leaving the attentional window. Nevertheless, by definition, either type of regression can only occur due to incomplete word recognition.

Clearly, neither SWIFT nor the E-Z Reader in their current implementation can account for regressive saccades due comprehension difficulties that arise from listening to intelligible speech. Arguably, such effects may be easier to implement in the E-Z Reader framework due to the added post-lexical integration stage (Reichle et al., 2009). However, in their simulations, Reichle et al. adopted the simplifying assumption that the integration failure can only result in regressions to the previous word in the sentence. Such an assumption would be questionable in the context of auditory distraction by intelligible speech, given that the increase in re-reading fixations in Chapter 3 was found to occur not only on the immediately preceding word, but also on the previous 3-4 words before that. Additionally, modelling the increase in regressions by intelligible speech would require a deeper theory about the exact nature of the online comprehension difficulty, which cannot be constructed based on the present data alone.

Modelling the effect in SWIFT may prove to be more challenging, as this would require a new mechanism that can explain regressions not only due to word identification problems, but also due to comprehension difficulties. This may require the introduction of a second (independent) criterion for saccade targeting that can be influenced by online comprehension difficulties. For example, a new assumption could be added to the model that saccades can be targeted not only to words that have accrued lexical activation, but also to words that have already been lexically identified and have later accrued "comprehension difficulty" activation in a manner predicted by the theory. In this way, SWIFT could potentially retain its ability to explain saccade targeting decisions based on lexical activation, while also making it possible to simulate regressive eye-movements in response to semantic interference and online comprehension difficulties. In summary, studying the role of regressions in auditory distraction can help us improve our understanding of their function in the reading process and how they can be implemented in computational models of reading.

5.3.1. Methodological Challenges in Preventing Re-reading

The research presented in this Chapter also raises the question of what methodological approach is most suitable for preventing regressions when studying the role of re-reading fixations in auditory distraction. Clearly, both methodologies that were employed in the present research (RSVP and gaze-contingent masking) come with their unique set of advantages and disadvantages. Because re-reading fixations appear to be an important marker of auditory distraction by intelligible speech, it is perhaps useful to briefly consider how the strengths and weaknesses of the different paradigms may influence the results from such studies.

Schotter et al.'s (2014) trailing mask paradigm is very useful because it prevents the acquisition of any useful information during regressions while still keeping the reading conditions similar to natural reading. Of course, one technical limitation of this method is that it is more difficult to apply to larger pieces of text that also require tracking the vertical position of the eye. Due to the technical limitations of current eye-tracking systems, such applications inevitably require some additional checks to ensure that the trailing mask is accurately triggered in multiple-line experiments. Experiment 2 was one the first attempts to extend the trailing mask paradigm by Schotter et al. (2014) to multiple-line reading. The results suggested that this limitation can be overcome by adding gaze-contingent checks on each line. As participants could quickly adapt to these additional checks and generally did not report any difficulties with the reading task, this method could prove to be useful in future research. It should be noted that Olkoniemi, Johander, and Kaakinen (2018) have also recently used the trailing mask paradigm in a paragraph-reading study. However, in their experiment the trailing mask was triggered at the sentence level and not word-by-word as in the present research.

Another potential limitation that has not been thoroughly investigated until now is whether the trailing mask in itself may alter first-pass reading behaviour in some way. Although Schotter et al. (2014) considered this possibility in their original experiment, they only reported results for gaze durations on a single target word in the sentence. This in turn makes it difficult to assess how first-pass reading was affected more globally. To understand why the trailing mask may potentially affect first-pass reading behaviour, it might be helpful to compare it to another manipulation that involves a similar type of visual masking- the moving window paradigm (McConkie & Rayner, 1975; Rayner, 2014). In the moving window paradigm, the text that falls outside a pre-defined “window” around the point of gaze is visually masked by letters. This ensures that readers can process only the text that falls within this window. Although both manipulations involve considerable masking of the text, participants in the moving window paradigm still retain control over how they choose to read the sentence and the words are only temporarily masked and can be revealed again later on as long as they fall within the pre-defined gaze-contingent window. In contrast, words are permanently masked in the trailing mask paradigm and participants arguably become quickly aware of the fact that any saccade to the right will result in their permanent inability to further process visually any of the preceding words. This in turn may influence how participants approach the first-pass reading of the text and lead to changes in their reading behaviour.

In fact, this is exactly what was found in Experiment 2: first-pass reading changed in the trailing mask condition as participants made fewer but longer fixations compared to the normal reading condition. This indicates that there was a shift in the reading strategy adopted by participants, which may have occurred because they became more cautious in their first-pass reading of the text as to avoid accidentally masking any words before they had fully processed them. Given that any new fixation has the potential to unintentionally mask words before participants are ready to move on to the unexplored text, it might be advantageous to make fewer but longer fixations to minimise this risk while still allowing for enough time to process the fixated words.

Because Experiment 2 made critical predictions only with respect to second-pass reading and comprehension accuracy, this change in first-pass reading behaviour does not compromise the conclusions from this experiment. However, future studies using a similar implementation of the trailing mask paradigm may need to take this into account if their hypotheses also make predictions about first-pass reading. A similar change in first-pass reading behaviour may not have necessarily occurred in Schotter et al.'s (2014) experiment due to the shorter text stimuli that were used. Further research (or a re-analysis of their data) is needed to test if this is the case.

RSVP presentation has the advantage that it is easy to implement and does not require the use of expensive eye-tracking equipment. However, as it was already mentioned in the Discussion section of Experiment 1, this method has some disadvantages, such as the different spatial presentation of the text and the uniform forced-fixation time that is used for all words. Nevertheless, it is worth mentioning that the RSVP technique could potentially be modified to make it more similar to how readers normally process words. For example, it would be possible to adjust the fixation time for each individual word based on its psycholinguistic variables such as word frequency and word predictability, while still maintaining the same overall reading speed of the text. Given that computational models of eye-movement control can successfully simulate fixation durations based on these variables (Engbert et al., 2002, 2005, Reichle et al., 1998, 2003, 2009; Schad & Engbert, 2012), it seems likely that this may make the reading task more natural. In fact, there is some evidence indicating that adjusting the exposure time of words based on their length or lexical frequency can improve the readability of the text (Öquist & Lundin, 2007; Öquist, Sågvall-Hein, Ygge, & Goldstein, 2004).

When using an RSVP mode of presentation, it is important to maintain reading speed that is similar to that of normal reading. For example, in Experiment 1 this was achieved by taking the average speed of readers on the same materials and from the same population, and then adjusting it for the lack of parafoveal preview of the upcoming word (which would be available under normal reading conditions). While the RSVP method of reading has often been criticised on the grounds that it negatively affects text comprehension (e.g., Acklin & Papesh, 2017; Benedetto et al., 2015; Rayner, Schotter, Masson, Potter, & Treiman, 2016), this decrease in comprehension largely depends on the reading speed that is being used. It is well known that comprehension during RSVP reading decreases as a function of increasing reading speed (Juola, Ward, & McNamara, 1982; Masson, 1983; Ricciardi & Di Nocera, 2017), which demonstrates that there is a speed-accuracy trade-off in this mode of presentation (Rayner et al., 2016). Therefore, when using RSVP to study auditory distraction, it is important to use a reading speed that has been equated to normal reading for the stimuli and participants under investigation.

Although not explicitly considered in this chapter, *self-paced reading* (e.g., Aaronson & Scarborough, 1976; Jegerski, 2014; Mitchell & Green, 1978) is another method that could be used to prevent regressions in a similar way to the RSVP presentation in Experiment 1. In self-paced reading, the first word of the sentence is revealed on the screen and participants press a button when they are ready to reveal the next word. This procedure is then repeated until the whole sentence has been presented (Jegerski, 2014). Typically, there are two ways in which the text can be displayed: 1) *cumulative* self-paced reading, in which previously-read words remain visible for the whole duration of the trial; and 2) *non-cumulative* self-paced reading, in which only the currently-read word is visible and all previous words are masked after they have been processed (Jegerski, 2014). The non-cumulative method prevents re-reading behaviour in a similar way to the trailing mask paradigm as participants cannot obtain any useful information from the masked words even if they are re-fixated later on during a regression.

Self-paced reading and RSVP presentation are similar to one another in the sense that both methods present new text one word at a time. However, in self-paced reading, participants manually choose when to move on to the next word by pressing a button. Because of this, they regain control over how long to fixate each individual word. In this way, self-paced reading avoids the problem of using a uniform forced fixation time for each word as in the RSVP mode of presentation. Additionally, it also allows the text to be displayed spatially on different lines, which is typical for everyday reading. While this method is very useful because it avoids some of the limitations of RSVP reading, it becomes increasingly less feasible the longer the reading stimuli are. Because a manual response is required on every word, presenting large pieces of text could potentially cause fatigue and task disengagement due to the large number of motor responses that would be needed. For example, even the passages from the present study (which were fairly short) would have required a few thousand button presses. Therefore, this method of preventing regressions would be less suitable when using large pieces of text, such as book chapters or long stories.

In summary, RSVP, self-paced reading, and Schotter et al.'s (2014) gaze-contingent trailing mask are all methods that can be used to prevent re-reading behaviour when studying auditory distraction. Since all of them have their own advantages and disadvantages, the choice of method may depend on the type of research question and the reading stimuli. Self-paced reading and the trailing mask paradigm would usually be the preferred methods of choice, particularly for text stimuli that are not excessively long. Additionally, depending on the type of research question, the trailing mask paradigm may be preferable out of the two as it provides richer data about fixation patterns during reading. Finally, RSVP presentation may also be useful in some cases, but only if reading speed is kept as closely as possible to normal reading, and if the spatial presentation of the text and the uniform forced fixation time are not deemed to be of critical importance.

5.3.2. Conclusion

Previous studies have demonstrated that intelligible speech leads to an increase in re-reading fixations but, until now, little was known about why this occurs. The present research demonstrated that the increase in re-reading fixations occurs because participants are actively trying to maintain comprehension of the text when reading under distracting conditions. Once participants' ability to selectively re-read the text was prevented, their immediate comprehension was negatively affected. This suggests that regressions and re-reading fixations play a key role in overcoming transient interference from the irrelevant speech and allow readers to resolve any comprehension difficulties before resuming the progressive reading of the text.

CHAPTER 6: DISTRACTION BY DEVIANT SOUNDS DURING READING

Oddball studies have shown that task-irrelevant sounds that unexpectedly differ from an otherwise structured or repeated sequence of sounds yield specific electrophysiological responses and behavioural distraction in an unrelated task (Berti, 2008; Berti & Schröger, 2001, 2003; Horváth, Roeber, Bendixen, & Schröger, 2008; Schröger, 1996). In the oddball paradigm, an irrelevant sound is presented before the appearance of a target stimulus on the screen. On most trials, the same sound is presented (standard), while on rare and unpredictable occasions it is replaced by a different sound (deviant). The typical finding from such studies is that deviant sounds delay responses in categorization tasks where participants must respond to target stimuli while ignoring task-irrelevant sounds (Ljungberg & Parmentier, 2012; Parmentier, 2014; Parmentier et al., 2008; Parmentier, Vasilev, & Andrés, 2018).

Previous research has shown that deviant sounds are distracting not because of their acoustic features per se, but rather because they violate the cognitive system's predictions (Bubic, von Cramon, Jacobsen, Schröger, & Schubotz, 2009; Parmentier, Elsley, Andrés, & Barceló, 2011). In fact, attentional distraction has been observed at the electrophysiological and behavioural level for both small pitch differences and larger spectral differences between the standard and the deviant sound (Parmentier et al., 2008; Schröger, 1996). Additionally, there is abundant evidence showing that deviance distraction does not depend on the specific identity of the sounds: it occurs regardless of whether sound A (e.g., a sinewave tone) is used as the standard and sound B (e.g., white noise) is used as the deviant, or vice versa (Leiva, Parmentier, et al., 2015b). This latter finding has also been shown to generalise to the tactile modality (Parmentier, Ljungberg, et al., 2011).

Interestingly, deviant sounds that convey meaning can also yield distraction because they undergo some automatic semantic evaluation (Parmentier & Kefauver, 2015; Parmentier, Pacheco-Unguetti, & Valero, 2018; Roye, Jacobsen, & Schröger, 2007; Schröger et al., 2000). For example, the deviant sounds “left” and “right” affect response times in a left/right arrow categorization task as a function of the relationship (congruent or incongruent) between the deviant words’ meaning and the visual arrows (Parmentier, 2008; Parmentier & Kefauver, 2015; Parmentier, Turner, & Elsley, 2011; Parmentier, Turner, & Perez, 2014). In addition, the semantics of deviant sounds can be processed even when the words’ meaning bears no connection to the primary task (Escera, Yago, Corral, Corbera, & Nuñez, 2003). For example, participants performing a digit categorization task in a state of hunger exhibit greater distraction by deviant words related to food compared to neutral words (Parmentier, Pacheco-Unguetti, et al., 2018). It is also worth noting that some studies have reported evidence for neural responses to the semantic content of unexpected sounds even when participants are passively exposed to such sounds (Czigler, Cox, Gyimesi, & Horváth, 2007; Frangos, Ritter, & Friedman, 2005; Friedman, Cycowicz, & Dziobek, 2003; Roye et al., 2007; Shtyrov, Hauk, & Pulvermuller, 2004; Shtyrov & Pulvermuller, 2003). Finally, deviance distraction can also be modulated by other factors, such as participants’ age. For example, deviant sounds cause greater behavioural distraction in old age under certain conditions (Leiva, Andrés, & Parmentier, 2015; Leiva, Parmentier, & Andrés, 2015a), although this does not appear to reflect age-related differences in the electrophysiological orienting response (Berti, Vossel, & Gamer, 2017).

The traditional explanation of deviance distraction is that it reflects an involuntary switch of attention away from the main task that is caused by the detection of subtle auditory changes in the human brain (Escera et al., 1998; Schröger, 1996). Therefore, deviant sounds likely trigger a neural system for monitoring the external sensory input; once this input exceeds a certain sensory threshold, attentional resources are obligatorily redirected towards the deviant stimulus. In this sense, deviance distraction is typically viewed as an orienting response (see Sokolov, 1963) that is characterized by a burst of arousal and a reflexive orienting of attention towards the eliciting stimulus (Näätänen, 1992). Deviant sounds are associated with a specific neurophysiogical signature that is shown by three distinct ERP components: 1) the early MMN component (Näätänen et al., 1978, 2007) that reflects the pre-attentive detection auditory changes in the brain (Berti & Schröger, 2001); 2) the P3a component that reflects the involuntary orienting of attention towards the deviant sound (Berti & Schröger, 2001; Escera et al., 2000), and 3) the RON component that reflects the refocusing of attention back to the main task (Berti, 2008; Schröger & Wolff, 1998a).

Interestingly, while deviance distraction has typically been regarded as an example of attentional distraction, recent work suggests that deviant sounds may also affect behaviour by triggering a temporary inhibition of motor cortical areas (Wessel, 2017; Wessel & Aron, 2013). For example, Wessel and Aron (2013) reported an experiment in which a 200 ms sound (either a standard or a novel one) was presented 300 ms before the appearance of a target letter on the screen. Participants' task in this study was to speak out the target letter presented on each trial. Critically, however, Wessel and Aron also administered transcranial magnetic stimulation (TMS) on the right hand locus of the motor cortex at varying time intervals after the onset of the sound (at 150, 175, or 200 ms). The effect of the TMS stimulation was then measured with motor-evoked potentials on participants' right hand (which was not relevant to the naming task).Wessel and Aron (2013) observed reduced corticospinal excitability of the hand following the presentation of novel sounds. This effect was found only when the TMS stimulation was administered 150 ms after the onset of the novel sound, but not in the later time intervals. Because corticospinal excitability of the hand was unrelated to participants' performance on the letter naming task, the authors argued that novel sounds induce global motor inhibition by activating the same neural circuits that are used for interrupting ongoing actions. A more recent study has extended this finding by showing that the reduction in corticospinal excitability by unexpected novel sounds is significantly and positively correlated (r= 0.45) with action-stopping behaviour in a Go/ NoGo task (Dutra, Waller, & Wessel, 2018).

Therefore, these recent results extend the traditional explanation of deviance distraction as an orienting response by suggesting that unexpected sounds may also induce global motor inhibition because they recruit the same neural circuits that are used to stop ongoing action plans. The purpose of this global inhibition is thought to be the temporary suspension of ongoing processes that may facilitate the effective and timely processing of the unexpected stimulus (Wessel, 2017; Wessel & Aron, 2017). In this sense, the orienting response and motor inhibition accounts are not necessarily mutually exclusive and both could be a consequence of encountering unexpected sounds in the environment.

The potential role of motor inhibition in deviance distraction is exciting as it suggests that deviant sounds may potentially affect a large range of activities, including those relying on relatively automatic motor processes. However, despite the potential impact of novelty distraction on everyday life situations, these studies have used simple laboratory tasks that bear little resemblance to more complex and ecologically-valid tasks. In the present study, we sought to address this issue by exploring for the first time whether deviant sounds may affect performance on one important and complex everyday task: reading. To do so, we developed a new method to measure the effect of deviant sounds on eye-movements during reading.

Reading is a theoretically interesting task for studying the effect of deviant sounds on human performance because it does not require any specific response from participants upon hearing the task-irrelevant sounds. Unlike categorization tasks where participants need to make a dichotomous response after the presentation of the sound (e.g., judging whether a number is odd or even; Parmentier et al., 2008), subjects simply have to read the text for comprehension and ignore the task-irrelevant sounds. This makes it possible to investigate how deviant sounds affect performance on a natural, everyday task that does not involve any response preparation or the need to act upon a specific stimulus after the sound is presented. Additionally, skilled adult reading is a fairly automatized process (LaBerge & Samuels, 1974), which involves the intricate coordination of oculomotor and cognitive processes that determines when and where to move the eyes next. As a result, it can yield valuable insights into how deviant sounds influence cognitive and oculomotor processes.

While the topic of auditory distraction during reading has a very long history, most studies have only considered the influence of continuous auditory distractors (e.g., irrelevant speech or music) on behavioural measures such as comprehension accuracy (see Chapters 1-2). However, recording participants' eye-movements makes it possible to investigate how irrelevant sounds affect the moment-to-moment decision of when and where to move the eyes next. Because eye-movements during reading are sensitive to the underlying cognitive processing of the text (Rayner, 1998, 2009), this method has the potential to detect transient auditory distraction effects that may not be captured by behavioural measures of comprehension (see Chapters 3-5).

It is currently not known whether eye-movements during reading are sensitive to discrete deviant sounds. However, recent work does indicate that they are affected by certain types of continuous sounds. For example, background music or unintelligible speech in a foreign language generally do not appear to affect fixation durations or fixation probabilities during reading (Cauchard et al., 2012; Hyönä & Ekholm, 2016; R. Johansson et al., 2012; although see Zhang et al., 2018 for conflicting evidence). However, intelligible background speech disrupts the ongoing reading process by prompting participants to make more re-reading fixations on previously-read words (Cauchard et al., 2012; Hyönä & Ekholm, 2016; Yan et al., 2017). This latter finding is interesting because it suggests that certain task-irrelevant sounds can have a direct influence on eye-movements and interfere with the ongoing cognitive processing of the text.

While discrete deviant sounds could also potentially influence fixation durations during reading, this is not expected to occur through the same mechanism that is responsible for distraction by intelligible speech. The available evidence suggests that intelligible speech causes distraction because readers process its semantic features, which in turn interferes with extracting the meaning of the written text (e.g., Hyönä & Ekholm, 2016; Martin, Wogalter, & Forlano, 1988; see also Marsh, Hughes, & Jones, 2008, 2009). On the other hand, we hypothesize that deviant sounds would cause distraction not because of semantic interference, but rather because they violate readers' expectation that another standard sound will be presented (Bubic et al., 2009; Parmentier, Elsley, et al., 2011). Therefore, the present study builds upon previous work on distraction by continuous sounds by exploring a different mechanism through which task-irrelevant sounds may influence eye-movements during reading.

6.1. Present Study

The available evidence from both electrophysiological (e.g., Escera et al., 2000; Schröger, 1996) and behavioural studies (e.g., Dalton & Hughes, 2014; Parmentier, 2014) has shown that task-irrelevant deviant sounds can distract participants from their main task. The goal of the present study was to investigate whether fixation durations during reading can also be influenced by discrete deviant sounds. To study deviance distraction during reading, a new manipulation was developed in which the presentation of task-irrelevant sounds was contingent on participants' eye-movements. While participants read single sentences for comprehension, short sounds were presented upon fixating five target words in each sentence. On most occasions, the sound was a sine-wave tone (this served as the standard sound), while on rare and unpredictable occasions this sound was replaced by a short burst of white noise (this served as the deviant sound).

Only few studies to date have used gaze-contingent auditory presentation in a reading task (Eiter & Inhoff, 2010; Inhoff, Connine, Eiter, Radach, & Heller, 2004; Inhoff, Connine, & Radach, 2002). For example, Inhoff et al. (2004) presented a spoken word once participants fixated a target word in the sentence. In their experiment, there were three types of spoken words that participants heard: 1) identical to the target word; 2) phonologically similar to the target word; and 3) phonologically dissimilar to the target word. To present the spoken word, Inhoff et al. (2004) utilized the gaze-contingent boundary paradigm (Rayner, 1975) that has been extensively used to study the perceptual span during reading (see Rayner, 1998, 2009). In the classical boundary paradigm, a target word in the sentence (e.g., "couch") is visually masked by a string of letters before participants directly fixate it (e.g., "couch" -> "xxxxx"). Once the gaze position of the eye crosses an invisible boundary located just before the target word, the mask is replaced by the target word itself ("xxxxx"-> "couch"). Most of the time, this change happens during the saccade towards the target word and it is not typically perceived by participants due to saccadic suppression (Schotter et al., 2012; although, see Angele, Slattery, & Rayner, 2016; Slattery, Angele, & Rayner, 2011; White, Rayner, & Liversedge, 2005).

By applying the same general method to the auditory modality, Inhoff et al. (2004) presented the spoken word in their experiment once participants' gaze position crossed the invisible boundary located just before the target word. The present experiment utilized a similar procedure in which an invisible boundary was inserted before each of the five target words in the sentence. Once participants' gaze crossed each of the five invisible boundaries, the irrelevant sound was presented (i.e., either the standard or the deviant one).

If deviant sounds in this task elicit an orienting response (e.g., Escera et al., 1998; Parmentier, 2014), this may happen either overtly or covertly. According to Posner (1980), *overt orienting* occurs when there is an eye-movement directed towards the eliciting stimulus. Conversely, *covert orienting* occurs when there is a shift of attention without any corresponding eye-movements. If deviant sounds trigger an overt shift of attention, this would manifest itself in shorter fixation durations after the presentation of the sound. This is because the current fixation will be interrupted and the orienting response (i.e., the eye-movement) will occur. However, because the deviant stimulus was in the auditory domain and because it was not linked to a specific location on the screen, it is more likely that the orienting response would occur covertly (i.e., without an eye-movement). In this case, attention would be redirected away from processing the words in the sentence and towards the deviant sound. This would be associated with an increase in fixation durations that could be due to a disruption in the lexical processing of the target words or, if the orienting response occurs later in time, the allocation of attention to other words in the sentence, or the post-lexical stages of sentence integration. Similarly, if deviant sounds elicit global motor inhibition (Wessel & Aron, 2013), this would also lead to longer fixation durations after the presentation of the sound, which would likely be due to a delay in saccade programming.

On the basis of previous work suggesting that deviant sounds delay the processing of target stimuli (Parmentier et al., 2008) and the notion that this might reflect the temporary suppression of ongoing actions (Wessel & Aron, 2013), we hypothesized that the deviant sounds would lead to an increase in fixation durations on the target words. Additionally, to determine if any potential increase in fixation durations is due to a disruption in the initial stages of lexical processing, we examined the time course of the effect and tested whether it is modulated by the corpus lexical frequency of target words.

6.2. Method

6.2.1. Participants

Forty-eight students from Bournemouth University participated for course credit (45 female)[17]. Their average age was 19.7 years (*SD*= 2.4 years; range 18-32 years). All

[17] Two more participants were tested, but they were excluded due to tracking problems.

participants were native speakers of British English, reported normal or corrected-to-normal vision, normal hearing and no prior diagnosis of reading disorders. Participants were naïve as to the purpose of the experiment. Ethical approval of the study was obtained from the Bournemouth University Research Ethics Committee (protocol No. 16999).

Because no previous studies of deviance distraction during reading exist, it was not feasible to conduct a formal statistical power analysis. However, the same number of subjects was tested as in the previous experiments in this Thesis. Additionally, the present experiment had more trials per sound condition compared to the sentence-reading experiment from Chapter 3 (40 vs. 32 trials), which also helped improve statistical power.

6.2.2. Materials and Design

6.2.2.1. Reading stimuli.

The reading materials consisted of 120 English sentences (see Appendix H for a complete list). Their average length was 14.3 words (range: 13-18 words). Each sentence contained five target words on which the sound stimuli were played (we use the term "target" to denote the words on which sounds were played, as opposed to the other words in the sentence where no sounds were played). This is illustrated in Figure 22a. The five target words were always the third, fifth, seventh, ninth, and eleventh word in the sentence. Their mean length was 6.75 letters (*SD*= 1.89 letters; range: 3-13 letters). The average lexical frequency of the target words was 180 counts per million in the SUBTLEX-UK database (Van Heuven, Mandera, Keuleers, & Brysbaert, 2014; *SD*= 786; range: 0.06 – 15450 counts per million). There was one non-target word between any two target words that served as a buffer to increase the sound inter-stimulus interval. Additionally, there were always at least two words following the last target word in the sentence in order to avoid artefacts due to sentence wrap-up effects. Short function words that are likely to be skipped during first-pass reading were not used as target words.

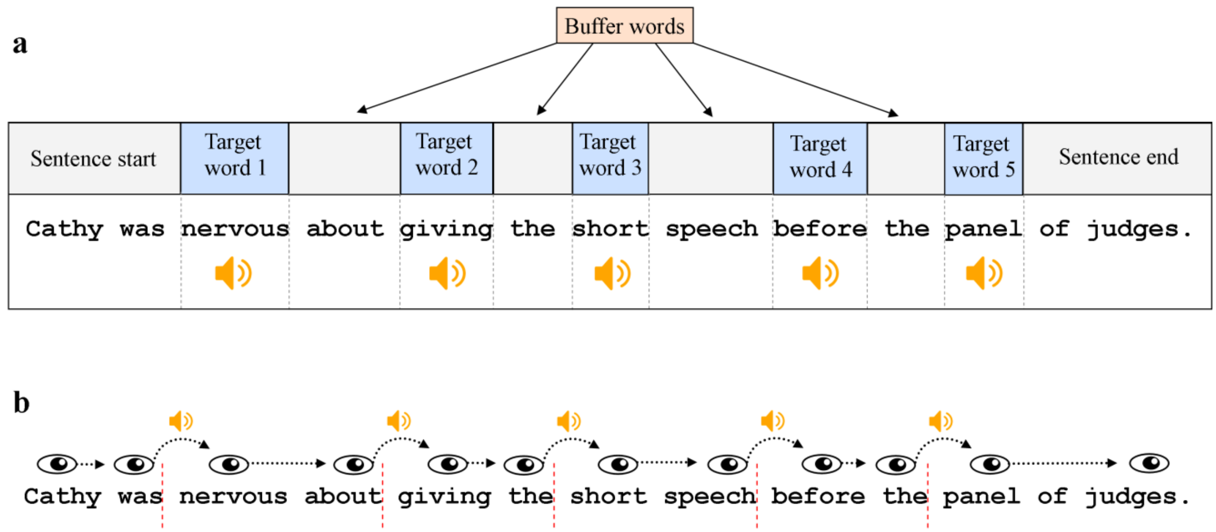

Figure 22. An illustration of the sound presentation in Chapter 6. Panel **a** shows an example sentence and the position of the target words. Panel **b** illustrates the gaze-contingent sound presentation. Once the eye moves to the right of each boundary, the sound is played. The invisible boundaries are shown by vertical red lines.

Because the reading stimuli were specifically developed for this experiment and were not part of a pre-existing corpus, a norming study was carried out to determine their difficulty and naturalness. Twelve participants (9 female; mean age= 29.1 years) who did not take part in the eye-tracking experiment were asked to read each sentence and rate it on a scale from 1 to 10 based on how difficult and how natural they thought it was (1= very difficult/ very unnatural; 10= very easy/ very natural). Participants were instructed that very difficult sentences (1) are those that they cannot understand at all and that very easy sentences (10) are those that they have no problems at all understanding. Likewise, they were instructed that very unnatural sentences (1) are those that sound very unusual to them and are not at all typical for normal English language use, and that very natural sentences (10) are those that sound very normal to them and are very typical for normal English language use. The results showed that the sentences were rated both as very natural (*M*= 8.84; *SD*= 1.53) and very easy to understand (*M*= 9.29; *SD*= 1.23). In summary, the reading stimuli in the present experiment sounded very natural to speakers of English and were fairly easy to understand.

6.2.2.2. Auditory stimuli.

The standard sound was a 400 Hz sine wave and the deviant sound was a burst of white noise. The sounds were generated in Matlab R2014a (MathWorks, 2014). Both sounds were 50 ms long and had a 10 ms fade-in and fade-out. The amplitude resolution of the sounds was 16 bits and the sampling frequency was 48 kHz.

There were two experimental blocks: one block of 40 sentences that was completed in silence, and another block of 80 sentences that contained the gaze-contingent sound presentation. The purpose of adding the silence block was to determine whether the presentation of five gaze-contingent standard sounds leads to a change in reading behaviour compared to reading in silence. This was important as previous studies have not presented multiple gaze-contingent sounds in a reading task. The gaze-contingent block contained two types of trials: 1) trials that contained five standard sounds; and 2) trials that contained four standard sounds and one deviant sound. Trials with standard-only sounds were used to limit participant's ability to predict the appearance of a deviant sound. This was because there is evidence that distraction can be reduced or even eliminated when the deviant sound is highly predictable (e.g., Parmentier et al., 2011; Vachon, Hughes, & Jones, 2012).

In trials with a deviant sound, the first sound (on target word 1) was also always a standard sound. This was done in order re-activate the representation of the standard sound at the beginning of the trial. The deviant sound was then presented on one of the four remaining target words (2 through 5) with equal probability across the experiment. The remaining three target word positions were again filled with standard sounds. In this way, there was a 50% probability that a trial in the gaze-contingent sentence block would contain a deviant sound. The experiment-wise occurrence of a deviant sound was kept low, with 10% of all sounds being deviant. The assignment of conditions to sentences, the position of the deviant sound and the order of the two experimental blocks were counter-balanced with a full Latin square design. The sentences within each block appeared in random order. The gaze-contingent block always started with three standard-only trials to establish the sine wave as the standard sound.

6.2.3. Apparatus

Participants' eye-movements were recorded with an EyeLink 1000 Tower Mount at a sampling frequency of 1000 Hz. The resolution noise was < 0.01° and the velocity noise was < 0.5° on average. Participants rested their chin on a headrest to minimise head-movement artefacts. Viewing was binocular, but only the right eye was recorded. The experiment was programmed in Matlab R2014a (MathWorks, 2014) by using the Psychophysics toolbox (Brainard, 1997; Pelli, 1997) and Eyelink libraries (Cornelissen et al., 2002).

The sentences were presented on a Cambridge Research Systems LCD++ monitor with a screen resolution of 1920 x 1080 pixels and a refresh rate of 100 Hz. The sentences were formatted in a Courier New 18 pt. font and appeared as black text over white background on a single line in the middle of the screen. The width of each letter was 14 pixels. The distance between the eye and the monitor was 80 cm. At this distance, each letter subtended approximately 0.34° per visual angle. The sounds were played on a Creative Labs Sound Blaster X-Fi SB0770 sound card by using the low-latency mode of presentation in the Psychophysics toolbox (Brainard, 1997; Pelli, 1997). The sound stimuli were presented binaurally at 65 dB(A) SPL through Bose QuietComfort 25 noise-cancelling headphones [18]. The experiment was run on a PC in a Windows 7 environment.

6.2.4. Procedure

Participants were tested individually in a session that lasted for about 30-40 minutes. Participants were instructed that they will sometimes hear short sounds while they are reading, but that they should try to ignore them and read as normally as possible. Before the start of the experiment, participants were calibrated on a 3-point calibration grid. A drift check was presented before each trial and participants were re-calibrated whenever that was necessary. The calibration error was kept at < 0.3°. All beeps during calibration and drift check were turned off. The experiment started with six practice trials where no sounds were presented. Each trial began with a black gaze box that appeared at 50 pixels from the left side of the screen. Once participants fixated the box for 100 ms, it disappeared and the sentence was presented on the screen, with the first letter appearing in the middle of the location where the box was.

The gaze-contingent sound manipulation is illustrated in Figure 22b. An invisible boundary (Rayner, 1975) was placed at the first pixel of the empty space before each of the

[18] While the sound intensity level was 5 dB(A) higher than the one used in the previous experiments, this change was done to ensure that the sound stimuli were still clearly audible. This was necessary as both sounds were very short (only 50 ms), whereas the sounds in the previous studies were played continuously for the duration of the whole trial.

five target words. Once the eye crossed a boundary, the command to play the sound was sent. This usually happened during the saccade towards the target word. The delay between sending the command to play the sound and the sound coming out from the headphones was measured to be 14 ms with the Black Box ToolKit 2 (Sheffield, UK). Because the command to play the sound was usually sent several milliseconds before the end of the saccade, participants heard the sound within several milliseconds of the fixation onset on the target word. In other words, the sound onset delay relative to the fixation onset was 14 ms minus the time difference between sending the command to play the sound and the start of the next fixation. The distribution of sound onset delays relative to the start of fixation is presented in Figure 23a. To avoid sound overlap when participants make a long saccade and trigger more than one boundary, a sound was played only when at least 10 ms had passed since the previous sound had stopped playing. Each sound was played only once when the target word was first fixated; the sound was not repeated if the target word was subsequently re-fixated during a regression. Participants pressed the left button of the mouse to terminate the trial. One third of the sentences were followed by a "Yes/No" comprehension question. For example, in the sentence "Cathy was nervous about giving the short speech before the panel of judges.", the comprehension question was "Was Cathy relaxed about giving her speech? Yes/No".

6.2.5. Data Analysis

The experiment had a within-subject design with one factor: sound type (silence, standard, deviant). Because the deviant sound is rare compared to both the standard sound and reading in silence, it was necessary to analyse a balanced dataset. Fixation durations during reading are influenced not only by the sound manipulation, but also by the visual and cognitive processing of the words on which they occur. Therefore, only fixations on target words 2-5 were analysed because these are the words on which all three sound conditions were presented. Additionally, in the silence and standard sound condition, only one target word was sampled per trial. This word was determined from the design matrix that was used to counter-balance the word on which the deviant sound was presented (i.e., a specific target word position was assigned to all sound conditions, although this had no special meaning for the silence and standard sound conditions). This made it possible to analyse a fully-balanced dataset with an equal number of observations per sound condition before the data pre-processing stage.

A few standard fixation duration measures were used as dependent variables: first fixation duration (FFD), which is the duration of the first fixation on the word; single fixation duration (SFD), which is the fixation duration when the word was fixated only once; gaze duration (GD), which is the sum of all fixations on the word before the eyes move on to another word; and total viewing time (TVT), which is the sum of all fixations on the word, including the ones made during a regression. Additionally, a few measures of global reading were also analysed: sentence reading time, fixation duration, number of fixations, and saccade length. The data were analysed with (Generalised) Linear Mixed Models ((G)LMMs) by using the lme4 package v. 1.1-12 (Bates et al., 2014) in R 3.3.0 (R CoreTeam, 2016). Sound type was entered as a fixed effect in the model. Random intercepts and random slopes for sound type were added for both participants and items[19] (Baayen,

[19] Only the models for GD and TVT converged with a random slope for items. In all remaining models, the slope for items was removed. When there was a convergence failure, we first tried to remove the random intercept before removing the random slope. If the model still did not converge, the intercept was retained, but the slope was removed. This was done following Barr et al.'s (2013) recommendation that, when the

Davidson, & Bates, 2008; Barr, Levy, Scheepers, & Tily, 2013). Fixation durations were log-transformed in all analyses. Treatment contrast coding was used where the standard sound was the baseline. This contrast coding made it possible to do the two key comparisons in the experiment: 1) the difference between silence and the standard sound to determine whether the presentation of gaze-contingent sounds in general influenced reading behaviour compared to reading in silence; and 2) the comparison between the deviant and the standard sound to test for the presence of a sound deviance effect. *P*-values were calculated with the lmerTest package v.2.0-33 (Kuznetsova et al., 2017). The results were considered statistically significant if the p values were ≤ 0.05. Standardized effect sizes in Cohen's d (Borenstein, 2009) are also reported for the significant results.

6.3. Results

The average comprehension accuracy was 95% (*SD*= 22%), thus indicating that participants understood the sentences (Silence: *M*= 95.1%; *SD*= 21.7%; Standard: *M*= 94.7%; *SD*= 22.4%; Deviant: *M*= 95.2%; *SD*= 21.3%). There were no differences in comprehension accuracy across the three sound conditions (all ps ≥ 0.62). Only two participants (4.1%) reported some awareness that the sounds were played depending on their eye-movements when asked after the experiment[20]. Trials with blinks were excluded from the data (6.2%). Additionally, trials in which the command to play the sound was sent after the start of fixation on the target word were also excluded (12.2%). Finally, trials with boundary "hooks" were also excluded (8.5%). A hook occurs when the eye crosses the

---

"maximal" model (in this case, a model with both a random slope and a random intercept for sound type) fails to converge, a model with a missing intercept is preferable over a model with a missing slope. This was because models without random slopes were found to be more anti-conservative in their analyses (i.e., they led to an increase in Type 1 error probability).

[20] Their data was retained since they were not completely sure of this and they reported that it did not influence how they read the sentences.

invisible boundary, thus triggering the sound, but then returns to the left of the boundary and lands on previous words. Fixations shorter than 80 ms that occurred within one letter of another fixation were combined with that fixation. Trials with fixation durations longer than 800 ms for FFD, 2000 ms for GD, and 4000 ms for TVT were excluded as outliers in all analyses (0.44 % of the data). This left 72.6% of the data for analysis (a total of 4206 trials).

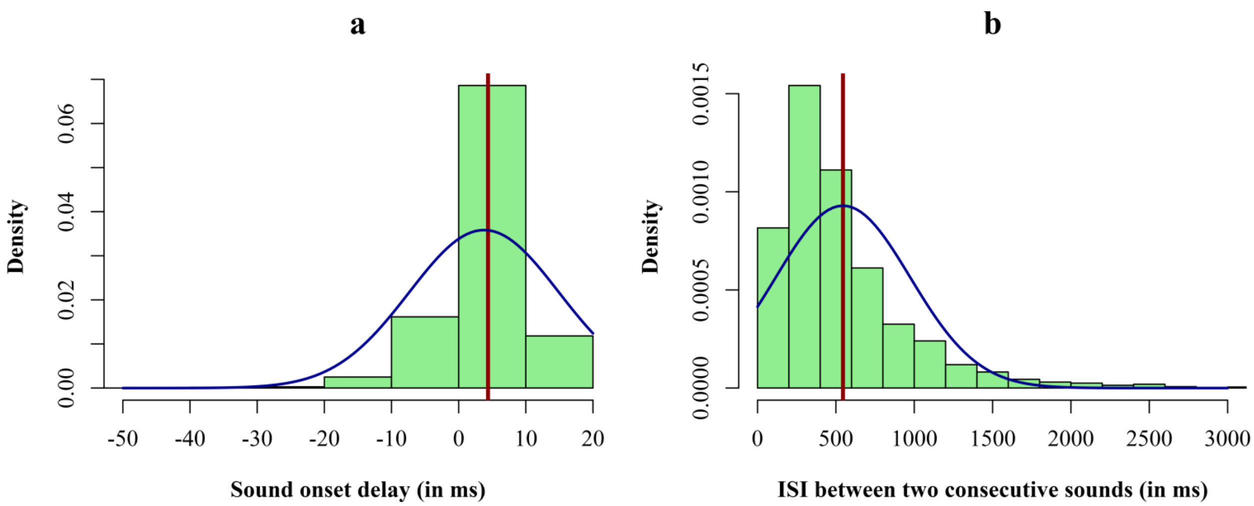

Figure 23. Timing of the sound presentation in Chapter 6. Panel **a** shows the distribution of sound onset delay relative to the fixation onset on the target word. The maximum delay was 14 ms. Panel **b** shows the distribution of the inter-stimulus interval (ISI) between two consecutive sounds. Vertical line shows the mean in both panels.

6.3.1. Global Reading

The descriptive statistics for global reading measures are shown in Table 17 and the LMM results are shown in Table 18. No significant differences were observed between the sound conditions on any of these measures: deviant sounds did not affect performance relative to the standard condition, and performance was comparable in the silence and standard conditions. Therefore, global reading was not disrupted by the deviant sound or the presence of gaze-contingent sounds in general.

Sound	Sentence reading time (in ms)	Fixation duration (in ms)	Number of fixations	Saccade length (in letters)
Silence	3670 (1781)	237 (114)	15.4 (6.25)	9.76 (9.44)
Standard	3580 (1715)	237 (109)	15.1 (5.81)	9.59 (9.04)
Deviant	3560 (1697)	236 (110)	15.1 (5.83)	9.51 (8.64)

Table 17. Means of global reading measures in Chapter 6 (SDs in parenthesis).

Effect	Sentence reading time[1]				Fixation duration			
	b	SE	t	p	b	SE	t	p
Intercept	8.09	0.04	185	**<.001**	5.38	0.02	282.5	**<.001**
Deviant vs Standard	< -0.01	0.01	-0.57	.56	< -0.01	< 0.01	-1.24	.22
Silence vs Standard	0.02	0.03	0.81	.42	< 0.01	< 0.01	0.09	.92

Effect	Number of fixations				Saccade length[1]			
	b	SE	t	p	b	SE	t	p
Intercept	15.08	0.56	26.75	**<.001**	2.03	0.02	95.82	**<.001**
Deviant vs Standard	-0.01	0.13	-0.08	.93	< -0.01	0.01	-0.57	.56
Silence vs Standard	0.35	0.42	0.83	.41	0.01	0.01	0.87	.39

Table 18. Results from LMMs for global reading measures in Chapter 6. Statistically significant *p* values are formatted in bold.

[1] The random intercept for subjects was removed due to a convergence failure.

6.3.2. Target Words

As illustrated in Figure 24, fixation durations were significantly longer immediately following the deviant sound compared to the standard sound (FFD: b= 0.03, SE= 0.02, t= 2.02, p= 0.05, d= 0.20; SFD: b= 0.04, SE= 0.02, t= 2.43, p= 0.02, d= 0.29; GD: b= 0.05, SE= 0.02, t= 2.8, p= 0.007, d= 0.21; TVT: b= 0.04, SE= 0.02, t= 2.16, p= 0.03, d= 0.18). However, fixation durations were comparable in the standard and silent conditions (FFD: b= 0.02, SE= 0.01, t= 1.35, p= 0.17; SFD: b= 0.03, SE= 0.02, t= 1.86, p= 0.07; GD: b= 0.01, SE= 0.02, t= 0.2, p= 0.84; TVT: b= 0.01, SE= 0.02, t= 0.55, p= 0.58). When the position of the target word was added as a fixed effect in the model, there were no significant interactions with the contrast between the standard and the deviant sound (all *p*s ≥ 0.079). This shows that the deviance effect was not modulated by the position of the target word in the sentence. Furthermore, there was a 3% greater probability of making a regressive saccade immediately after hearing the deviant sound compared to hearing the standard sound ($b= 0.25$, $SE= 0.13$, $z= 1.96$, $p= 0.049$, $d= 0.25$; Deviant: $M= 21.3\%$, $SD= 40.9\%$; Standard: $M= 17.9\%$, $SD= 38.3\%$). No such difference was observed between the standard sound and the silence condition ($b= 0.14$, $SE= 0.12$, $z= 1.23$, $p= 0.21$; Standard: $M= 17.9\%$, $SD= 38.3\%$; Silence: $M= 19.2\%$, $SD= 39.4\%$). These results clearly indicate that the deviant sound significantly and selectively affected fixation durations on the target word immediately after it was presented. However, as the deviant sound was played on only one word in the sentence, this effect does not appear in global reading measures due to the averaging over all words in the sentence, the vast majority of which were not affected by the sound.

Post-hoc analyses were carried out to test if the disruption by the deviant sound also affected fixation durations on the next word in the sentence (i.e., the post-target word). The results (presented in Appendix I) indicated that the deviant sound had no effect on fixation durations on the next word in the sentence, which suggests that the disruption occurred before the next word was fixated. Additionally, the magnitude of the deviance effect on the target words was not modulated by whether participants were slow or fast readers (see Appendix I).

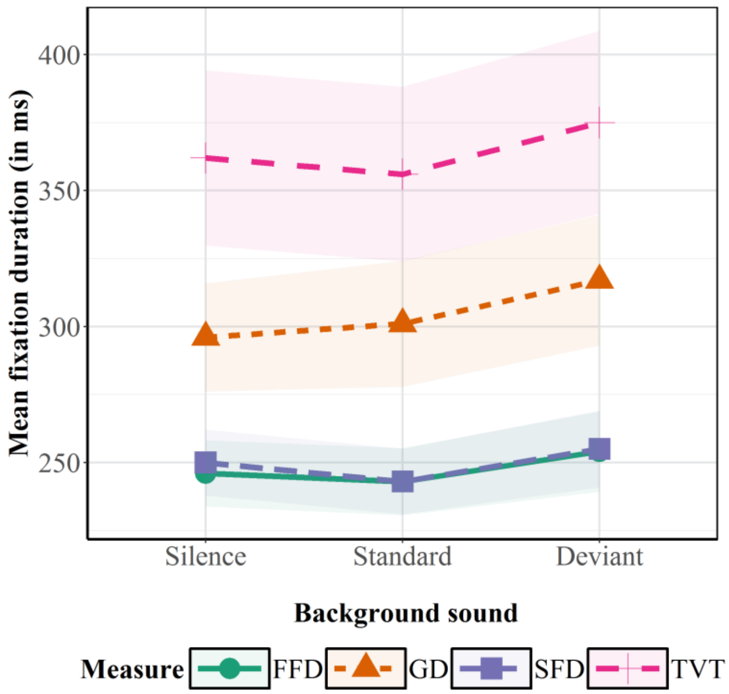

Figure 24. Mean fixation durations on the target words in Chapter 6. FFD: first fixation duration. SFD: single fixation duration. GD: gaze duration. TVT: total viewing time. Shading indicates the standard error.

To establish the point in time when the deviant sound first started to affect fixation durations, we used the Confidence Interval Divergence Point Analysis (CI-DPA) (Reingold & Sheridan, 2014, 2017)**.** This is a survival analysis technique that can determine the earliest point in time when the distributions of two experimental conditions begin to diverge (i.e., significantly differ from one another). The CI-DPA analysis was run with 10 000 bootstrap iterations on the FFD of the target words using the method described in Reingold and Sheridan (2017). The analysis indicated that the deviant sound first started to affect fixation durations at 180 ms (95% CI [167, 198]). This is illustrated in Figure 25.

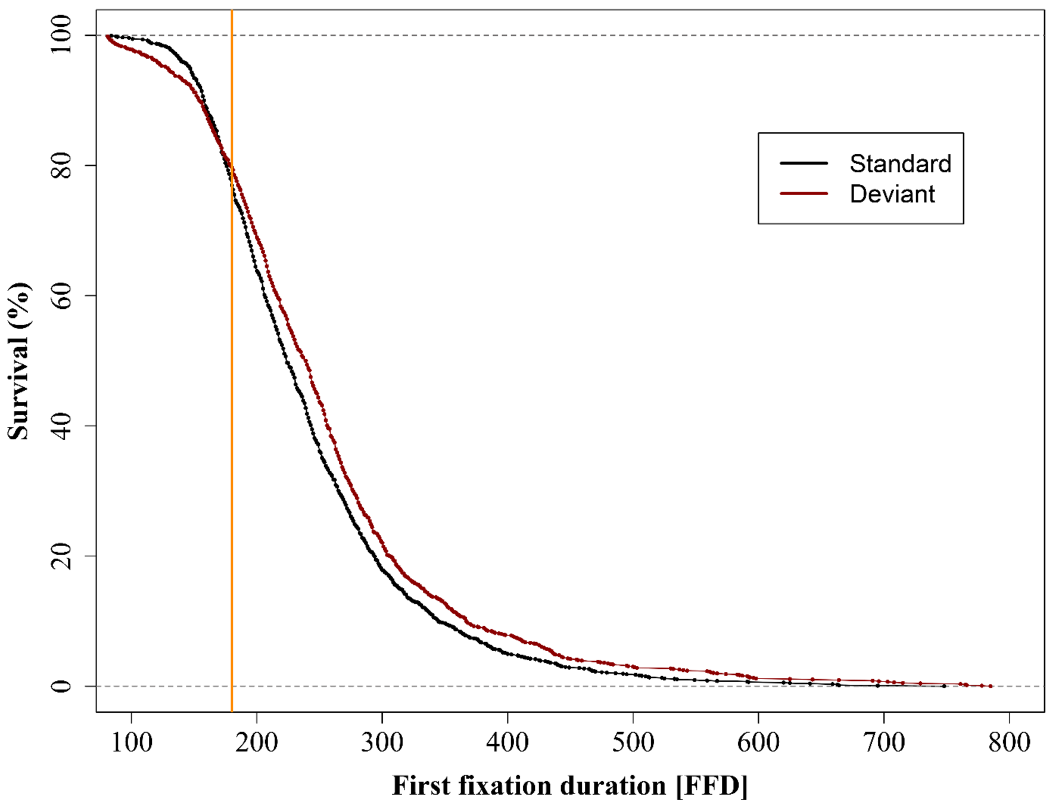

Figure 25. Survival curves of the first fixation duration during which participants heard the sound in Chapter 6. The divergence point (at 180 ms) is shown by the vertical orange line.

Finally, the corpus lexical frequency of target words was entered into a model with the fixation durations in the three sound conditions. If the deviant sound interfered with the lexical processing of the target words, we would expect to see an interaction between lexical frequency and the deviant condition. The results showed no significant interactions between lexical frequency and the deviant sound (see Table 19). This suggests that the deviant sound did not interfere with accessing the lexical representation of words.

Effect	FFD				SFD				GD			
	b	SE	t	p	b	SE	t	p	b	SE	t	p
Intercept	5.4	.02	254.2	**<.001**	5.4	.02	251.45	**<.001**	5.6	.03	191.2	**<.001**
Freq	-.02	.01	-2.55	**.01**	-.03	.01	-3.50	**<.001**	-.05	.01	-4.99	**<.001**
Deviant	.03	.02	2.02	**.05**	.04	.02	2.37	**.02**	.05	.02	2.86	**.006**
Standard	.02	.01	1.31	.19	.03	.01	1.86	.07	<.01	.02	.18	.85
Freq: Deviant	<.01	.01	.36	.71	.02	.01	1.52	.12	.02	.02	1.27	.20
Freq: Standard	-.01	.01	-.83	.40	-.01	.01	-.39	.69	-.01	.02	-.35	.72

Table 19. Interactions between fixation durations on the target words and corpus lexical frequency in Chapter 6. Freq: lexical frequency Statistically significant *p*-values are formatted in bold. Lexical frequency was log-transformed.

6.4. Discussion

The present results showed clear evidence of deviance distraction in eye-movements during reading. Indeed, fixation durations on the target words were longer after hearing the deviant sound compared to hearing the standard one. Interestingly, global reading measures were unaffected by either sound and the mere presence of gaze-contingent sounds did not influence how participants read the sentences. In other words, reading occurred normally in the presence of the standard sounds. However, deviant sounds selectively prolonged fixation durations on the currently-read word at the time of the sound’s presentation. As comprehension accuracy was not affected by the presentation of a deviant sound in the sentence, it is not likely that the observed longer fixation durations on the target words are related to comprehension difficulties. This is in line with previous studies demonstrating that irrelevant speech can disrupt fixation durations during reading without an associated disruption in comprehension (Chapters 3-4; Cauchard et al., 2012; Hyönä & Ekholm, 2016; Yan et al., 2017).

The time course analysis of deviance distraction and the absence of modulation by lexical frequency made it possible to further localise the source of the effect. For example, in the E-Z Reader model of eye-movement control during reading (Reichle, Pollatsek, Fisher, & Rayner, 1998; Reichle, Rayner, & Pollatsek, 2003; Reichle, Warren, & McConnell, 2009), word processing starts with an early visual processing stage during which the visual features of the word are propagated from the retina to the brain. This stage is then followed by two lexical processing stages: familiarity check (L1) and lexical access (L2). In this model, completion of familiarity check initiates the programming of the next saccade because the second stage of lexical processing (L2) is likely to be completed soon (Reichle et al., 1998). The programming of the next saccade also happens in two stages: a labile stage (M1) during which the saccade programme can be cancelled, and a non-labile stage (M2) during which the saccade programme can no longer be cancelled.

Our data suggest that the deviant sound did not interfere with the lexical processing stages for a few reasons. First, the CI-DPA analysis indicated that the earliest discernible effect of the deviant sound occurs somewhat late, at 180 ms. This exceeds the temporal estimates of lexical processing reported in the neurophysiological literature (127-172 ms on average; Reichle & Reingold, 2013). Second, Reingold and Sheridan (2014) used the CI-DPA analysis to estimate that the lexical frequency effect (i.e., the difference in fixation times between high and low frequency words) starts at 138 ms after fixation onset- that is, some 42 ms earlier than the sound deviance effect. Finally, deviance distraction was not modulated by target word lexical frequency. Taken together, these findings suggest that deviant sounds did not increase fixation durations on the target words because of delayed lexical processing. This conclusion is similar to the results from Chapter 3, which also showed that the semantic properties of irrelevant speech did not interfere with accessing the lexical representation of words.

Since the deviance effect appears to occur after the lexical processing stages of reading, we hypothesize that the interference is likely due to saccadic inhibition during the programming of the next saccade. This conclusion is generally consistent with Wessel and Aron's (2013) proposition of a general action suppression upon the presentation of a deviant sound. In their study, action inhibition took the form of a reduced corticospinal excitability of the hand following TMS stimulation of the corresponding motor cortex some 150 ms after the deviant sound’s onset. The time course of this effect is generally similar to the time line of the deviant distraction effect found in the present experiment. Because the programming and execution of saccades involves subcortical structures, such as the superior colliculus, cerebellum, and the brainstem (Munoz, 2002b), our results in fact extend Wessel and Aron’s (2013) by suggesting that deviant sounds may also inhibit subcortical brain areas. Of course, the present findings do not make it possible to pinpoint exactly where in the brain this effect is originating from and this question needs to be addressed by future research.

Further evidence in support of this saccadic inhibition account is the finding that the deviant effect was not modulated by the position of the target word in the sentence. Because inhibition of the oculomotor system should be independent of any underlying word identification or syntactic processes, the effect would be expected to be invariant with respect to where it occurs in the sentence. Therefore, the lack of modulation by target word position is consistent with the notion that this effect is likely oculomotor in nature. Even though deviant sounds had no effect on spatial measures of saccades (e.g., saccade length), this is not inconsistent with the saccadic inhibition explanation because the hypothesized inhibition occurs during the planning stages rather than during the execution of the next saccade. This effect is similar to Reingold and Stampe's (1999, 2002, 2004) finding that transient visual changes, such as replacing the text with a black screen for 33 ms, leads to saccadic inhibition some 60-70 ms after the onset of the display change. In a similar fashion, the deviant sound in the present experiment also likely triggered saccadic inhibition, although this occurred at a later point in time following the onset of the distractor.

Similar evidence for saccadic inhibition was also found in a study by Graupner, Velichkovsky, Pannasch, and Marx (2007) where a sequence of auditory distractors (which also included a deviant) was presented in a picture viewing task. Graupner et al. (2007) reported that the deviant distractor reduced the proportion of terminated fixations, first at around 90 ms and then at around 150 ms. They interpreted this finding as “first” and “second” saccadic inhibition, respectively. The first one was assumed to reflect a fast inhibition process similar to the one observed by Reingold and Stampe (1999, 2000, 2004) in response to transient visual distractors. On the other hand, the second one was assumed to represent a secondary wave of inhibition that potentially originates from a neural network that includes the amygdala. While there was no evidence for first saccadic inhibition in the present experiment, the onset of the deviance distraction effect is generally consistent with the secondary wave of inhibition that the authors reported in their study.

Our results suggest that the sound deviance effect was likely due to saccadic inhibition. This is in line with the notion that deviant sounds capture attention away from ongoing processing (e.g., Escera et al., 1998; Schröger, 1996) and inhibit motor processes (Wessel & Aron, 2013). The present experiment contributes to our understanding of the time course of this effect by suggesting that a covert orienting of attention to the deviant sound does not occur during the lexical processing stages of the fixated word (as indicated by the lack of modulation of the effect by lexical frequency) but at a subsequent stage (namely, the preparation of the next saccade). In oddball tasks, the effect of deviant sounds on electrophysiological measures of the orienting response is traceable from about 150 ms to 600 ms from the sound's onset (Berti & Schröger, 2003; Escera et al., 1998), while their behavioural effect spans further to the production of a response in the current trial. So far, the temporal dynamics of the motor inhibition yielded by deviant sounds remain to be established, but early evidence places it at around 150 ms after the sound's onset (Wessel, 2017; Wessel & Aron, 2013). Our experiment does not allow us to disentangle the potential contributions of the orienting response and the motor inhibition to the effect. Exploring this issue could be an objective for future research. For example, the precise time course of the orienting response in our task could be studied by using EEG and eye-tracking co-registration (e.g., Dimigen, Sommer, Hohlfeld, Jacobs, & Kliegl, 2011; Plöchl, Ossandón, & König, 2012), capitalizing on the well-known ERP signature of the orienting response (i.e., the P3a component; Berti & Schröger, 2001) to study its potential correlation with the observed disruption in eye-movements and to better understand its influence on the reading process.

It should be noted that the size of the deviance distraction effect in the present experiment was relatively small: it ranged from 11 ms for FFD to 19 ms for TVT (*M*= 14.5 ms). However, it is comparable to the numerical size of the distraction effect by deviant and novel sounds on reaction times in previous experiments using the auditory-visual cross-modal task, which ranged from 13 to 24 ms (*M*= 17 ms) (Andrés, Parmentier, & Escera, 2006; Escera et al., 1998; Ljungberg, Parmentier, Leiva, & Vega, 2012; Parmentier, 2016;

Parmentier et al., 2008; Parmentier, Elsley, & Ljungberg, 2010; Parmentier, Vasilev, et al., 2018; Wetzel, Schröger, & Widmann, 2013). Nevertheless, the standardised effect size measured in Cohen's *d* was considerably smaller than that of previous cross-modal oddball studies (an average of d= 0.22 in the present experiment compared to an average of d= 0.65 in oddball experiments). This discrepancy between the numerical and standardized effect size is most likely due to the greater variability in eye-movement responses compared to behavioural reaction time responses. The small effect sizes in the present study are not necessarily surprising: for example, the meta-analysis of auditory distraction by continuous sounds such as speech, noise, and music in Chapter 2 found that the standardised effect sizes ranged between 0.06-0.35. We speculate that the smaller effects may be due to the fact that, unlike categorization tasks, our reading task did not require a specific response from participants upon the presentation of the irrelevant sounds. This in turn may have introduced more variability into the data.

Traditionally, the majority of research on deviance distraction has been conducted using categorisation tasks, such as judging the parity of numbers presented on the screen (e.g., Parmentier et al., 2008) or judging the duration of the irrelevant sound (short vs long; e.g., Schröger & Wolff, 1998b). Perhaps one of the most important ways in which the present experiment differs from such studies is that it did not require any type of response from participants after the presentation of the irrelevant sounds. Rather, participants simply needed to do well overall on the task (i.e., comprehend the sentences) and ignore the sounds. In contrast, participants in categorization tasks typically need to make a binary response after each sound. As a result, the sound may serve as an unspecific warning signal that alerts them to the imminent presentation of a target stimulus that requires a response (Parmentier, 2014).

Although it has been recently shown that deviance distraction in oddball tasks occurs even when the sound does not fulfil this unspecific warning function (Parmentier, 2016), the evidence for this again comes from studies that required a response on at least some of the trials (typically on half of them; e.g., see Ljungberg et al., 2012; Parmentier et al., 2010; Wetzel et al., 2013). Therefore, the present experiment contributes to our theoretical understanding of deviance distraction by showing that deviant sounds disrupt ongoing processes even when the sounds are not relevant in any way to the main task and do not constitute cues for goal-directed behaviour.

While the size of the deviance effect in the present experiment was relatively small, it may be possible to increase its magnitude in future studies. One possible way to do this could be to use novel sounds that are not repeated throughout the experiment instead of a single deviant sound. For example, there is at least some evidence indicating that novel sounds can lead to numerically larger effect sizes in reaction times compared to a single deviant sound (Berti, 2012; Escera et al., 1998; Wetzel, Schröger, & Widmann, 2016; although see Wetzel et al., 2013). The higher propensity of novel sounds to cause distraction may be due to a number of factors, such as their greater spectral complexity, meaning, and novelty (Wetzel et al., 2016). Additionally, participants would be slower to habituate to novel sounds since they are repeated fewer times than deviant sounds.

More broadly, the present results are also relevant for understanding how the external auditory environment may influence readers' eye-movement behaviour. While a lot of progress has been made in understanding how different cognitive and oculomotor processes influence eye-movement control during reading (see Rayner, 1998, 2009), most of this research has been conducted in a quiet and well-controlled environment that is not very typical of everyday life. Interestingly, recent work has suggested that certain sounds, such as intelligible speech, that are present in the readers' environment can result in attentional distraction that is detectible at the level of eye fixations, while other sounds, such as background music or unintelligible speech, generally have no influence on eye-movement behaviour (Cauchard et al., 2012; Hyönä & Ekholm, 2016; R. Johansson et al., 2012; Yan et al., 2017; but see Zhang et al., 2018 for conflicting evidence). These results suggest that certain auditory environments can interfere with readers' ability to maintain sustained attention on the task and process the text in an efficient way. The present experiment builds upon these findings by demonstrating that deviant sounds that violate readers' expectations about what type of sound they will hear next also result in an immediate disruption of eye-movement behaviour. However, unlike continuous distractors such as intelligible speech that typically result in an increase in re-reading behaviour (e.g., Hyönä & Ekholm, 2016; Yan et al., 2017), the present results suggest that this disruption may occur due to motor inhibition that interferes with the programming of the next saccade.

Another interesting finding in the present experiment was that the standard sound did not lead to significantly longer fixation durations compared to silence. This is in contrast to some evidence showing that the presentation of a short auditory distractor similar to the standard sound used in the present study can lead to an increase in fixation durations (e.g., Pannasch, Dornhoefer, Unema, & Velichkovsky, 2001; Pannasch & Velichkovsky, 2009, Experiment 3; but see Reingold & Stampe, 2004, Experiment 1). There are a few possible reasons why the standard sound did not lead to significantly longer fixation durations compared to the silence baseline. First, the standard sound in the present experiment was presented only for 50 ms, which may not be long enough for it to cause any meaningful distraction. Second, the standard sound was presented very frequently and in quick succession throughout the experiment (overall, 360 standard sounds were presented and the inter-stimulus interval between any two sounds in a trial was only 545 ms on average). Therefore, it is likely that participants rapidly habituated to the standard sound, which would help explain why there was no increase in fixation durations compared to reading in silence. In fact, there is evidence showing that participants can habituate to distracting sounds such as irrelevant speech if they are exposed to them prior to the experiment (Banbury & Berry, 1997; Bell, Röer, Dentale, & Buchner, 2012; N. Morris & Jones, 1990). This further suggests that increased exposure to task-irrelevant sounds makes them less distracting.

One limitation of the present study was that it did not experimentally manipulate the psycholinguistic properties of target words, such as their lexical frequency or predictability given the preceding sentence context. Rather, the modulation of the effect by lexical frequency was tested using the corpus frequency of the target words. Therefore, future studies should include a lexical frequency or a predictability manipulation on the target words in order to replicate and extend the present results. If the sound deviance effect is due to inhibition during the programming of the next saccade, as the present findings have suggested, it should fail to interact with these variables and it should lead to an increase in fixation durations that is independent of the cognitive processing of the word. An even stronger test of the saccadic inhibition hypothesis would be to use a task that resembles reading, but that does not require any cognitive processing of the words in the sentence (e.g., z-string reading; Rayner & Fischer, 1996; Vitu, O'Regan, Inhoff, & Topolski, 1995). In this case, a very similar saccadic inhibition effect to that found in present study should be observed.

It should also be noted that the present data do not allow us to determine with certainty the specific component of saccade programming that may be affected by the deviant sound. For example, both the E-Z Reader (Reichle et al., 1998) and SWIFT (Engbert et al., 2005) models of eye-movement control during reading assume that saccade programming occurs in two stages: 1) a labile stage that can be cancelled; and 2) a non-labile stage that no longer can be cancelled. While at present it is not known which of the two stages may be affected, we believe that the present experiment provides the first step in studying this topic that will incentivize further research into this issue. Additionally, if this effect is due to a general motor inhibition, as Wessel and Aron (2013) have argued, it can be speculated that the resulting inhibition should be generally similar regardless of the exact saccade programming stage in which it occurs.

6.4.1. Conclusion

In summary, the present experiment introduced a new method to study deviance distraction during reading by utilizing a gaze-contingent presentation of task-irrelevant sounds (e.g., Inhoff et al., 2002). The results showed that deviant sounds lead to an increase in fixation durations, which is most likely due to saccadic inhibition. This finding contributes to our growing understanding of how task-irrelevant sounds influence eye-movements during reading by showing that unexpected sounds can have an immediate effect on the programming of the next saccade. More broadly, the present study also raises the possibility that unexpected sounds may inhibit ongoing motor processes in everyday tasks similar to reading that rely on relatively automatic motor control. Finally, as the presentation of multiple gaze-contingent sounds did not in itself affect global reading behaviour, the present study also demonstrates that this method of administering discrete sound stimuli can be successfully used to study auditory distraction during reading.

CHAPTER 7: GENERAL DISCUSSION

Skilled adult reading is a fairly automatized process that may look almost effortless to the naïve observer, but it involves a complex coordination of oculomotor and cognitive control that guides the readers' eyes across the page. While a lot has been learned about the different cognitive and oculomotor processes that affect the moment-to-moment decision of when and where to move the eyes next (see Rayner, 1998, 2009), less is known about how external auditory distractors may influence the ongoing reading process. It is only very recently that researchers have attempted to answer the question of how different auditory environments may influence eye movements during reading (Cauchard et al., 2012; Hyönä & Ekholm, 2016; R. Johansson et al., 2012; Yan et al., 2017; Zhang et al., 2018).

One of the main goals of the present Thesis was to gain a better understanding of what properties of background sounds give rise to distraction in eye-movements during reading. This was accomplished by investigating whether distraction by intelligible speech in eye-movements during reading arises from the semantic (Marsh et al., 2008, 2009; Martin et al., 1988) or phonological information (Salamé & Baddeley, 1982, 1989) of the speech sound, or some combination of the two. Additionally, the present Thesis made the novel contribution of testing whether short sounds that violate participants' expectations can also cause distraction. While deviance distraction has been well-established in the electrophysiological (e.g., Escera et al., 1998; Näätänen et al, 2007; Schröger, 1996) and behavioural literature (e.g., P. Dalton & Hughes, 2014; Parmentier, 2014), little is known about how it may affect complex cognitive tasks such as reading. Therefore, to test how deviant sounds may affect eye-movement control during reading, a new method was developed that involved presenting multiple gaze-contingent sounds in a trial.

The other main goal of this Thesis was to investigate which aspects of the reading process are disrupted by irrelevant sounds, and how the observed disruption may be related to ongoing cognitive, oculomotor, and comprehension processes. This was accomplished in a few ways. First, one experiment tested whether intelligible speech may interfere with the lexical processing of words in readers of alphabetical languages (e.g., see Yan et al., 2017). Second, another experiment tested whether intelligible speech may affect the integration of information across sentences and whether it may disrupt different levels of text comprehension. Third, the present research also investigated how the disruption in eye-movements during reading by intelligible speech may be related to participants' comprehension of the text. Finally, the present research also considered which aspects of the reading process may be affected by deviant sounds.

The present Thesis contained four empirical investigations and one meta-analysis of previous findings. In the remainder of this Chapter, the main results from these studies will be first briefly summarised. Then, the results will be considered in their wider context of research on auditory distraction and eye-movement control during reading. Finally, the theoretical and practical implications of the present results will be considered, along with suggestions for future research.

7.1. Summary of Main Findings

7.1.1. Meta-Analysis of Previous Findings (Chapter 2)

The first empirical investigation of this Thesis was a meta-analytical synthesis of previous findings on auditory distraction during reading. This opening investigation was the necessary first step in understanding the nature of auditory distraction in behavioural measures such as comprehension accuracy because the research literature has been undermined by a number of inconsistent findings and by the limited theoretical understanding of what makes background sounds distracting. One of the key findings to emerge from this study was that background speech, noise, and music all have small but reliably detrimental effect on comprehension accuracy. Intelligible speech was found to result in the biggest amount of distraction in comprehension accuracy. Interestingly, lyrical music, which also contains intelligible language in the form of sung lyrics, was found to be just as distracting as intelligible speech. This suggests that the presence of intelligible language in background sounds is the strongest predictor of auditory distraction. Consistent with theories of semantic distraction (Marsh et al., 2008, 2009; Martin et al., 1988), the meta-regression results indicated that intelligible speech was more distracting than unintelligible speech. Background noise was also found to lead to a small decrease in comprehension accuracy and there was partial support for the changing-state hypothesis (Beaman & Jones, 1997; Jones et al., 1992), which predicts that sounds exhibiting greater acoustic variation are more distracting that steady-state sounds such as acoustical noise.

7.1.2. The Effect of Intelligible Speech on Lexical Processing (Chapter 3)

The first eye-tracking study followed-up on the results from the meta-analysis by investigating whether the disruption by intelligible speech in eye-movements during reading is semantic or phonological in nature (or a combination of the two). Additionally, similar to Yan et al.'s (2017) study, it also tested whether intelligible speech affects the lexical access of words in English readers. In this experiment, participants' eye-movements were recorded while they read single sentences, each of which contained a target word with a lexical frequency manipulation. The results supported most strongly the hypothesis that auditory distraction effects by intelligible speech are entirely semantic in nature (Marsh et al., 2008, 2009; Martin et al., 1988). Importantly, intelligible speech did not affect the lexical access of words. However, consistent with previous studies (Cauchard et al., 2012; Hyönä & Ekholm, 2016; Yan et al., 2017), it resulted in an increase in re-reading behaviour, which was characterised by making more regression and more re-reading fixations on previous words. This suggests that intelligible speech likely made it more difficult to integrate the meaning of individual words in order to form the meaning of the whole sentence. Interestingly, however, there was no associated disruption in comprehension accuracy by intelligible speech.

7.1.3. The Effect of Intelligible Speech on Comprehension and Integration Processes (Chapter 4)

Chapter 4 extended the results from the first eye-tracking study by investigating how intelligible speech affects the immediate comprehension of short passages and the integration of information across sentences. More specifically, it focused on testing whether intelligible speech affects comprehension accuracy only when the questions are more difficult to answer and reflect a deeper level of text understanding. In this experiment, participants answered either easy questions that could typically be answered by recognising words or phrases from the text, or more difficult questions that required understanding the meaning of the whole paragraph to answer. The results replicated the increase in second-pass reading in response to intelligible speech. However, there was no increase in the re-reading of previous sentences, which suggested that intelligible speech does not disrupt the integration of meaning across multiple sentences. Nevertheless, the magnitude of the disruption effects in second-pass reading measures was considerably larger than that of Chapter 3, thus suggesting that intelligible speech was more disruptive in a paragraph-reading compared to a single-sentence reading task. Importantly, however, comprehension remained unaffected regardless of whether participants were answering easy or difficult questions. The results were again best explained by the semantic distraction account (Marsh et al., 2008, 2009; Martin et al., 1988) and there was very limited evidence for contribution of phonology in auditory distraction (Salamé & Baddeley, 1982, 1987).

7.1.4. The Role of Re-reading Behaviour in Distraction by Intelligible Speech (Chapter 5)

In Chapter 5, two experiments tested the distraction re-reading hypothesis, which predicted that the increase in second-pass reading in response to intelligible speech is due to participants' attempt to maintain an accurate comprehension of the text under the distracting conditions. Re-reading of previous words was prevented with the RSVP method (K. I. Forster, 1970) in Experiment 1 and with Schotter et al.'s (2014) trailing mask paradigm in Experiment 2. The results from both experiments indicated that participants' comprehension of the text was significantly lower when re-reading of previous words was prevented. This suggests that the increase in re-reading behaviour in response to intelligible speech is, at least in part, related to maintaining an accurate comprehension of the text. Additionally, the eye-movement data from Experiment 2 replicated the distraction effects of intelligible speech on measures of second-pass reading from Chapter 4.

7.1.5. Distraction by Deviant Sounds during Reading (Chapter 6)

The final experiment in this Thesis investigated whether discrete deviant sounds that violate participants' expectations can also lead to distraction. In this experiment, a new gaze-contingent paradigm was developed in which five shorts sounds were presented upon the fixation of five target words in the sentence. On most occasions, the same sounds was presented (standard), while on rare and unpredictable occasions it was replaced by a different sound (deviant). The results indicated that the deviant sound prolonged fixation durations on the target words immediately after the sound's presentation, but it had no influence on global reading measures. Additionally, the results indicated that the deviant sound likely did not interfere with the lexical processing of the fixated word, but that it likely inhibited the programming of the next saccade. This last finding is in line with the recent proposition that deviant sounds evoke a global motor inhibition some 150 ms after the onset of the deviant sound (Wessel & Aron, 2013).

7.2. Distraction by Task-Irrelevant Sounds: Discussion and Theoretical Implications

7.2.1. Intelligible speech.

The main type of auditory distraction studied in this Thesis was that by intelligible speech. The distraction effects by intelligible speech on eye-movements during reading provided strong support for theories of semantic distraction (Marsh et al., 2008, 2009; Martin et al., 1988). As the summary of key comparisons in Table 20 shows, evidence for semantic distraction in the comparison between English and Mandarin was consistently found in measures of second-pass reading (with the exception of inter-sentence regression probability). Additionally, there was also some evidence for disruption in first-pass reading measures and saccade length in a paragraph-reading paradigm (Chapters 4-5). In contrast, there was no support for the strong form of the phonological disruption account (Salamé & Baddeley, 1982, 1987), according to which any speech sound should be equally distracting because it gains access to the phonological store of working memory.

Measure	English vs Silence			English vs Mandarin	
	Chapter 3	Chapter 4	Chapter 5	Chapter 3	Chapter 4
First-pass reading					
FFD	×	✓	✓	×	×
GD	×	✓	✓	×	✓
Number of 1st-pass fixations	×	×	×	×	✓
Second-pass reading					
TVT	✓	✓	✓	✓	✓
Intra-sentence regression	✓	✓	✓	✓	✓
Inter-sentence regression	N/A	×	×	N/A	×
Number of 2nd-pass fixations	✓	✓	✓	✓	✓
All reading					
SRT/ PRT	✓	✓	✓	✓	✓
Saccade length	×	×	✓	×	✓
Saccade landing position	×	×	×	×	×

Table 20. A summary of auditory disruption effects by intelligible speech in the first three eye-tracking experiments. A tick sign (✓) indicates that a significant difference between the two conditions was observed (English vs Silence or English vs Mandarin), whereas a cross sign (×) indicates that there was no significant difference. FFD: first fixation duration. GD: gaze duration. TVT: total viewing time. SRT: sentence reading time. PRT: paragraph reading time. Chapter 3 contained a sentence-reading study, whereas Chapters 4 and 5 contained a paragraph-reading study.

Nevertheless, two effects suggested a possible contribution of phonology in distraction by intelligible speech. In Chapter 3, unintelligible speech (Mandarin) resulted in more second-pass fixations compared to noise, and in Chapter 4 unintelligible speech resulted in more regressions within the currently-read sentence compared to noise. It is worth considering these two findings in more detail in order to assess the possible role of phonology in auditory distraction. First, the effect from Chapter 3 was partially driven by the fact that participants made fewer second-pass fixations in noise compared to silence. This was confirmed by the lack of significant difference between Mandarin and silence (p= 0.72), which suggests that the effect reached significance because the means in the Mandarin and Noise condition were going in the opposite direction in relation to the silence baseline. Additionally, this effect was not replicated in Chapter 4, which further raises questions about its generalizability across different types of reading materials.

Furthermore, even though there was a significant difference in intra-sentence regression probability between Mandarin and Noise in Chapter 4, the lack of increase in number of second-pass fixations suggests that participants did not actually spend more time re-reading words in the sentence (this was also confirmed by a lack of difference in sentence re-reading time between Mandarin and Noise in Chapter 4). In other words, participants in Chapter 4 were more likely to regress back within the current sentence in Mandarin speech compared to Noise, but they did not actually spend more time processing words again. To some extent, this argues against an explanation of disrupted word processing or sentence integration by Mandarin speech because participants would have likely made more re-reading fixations in order to recover from the disruption (as was the case when they listened to English speech). However, the increase in regression probability without an associated increase in re-reading fixations could suggest that the unfamiliar Mandarin speech may have elicited some type of attention orienting response (e.g., Sokolov, 2001). This could be either due to its perceptual novelty or to some unexpected prosodic features that were present in the speech. At present, this remains as a speculation that needs to be tested by future research.

Because the unintelligible Mandarin speech in the present studies contained distinct tones that are not present in English speech (Duanmu, 2006), it is also possible that the two effects above may be due to differences in pitch. The present research cannot exclude this possibility and further work is required to rule out this alternative explanation. Nevertheless, it should be noted that this explanation is at odds with the common finding that native speakers of atonal languages such as English often have difficulties in distinguishing between Mandarin tones (e.g., Kiriloff, 1969; Morett & Chang, 2015; see also Wang, Spence, Jongman, & Sereno, 1999). In summary, the two significant differences between Mandarin and Noise present, at best, only limited evidence for a partial contribution of phonology in distraction by intelligible speech in eye-movements. This conclusion is largely in agreement with the meta-regression results from Chapter 2 and with the findings from Hyönä and Ekholm's (2016) Experiment 1.

While the present findings are consistent at the basic level with the semantic disruption accounts of Martin et al. (1988) and Marsh et al. (2008, 2009), these theories do not make specific predictions about how intelligible speech affects eye-movements during reading. Therefore, the present investigation provides a more detailed account of how the semantic properties of background speech affect the decision of when and where to move the eyes next. One of the key findings was that the semantic properties of background speech did not disrupt the initial lexical identification of words in the sentence. This finding points to the fact that intelligible speech affects only the post-lexical stages of language processing. While there was evidence for a general slowing down of language processing that was shown by the longer first-pass fixation durations (Chapters 4-5), progressive reading behaviour remained relatively unaffected. This was evidenced by the lack of disruption in oculomotor measures, such as saccade landing position. Even though there was some evidence for disruption in saccade length in Chapters 4 and 5, the magnitude of the effects was quite weak (*d*s ranging between 0.02-0.03), which suggests that this disruption may be of limited practical significance (Kirk, 1996). Therefore, the present results point to the fact that participants likely did not experience great difficulty in progressing through the text and reading new words. Instead, the semantic properties of the irrelevant speech likely created a temporary difficulty in constructing the semantic meaning of the sentence. This in turn may have prompted participants to make more regressions in order to resolve the difficulty before they continue reading new words.

The present results also provide insights into how the disruption by intelligible speech could potentially be simulated in computational models of eye-movement control during reading. For example, a recent version of the E-Z Reader model (Reichle et al., 2009) has attempted to simulate effects of higher-level language processing on eye-movements. Reichle et al. (2009) introduced a new post-lexical integration stage that reflects the processing associated with integrating the currently fixated word into higher-level language representations, such as the syntactic structure of the sentence. In this framework, the present results could be modelled by implementing a parameter that checks for interference by intelligible speech, and which then prompts a regression back to the word where the interference occurred. Therefore, the detection of such disruption by intelligible speech would be associated with greater probability of making a regression to previous words within the currently-read sentence.

A similar principle for modelling this type of disruption by intelligible speech could, at least in theory, also work in parallel-attention models such as SWIFT (Engbert et al., 2002, 2005). The reason for this is that the distinction between serial- and parallel-attention models largely relates to differences in the allocation of attention during the first-pass reading of words, and the effects by intelligible speech occurred in the post-lexical, second-pass reading of words. Therefore, the implementation of this type of interference would likely depend more on the model's computational architecture rather than its assumptions about the allocation of attention during first-pass reading. So far, there have been no attempts to implement any mechanisms that would allow SWIFT to explain higher-level language processing effects in the same way that more recent versions of the E-Z Reader model (e.g., Reichle et al., 2009) have done.

Nevertheless, as mentioned in Chapter 5, one potential way to simulate the increase in regressions and re-reading fixations in response to semantic interference could be to add a second saccade targeting mechanism for detecting interference or difficulty in constructing the sentence meaning. For example, words in SWIFT could also accrue "comprehension difficulty" activation after they have been identified lexically. An increase in this activation for a word would then be associated with a greater probability that this word will be selected as the next saccadic target during a regressive saccade. In this way, the model could potentially be able to interrupt the progressive reading of the text until the difficulty induced by the semantic interference can be successfully resolved. Therefore, implementing the key disruption effects by intelligible speech in both models seems plausible, although actual modelling work is required to test whether this is the case.

7.2.2. Deviant sounds.

The second type of distraction studied in the present Thesis was that by unexpected deviant sounds. This type of distraction differs from the one by intelligible speech because deviant sounds are typically short and discrete stimuli that are only infrequently played throughout the experiment. Additionally, there is also one important difference in the underlying cause of distraction in the two cases: while intelligible speech is distracting due to inherent properties of the speech sound itself (i.e., its semantic content; Martin et al., 1988), deviant sounds cause distraction because they violate the predictions of the cognitive system (Bubic et al., 2009; Parmentier, Elsley, et al., 2011). In fact, deviance distraction can be eliminated if the presentation of the deviant sound is made predictable (Horváth et al., 2011; Sussman et al., 2003).

The traditional explanation of deviance distraction has been in terms of an involuntary reorienting of attention away from the task at hand and towards the deviant sound (e.g., Escera et al., 1998; Parmentier, 2014; Schröger, 1996). A more recent addition to this explanation is that deviant sounds may also induce global motor inhibition some 150 ms after their presentation (Dutra et al., 2018; Wessel, 2017; Wessel & Aron, 2013). The present results suggested that deviant sounds likely inhibit the programming of the next saccade. This finding is consistent with both the orienting response and motor inhibition explanations. The global motor inhibition account is relatively recent and has not been developed theoretically in great detail yet. At present, it seems possible that the orienting response and motor inhibition accounts may not be completely independent from each other. For example, motor inhibition may be the first step in deviance distraction that aims to stop ongoing actions, which could then facilitate the subsequent re-orienting of attention towards the deviant sound (Wessel, 2017; Wessel et al., 2016; see also Wessel & Aron, 2017). Alternatively, the orienting and motor inhibition effects may both be part of a complex response to unexpected sound in the auditory environment. At any rate, the present results point towards a possible role of motor inhibition in deviance distraction and would likely be useful in guiding future research into this issue.

Because the deviance distraction effect was consistent with a saccadic inhibition explanation, a potential integration of this effect in existing models of eye-movement control during reading would require it to be added to the saccadic programming stages of these models. However, as SWIFT (Engbert et al., 2005) has adopted the same general assumptions as the E-Z Reader model (Reichle et al., 1998) about the stages in which saccade programming occurs, this could potentially make the cross-model integration of this effect easier. In both models, programming of the next saccade occurs in two stages: 1) a labile stage that can be cancelled by a concurrent saccadic programme, and 2) a non-labile stage that can no longer be cancelled. One possibility for modelling this effect would be to add a saccadic inhibition parameter that adds to the time needed to complete the current saccadic programming stage. This assumption may have some biological plausibility if the inhibition effect by deviant sounds reduces the firing of neurons that are involved in saccadic programming (e.g., see Munoz, 2002 for an overview of the neural circuitry). The present results do not make it possible to pinpoint the exact saccadic programming stage during which such inhibition may take place. However, if this effect is due to a general inhibition of the motor system (Wessel & Aron, 2013), it could be the case that the resulting inhibition is functionally comparable in the two stages.

Given that both intelligible speech and deviant sounds resulted in distraction in eye-movements during reading, it may be interesting to consider whether a single theory could account for both types of auditory distraction that were studied in the present Thesis. One such account that was outlined in Chapter 1 is the duplex theory of auditory distraction (Hughes, 2014; Hughes et al., 2005, 2007). According to this theory, auditory distraction occurs in two distinct forms: 1) interference-by-process (Marsh et al., 2008, 2009); and 2) attentional capture (Hughes et al., 2005; Vachon et al., 2012). As noted previously, interference-by-process occurs when the irrelevant sound interferes with a process that is important for the main task (e.g., semantic processing). On the other hand, the attentional capture account can be considered to be equivalent to the orienting response explanation of deviance distraction discussed in Chapter 6 and also above.

The present research is generally consistent with the duplex theory because there was evidence for interference-by-process distraction in Chapters 3-5 and the deviance distraction effect from Chapter 6 is also in principle consistent with the attention capture explanation. However, even so, one limitation of the duplex theory is that its predictions are just a combination of two already existing theoretical accounts. In other words, the duplex account on its own does not add much to our theoretical understanding of auditory distraction beyond the separate contributions of the interference-by-process and attentional capture accounts[21]. Ideally, further evidence would be required from a reading task that can show a unique contribution of the duplex theory that goes beyond the two existing accounts. Therefore, at present, separate accounts for distraction by intelligible speech and distraction by deviant sounds in eye-movements during reading seems to be just as good of an explanation.

7.3. Practical Implications

The present research also has some practical implications for educational and work settings where reading may be accompanied by irrelevant speech. For example, intelligible speech is a common problem in open-plan offices and other shared work areas because they often have poor acoustic privacy (Haapakangas et al., 2017, 2014; Schlittmeier & Liebl, 2015). As a result, irrelevant speech from nearby workers or phone conversations can have a

[21] I thank Fabrice Parmentier for bringing this to my attention.

negative impact on reading and other office tasks that rely on processing the meaning of written text (e.g., proofreading or copying written information). The present results suggest that intelligible speech in the background would likely result in slower reading of the text due to the need for greater re-reading of previous words. This has direct implications for job performance as workers will generally need more time to complete reading tasks if intelligible speech is present in the background. Additionally, deficits in text comprehension could also potentially occur if workers do not have enough time to engage in effective re-reading of previous text in order to compensate for the experienced distraction. More research in applied settings is required to test directly the magnitude of disruption in reading performance among workers in open-plan offices.

Importantly, listening to music is also a common habit among office workers (Haake, 2006) and students who are studying or doing homework (Calderwood et al., 2014; David et al., 2015). If the background music contains lyrics, it could also potentially have a negative impact on reading efficiency much in the same way that intelligible speech does. Previous research has not directly tested whether there is a difference in the magnitude of auditory distraction in eye-movements during reading between intelligible speech and lyrical music[22]. The results from the meta-analysis of previous findings in Chapter 2 suggested that intelligible speech is just as distracting as lyrical music in measures of reading comprehension. However, this finding may not necessarily extend to eye-movement measures of second-pass reading where robust disruption effects by intelligible speech were observed in Chapters 3-5. Therefore, it remains to be seen whether lyrical music leads to the

---

[22] While Cauchard et al.'s (2012) study had both a background music and background speech manipulation, the music stimuli were entirely instrumental and therefore had no verbal component. Additionally, Zhang et al.'s (2018) study had a background music, but not background speech manipulation.

same increase in re-reading behaviour and whether it has similar practical implications for office workers and students who read in such auditory conditions.

Distraction by deviant or unexpected sounds also has practical implications for real-world situations. However, unlike continuous distractors such as intelligible speech, deviant sounds arguably have a lesser propensity to yield sustained distraction over a long period of time. The reason for this is that the underlying cause of deviance distraction appears to be the violation of regularity and the expectations of the cognitive system, rather than the acoustic deviance of the sound itself (Bubic et al., 2009; Parmentier, Elsley, et al., 2011). Therefore, if a deviant sound is first heard in the environment, it may likely yield distraction as the findings from Chapter 6 and previous research (e.g., Parmentier, 2014) have demonstrated. However, if the deviant sound is then continuously repeated afterwards, it would eventually cease to be distracting as the appearance of the sound would gradually become more predictable (and therefore less distracting).

Nevertheless, even if deviant sounds may not result in severe disruption of the ongoing reading process, the present results still suggest that they would likely lead to a mild but immediate inhibition of saccade programming when the sound is first heard. Additionally, in real-world situations, it is possible that deviant sounds may also elicit an orientating response of the head towards the source of the sound (see Sokolov, 1963, 2001). Obviously, this was not possible in the experiment from Chapter 6 as the sound was presented centrally through the headphones and participants' head was firmly fixed on the headrest in order to prevent artefacts in the eye-movement data. However, if participants are free to move their head and if the spatial location of the sound is not always the same, this may result in an orientation reflex of the head. This last point is just a speculation at the present moment, but it is an interesting question to study in the future.

7.4. Future Research

One interesting overall finding from the present research was that auditory distraction effects in eye-movements during reading were relatively small in magnitude. In fact, many of the significant results would be considered as "small" effects ($d \leq 0.20$) in J. Cohen's (1988) classification, and the remaining ones would be considered as "medium" effects ($d \leq 0.50$). This is consistent with the meta-analysis of previous findings in Chapter 2, which also found a very similar range of effects in behavioural measures of reading performance. The small distraction effects are clearly a testament to how adaptable the reading system is under different environmental conditions and how skilled readers can, for the most part, successfully maintain sustained attention on the reading task while ignoring task-irrelevant sounds. Therefore, even though background sounds such as intelligible speech are subjectively judged to be fairly distracting and annoying (Haapakangas et al., 2011; Haka et al., 2009; Landström et al., 2002), they do not lead to a complete breakdown of the reading process. Rather, they only appear to result in transient episodes of (mild) distraction that readers can overcome and still attain accurate comprehension when reading single sentences or short paragraphs.

Nevertheless, it should be mentioned that the observed effects could potentially be larger in certain participant populations. For example, children may show greater auditory distraction effects due to their poorer control of attention and limited ability to filter out task-irrelevant stimuli (A.-B. Doyle, 1973; Gomes, Molholm, Christodoulou, Ritter, & Cowan, 2000; Plude, Enns, & Brodeur, 1994). While no studies to date have compared distraction in eye-movements during reading between adults and children, there is some evidence from behavioural studies suggesting that children may be distracted more than adults by certain irrelevant sounds. For instance, novel sounds generally lead to a greater impairment of task performance in younger children compared to older children or adults (Gumenyuk et al., 2001; Gumenyuk, Korzyukov, Alho, Escera, & Näätanen, 2004; Wetzel, Scharf, & Widmann, 2018; Wetzel, Schröger, & Widmann, 2016; but see Leiva, Andrés, Servera, Verbruggen, & Parmentier, 2016). Additionally, there is also at least some evidence indicating that children show a larger irrelevant-speech effect in serial-recall and serial-recognition tasks compared to adults (Elliott, 2002; Elliott et al., 2016, Experiment 1; Elliott & Briganti, 2012), although it should be noted that a number of studies that have failed to find evidence for such age-related differences (Klatte, Lachmann, Schlittmeier, & Hellbruck, 2010; Röer, Bell, Körner, & Buchner, 2018; Schwarz et al., 2015). Therefore, future studies should investigate whether the magnitude of auditory distraction in eye-movements during reading is modulated by participants' age.

Interestingly, skilled adult readers may also not be a completely homogenous group in terms of their susceptibility to distraction by irrelevant sounds. In other words, even in a sample of young, college-age adults, there may be considerable variability among participants, with some of them showing very strong distraction effects, while others showing mild effects or no distraction at all. For example, S. Forster and Lavie (2016) have recently argued that an "attention-distractibility trait" may exist in the general population, which confers greater vulnerability to distraction by task-irrelevant stimuli to some people. The authors made this conclusion based on their finding that attention-deficit/hyper-activity disorder (ADHD) symptoms in a non-clinical sample of young adults were significantly correlated with the disruption of task performance by irrelevant visual distractors. This suggests that there may be certain traits that make some individuals more susceptible to distraction than others. Additionally, there is some evidence indicating that not every participant may show an irrelevant-speech effect, at least in a serial recall memory task. More specifically, Ellermeier and Zimmer (1997) found that approximately one-eighth of their participants (12.5%) either showed no irrelevant speech effect at all or the affect was in the opposite direction. Furthermore, the effect sizes of individual participants also varied considerably, sometimes even by over 300 %. It is currently not known how much variability there is in distraction by intelligible speech or deviant sounds in eye-movements during reading, or what factors may influence participants' susceptibility to distraction. However, these are all interesting questions that would be worth investigating in the future.

Future research may also benefit from studying other factors that could potentially modulate distraction by intelligible speech. For example, there is evidence from behavioural studies that increased task engagement (e.g., reading a visually degraded text or reading a text in an unfamiliar font) reduces distraction by intelligible speech, presumably because it is easier for participants to filter out the irrelevant speech sound when the focal task is more demanding (Halin, 2016; Halin, Marsh, Haga, et al., 2014; Halin, Marsh, Hellman, et al., 2014; Sörqvist & Marsh, 2015). Similarly, working memory and participants' ability to suppress the irrelevant speech sound may also modulate distraction by intelligible speech (see Sörqvist, Halin, et al., 2010; Sörqvist, Ljungberg, & Ljung, 2010). However, as these results have been found in behavioural measures such as comprehension accuracy, it is not known yet whether they can also be extended to distraction in eye-movements during reading. Finally, as Chapter 6 has illustrated, gaze-contingent presentation of auditory distractors can also be an effective tool in studying the immediate effect of different distractors on the ongoing reading process. Therefore, this type of methodology also holds promise in gaining a better understanding of auditory distraction in eye-movements during reading.

7.5. Summary and Conclusion

Auditory distraction during reading has been a topic of interest for more than eight decades and it will likely continue to be an active area of research in the future. While a lot has been learned about the type of auditory environments that can give rise to distraction, most of this literature has utilised behavioural indices of task performance, such as comprehension accuracy, which only measure the end product of the reading process. The research presented in this Thesis measured eye-movements during reading in an attempt to find out what makes task-irrelevant sounds distracting and how they influence online reading behaviour. The experiments focused on two types of distraction: distraction by intelligible speech and distraction by deviant sounds.

Intelligible speech was found to disrupt online reading behaviour as a result of semantic interference between the meaning of the speech sound and processing the meaning of the written text. The distraction did not influence the lexical identification of words, but it led to an increase in re-reading behaviour, which pointed towards an integration difficulty in constructing the meaning of the sentence. However, this difficulty was limited only to the currently read sentence and did not affect the integration of meaning across multiple sentences. Additionally, the increase in re-reading behaviour was found to be important for maintaining the immediate comprehension of short passages. Deviant sounds were also found to cause distraction immediately after their presentation, but this was likely due to motor inhibition that slowed down the programming of the next saccade. Because there was nothing special about deviant sounds other than the fact that they violated participants' expectations, this suggests that distraction in eye-movements during reading is not only limited to sounds that can be processed semantically. Research in this area is still in its infancy and there are many important questions that are open for investigation. It is hoped that the present research will serve as a stepping stone in addressing these questions.

REFERENCES

References marked with an asterisk indicate studies included in the meta-analysis

***** Aaronson, D., & Scarborough, H. S. (1976). Performance theories for sentence coding: Some qualitative observations. *Journal of Experimental Psychology: Human Perception and Performance*, *2*(1), 42–55. http://doi.org/10.1037/0096-1523.2.1.42

Abrams, S. G., & Zuber, B. L. (1972). Some temporal characteristics of information processing during reading. *Reading Research Quarterly*, *8*(1), 40–51. Retrieved from http://www.jstor.org/stable/746979

Acklin, D., & Papesh, M. H. (2017). Modern speed-reading apps do not foster reading comprehension. *American Journal of Psychology*, *130*(2), 183–199. http://doi.org/10.5406/amerjpsyc.130.2.0183

Ahuja, J. K. (2016). *The effect of background music on cognitive performance* (Unpiblished dissertation). The University of Arizona, Tucson, AZ, USA.

* Anderson, S. A., & Fuller, G. B. (2010). Effect of music on reading comprehension of junior high school students. *School Psychology Quarterly*, *25*(3), 178–187. http://doi.org/10.1037/a0021213

Andrés, P., Parmentier, F. B. R., & Escera, C. (2006). The effect of age on involuntary capture of attention by irrelevant sounds: A test of the frontal hypothesis of aging. *Neuropsychologia*, *44*(12), 2564–2568. http://doi.org/10.1016/j.neuropsychologia.2006.05.005

Andriessen, J. J., & de Voogd, A. H. (1973). Analysis of eye movement patterns in silent reading. *IPO Annual Progress Report*, *8*, 29–34. Retrieved from http://alexandria.tue.nl/tijdschrift/IPO 8.pdf

Angele, B., Slattery, T. J., & Rayner, K. (2016). Two stages of parafoveal processing during reading: Evidence from a display change detection task. *Psychonomic Bulletin & Review*, *23*(4), 1241–1249. http://doi.org/10.3758/s13423-015-0995-0

* Armstrong, G. B., Boiarsky, G. A., & Mares, M. (1991). Background television and reading performance. *Communication Monographs*, *58*(3), 235–253. http://doi.org/10.1080/03637759109376228

* Armstrong, G. B., & Chung, L. (2000). Background television and reading memory in context: Assessing TV interference and facilitative context effects on encoding versus retrieval processes. *Communication Research*, *27*(3), 327–352. http://doi.org/10.1177/009365000027003003

* Avila, C., Furnham, A., & McClelland, A. (2012). The influence of distracting familiar vocal music on cognitive performance of introverts and extraverts. *Psychology of Music*, *40*(1), 84–93. http://doi.org/10.1177/0305735611422672

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*(4), 390–412. http://doi.org/10.1016/j.jml.2007.12.005

Baddeley, A. D., & Hitch, G. J. (1974). Working memory. *Psychology of Learning and Motivation*, *8*, 47–89. http://doi.org/10.1016/S0079-7421(08)60452-1

Baddeley, A. D., & Hitch, G. J. (1994). Developments in the concept of working memory. *Neuropsychology*, *8*(4), 485–493. http://doi.org/10.1037/0894-4105.8.4.485

Baddeley, A. D., & Salamé, P. (1986). The unattended speech effect: Perception or memory? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *12*(4), 525–529. http://doi.org/10.1037/0278-7393.12.4.525

* Baker, R. W., & Madell, T. O. (1965). A continued investigation of susceptibility to distraction in academically underachieving and achieving male college students. *Journal of Educational Psychology*, *56*(5), 254–258. http://doi.org/10.1037/h0022467

Balota, D. A., Pollatsek, A., & Rayner, K. (1985). The interaction of contextual constraints and parafoveal visual information in reading. *Cognitive Psychology*, *17*(3), 364–390. http://doi.org/10.1016/0010-0285(85)90013-1

Banbury, S., & Berry, D. C. (1997). Habituation and dishabituation to speech and office noise. *Journal of Experimental Psychology: Applied*, *3*(3), 181–195. http://doi.org/10.1037//1076-898X.3.3.181

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*(3), 255–278. http://doi.org/10.1016/j.jml.2012.11.001

Bates, D. M., Machler, M., Bolker, B. M., & Walker, S. C. (2014). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. http://doi.org/10.18637/jss.v067.i01

Beaman, C. P. (2005). Auditory distraction from low-intensity noise: A review of the consequences for learning and workplace environments. *Applied Cognitive Psychology*, *19*(8), 1041–1064. http://doi.org/10.1002/acp.1134

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*(2), 459–471. http://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*(6), 1542–1557. http://doi.org/10.1037/a0028459

Bench, J., Kowal, Å., & Bamford, J. (1979). The BKB (Bamford-Kowal-Bench) sentence lists for partially-hearing children. *British Journal of Audiology*, *13*(3), 108–112.

http://doi.org/10.3109/03005367909078884

Benedetto, S., Carbone, A., Pedrotti, M., Le Fevre, K., Bey, L. A. Y., & Baccino, T. (2015). Rapid serial visual presentation in reading: The case of Spritz. *Computers in Human Behavior*, *45*, 352–358. http://doi.org/10.1016/j.chb.2014.12.043

Berti, S. (2008). Cognitive control after distraction: Event-related brain potentials (ERPs) dissociate between different processes of attentional allocation. *Psychophysiology*, *45*(4), 608–620. http://doi.org/10.1111/j.1469-8986.2008.00660.x

Berti, S. (2012). Automatic processing of rare versus novel auditory stimuli reveal different mechanisms of auditory change detection. *Neuroreport*, *23*(7), 441–6. http://doi.org/10.1097/WNR.0b013e32835308b5

Berti, S., & Schröger, E. (2001). A comparison of auditory and visual distraction effects: Behavioral and event-related indices. *Cognitive Brain Research*, *10*(3), 265–273. http://doi.org/10.1016/S0926-6410(00)00044-6

Berti, S., & Schröger, E. (2003). Working memory controls involuntary attention switching: Evidence from an auditory distraction paradigm. *European Journal of Neuroscience*, *17*(5), 1119–1122. http://doi.org/10.1046/j.1460-9568.2003.02527.x

Berti, S., Vossel, G., & Gamer, M. (2017). The orienting response in healthy aging: Novelty P3 indicates no general decline but reduced efficacy for fast stimulation rates. *Frontiers in Psychology*, *8*, 1–14. http://doi.org/10.3389/fpsyg.2017.01780

Bicknell, K., & Levy, R. (2011). Why readers regress to previous words: A statistical analysis. *Proceedings of the 33rd Annual Meeting of the Cognitive Science Society*, 931–936.

Blanchard, H. E., & Iran-Nejad, A. (1987). Comprehension processes and eye movement patterns in the reading of surprise-ending stories. *Discourse Processes*, *10*(1), 127–138. http://doi.org/10.1080/01638538709544663

Boman, E. (2004). The effects of noise and gender on children's episodic and semantic memory. *Scandinavian Journal of Psychology*, *45*(5), 407–416. http://doi.org/10.1111/j.1467-9450.2004.00422.x

Booth, R. W., & Weger, U. W. (2013). The function of regressions in reading: Backward eye movements allow rereading. *Memory & Cognition*, *41*(1), 82–97. http://doi.org/10.3758/s13421-012-0244-y

Borenstein, M. (2009). Effect sizes for continuous data. In H. M. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (2nd ed., pp. 221–235). New York, USA: Russel Sage Foundation.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*(4), 433–436. http://doi.org/10.1163/156856897X00357

Bronzaft, A. L., & McCarthy, D. P. (1975). The effect of elevated train noise on reading ability. *Environment and Behavior*, *7*(4), 517–528. http://doi.org/10.1177/001391657500700406

Bubic, A., von Cramon, D. Y., Jacobsen, T., Schröger, E., & Schubotz, R. I. (2009).

Violation of expectation: Neural correlates reflect bases of prediction. *Journal of Cognitive Neuroscience*, *21*(1), 155–168. http://doi.org/10.1162/jocn.2009.21013

Calderwood, C., Ackerman, P. L., & Conklin, E. M. (2014). What else do college students "do" while studying? An investigation of multitasking. *Computers and Education*, *75*(2014), 19–29. http://doi.org/10.1016/j.compedu.2014.02.004

Cassel, E. E., & Dallenbach, K. M. (1918). The effect of auditory distraction upon the sensory reaction. *American Journal of Psychology*, *29*(2), 129–143. http://doi.org/10.2307/1413558

* 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*(3), 381–390. http://doi.org/10.1002/acp.1837

Chew, A. S.-Q., Yu, Y.-T., Chua, S.-W., & Gan, S. K.-E. (2016). The effects of familiarity and language of background music on working memory and language tasks in Singapore. *Psychology of Music*, *44*(6), 1431–1438. http://doi.org/10.1177/0305735616636209

Clark, C., & Sörqvist, P. (2012). A 3-year update on the influence of noise on performance and behavior. *Noise & Health*, *14*(61), 292–296.

Clark, H. H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*, *12*(4), 335–359. http://doi.org/10.1016/S0022-5371(73)80014-3

Clarke, M., & Clarke, T. (2000). A study of the references used in Cochrane protocols and reviews: Three bibles, three dictionaries, and nearly 25,000 other things. *International Journal of Technology Assessment in Health Care*, *16*(3), 907–909.

Clifton Jr., C., Staub, A., & Rayner, K. (2007). Eye movements in reading words and sentences. In R. P. G. van Gompel, M. H. Fischer, W. S. Murray, & R. L. Hill (Eds.), *Eye movements: A window on mind and brain* (pp. 341–371). Oxford, UK: Elsevier. http://doi.org/10.1016/B978-008044980-7/50017-3

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Cohen, S., Evans, G. W., Krantz, D. S., & Stokols, D. (1980). Physiological, motivational, and cognitive effects of aircraft noise on children: Moving from the laboratory to the field. *The American Psychologist*, *35*(3), 231–243. http://doi.org/10.1037/0003-066X.35.3.231

Cohen, S., Glass, D. C., & Singer, J. E. (1973). Apartment noise, auditory discrimination, and reading ability in children. *Journal of Experimental Social Psychology*, *9*(5), 407–422. http://doi.org/10.1016/S0022-1031(73)80005-8

Coleman, E. B. (1964). Generalizing to a language population. *Psychological Reports*, *14*, 219–226. http://doi.org/10.2466/pr0.1964.14.1.219

Colle, H. A. (1980). Auditory encoding in visual short-term recall: Effects of noise intensity and spatial location. *Journal of Verbal Learning and Verbal Behavior*, *19*(6), 722–735.

http://doi.org/10.1016/S0022-5371(80)90403-X

Colle, H. A., & Welsh, A. (1976). Acoustic masking in primary memory. *Journal of Verbal Learning and Verbal Behavior*, *15*(1), 17–31. http://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*(1), 204–256. http://doi.org/10.1037/0033-295X.108.1.204

* Cool, V. A., Yarbrough, D. B., Patton, J. E., Runde, R., & Keith, T. Z. (1994). Experimental effects of radio and television distractors on children's performance on mathematics and reading assignments. *The Journal of Experimental Education*, *62*(3), 181–194. http://doi.org/10.1080/00220973.1994.9943839

Cornelissen, F. W., Peters, E. M., & Palmer, J. (2002). The Eyelink Toolbox: Eye tracking with MATLAB and the Psychophysics Toolbox. *Behavior Research Methods, Instruments, & Computers*, *34*(4), 613–617. http://doi.org/10.3758/BF03195489

Czigler, I., Cox, T. J., Gyimesi, K., & Horváth, J. (2007). Event-related potential study to aversive auditory stimuli. *Neuroscience Letters*, *420*(3), 251–256. http://doi.org/10.1016/j.neulet.2007.05.007

Dalton, B. H., & Behm, D. G. (2007). Effects of noise and music on human and task performance: A systematic review. *Occupational Ergonomics*, *7*(3), 143–152. http://doi.org/10.4103/1463-1741.169709

Dalton, P., & Hughes, R. W. (2014). Auditory attentional capture: Implicit and explicit approaches. *Psychological Research*, *78*(3), 313–320. http://doi.org/10.1007/s00426-014-0557-5

* Daoussis, L., & Mc Kelvie, S. J. (1986). Musical preferences and effects of music on a reading comprehension test for extraverts and introverts. *Perceptual and Motor Skills*, *62*(1), 283–289. http://doi.org/10.2466/pms.1986.62.1.283

David, P., Kim, J.-H., Brickman, J. S., Ran, W., & Curtis, C. M. (2015). Mobile phone distraction while studying. *New Media & Society*, *17*(10), 1661–1679. http://doi.org/10.1177/1461444814531692

Davis, C. J. (2005). N-watch: A program for deriving neighborhood size and other psycholinguistic statistics. *Behavior Research Methods*, *37*(1), 65–70. http://doi.org/10.3758/BF03206399

Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. *Frontiers in Psychology*, *5*, 781. http://doi.org/10.3389/fpsyg.2014.00781

Dienes, Z. (2016). How Bayes factors change scientific practice. *Journal of Mathematical Psychology*, *72*, 78–89. http://doi.org/10.1016/j.jmp.2015.10.003

Dimigen, O., Sommer, W., Hohlfeld, A., Jacobs, A. M., & Kliegl, R. (2011). Coregistration of eye movements and EEG in natural reading: Analyses and review. *Journal of Experimental Psychology: General*, *140*(4), 552–572. http://doi.org/10.1037/a0023885

* Dockrell, J. E., & Shield, B. M. (2006). Acoustical barriers in classrooms: The impact of

noise on performance in the classroom. *British Educational Research Journal*, *32*(3), 509–525. http://doi.org/10.1080/01411920600635494

* Dove, M. K. (2009). *The relationship of rhytmic and melodic perception with background music distraction in college level students* (Unpublished doctoral dissertation). University of Kansas, Kansas City.

Doyle, A.-B. (1973). Listening to distraction: A developmental study of selective attention. *Journal of Experimental Child Psychology*, *15*(1), 100–115. http://doi.org/http://dx.doi.org/10.1016/0022-0965(73)90134-3

* Doyle, M., & Furnham, A. (2012). The distracting effects of music on the cognitive test performance of creative and non-creative individuals. *Thinking Skills and Creativity*, *7*(1), 1–7. http://doi.org/10.1016/j.tsc.2011.09.002

Duanmu, S. (2006). Chinese (Mandarin): Phonology. In *Encyclopedia of Language & Linguistics* (pp. 351–355). Elsevier Science.

Duchowski, A. T. (2007). *Eye tracking methodology* (2nd ed.). London: Springer. http://doi.org/10.1007/978-3-319-57883-5

Dutra, I., Waller, D. A., & Wessel, J. R. (2018). Perceptual surprise improves action stopping by non-selectively suppressing motor activity via a neural mechanism for motor inhibition. *The Journal of Neuroscience*, 3091–3117. http://doi.org/10.1523/JNEUROSCI.3091-17.2017

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ*, *315*(7109), 629–634. http://doi.org/10.1136/bmj.315.7109.629

Eiter, B. M., & Inhoff, A. W. (2010). Visual word recognition during reading is followed by subvocal articulation. *Journal of Experimental Psychology: Learning Memory and Cognition*, *36*(2), 457–470. http://doi.org/10.1037/a0018278

Ellermeier, W., & Zimmer, K. (1997). Individual differences in susceptibility to the "irrelevant speech effect." *The Journal of the Acoustical Society of America*, *102*(4), 2191–2199. http://doi.org/10.1121/1.419596

Elliott, E. M. (2002). The irrelevant-speech effect and children: Theoretical implications of developmental change. *Memory & Cognition*, *30*(3), 478–487. http://doi.org/10.3758/BF03194948

Elliott, E. M., & Briganti, A. M. (2012). Investigating the role of attentional resources in the irrelevant speech effect. *Acta Psychologica*, *140*(1), 64–74. http://doi.org/10.1016/j.actpsy.2012.02.009

Elliott, E. M., Hughes, R. W., Briganti, A., Joseph, T. N., Marsh, J. E., & Macken, B. (2016). Distraction in verbal short-term memory: Insights from developmental differences. *Journal of Memory and Language*, *88*, 39–50. http://doi.org/10.1016/j.jml.2015.12.008

Emberson, L. L., Lupyan, G., Goldstein, M. H., & Spivey, M. J. (2010). Overheard cell-phone conversations: When less speech is more distracting. *Psychological Science*, *21*(10), 1383–1388. http://doi.org/10.1177/0956797610382126

Engbert, R., Longtin, A., & Kliegl, R. (2002). A dynamical model of saccade generation in reading based on spatially distributed lexical processing. *Vision Research*, *42*(5), 621–636. http://doi.org/10.1016/S0042-6989(01)00301-7

Engbert, R., Nuthmann, A., Richter, E. M., & Kliegl, R. (2005). SWIFT: A dynamical model of saccade generation during reading. *Psychological Review*, *112*(4), 777–813. http://doi.org/10.1037/0033-295X.112.4.777

Escera, C., Alho, K., Schröger, E., & Winkler, I. (2000). Involuntary attention and distractibility as evaluated with event-related brain potentials. *Audiology and Neurotology*, *5*(3–4), 151–166. http://doi.org/10.1159/000013877

Escera, C., Alho, K., Winkler, I., & Näätänen, R. (1998). Neural mechanisms of involuntary attention to acoustic novelty and change. *Journal of Cognitive Neuroscience*, *10*(5), 590–604. http://doi.org/10.1162/089892998562997

Escera, C., Yago, E., Corral, M. J., Corbera, S., & Nuñez, M. I. (2003). Attention capture by auditory significant stimuli: Semantic analysis follows attention switching. *European Journal of Neuroscience*, *18*(8), 2408–2412. http://doi.org/10.1046/j.1460-9568.2003.02937.x

* Etaugh, C., & Michals, D. (1975). Effects on reading comprehension of preferred music and frequency of studying to music. *Perceptual and Motor Skills*, *41*(2), 553–554. http://doi.org/10.2466/pms.1975.41.2.553

* Etaugh, C., & Ptasnik, P. (1982). Effects of studying to music and post-study relaxation on reading comprehension. *Perceptual and Motor Skills*, *55*(1), 141–142. http://doi.org/10.2466/pms.1982.55.1.141

Evans, G. W., Hygge, S., & Bullinger, M. (1995). Chronic noise and psychological stress. *Psychological Science*, *6*(6), 333–338. http://doi.org/10.1111/j.1467-9280.1995.tb00522.x

Evans, G. W., & Maxwell, L. (1997). Chronic noise exposure and reading deficits. *Environment and Behavior*, *29*(5), 638–656. http://doi.org/10.1177/0013916597295003

Eysenck, H. (1967). *The biological basis of personality*. Springfield, IL: Thomas.

* Falcon, E. (2017). *The relationship between background classical music and reading comprehension on 7th and 8th grade students* (Unpublished doctoral dissertation). St. Thomas University, Florida, USA.

* Fendrick, P. (1937). The influence of music distraction upon reading efficiency. *Journal of Educational Research*, *31*(4), 264–271. http://doi.org/10.1007/s13398-014-0173-7.2

* Fogelson, S. (1973). Music as a distractor on reading-test performance of eighth grade students. *Perceptual and Motor Skills*, *36*, 1265–1266. http://doi.org/10.2466/pms.1973.36.3c.1265

Forster, K. I. (1970). Visual perception of rapidly presented word sequences of varying complexity. *Perception & Psychophysics*, *8*(4), 215–221. http://doi.org/10.3758/BF03210208

Forster, S., & Lavie, N. (2016). Establishing the attention-distractibility trait. *Psychological Science*, *27*(2), 203–212. http://doi.org/10.1177/0956797615617761

Frangos, J., Ritter, W., & Friedman, D. (2005). Brain potentials to sexually suggestive whistles show meaning modulates the mismatch negativity. *Neuroreport*, *16*(12), 1313–1317. http://doi.org/10.1097/01.wnr.0000175619.23807.b7

Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology*, *14*(2), 178–210. http://doi.org/10.1016/0010-0285(82)90008-1

* Freeburne, C. M., & Fleischer, M. S. (1952). The effect of music distraction upon reading rate and comprehension. *Journal of Educational Psychology*, *43*, 101–109. http://doi.org/10.1037/h0054219

Friedman, D., Cycowicz, Y. M., & Dziobek, I. (2003). Cross-form conceptual relations between sounds and words: Effects on the novelty P3. *Cognitive Brain Research*, *18*(1), 58–64. http://doi.org/10.1016/j.cogbrainres.2003.09.002

* Furnham, A., & Allass, K. (1999). The influence of musical distraction of varying complexity on the cognitive performance of extroverts and introverts. *European Journal of Personality*, *13*(1), 27–38. http://doi.org/10.1002/(SICI)1099-0984(199901/02)13:1<27::AID-PER318>3.0.CO;2-R

* Furnham, A., & Bradley, A. (1997). Music while you work: The differential distraction of background music on the cognitive test performance of introverts and extraverts. *Applied Cognitive Psychology*, *11*(5), 445–455. http://doi.org/10.1002/(SICI)1099-0720(199710)11:5<445::AID-ACP472>3.0.CO;2-R

* Furnham, A., Gunter, B., & Peterson, E. (1994). Television distraction and the performance of introverts and extroverts. *Applied Cognitive Psychology*, *8*, 705–711. http://doi.org/10.1002/acp.2350080708

Furnham, A., & Stephenson, R. (2007). Musical distracters, personality type and cognitive performance in school children. *Psychology of Music*, *35*(3), 403–420. http://doi.org/10.1177/0305735607072653

* Furnham, A., & Strbac, L. (2002). Music is as distracting as noise: The differential distraction of background music and noise on the cognitive test performance of introverts and extraverts. *Ergonomics*, *45*(3), 203–217. http://doi.org/10.1080/00140130210121932

* Furnham, A., Trew, S., & Sneade, I. (1999). The distracting effects of vocal and instrumental music on the cognitive test performance of introverts and extraverts. *Personality and Individual Differences*, *27*(2), 381–392. http://doi.org/10.1016/S0191-8869(98)00249-9

Gawron, V. J. (1984). Noise: Effect and aftereffect. *Ergonomics*, *27*(1), 5–18. http://doi.org/10.1080/00140138408963460

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, *7*(4), 457–472. http://doi.org/10.2307/2246093

Gernsbacher, M. A., & Foertsch, J. A. (2000). Three models of discourse comprehension. In S. Garrod & M. J. Pickering (Eds.), *Language processing* (pp. 283–299). East Sussex, UK: Psychology Press.

Gillis, A. (2010). *Effect of background music on reading comprehension and self-report of college students* (Unpublished Master's thesis). The Florida State University.

Gomes, H., Molholm, S., Christodoulou, C., Ritter, W., & Cowan, N. (2000). The development of auditory attention in children. *Frontiers in Bioscience*, *5*, 108–120. http://doi.org/10.1093/toxsci/kfs057

Graupner, S. T., Velichkovsky, B. M., Pannasch, S., & Marx, J. (2007). Surprise, surprise: Two distinct components in the visually evoked distractor effect. *Psychophysiology*, *44*(2), 251–261. http://doi.org/10.1111/j.1469-8986.2007.00504.x

Green, K. B., Pasternack, B. S., & Shore, R. E. (1982). Effects of aircraft noise on reading ability of school–age children. *Archives of Environmental Health: An International Journal*, *37*(1), 141–145. http://doi.org/10.1080/00039896.1982.10667528

Greenhouse, J. B., & Iyengar, S. (2009). Sensitivity analysis and diagnostics. In H. M. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (pp. 418–434). New York, USA: Russell Sage Foundation.

Griffiths, T. L., Steyvers, M., & Tenenbaum, J. B. (2007). Topics in semantic representation. *Psychological Review*, *114*(2), 211–244. http://doi.org/10.1037/0033-295X.114.2.211

Gumenyuk, V., Korzyukov, O., Alho, K., Escera, C., & Näätänen, R. (2004). Effects of auditory distraction on electrophysiological brain activity and performance in children aged 8-13 years. *Psychophysiology*, *41*(1), 30–36. http://doi.org/10.1111/1469-8986.00123

Gumenyuk, V., Korzyukov, O., Alho, K., Escera, C., Schröger, E., Ilmoniemi, R. J., & Näätänen, R. (2001). Brain activity index of distractibility in normal school-age children. *Neuroscience Letters*, *314*(3), 147–150. http://doi.org/10.1016/S0304-3940(01)02308-4

Haake, A. B. (2006). Music listening practices in workplace settings in the UK: An exploratory survey of office-based settings. In *Proceedings of the Ninth International Conference on Music Perception and Cognition*.

Haapakangas, A., Hongisto, V., Eerola, M., & Kuusisto, T. (2017). Distraction distance and perceived disturbance by noise-An analysis of 21 open-plan offices. *The Journal of the Acoustical Society of America*, *141*(1), 127–136. http://doi.org/10.1121/1.4973690

Haapakangas, A., Hongisto, V., Hyönä, J., Kokko, J., & Keränen, J. (2014). Effects of unattended speech on performance and subjective distraction: The role of acoustic design in open-plan offices. *Applied Acoustics*, *86*, 1–16. http://doi.org/10.1016/j.apacoust.2014.04.018

* Haapakangas, A., Kankkunen, E., Hongisto, V., Virjonen, P., Oliva, D., & Keskinen, E. (2011). Effects of five speech masking sounds on performance and acoustic satisfaction. Implications for open-plan offices. *Acta Acustica United with Acustica*, *97*(4), 641–655. http://doi.org/10.3813/AAA.918444

Haines, M. M., Stansfeld, S. A., Brentnall, S., Head, J., Berry, B., Jiggins, M., & Hygge, S. (2001). The West London Schools study: The effects of chronic aircraft noise exposure on child health. *Psychological Medicine*, *31*(8), 1385–1396. http://doi.org/10.1017/S003329170100469X

Haines, M. M., Stansfeld, S. A., Head, J., & Job, R. F. S. (2002). Multilevel modelling of aircraft noise on performance tests in schools around Heathrow Airport London. *Journal of Epidemiology and Community Health*, *56*(2), 139–144. http://doi.org/10.1136/jech.56.2.139

Haines, M. M., Stansfeld, S. a, Job, R. F., Berglund, B., & Head, J. (2001). Chronic aircraft noise exposure, stress responses, mental health and cognitive performance in school children. *Psychological Medicine*, *31*(2), 265–277. http://doi.org/10.1017/S0033291701003282

Haka, M., Haapakangas, A., Keränen, J., Hakala, J., Keskinen, E., & Hongisto, V. (2009). Performance effects and subjective disturbance of speech in acoustically different office types- A laboratory experiment. *Indoor Air*, *19*(6), 454–467. http://doi.org/10.1111/j.1600-0668.2009.00608.x

* Halin, N. (2016). Distracted while reading? Changing to a hard-to-read font shields against the effects of environmental noise and speech on text memory. *Frontiers in Psychology*, *7*, 1–6. http://doi.org/10.3389/fpsyg.2016.01196

* 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*(1), 69–80. http://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*(1), 31–36. http://doi.org/10.1016/j.jarmac.2014.01.003

Hall, J. C. (1952). The effect of background music on the reading comprehension of 278 eighth and ninth grade students. *The Journal of Educational Research*, *45*(6), 451–458. http://doi.org/10.1080/00220671.1952.10881962

Hansen, W. B., & Collins, L. M. (1994). Seven ways to increase power without increasing N. *NIDA Research Monograph*, *142*, 184–95. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9243537

Hedges, L. V, & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.

* Henderson, M. T., Crew, A., & Barlow, J. (1945). A study of the effect of music distraction on reading efficiency. *Journal of Applied Psychology*, *29*(4), 313–317. http://doi.org/10.1037/h0056128

Hilliard, O. M., & Tolin, P. (1979). Effect of familiarity with background music on performance of simple and difficult reading comprehension tasks. *Perceptual and Motor Skills*, *49*(3), 713–714. http://doi.org/10.2466/pms.1979.49.3.713

Hofmeister, J., Heller, D., & Radach, R. (1999). The return sweep in reading. In *Current*

Oculomotor Research (pp. 349–357). Boston, MA: Springer US. http://doi.org/10.1007/978-1-4757-3054-8_49

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, *6*(6), 65–70. http://doi.org/10.2307/4615733

Hongisto, V. (2005). A model predicting the effect of speech of varying intelligibility on work performance. *Indoor Air*, *15*(6), 458–468. http://doi.org/10.1111/j.1600-0668.2005.00391.x

Horváth, J., Roeber, U., Bendixen, A., & Schröger, E. (2008). Specific or general? The nature of attention set changes triggered by distracting auditory events. *Brain Research*, *1229*, 193–203. http://doi.org/10.1016/j.brainres.2008.06.096

Horváth, J., Sussman, E., Winkler, I., & Schröger, E. (2011). Preventing distraction: Assessing stimulus-specific and general effects of the predictive cueing of deviant auditory events. *Biological Psychology*, *87*(1), 35–48. http://doi.org/10.1016/j.biopsycho.2011.01.011

Horváth, J., Winkler, I., & Bendixen, A. (2008). Do N1/MMN, P3a, and RON form a strongly coupled chain reflecting the three stages of auditory distraction? *Biological Psychology*, *79*(2), 139–147. http://doi.org/10.1016/j.biopsycho.2008.04.001

Hughes, R. W. (2014). Auditory distraction: A duplex-mechanism account. *PsyCh Journal*, *3*(1), 30–41. http://doi.org/10.1002/pchj.44

Hughes, R. W., & Jones, D. M. (2001). The intrusiveness of sound: Laboratory findings and their implications for noise abatement. *Noise & Health*, *4*(13), 51–70.

Hughes, R. W., Vachon, F., & Jones, D. M. (2005). Auditory attentional capture during serial recall: Violations at encoding of an algorithm-based neural model? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(4), 736–749. http://doi.org/10.1037/0278-7393.31.4.736

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*(6), 1050–1061. http://doi.org/10.1037/0278-7393.33.6.1050

Hygge, S., Evans, G. W., & Bullinger, M. (2002). A prospective study of some effects of aircraft noise on cognitive performance in schoolchildren. *Psychological Science, 13*(5), 469–474. http://doi.org/10.1111/1467-9280.00483

Hyönä, J. (1995). An eye movement analysis of topic-shift effect during repeated reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(5), 1365–1373. http://doi.org/10.1037/ 0278-7393.21.5.1365

* Hyönä, J., & Ekholm, M. (2016). Background speech effects on sentence processing during reading: An eye movement study. *PloS One*, *11*(3), e0152133. http://doi.org/10.1371/journal.pone.0152133

Hyönä, J., Lorch, R. F. J., & Rinck, M. (2003). Eye movement measures 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. http://doi.org/10.1016/B978-044451020-4/50018-9

Hyönä, J., & Niemi, P. (1990). Eye movements during repeated reading of a text. *Acta Psychologica*, *73*(3), 259–280. http://doi.org/10.1016/0001-6918(90)90026-C

Inhoff, A. W., Connine, C., Eiter, B., Radach, R., & Heller, D. (2004). Phonological representation of words in working memory during sentence reading. *Psychonomic Bulletin & Review*, *11*(2), 320–325. http://doi.org/10.3758/BF03196577

Inhoff, A. W., Connine, C., & Radach, R. (2002). A contingent speech technique in eye movement research on reading. *Behavior Research Methods, Instruments, and Computers*, *34*(4), 471–480. http://doi.org/10.3758/BF03195476

Inhoff, A. W., Greenberg, S. N., Solomon, M., & Wang, C.-A. (2009). Word integration and regression programming during reading: A test of the E-Z Reader 10 model. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(5), 1571–1584. http://doi.org/10.1037/a0014250

Inhoff, A. W., & Rayner, K. (1986). Parafoveal word processing during eye fixations in reading: Effects of word frequency. *Perception & Psychophysics*, *40*(6), 431–439. http://doi.org/10.3758/BF03208203

Inhoff, A. W., & Weger, U. W. (2005). Memory for word location during reading: Eye movements to previously read words are spatially selective but not precise. *Memory and Cognition*, *33*(3), 447–461. http://doi.org/10.3758/BF03193062

Jäger, L. A., Engelmann, F., & Vasishth, S. (2017). Similarity-based interference in sentence comprehension: Literature review and Bayesian meta-analysis. *Journal of Memory and Language*, *94*, 316–339. http://doi.org/10.1016/j.jml.2017.01.004

Jahncke, H., Björkeholm, P., Marsh, J. E., Odelius, J., & Sörqvist, P. (2016). Office noise: Can headphones and masking sound attenuate distraction by background speech? *Work*, *55*(3), 505–513. http://doi.org/10.3233/WOR-162421

Jahncke, H., Hygge, S., Halin, N., Green, A. M., & Dimberg, K. (2011). Open-plan office noise: Cognitive performance and restoration. *Journal of Environmental Psychology*, *31*(4), 373–382. http://doi.org/10.1016/j.jenvp.2011.07.002

Jeffreys, H. (1961). *Theory of probability* (3rd ed.). Oxford, UK: Oxford University Press.

Jegerski, J. (2014). Self-paced reading. In J. Jegerski & B. VanPatten (Eds.), *Research methods in second language psycholinguistics* (pp. 20–49). New York, USA: Routledge.

* Johansson, C. R. (1983). Effects of low intensity, continuous and intermittent noise on mental performance and writing pressure of children with different intelligence and personality characteristics. *Ergonomics*, *26*(3), 275–288. http://doi.org/10.1080/00140138308963341

* 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. http://doi.org/10.1177/0305735610387777

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*(2), 369–381. http://doi.org/10.1037/0278-7393.19.2.369

Jones, D. M., Madden, C., & Miles, C. (1992). Privileged access by irrelevant speech to short-term memory: The role of changing state. *The Quarterly Journal of Experimental Psychology*, *44*(4), 645–669. http://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*(4), 905–22. http://doi.org/10.1037/a0026781

Jones, D. M., Miles, C., & Page, J. (1990). Disruption of proofreading by irrelevant speech: Effects of attention, arousal or memory? *Applied Cognitive Psychology*, *4*(2), 89–108. http://doi.org/10.1002/acp.2350040203

Juhasz, B. J. (2005). Age-of-acquisition effects in word and picture identification. *Psychological Bulletin*, *131*(5), 684–712. http://doi.org/10.1037/0033-2909.131.5.684

Juhasz, B. J., & Rayner, K. (2003). Investigating the effects of a set of intercorrelated variables on eye fixation durations in reading. *Journal of Experimental Psychology: Learning Memory and Cognition*, *29*(6), 1312–1318. http://doi.org/10.1037/0278-7393.29.6.1312

Juhasz, B. J., & Rayner, K. (2006). The role of age of acquisition and word frequency in reading: Evidence from eye fixation durations. *Visual Cognition*, *13*(7–8), 846–863. http://doi.org/10.1080/13506280544000075

Juola, J. F., Ward, N. J., & McNamara, T. (1982). Visual search and reading of rapid serial presentations of letter strings, words, and text. *Journal of Experimental Psychology: General*, *111*(2), 208–227. http://doi.org/10.1037//0096-3445.111.2.208

Just, M. A., Carpenter, P. A., & Woolley, J. D. (1982). Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General*, *111*(2), 228–238. http://doi.org/10.1037/0096-3445.111.2.228

Kaakinen, J. K., & Hyönä, J. (2007). Perspective effects in repeated reading: An eye movement study. *Memory and Cognition*, *35*(6), 1323–1336. http://doi.org/10.3758/BF03193604

Kaakinen, J. K., & Hyönä, J. (2010). Task effects on eye movements during reading. *Journal of Experimental Psychology: Learning Memory and Cognition*, *36*(6), 1561–1566. http://doi.org/10.1037/a0020693

Kallinen, K. (2002). Reading news from a pocket computer in a distracting environment: Effects of the tempo of background music. *Computers in Human Behavior*, *18*(5), 537–551. http://doi.org/10.1016/S0747-5632(02)00005-5

Kämpfe, J., Sedlmeier, P., & Renkewitz, F. (2011). The impact of background music on adult listeners: A meta-analysis. *Psychology of Music*, *39*(4), 424–448. http://doi.org/10.1177/0305735610376261

Kass, R. E., Carlin, B. P., Gelman, A., Neal, R. M., & Carlin, P. B. P. (1998). Markov Chain Monte Carlo in practice: A roundtable discussion. *The American Statistician*, *52*(2), 93–100. http://doi.org/10.1198/000313007X188252

Kehler, A. (2004). Discourse coherence. In L. R. Horn & G. Ward (Eds.), *The handbook of pragmatics* (pp. 241–265). Oxford, UK: Blackwell Publishing.

Kelly, S. N. (1994). A comparison of the effects of background music on the reading comprehension of university undergraduate music majors and nonmusic majors. *Southeastern Journal of Music Education*, *5*, 86–97.

* Kiger, D. M. (1989). Effects of music information load on a reading comprehension task. *Perceptual and Motor Skills*, *69*(2), 531–534. http://doi.org/10.2466/pms.1989.69.2.531

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. USA: Cambridge University Press.

Kintsch, W., & Rawson, K. A. (2005). Comprehension. In M. J. Snowling & C. Hulme (Eds.), *The science of reading: A handbook* (pp. 209–226). United Kingdom: Blackwell Publishing.

Kiriloff, C. (1969). On the auditory perception of tones in Mandarin. *Phonetica*, *20*(2–4), 63–67. http://doi.org/10.1159/000259274

Kirk, R. E. (1996). Practical significance: A concept whose time has come. *Educational and Psychological Measurement*, *56*(5), 746–759. http://doi.org/10.1177/0013164496056005002

Klatte, M., Bergström, K., & Lachmann, T. (2013). Does noise affect learning? A short review on noise effects on cognitive performance in children. *Frontiers in Psychology*, *4*, 578. http://doi.org/10.3389/fpsyg.2013.00578

Klatte, M., Lachmann, T., Schlittmeier, S., & Hellbruck, J. (2010). The irrelevant sound effect in short-term memory: Is there developmental change? *European Journal of Cognitive Psychology*, *22*(8), 1168–1191. http://doi.org/10.1080/09541440903378250

Klatte, M., Lee, N., & Hellbruck, J. (2002). Effects of irrelevant speech and articulatory suppression on serial recall of head and read stimuli. *Psychologische Beiträge*, *44*, 166–186.

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*(1–2), 262–284. http://doi.org/10.1080/09541440340000213

Kliegl, R., Nuthmann, A., & Engbert, R. (2006). Tracking the mind during reading: the influence of past, present, and future words on fixation durations. *Journal of Experimental Psychology. General*, *135*(1), 12–35. http://doi.org/10.1037/0096-3445.135.1.12

Knez, I., & Hygge, S. (2002). Irrelevant speech and indoor lighting: Effects on cognitive performance and self-reported affect. *Applied Cognitive Psychology*, *16*(6), 709–718. http://doi.org/10.1002/acp.829

* Kou, S., McClelland, A., & Furnham, A. (2018). The effect of background music and

noise on the cognitive test performance of Chinese introverts and extraverts. *Psychology of Music*, *46*(1), 125–135. http://doi.org/10.1177/0305735617704300

Kruschke, J. K. (2015). *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan* (2nd ed.). Burlington, MA: Academic Press / Elsevier.

Kruschke, J. K., & Liddell, T. M. (2018). The Bayesian New Statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective. *Psychonomic Bulletin & Review*, *25*(1), 178–206. http://doi.org/10.3758/s13423-016-1221-4

Kuo, Y.-C. (2006). *Cochlear implants in a tone language: Mandarin Chinese (upublished doctoral thesis)*. University College London, United Kingdom.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13). http://doi.org/10.18637/jss.v082.i13

LaBerge, D., & Samuels, S. J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology*, *6*(2), 293–323. http://doi.org/10.1016/0010-0285(74)90015-2

Lajeunesse, M. J. (2013). Recovering missing or partial data from studies: A survey of conversions and imputations for meta-analysis. In J. Koricheva, J. Gurevitch, & K. Mengersen (Eds.), *The handbook of meta-analysis in ecology and evolution* (pp. 195–206). Princeton, USA: Princeton University Press.

Landström, U., Söderberg, L., Kjellberg, A., & Nordström, B. (2002). Annoyance and performance effects of nearby speech. *Acta Acustica United with Acustica*, *88*(4), 549–553.

LeCompte, D. C. (1995). An irrelevant speech effect with repeated and continuous background speech. *Psychonomic Bulletin & Review*, *2*(3), 391–397. http://doi.org/10.3758/BF03210978

Lefebvre, C., Manheimer, E., & Glanville, J. (2008). Searching for studies. In J. P. T. Higgins & S. Green (Eds.), *Cochrane handbook for systematic reviews of interventions: Cochrane book series* (pp. 95–150). Chichester, England: Wiley-Blackwell.

Leiva, A., Andrés, P., & Parmentier, F. B. R. (2015). When aging does not increase distraction: Evidence from pure auditory and visual oddball tasks. *Journal of Experimental Psychology: Human Perception and Performance*, *41*(6), 1612–1621. http://doi.org/10.1037/xhp0000112

Leiva, A., Andrés, P., Servera, M., Verbruggen, F., & Parmentier, F. B. R. (2016). The role of age, working memory, and response inhibition in deviance distraction: A cross-sectional study. *Developmental Psychology*, *52*(9), 1381–1393. http://doi.org/10.1037/dev0000163

Leiva, A., Parmentier, F. B. R., & Andrés, P. (2015a). Aging increases distraction by auditory oddballs in visual, but not auditory tasks. *Psychological Research*, *79*(3), 401–410. http://doi.org/10.1007/s00426-014-0573-5

Leiva, A., Parmentier, F. B. R., & Andrés, P. (2015b). Distraction by deviance: Comparing the effects of auditory and visual deviant stimuli on auditory and visual target processing. *Experimental Psychology*, *62*(1), 54–65. http://doi.org/10.1027/1618-3169/a000273

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: SAGE Publications, Inc.

Liversedge, S. P., Paterson, K. B., & Pickering, M. J. (1998). Eye movements and measures of reading time. In G. Underwood (Ed.), *Eye guidance in reading and scene perception* (pp. 55–75). Oxford: Elsevier. http://doi.org/10.1016/B978-008043361-5/50004-3

* Ljung, R., Sörqvist, P., & Hygge, S. (2009). Effects of road traffic noise and irrelevant speech on children's reading and mathematical performance. *Noise & Health*, *11*(45), 194–198. http://doi.org/10.4103/1463-1741.56212

Ljungberg, J. K., & Parmentier, F. B. R. (2012). Cross-modal distraction by deviance: Functional similarities between the auditory and tactile modalities. *Experimental Psychology*, *59*(6), 355–363. http://doi.org/10.1027/1618-3169/a000164

Ljungberg, J. K., Parmentier, F. B. R., Leiva, A., & Vega, N. (2012). The informational constraints of behavioral distraction by unexpected sounds: The role of event information. *Journal of Experimental Psychology: Learning Memory and Cognition*, *38*(5), 1461–1468. http://doi.org/10.1037/a0028149

Lukas, J. S., & DuPree, R. B. (1980). Effects of freeway noise on academic achievement of elementary school children. *The Journal of the Acoustical Society of America*, *68*(S1), S90–S90. http://doi.org/10.1121/1.2004993

Lynch, S. M. (2007). *Introduction to applied Bayesian statistics and estimation for social scientists*. Berlin: Springer.

MacLeod, A., & Summerfield, Q. (1990). A procedure for measuring auditory and audiovisual speech-reception thresholds for sentences in noise: Rationale, evaluation, and recommendations for use. *British Journal of Audiology*, *24*, 29–43. http://doi.org/10.3109/03005369009077840

* Madsen, C. K. (1987). Background music: Competition for focus of attention. In C. K. Madsen & C. A. Prickett (Eds.), *Applications of research in music behavior* (pp. 315–325). Tuscaloosa, USA: The University of Alabama Press.

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*(3), 682–700. http://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*(1), 23–38. http://doi.org/10.1016/j.cognition.2008.08.003

Marsh, J. E., & Jones, D. M. (2010). Cross-modal distraction by background speech: What role for meaning? *Noise & Health*, *12*(49), 210–216. http://doi.org/10.4103/1463-1741.70499

Marsh, J. E., Ljung, R., Jahncke, H., MacCutcheon, D., Pausch, F., Ball, L. J., & Vachon, F. (2018). Why are background telephone conversations distracting? *Journal of Experimental Psychology: Applied*, *24*(2), 222–235. http://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. http://doi.org/10.3758/s13421-014-0438-6

Marsh, J. E., Sörqvist, P., Hodgetts, H. M., Beaman, C. P., & Jones, D. M. (2015). Distraction control processes in free recall: Benefits and costs to performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(1), 118–133. http://doi.org/10.1037/a0037779

Marsman, M., Schönbrodt, F. D., Morey, R. D., Yao, Y., Gelman, A., & Wagenmakers, E.-J. (2017). A Bayesian bird's eye view of "Replications of important results in social psychology." *Royal Society Open Science*, *4*(1), 160426. http://doi.org/10.1098/rsos.160426

* Martin, R. C., Wogalter, M. S., & Forlano, J. G. (1988). Reading comprehension in the presence of unattended speech and music. *Journal of Memory and Language*, *27*(4), 382–398. http://doi.org/10.1016/0749-596X(88)90063-0

Masson, M. E. J. (1983). Conceptual processing of text during skimming and rapid sequential reading. *Memory & Cognition*, *11*(3), 262–74. http://doi.org/10.3758/BF03196973

MathWorks. (2014). Matlab R2014a [Computer software]. Natick, Massachusetts, USA.

McConkie, G. W., Kerr, P. W., Reddix, M. D., & Zola, D. (1988). Eye movement control during reading: I . The location of initial eye fixations on words. *Vision Research*, *28*(10).

McConkie, G. W., & Rayner, K. (1975). The span of the effective stimulus during a fixation in reading. *Perception & Psychophysics*, *17*(6), 578–586. http://doi.org/10.3758/BF03203972

Meseguer, E., Carreiras, M., & Clifton, C. (2002). Overt reanalysis strategies and eye movements during the reading of mild garden path sentences. *Memory & Cognition*, *30*(4), 551–561. http://doi.org/10.3758/BF03194956

Miles, C., Jones, D. M., & Madden, C. a. (1991). Locus of the irrelevant speech effect in short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*(3), 578–584. http://doi.org/10.1037/0278-7393.17.3.578

* Miller, C. (2014). *The differentiated effects of lyrical and non-lyrical music on reading comprehension (Unpublished Master's thesis)*. Rowan University, New Jersey, USA.

* Miller, L. K., & Schyb, M. (1989). Facilitation and interference by background music. *Journal of Music Therapy*, *26*(1), 42–54. http://doi.org/10.1093/jmt/26.1.42

Miller, L. R. (1947). Some effects of radio-listening on the efficiency of reading-type study activities. *Journal of Educational Psychology*, *38*(2), 105–118. http://doi.org/10.1037/h0062228

* Mitchell, A. H. (1949). The effect of radio programs on silent reading achievement of ninety-one sixth grade students. *The Journal of Educational Research*, *42*(6), 460–470. http://doi.org/10.1080/00220671.1949.10881709

Mitchell, D. C., & Green, D. W. (1978). The effects of context and content on immediate processing in reading. *Quarterly Journal of Experimental Psychology*, *30*(4), 609–636. http://doi.org/10.1080/14640747808400689

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., Antes, G., Atkins, D., … Tugwell, P. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine*, *6*(7), 1–6. http://doi.org/10.1371/journal.pmed.1000097

Morett, L. M., & Chang, L. Y. (2015). Emphasising sound and meaning: pitch gestures enhance Mandarin lexical tone acquisition. *Language, Cognition and Neuroscience*, *30*(3), 347–353. http://doi.org/10.1080/23273798.2014.923105

Morey, R. D., Hoekstra, R., Rouder, J. N., Lee, M. D., & Wagenmakers, E. J. (2016). The fallacy of placing confidence in confidence intervals. *Psychonomic Bulletin and Review*, *23*(1), 103–123. http://doi.org/10.3758/s13423-015-0947-8

Morey, R. D., Rouder, J. N., & Jamil, T. (2015). BayesFactor: Computation of Bayes Factors for Common Designs [R package version 0.9.12-2]. Retrieved from https://cran.r-project.org/package=BayesFactor

Morgan, J. B. (1917). The effect of sound distraction upon memory. *The American Journal of Psychology*, *28*(2), 191–208.

Morris, N., & Jones, D. M. (1990). Habituation to irrelevant speech: Effects on a visual short-term memory task. *Perception & Psychophysics*, *47*(3), 291–297. http://doi.org/10.3758/BF03205003

Morris, S. B., & DeShon, R. P. (2002). Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. *Psychological Methods*, *7*(1), 105–125. http://doi.org/10.1037//1082-989X.7.1.105

Morrison, R. E., & Rayner, K. (1981). Saccade size in reading depends upon character spaces and not visual angle. *Perception & Psychophysics*, *30*(4), 395–396. http://doi.org/10.3758/BF03206156

Morton, J. (1969). Interaction of information in word recognition. *Psychological Review*, *76*(2), 165–178. http://doi.org/10.1037/h0027366

* Mulliken, C., & Henk, W. A. (1985). Using music as a background for reading: An exploratory study. *Journal of Reading*, *28*(4), 353–358.

Munoz, D. P. (2002a). Commentary: Saccadic eye movements: Overview of neural circuitry. In *Progress in Brain Research* (Vol. 140, pp. 89–96). http://doi.org/10.1016/S0079-6123(02)40044-1

Munoz, D. P. (2002b). Saccadic eye movements: Overview of neural circuitry. *Progress in Brain Research*, *140*, 89–96.

Näätänen, R. (1992). *Attention and brain function*. Hillsdale, NJ: Lawrence Erlbaum

Associates.

Näätänen, R. (1995). The mismatch negativity: A powerful tool for cognitive neuroscience. *Ear & Hearing*, *16*(1), 6–18. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/7774770

Näätänen, R., Gaillard, A. W. K., & Mäntysalo, S. (1978). Early selective-attention effect on evoked potential reinterpreted. *Acta Psychologica*, *42*(4), 313–329. http://doi.org/10.1016/0001-6918(78)90006-9

Näätänen, R., & Michie, P. T. (1979). Early selective-attention effects on the evoked potential: A critical review and reinterpretation. *Biological Psychology*, *8*(2), 81–136. http://doi.org/10.1016/0301-0511(79)90053-X

Näätänen, R., Paavilainen, P., Rinne, T., & Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: A review. *Clinical Neurophysiology*, *118*(12), 2544–2590. http://doi.org/10.1016/j.clinph.2007.04.026

Neely, C. B., & LeCompte, D. C. (1999). The importance of semantic similarity to the irrelevant speech effect. *Memory & Cognition*, *27*(1), 37–44. http://doi.org/10.3758/BF03201211

Nicenboim, B., & Vasishth, S. (2016). Statistical methods for linguistic research: Foundational Ideas—Part II. *Linguistics and Language Compass*, *10*(11), 591–613. http://doi.org/10.1111/lnc3.12207

Noble, D. W. A., Lagisz, M., O'dea, R. E., & Nakagawa, S. (2017). Nonindependence and sensitivity analyses in ecological and evolutionary meta-analyses. *Molecular Ecology*, 2410–2425. http://doi.org/10.1111/mec.14031

Öquist, G., & Lundin, K. (2007, December). Eye movement study of reading text on a mobile phone using paging, scrolling, leading, and RSVP. *In Proceedings of the 6th international conference on Mobile and ubiquitous multimedia* (pp. 176-183). ACM.

Öquist, G., Sågvall-Hein, A., Ygge, J., and Goldstein, M. (2004). Eye movement study of reading text on a mobile device using the traditional page and the dynamic RSVP format. *In Proceedings of Mobile HCI 2004* (Glasgow, Scotland). Berlin: Springer. 108-119.

Olkoniemi, H., Johander, E., & Kaakinen, J. K. (2018). The role of look-backs in the processing of written sarcasm. *Memory & Cognition*, 1-19. Advance online publication. http://doi.org/10.3758/s13421-018-0852-2

O'Brien, E. J., & Cook, A. E. (2015). Models of discourse comprehension. In A. Pollatsek & R. Treiman (Eds.), *The Oxford handbook of reading* (pp. 217–231). New York, USA: Oxford University Press.

O'Regan, J. K. (1983). Elementary perceptual and eye movement control processes in reading. In K. Rayner (Ed.), *Eye Movements in Reading* (pp. 121–139). New York, USA: Academic Press. http://doi.org/10.1016/B978-0-12-583680-7.50012-6

Pannasch, S., Dornhoefer, S. M., Unema, P. J. A., & Velichkovsky, B. M. (2001). The omnipresent prolongation of visual fixations: Saccades are inhibited by changes in situation and in subject's activity. *Vision Research*, *41*(25–26), 3345–3351. http://doi.org/10.1016/S0042-6989(01)00207-3

Pannasch, S., & Velichkovsky, B. M. (2009). Distractor effect and saccade amplitudes: Further evidence on different modes of processing in free exploration of visual images. *Visual Cognition*, *17*(6–7), 1109–1131. http://doi.org/10.1080/13506280902764422

Papanikolaou, M., Skenteris, N., & Piperakis, S. M. (2014). Effect of external classroom noise on schoolchildren's reading and mathematics performance: Correlation of noise levels and gender. *International Journal of Adolescent Medicine and Health*, *27*(1), 25–29. http://doi.org/10.1515/ijamh-2014-0006

Parmentier, F. B. R. (2008). Towards a cognitive model of distraction by auditory novelty: The role of involuntary attention capture and semantic processing. *Cognition*, *109*(3), 345–362. http://doi.org/10.1016/j.cognition.2008.09.005

Parmentier, F. B. R. (2014). The cognitive determinants of behavioral distraction by deviant auditory stimuli: A review. *Psychological Research*, *78*(3), 321–338. http://doi.org/10.1007/s00426-013-0534-4

Parmentier, F. B. R. (2016). Deviant sounds yield distraction irrespective of the sounds' informational value. *Journal of Experimental Psychology: Human Perception and Performance*, *42*(6), 837–846. http://doi.org/10.1037/xhp0000195

Parmentier, F. B. R., Elford, G., Escera, C., Andrés, P., & Miguel, I. S. (2008). The cognitive locus of distraction by acoustic novelty in the cross-modal oddball task. *Cognition*, *106*(1), 408–432. http://doi.org/10.1016/j.cognition.2007.03.008

Parmentier, F. B. R., Elsley, J. V., Andrés, P., & Barceló, F. (2011). Why are auditory novels distracting? Contrasting the roles of novelty, violation of expectation and stimulus change. *Cognition*, *119*(3), 374–380. http://doi.org/10.1016/j.cognition.2011.02.001

Parmentier, F. B. R., Elsley, J. V., & Ljungberg, J. K. (2010). Behavioral distraction by auditory novelty is not only about novelty: The role of the distracter's informational value. *Cognition*, *115*(3), 504–511. http://doi.org/10.1016/j.cognition.2010.03.002

Parmentier, F. B. R., & Kefauver, M. (2015). The semantic aftermath of distraction by deviant sounds: Crosstalk interference is mediated by the predictability of semantic congruency. *Brain Research*, *1626*, 247–257. http://doi.org/10.1016/j.brainres.2015.01.034

Parmentier, F. B. R., Ljungberg, J. K., Elsley, J. V., & Lindkvist, M. (2011). A behavioral study of distraction by vibrotactile novelty. *Journal of Experimental Psychology: Human Perception and Performance*, *37*(4), 1134–1139. http://doi.org/10.1037/a0021931

Parmentier, F. B. R., Pacheco-Unguetti, A. P., & Valero, S. (2018). Food words distract the hungry: Evidence of involuntary semantic processing of task-irrelevant but biologically-relevant unexpected auditory words. *Plos One*, *13*(1), e0190644. http://doi.org/10.1371/journal.pone.0190644

Parmentier, F. B. R., Turner, J., & Elsley, J. V. (2011). Distraction by auditory novelty: The course and aftermath of novelty and semantic effects. *Experimental Psychology*, *58*(2), 92–101. https://doi.org/10.1027/1618-3169/a000072.

Parmentier, F. B. R., Turner, J., & Perez, L. (2014). A dual contribution to the involuntary semantic processing of unexpected spoken words. *Journal of Experimental Psychology: General*, *143*(1), 38–45. http://doi.org/10.1037/a0031550

Parmentier, F. B. R., Vasilev, M. R., & Andrés, P. (2018). Surprise as an explanation to auditory novelty distraction and post-error slowing. *Journal of Experimental Psychology: General*. http://doi.org/10.1037/xge0000497

Peirce, J. W. (2007). PsychoPy-Psychophysics software in Python. *Journal of Neuroscience Methods*, *162*(1–2), 8–13. http://doi.org/10.1016/j.jneumeth.2006.11.017

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*(4), 437–442. http://doi.org/10.1163/156856897X00366

* Perham, N., & Currie, H. (2014). Does listening to preferred music improve reading comprehension performance? *Applied Cognitive Psychology*, *28*(2), 279–284. http://doi.org/10.1002/acp.2994

Plöchl, M., Ossandón, J. P., & König, P. (2012). Combining EEG and eye tracking: identification, characterization, and correction of eye movement artifacts in electroencephalographic data. *Frontiers in Human Neuroscience*, *6*(October), 1–23. http://doi.org/10.3389/fnhum.2012.00278

Plude, D. J., Enns, J. T., & Brodeur, D. (1994). The development of selective attention: A life-span overview. *Acta Psychologica*, *86*(2–3), 227–272. http://doi.org/10.1016/0001-6918(94)90004-3

Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In *Proceedings of the 3rd international workshop on distributed statistical computing* (p. 125). Technische Universität at Wien.

Pollatsek, A., Reichle, E. D., & Rayner, K. (2006). Tests of the E-Z Reader model: Exploring the interface between cognition and eye-movement control. *Cognitive Psychology*, *52*(1), 1–56. http://doi.org/10.1016/j.cogpsych.2005.06.001

* Pool, M. M., Van der Voort, T. H. A., Beentjes, J. W. J., & Koolstra, C. M. (2000). Background television as an inhibitor of performance on easy and difficult homework assignments. *Communication Research*, *27*(3), 293–326. http://doi.org/10.1177/009365000027003002

Posner, M. I. (1980). Orienting of attention. *The Quarterly Journal of Experimental Psychology*, *32*(July 2015), 3–25. http://doi.org/10.1080/00335558008248231

Posner, M. I. (2016). Orienting of attention: Then and now. *Quarterly Journal of Experimental Psychology*, *69*(10), 1864–1875. http://doi.org/10.1080/17470218.2014.937446

R Core Team (2016). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.r-project.org/

Radach, R., Huestegge, L., & Reilly, R. (2008). The role of global top-down factors in local eye-movement control in reading. *Psychological Research*, *72*(6), 675–688. http://doi.org/10.1007/s00426-008-0173-3

Raney, G. E., & Rayner, K. (1995). Word frequency effects and eye movements during two readings of a text. *Canadian Journal of Experimental Psychology*, *49*(2), 151–172. http://doi.org/10.1037/1196-1961.49.2.151

Ravaja, N., & Kallinen, K. (2004). Emotional effects of startling background music during reading news reports: The moderating influence of dispositional BIS and BAS sensitivities. *Scandinavian Journal of Psychology*, *45*(3), 231–238. http://doi.org/10.1111/j.1467-9450.2004.00399.x

Rayner, K. (1975). The perceptual span and peripheral cues in reading. *Cognitive Psychology*, *81*(7), 65–81. http://doi.org/10.1016/0010-0285(75)90005-5

Rayner, K. (1978). Eye movements in reading and information processing. *Psychological Bulletin*, *85*(3), 618–660. http://doi.org/10.1037/0033-2909.85.3.618

Rayner, K. (1979). Eye guidance in reading: Fixaiton location in words. *Perception*, *8*, 21–30. http://doi.org/10.1068/p080021

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*(3), 372–422. http://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*(8), 1457–1506. http://doi.org/10.1080/17470210902816461

Rayner, K. (2014). The gaze-contingent moving window in reading: Development and review. *Visual Cognition*, *22*(3), 242–258. http://doi.org/10.1080/13506285.2013.879084

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*(3), 241–255. http://doi.org/10.1207/s1532799xssr1003

Rayner, K., & Duffy, S. A. (1986). Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory & Cognition*, *14*(3), 191–201. http://doi.org/10.3758/BF03197692

Rayner, K., & Fischer, M. H. (1996). Mindless reading revisited: Eye movements during reading and scanning are different. *Perception and Psychophysics*, *58*(5), 734–747. http://doi.org/10.3758/BF03213106

Rayner, K., Li, X., & Juhasz, B. J. (2005). The effect of word predictability on the eye movements of Chinese readers. *Psychonomic Bulletin & Review*, *12*(6), 1089–1093. http://doi.org/10.3758

Rayner, K., & McConkie, G. W. (1976). What guides a reader's eye movements? *Vision Research*, *16*(1958), 829–837. http://doi.org/10.1016/0042-6989(76)90143-7

Rayner, K., Pollatsek, A., Ashby, J., & Clifton, C. J. (2012). *Psychology of reading* (2nd ed.). New York, USA: Psychology Press.

Rayner, K., Raney, G. E., & Pollatsek, A. (1995). Eye movements and discourse processing. In R. F. J. Lorch & E. J. O'Brien (Eds.), *Sources of coherence in reading* (pp. 9–35). Hillsdale, NJ: Lawrence Erlbaum Associates.

Rayner, K., & Reingold, E. M. (2015). Evidence for direct cognitive control of fixation durations during reading. *Current Opinion in Behavioral Sciences*, *1*(1), 107–112. http://doi.org/10.1016/j.cobeha.2014.10.008

Rayner, K., Schotter, E. R., Masson, M. E. J., Potter, M. C., & Treiman, R. (2016). So much to read, so little time: How do we read, and can speed reading help? *Psychological Science in the Public Interest*, *17*(1), 4–34. http://doi.org/10.1177/1529100615623267

Rayner, K., Sereno, S. C., Morris, R. K., Schmauder, A. R., & Clifton, C. (1989). Eye movements and on-line language comprehension processes. *Language and Cognitive Processes*, *4*(3–4), SI21-SI49. http://doi.org/10.1080/01690968908406362

Rayner, K., Slattery, T. J., Drieghe, D., & Liversedge, S. P. (2011). Eye movements and word skipping during reading: Effects of word length and predictability. *Journal of Experimental Psychology: Human Perception and Performance*, *37*(2), 514–528. http://doi.org/10.1037/a0020990

Rayner, K., & Well, A. D. (1996). Effects of contextual constraint on eye movements in reading: A further examination. *Psychonomic Bulletin & Review*, *3*(4), 504–509. http://doi.org/10.3758/BF03214555

Reichle, E. D., Pollatsek, A., Fisher, D. L., & Rayner, K. (1998). Toward a model of eye movement control in reading. *Psychological Review*, *105*(1), 125–57. http://doi.org/10.1037/0033-295X.105.1.125

Reichle, E. D., Rayner, K., & Pollatsek, A. (2003). The E-Z reader model of eye-movement control in reading: Comparisons to other models. *The Behavioral and Brain Sciences*, *26*(4), 445–76. http://doi.org/10.1017/S0140525X03430100

Reichle, E. D., & Reingold, E. M. (2013). Neurophysiological constraints on the eye-mind link. *Frontiers in Human Neuroscience*, *7*(July), 361. http://doi.org/10.3389/fnhum.2013.00361

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), 1–21. http://doi.org/10.3758/PBR.16.1.1

Reingold, E. M., & Sheridan, H. (2014). Estimating the divergence point: A novel distributional analysis procedure for determining the onset of the influence of experimental variables. *Frontiers in Psychology*, *5*(DEC), 1–10. http://doi.org/10.3389/fpsyg.2014.01432

Reingold, E. M., & Sheridan, H. (2017). On using distributional analysis techniques for determining the onset of the influence of experimental variables. *The Quarterly Journal of Experimental Psychology*, *5*(DEC), 1–32. http://doi.org/10.1080/17470218.2017.1310262

Reingold, E. M., & Stampe, D. M. (1999). Saccadic inhibition in complex visual tasks. In W. Becker, H. Deubel, & T. Mergner (Eds.), *Current oculomotor research: Physiological and psychological aspects* (pp. 249–255). Boston, MA: Springer US. http://doi.org/10.1007/978-1-4757-3054-8

Reingold, E. M., & Stampe, D. M. (2000). Saccadic inhibition and gaze contingent research paradigms. In A. Kenedy, R. Radach, D. Heller, & J. Pynte (Eds.), *Reading as a perceptual process* (pp. 119–145). Amsterdam: Elsevier. http://doi.org/10.1016/B978-008043642-5/50008-5

Reingold, E. M., & Stampe, D. M. (2002). Saccadic inhibition in voluntary and reflexive saccades. *Journal of Cognitive Neuroscience*, *14*(3), 371–388. http://doi.org/10.1162/089892902317361903

Reingold, E. M., & Stampe, D. M. (2004). Saccadic inhibition in reading. *Journal of Experimental Psychology: Human Perception and Performance*, *30*(1), 194–211. http://doi.org/10.1037/0096-1523.30.1.194

Ricciardi, O., & Di Nocera, F. (2017). Not so fast: A reply to Benedetto et al. (2015). *Computers in Human Behavior*, *69*, 381–385. http://doi.org/10.1016/j.chb.2016.12.047

Risse, S., Hohenstein, S., Kliegl, R., & Engbert, R. (2014). A theoretical analysis of the perceptual span based on SWIFT simulations of the n + 2 boundary paradigm. *Visual Cognition*, *22*(3–4), 283–308. http://doi.org/10.1080/13506285.2014.881444

Risse, S., & Kliegl, R. (2014). Dissociating preview validity and preview difficulty in parafoveal processing of word n +1 during reading. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(2), 653–668. http://doi.org/10.1037/a0034997

Röer, J. P., Bell, R., Körner, U., & Buchner, A. (2018). Equivalent auditory distraction in children and adults. *Journal of Experimental Child Psychology*, *172*, 41–58. http://doi.org/10.1016/j.jecp.2018.02.005

Rohatgi, A. (2015). digitizeR: Tool to extract numerical data from images of plots, maps, etc. Retrieved from http://github.com/ankitrohatgi/digitizeR

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, *86*(3), 638–641. http://doi.org/10.1037/0033-2909.86.3.638

Rouder, J. N., & Morey, R. D. (2011). A Bayes factor meta-analysis of Bem's ESP claim. *Psychonomic Bulletin and Review*, *18*(4), 682–689. http://doi.org/10.3758/s13423-011-0088-7

Rouder, J. N., & Morey, R. D. (2012). Default bayes factors for model selection in regression. *Multivariate Behavioral Research*, *47*(6), 877–903. http://doi.org/10.1080/00273171.2012.734737

Rouder, J. N., Morey, R. D., & Province, J. M. (2013). A Bayes Factor meta-analysis of recent extrasensory perception experiments: Comment on Storm, Tressoldi, and Di Risio (2010). *Psychological Bulletin*, *139*(1), 241–247. http://doi.org/10.1037/a0029008

Roye, A., Jacobsen, T., & Schröger, E. (2007). Personal significance is encoded automatically by the human brain: An event-related potential study with ringtones. *European Journal of Neuroscience*, *26*(3), 784–790. http://doi.org/10.1111/j.1460-9568.2007.05685.x

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*(2), 150–164. http://doi.org/10.1016/S0022-5371(82)90521-7

Salamé, P., & Baddeley, A. D. (1987). Noise, unattended speech and short-term memory. *Ergonomics*, *30*(8), 1185–1194. http://doi.org/10.1080/00140138708966007

Salamé, P., & Baddeley, A. D. (1989). Effects of background music on phonological short-term memory. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*, *41*(1–A), 107–122. http://doi.org/10.1080/14640748908402355

Schad, D. J., & Engbert, R. (2012). The zoom lens of attention: Simulating shuffled versus normal text reading using the SWIFT model. *Visual Cognition*, *20*(4–5), 391–421. http://doi.org/10.1080/13506285.2012.670143

Schilling, H. E. H., Rayner, K., & Chumbley, J. I. (1998). Comparing naming, lexical decision, and eye fixation times: Word frequency effects and individual differences. *Memory & Cognition*, *26*(6), 1270–1281. http://doi.org/10.3758/BF03201199

Schlittmeier, S. J., & Liebl, A. (2015). The effects of intelligible irrelevant background speech in offices- Cognitive disturbance, annoyance, and solutions. *Facilities*, *33*(1/2), 61–75. http://doi.org/10.1108/F-05-2013-0036

Schmid, C. H., & Mengersen, K. (2013). Bayesian meta-analysis. In J. Koricheva, J. Gurevitch, & K. Mengersen (Eds.), *The handbook of meta-analysis in ecology and evolution* (pp. 145–173). Princeton, USA: Princeton University Press.

Schotter, E. R., Angele, B., & Rayner, K. (2012). Parafoveal processing in reading. *Attention, Perception & Psychophysics*, *74*(1), 5–35. http://doi.org/10.3758/s13414-011-0219-2

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), 1–27. http://doi.org/10.1016/j.cognition.2013.11.018

Schotter, E. R., Lee, M., Reiderman, M., & Rayner, K. (2015). The effect of contextual constraint on parafoveal processing in reading. *Journal of Memory and Language*, *83*, 118–139. http://doi.org/10.1016/j.jml.2015.04.005

Schotter, E. R., Reichle, E. D., & Rayner, K. (2014). Rethinking parafoveal processing in reading: Serial-attention models can explain semantic preview benefit and N +2 preview effects. *Visual Cognition*, *22*(3–4), 309–333. http://doi.org/10.1080/13506285.2013.873508

Schotter, E. R., Tran, R., & Rayner, K. (2014). Don't believe what you read (only once): Comprehension is supported by regressions during reading. *Psychological Science*, *25*(6), 1218–1226. http://doi.org/10.1177/0956797614531148

Schröger, E. (1996). A neural mechanism for involuntary attention shifts to changes in auditory stimulation. *Journal of Cognitive Neuroscience*, *8*(6), 527–539. http://doi.org/10.1162/jocn.1996.8.6.527

Schröger, E., Giard, M. H., & Wolff, C. (2000). Auditory distraction: Event-related potential and behavioral indices. *Clinical Neurophysiology*, *111*(8), 1450–1460. http://doi.org/10.1016/S1388-2457(00)00337-0

Schröger, E., & Wolff, C. (1998a). Attentional orienting and reorienting is indicated by human event- related brain potentials. *NeuroReport*, *9*(15), 3355–3358. http://doi.org/10.1097/00001756-199810260-00003

Schröger, E., & Wolff, C. (1998b). Behavioral and electrophysiological effects of task-irrelevant sound change: A new distraction paradigm. *Cognitive Brain Research*, *7*(1), 71–87. http://doi.org/10.1016/S0926-6410(98)00013-5

Schwarz, H., Schlittmeier, S., Otto, A., Persike, M., Klatte, M., Imhof, M., & Meinhardt-Injac, B. (2015). Age differences in the irrelevant sound effect: A serial recognition paradigm. *Psihologija*, *48*(1), 35–43. http://doi.org/10.2298/PSI1501035S

Schwarzer, G. (2007). Meta: An R package for meta-analysis. *R News*, *7*(3), 40–45.

Shield, B. M., & Dockrell, J. E. (2003). The effects of noise on children at school: A review. *Building Acoustics*, *10*(2), 97–116. http://doi.org/10.1260/135101003768965960

Shield, B. M., & Dockrell, J. E. (2008). The effects of environmental and classroom noise on the academic attainments of primary school children. *The Journal of the Acoustical Society of America*, *123*(1), 133–144. http://doi.org/10.1121/1.2812596

Shtyrov, Y., Hauk, O., & Pulvermuller, F. (2004). Distributed neuronal networks for encoding category-specific semantic information: the mismatch negativity to action words. *European Journal of Neuroscience*, *19*(4), 1083–1092. http://doi.org/10.1111/j.1460-9568.2004.03126.x

Shtyrov, Y., & Pulvermuller, F. (2003). Distributed neuronal networks for semantic information: Mismatch negativity to action words. *Psychophysiology*, *19*, S79–S79. http://doi.org/10.1111/j.1460-9568.2004.03126.x

Slater, B. R. (1968). Effects of noise on pupil performance. *Journal of Educational Psychology*, *59*(4), 239–43. http://doi.org/10.1037/h0026025

Slattery, T. J., Angele, B., & Rayner, K. (2011). Eye movements and display change detection during reading. *Journal of Experimental Psychology: Human Perception and Performance*, *37*(6), 1924–1938. http://doi.org/10.1037/a0024322

* Smith-Jackson, T. L., & Klein, K. W. (2009). Open-plan offices: Task performance and mental workload. *Journal of Environmental Psychology*, *29*(2), 279–289. http://doi.org/10.1016/j.jenvp.2008.09.002

Sokolov, E. N. (1963). Higher nervous functions: The orienting reflex. *Annual Review of Physiology*, *25*(1), 545–580. http://doi.org/10.1146/annurev.ph.25.030163.002553

Sokolov, E. N. (2001). Orienting response. In *International Encyclopedia of the Social & Behavioral Sciences* (pp. 10978–10981). Elsevier Science Ltd.

http://doi.org/10.1016/B0-08-043076-7/03536-1

* Sörqvist, P. (2010a). Effects of aircraft noise and speech on prose memory: What role for working memory capacity? *Journal of Environmental Psychology, 30*(1), 112–118. http://doi.org/10.1016/j.jenvp.2009.11.004

Sörqvist, P. (2010b). High working memory capacity attenuates the deviation effect but not the changing-state effect: Further support for the duplex-mechanism account of auditory distraction. *Memory & Cognition*, *38*(5), 651–658. http://doi.org/10.3758/MC.38.5.651

* Sörqvist, P., Halin, N., & Hygge, S. (2010). Individual differences in susceptibility to the effect of speech on reading comprehension. *Applied Cognitive Psychology*, *24*(1), 67–76. http://doi.org/10.1002/acp.1543

* Sörqvist, P., Ljungberg, J. K., & Ljung, R. (2010). A sub-process view of working memory capacity: Evidence from effects of speech on prose memory. *Memory*, *18*(3), 310–326. http://doi.org/10.1080/09658211003601530

Sörqvist, P., & Marsh, J. E. (2015). How concentration shields against distraction. *Current Drections in Psychological Science*, *24*(4), 267–272. http://doi.org/10.1177/0963721415577356

Sperlich, A., Meixner, J., & Laubrock, J. (2016). Development of the perceptual span in reading: A longitudinal study. *Journal of Experimental Child Psychology*, *146*, 181–201. http://doi.org/10.1016/j.jecp.2016.02.007

Stansfeld, S., Berglund, B., Clark, C., Lopez-Barrio, I., Fischer, P., Öhrström, E., … Berry, B. (2005). Aircraft and road traffic noise and children's cognition and health: A cross-national study. *The Lancet*, *365*(9475), 1942–1949. http://doi.org/10.1016/S0140-6736(05)66660-3

Staub, A. (2007). The return of the repressed: Abandoned parses facilitate syntactic reanalysis. *Journal of Memory and Language*, *57*(2), 299–323. http://doi.org/10.1016/j.jml.2006.09.001

Staub, A., & Rayner, K. (2007). Eye movements and on-line language comprehension processes. In M. G. Gaskell (Ed.), *The Oxford handbook of psycholinguistics* (pp. 327–342). Oxford, UK: Oxford University Press. http://doi.org/10.1093/oxfordhb/9780198568971.013.0019

Sterne, J. A. C., Sutton, A. J., Ioannidis, J. P. A., Terrin, N., Jones, D. R., Lau, J., … Higgins, J. P. T. (2011). Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *BMJ*, *343*(7818), 1–8. http://doi.org/10.1136/bmj.d4002

Stracuzzi, D. J. (2004). EyeTrack (Version 0.7.10h) [Computer software]. Retrieved from http://blogs.umass.edu/eyelab

Stracuzzi, D. J., & Kinsey, J. D. (2009). EyeDoctor (Version 0.6.5) [Computer Software]. Retrieved from http://blogs.umass.edu/eyelab

* Sukowski, H., & Romanus, E. (2016). Effects of background speech on reading

performance in adults. In *Proceedings of Meetings on Acoustics 22ICA* (p. 50002). http://doi.org/10.1121/2.0000360

Sussman, E., Winkler, I., & Schröger, E. (2003). Top-down control over involuntary attention switching in the auditory modality. *Psychonomic Bulletin & Review*, *10*(3), 630–637. http://doi.org/10.3758/BF03196525

Sutton, A. J., & Abrams, K. R. (2001). Bayesian methods in meta-analysis and evidence synthesis. *Statistical Methods in Medical Research*, *10*(4), 277–303. http://doi.org/10.1191/096228001678227794

Sutton, A. J., Abrams, K. R., Jones, D. R., Sheldon, T. A., & Song, F. (2000). *Methods for meta-analysis in medical research*. Chichester, UK: John Wiley & Sons.

Szalma, J. L., & Hancock, P. A. (2011). Noise effects on human performance: A meta-analytic synthesis. *Psychological Bulletin*, *137*(4), 682–707. http://doi.org/10.1037/a0023987

Taylor, W. L. (1953). "Cloze procedure": A new tool for measuring readability. *Journalism Bulletin*, *30*(4), 415–433. http://doi.org/10.1177/107769905303000401

Thompson, S. G., & Higgins, J. P. T. (2002). How should meta-regression analyses be undertaken and interpreted? *Statistics in Medicine*, *21*(11), 1559–1573. http://doi.org/10.1002/sim.1187

Thompson, W. F., Schellenberg, E. G., & Letnic, A. K. (2012). Fast and loud background music disrupts reading comprehension. *Psychology of Music*, *40*(6), 700–708. http://doi.org/10.1177/0305735611400173

Tiitinen, H., May, P., Reinikainen, K., & Näätänen, R. (1994). Attentive novelty detection in humans is governed by pre-attentive sensory memory. *Nature*, *372*(6501), 90–92. http://doi.org/10.1038/372090a0

* Tucker, A., & Bushman, B. J. (1991). Effects of rock and roll music on mathematical, verbal, and reading comprehension performance. *Perceptual and Motor Skills*, *72*(3), 942–942. http://doi.org/10.2466/pms.1991.72.3.942

Vachon, F., Hughes, R. W., & Jones, D. M. (2012). Broken expectations: Violation of expectancies, not novelty, captures auditory attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(1), 164–177. http://doi.org/10.1037/a0025054

Van Heuven, W. J. B., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). SUBTLEX-UK: A new and improved word frequency database for British English. *Quarterly Journal of Experimental Psychology*, *67*(6), 1176–1190. http://doi.org/10.1080/17470218.2013.850521

Vasilev, M. R., & Angele, B. (2017). Parafoveal preview effects from word N + 1 and word N + 2 during reading: A critical review and Bayesian meta-analysis. *Psychonomic Bulletin & Review*, *24*(3), 666–689. http://doi.org/10.3758/s13423-016-1147-x

* Vasilev, M. R., Liversedge, S. P., Rowan, D., Kirkby, J. A., & Angele. (2017). Auditory distraction by meaningful background speech during reading. In *19th European*

Conference on Eye Movements. Wuppertal, Germany.

Vasishth, S. (2015). *A meta-analysis of relative clause processing in Mandarin Chinese using bias modelling (Unpublished Master's thesis)*. The University of Sheffield, United Kingdom.

Vasishth, S., Chen, Z., Li, Q., & Guo, G. (2013). Processing Chinese relative clauses: Evidence for the subject-relative advantage. *PLoS One*, *8*(10), 1–15. http://doi.org/10.1371/journal.pone.0077006

Veitch, J. A. (1990). Office noise and illumination effects on reading comprehension. *Journal of Environmental Psychology*, *10*(3), 209–217. http://doi.org/10.1016/S0272-4944(05)80096-9

Venetjoki, N., Kaarlela-Tuomaala, A., Keskinen, E., & Hongisto, V. (2006). The effect of speech and speech intelligibility on task performance. *Ergonomics*, *49*(11), 1068–1091. http://doi.org/10.1080/00140130600679142

Viechtbauer, W. (2010). Conducting Meta-Analyses in R with the metafor Package. *Journal of Statistical Software*, *36*(3), 1–48. http://doi.org/10.18637/jss.v036.i03

Vitu, F. (2005). Visual extraction processes and regressive saccades in reading. In G. Underwood (Ed.), *Cognitive processes in eye guidance* (pp. 1–32). Oxford, UK: Oxford University Press.

Vitu, F., & McConkie, G. W. (2000). Regressive saccades and word perception in adult reading. In A. Kennedy, D. Heller, J. Pynte, & R. Radach (Eds.), *Reading as a perceptual process* (pp. 301–326). Oxford, UK: Elsevier.

Vitu, F., O'Regan, J. K., Inhoff, A. W., & Topolski, R. (1995). Mindless reading: Eye-movement characteristics are similar in scanning letter strings and reading texts. *Perception & Psychophysics*, *57*(3), 352–364. http://doi.org/10.3758/BF03213060

Wallisch, P. (2015). Brighter than the sun: Powerscape visualizations illustrate power needs in neuroscience and psychology. Retrieved from http://arxiv.org/abs/1512.09368

Wang, Y., Spence, M. M., Jongman, A., & Sereno, J. A. (1999). Training American listeners to perceive Mandarin tones. *The Journal of the Acoustical Society of America*, *106*(6), 3649–3658. http://doi.org/10.1121/1.428217

* Weinstein, N. D. (1974). Effect of noise on intellectual performance. *Journal of Applied Psychology*, *59*(5), 548–554. http://doi.org/10.1037/h0037338

* Weinstein, N. D. (1977). Noise and intellectual performance: A confirmation and extension. *The Journal of Applied Psychology*, *62*(1), 104–107. http://doi.org/10.1037/0021-9010.62.1.104

Weiss, A. F., Kretzschmar, F., Schlesewsky, M., Bornkessel-Schlesewsky, I. D., & Staub, A. (2017). Comprehension demands modulate re-reading, but not first pass reading behavior. *The Quarterly Journal of Experimental Psychology*, 1–37. http://doi.org/10.1080/17470218.2017.1307862

Welton, N. J., Sutton, A. J., & Cooper, N. (2012). *Statistics in practice: Evidence synthesis for decision making in healthcare*. Somerset: Wiley.

Wessel, J. R. (2017). Perceptual surprise aides inhibitory motor control. *Journal of Experimental Psychology: Human Perception and Performance*, *43*(9), 1585–1593. http://doi.org/10.1037/xhp0000452

Wessel, J. R., & Aron, A. R. (2013). Unexpected events induce motor slowing via a brain mechanism for action-stopping with global suppressive effects. *Journal of Neuroscience*, *33*(47), 18481–18491. http://doi.org/10.1523/JNEUROSCI.3456-13.2013

Wessel, J. R., & Aron, A. R. (2017). On the globality of motor suppression: Unexpected events and their influence on behavior and cognition. *Neuron*, *93*(2), 259–280. http://doi.org/10.1016/j.neuron.2016.12.013

Wessel, J. R., Jenkinson, N., Brittain, J. S., Voets, S. H. E. M., Aziz, T. Z., & Aron, A. R. (2016). Surprise disrupts cognition via a fronto-basal ganglia suppressive mechanism. *Nature Communications*, *7*, 1–10. http://doi.org/10.1038/ncomms11195

Westfall, J. (2015). PANGEA: Power analysis for general anova designs. *Unpublished Manuscript*. Retrieved from http://jakewestfall.org/publications/pangea.pdf

Wetzel, N., Scharf, F., & Widmann, A. (2018). Can't ignore- Distraction by task-irrelevant sounds in early and middle childhood. *Child Development*, 1–12. http://doi.org/10.1111/cdev.13109

Wetzel, N., Schröger, E., & Widmann, A. (2013). The dissociation between the P3a event-related potential and behavioral distraction. *Psychophysiology*, *50*(9), 920–930. http://doi.org/10.1111/psyp.12072

Wetzel, N., Schröger, E., & Widmann, A. (2016). Distraction by novel and pitch-deviant sounds in children. *Frontiers in Psychology*, *7*, 1–12. http://doi.org/10.3389/fpsyg.2016.01949

Wetzels, R., Matzke, D., Lee, M. D., Rouder, J. N., Iverson, G. J., & Wagenmakers, E.-J. (2011). Statistical evidence in experimental psychology: An empirical comparison using 855 t tests. *Perspectives on Psychological Science*, *6*(3), 291–298. http://doi.org/10.1177/1745691611406923

White, S. J., Rayner, K., & Liversedge, S. P. (2005). Eye movements and the modulation of parafoveal processing by foveal processing difficulty: A reexamination. *Psychonomic Bulletin & Review*, *12*(5), 891–896. http://doi.org/10.3758/BF03196782

Wolf, F., & Gibson, E. (2005). Representing discourse coherence: A corpus-based study. *Computational Linguistics*, *31*(2), 249–287. http://doi.org/10.1162/0891201054223977

Wotschack, C., & Kliegl, R. (2013). Reading strategy modulates parafoveal-on-foveal effects in sentence reading. *Quarterly Journal of Experimental Psychology*, *66*(3), 548–562. http://doi.org/10.1080/17470218.2011.625094

* Yan, G., Meng, Z., Liu, N., He, L., & Paterson, K. B. (2017). Effects of irrelevant background speech on eye movements during reading. *The Quarterly Journal of Experimental Psychology*, 1–20. http://doi.org/10.1080/17470218.2017.1339718

Ylias, G., & Heaven, P. C. L. (2003). The influence of distraction on reading comprehension: A big five analysis. *Personality and Individual Differences*, *34*(6), 1069–1079. http://doi.org/10.1016/S0191-8869(02)00096-X

Zhang, H., Miller, K., Cleveland, R., & Cortina, K. (2018). How listening to music affects reading: Evidence from eye tracking. *Journal of Experimental Psychology: Learning Memory and Cognition*. http://doi.org/10.1037/xlm0000544

Appendix A

Chapter 2: Meta-analysis Study Inclusion Criteria

- The study investigated the effect of experimental exposure to background noise, speech, or music in a reading/ proofreading task.
- Only studies investigating the immediate effect of background sounds on reading/ proofreading were included. Experiments that studied the effect of long-term exposure to music as an intervention for reading were excluded. Studies that investigated the effects of chronic exposure to traffic noise were also excluded.
- The study contained a condition of reading in silence. This served as the baseline to which background sound manipulations were compared. Studies without a silence baseline were excluded.
- The study had appropriate randomization and counter-balancing of the sound conditions.
- Participants were native speakers of the language in which they were reading.
- The study was done with healthy, typically-developing participants (children or adults).
- The external environment or any additional manipulations did not introduce confounds.
- Participants were not tested on the contents of the sound that they were listening to (e.g. speech).
- The assessment task emphasized comprehension of the text rather than reproducing the text from memory as accurately as possible.

- The comprehension assessment did not occur too long after the reading phase (usually within 10-15 minutes).
- The comprehension assessment was done in silence.

Appendix B

Information about the studies included in the meta-analysis in Chapter 2.

Study	N_C	N_E	Sample	Design	DV	Sound	Sound type	dB(A)	g	var
Sörqvist et al. 2010	40		A	W	RC	speech	native	72.5	-0.24	0.01
Sörqvist et al. 2010	40		A	W	RS	speech	native	72.5	-0.05	0.01
Ljung et al. 2009	70	50	C	B	RC	noise	traffic	62	-0.16	0.03
Ljung et al. 2009	70	50	C	B	RS	noise	traffic	62	0.71	0.04
Ljung et al. 2009	70	66	C	B	RC	speech	babble	62	0.17	0.03
Ljung et al. 2009	70	66	C	B	RS	speech	babble	62	0.21	0.03
Fogelson 1973	14	14	C	B	RC	music	pop	-	-0.42	0.14
Tucker & Bushm. 1991	75	76	A	B	RC	music	rock & roll	80	0.00	0.03
Daoussis & Mc K. 1986	24	24	A	B	RC	music	rock	50	-0.52	0.08
Etaugh & Michals 1975	32		A	W	RC	music	preferred	-	-0.08	0.02
Etaugh & Ptasnik 1982	20	20	A	B	RC	music	preferred	-	-0.74	0.10
Kiger 1989	18	18	C	B	RC	music	low load	-	3.50	0.28
Kiger 1989	18	18	C	B	RC	music	high load	-	-0.69	0.11
Miller & Schyb 1989	49	49	A	B	RC	music	classical	47.5	0.11	0.04
Miller & Schyb 1989	49	49	A	B	RC	music	pop	47.5	0.23	0.04
Miller & Schyb 1989	49	49	A	B	RC	music	vocal	47.5	-0.46	0.04
Doyle & Furnham 2012	56		A	W	RC	music	vocal	-	0.10	0.01
Anderson & Fuller 2010	334		C	W	RC	music	lyrical	75	-0.28	0.00
Furnham & Strbac 2002	76		C	W	RC	noise	office	-	-0.78	0.01
Furnham & Strbac 2002	76		C	W	RC	music	vocal/unfam.	-	-0.83	0.01
Mullikin & Henk 1985	45		C	W	RC	music	classical	-	0.39	0.01
Mullikin & Henk 1985	45		C	W	RC	music	rock	-	-0.33	0.01
Avila et al. 2011	19	20	C	B	RC	music	vocal/ familiar	-	-1.61	0.13
Avila et al. 2011	19	19	C	B	RC	music	Instr./ familiar	-	-1.93	0.15
Freeburne & Fleis. 1952	43	46	A	B	RC	music	classical	-	0.02	0.04
Freeburne & Fleis. 1952	43	46	A	B	RS	music	classical	-	-0.35	0.04
Freeburne & Fleis. 1952	43	42	A	B	RC	music	pop	-	0.04	0.05
Freeburne & Fleis. 1952	43	42	A	B	RS	music	pop	-	-0.40	0.05
Freeburne & Fleis. 1952	43	40	A	B	RC	music	semi-classical	-	-0.08	0.05
Freeburne & Fleis. 1952	43	40	A	B	RS	music	semi-classical	-	-0.36	0.05
Freeburne & Fleis. 1952	43	37	A	B	RC	music	jazz	-	-0.17	0.05
Freeburne & Fleis. 1952	43	37	A	B	RS	music	jazz	-	-0.61	0.05
Fendrick 1937	61	62	A	B	RC	music	semi-classical	-	-0.47	0.03
Henderson et al. 1945	19	17	A	B	RC	music	classical	-	-0.12	0.11
Henderson et al. 1945	19	14	A	B	RC	music	pop	-	-1.07	0.14
Miller 2014	13	13	A	B	RC	music	classical lyrical	-	-0.84	0.16
Miller 2014	13	17	A	B	RC	music	classical instr.	-	0.13	0.13
Miller 2014	13	11	A	B	RC	music	rock lyrical	-	-0.38	0.16
Miller 2014	13	18	A	B	RC	music	rock instr.	-	-0.45	0.13
Furnham & Allass 1999	16	16	A	B	RC	music	complex	-	-0.02	0.12

Study	N_C	N_E	Sample	Design	DV	Sound	Sound type	dB(A)	g	var
Furnham & Allass 1999	16	16	A	B	RC	music	simple	-	-0.05	0.12
Furnham & Bradl. 1997	10	10	A	B	RC	music	pop	-	-0.97	0.21
Furnham at al. 1999	43	49	C	B	RC	music	instrumental	-	-0.12	0.04
Furnham at al. 1999	43	47	C	B	RC	music	vocal	-	-0.07	0.04
Perham & Currie 2014	30		A	W	RC	music	disliked lyrical	70	-0.71	0.02
Perham & Currie 2014	30		A	W	RC	music	non-lyrical	70	-0.16	0.02
Perham & Currie 2014	30		A	W	RC	music	liked lyrical	70	-0.60	0.02
Kelly 1994	13	12	A	B	RC	music	pop	65	-0.74	0.16
Dove 2009	28	28	A	B	RC	music	sedat. classical	62.5	0.10	0.07
Dove 2009	28	28	A	B	RC	music	stimul. classical	62.5	0.81	0.08
Dove 2009	28	28	A	B	RS	music	sedat. classical	62.5	-0.07	0.07
Dove 2009	28	28	A	B	RS	music	stimul. classical	62.5	-0.51	0.07
Furnham et al. 1994	20		A	W	RC	speech	TV drama	-	-0.45	0.03
Johansson 1983	22	22	C	B	RC	noise	continuous	51	0.28	0.09
Johansson 1983	22	22	C	B	RC	noise	intermittent	67.4	0.21	0.09
Halin 2016	28		A	W	RC	speech	native (easy)	60	-0.89	0.03
Halin 2016	28		A	W	RC	speech	native (diff)	60	-0.16	0.02
Halin 2016	28		A	W	RC	noise	traffic (easy)	60	-0.35	0.02
Halin 2016	28		A	W	RC	noise	traffic (diff)	60	-0.01	0.02
Halin 2016	28		A	W	RC	noise	aircraft (easy)	60	-0.23	0.02
Halin 2016	28		A	W	RC	noise	aircraft (diff)	60	-0.01	0.02
Smith-J. & Klein 2009	54		A	W	PR	speech	native	65	-0.04	0.01
Cauchard et al. 2012	30		A	W	RC	music	instrumental	65	0.18	0.02
Cauchard et al. 2012	30		A	W	RC	speech	native	65	-0.17	0.02
Cauchard et al. 2012	30		A	W	RS	music	instrumental	65	0.01	0.02
Cauchard et al. 2012	30		A	W	RS	speech	native	65	-0.20	0.02
Johansson et al. 2012	24		A	W	RC	music	preferred	65	-0.34	0.02
Johansson et al. 2012	24		A	W	RC	music	non-preferred	65	-0.67	0.03
Johansson et al. 2012	24		A	W	RC	noise	cafe	65	-0.31	0.02
Johansson et al. 2012	24		A	W	RS	music	preferred	65	-0.14	0.02
Johansson et al. 2012	24		A	W	RS	music	non-preferred	65	-0.10	0.02
Johansson et al. 2012	24		A	W	RS	noise	cafe	65	-0.07	0.02
Weinstein 1974	15	18	A	B	PR†	noise	teletype	70	-0.56	0.12
Weinstein 1974	15	18	A	B	PR‡	noise	teletype	70	-1.26	0.14
Weinstein 1977	29		A	W	PR†	speech	native	68	-0.03	0.02
Weinstein 1977	29		A	W	PR‡	speech	native	68	-0.29	0.02
Martin et al. 1988, E1	36		A	W	RC	speech	native	82	-0.20	0.01
Martin et al. 1988, E1	36		A	W	RC	speech	random	82	-0.18	0.01
Martin et al. 1988, E1	36		A	W	RC	music	instrumental	82	0.00	0.01
Martin et al. 1988, E1	36		A	W	RC	music	random tones	82	-0.11	0.01
Martin et al. 1988, E1	36		A	W	RC	noise	white	82	-0.04	0.01
Martin et al. 1988, E2	36		A	W	RC	music	instrumental	82	0.02	0.01
Martin et al. 1988, E2	36		A	W	RC	music	lyrical	82	-0.08	0.01
Martin et al. 1988, E4	48		A	W	RC	noise	white	82	-0.11	0.01

Study	N_C	N_E	Sample	Design	DV	Sound	Sound type	dB(A)	g	var
Martin et al. 1988, E4	48		A	W	RC	speech	native	82	-0.31	0.01
Martin et al. 1988, E4	48		A	W	RC	speech	foreign	82	-0.15	0.01
Martin et al. 1988, E5	48		A	W	RC	noise	white	82	-0.21	0.01
Martin et al. 1988, E5	48		A	W	RC	speech	non-word	82	-0.20	0.01
Martin et al. 1988, E5	48		A	W	RC	speech	random words	82	-0.33	0.01
Cool et al. 1994, E2	9		C	W	RS	music	radio/ generic	-	0.13	0.05
Cool et al. 1994, E2	9		C	W	RS	speech	movies	-	0.20	0.05
Cool et al. 1994, E2	9		C	W	RC	music	radio/ generic	-	-0.12	0.05
Cool et al. 1994, E2	9		C	W	RC	speech	movies	-	-0.22	0.05
Mitchell 1949	91		C	W	RTS	music	radio/ generic	-	-0.01	0.01
Armstrong et al. 1991	33	30	A	B	RTS	speech	TV ads	-	-0.63	0.07
Armstrong et al. 1991	33	32	A	B	RTS	speech	TV drama	-	-0.48	0.06
Pool et al. 2000, E1	30	30	C	B	RC	speech	TV soap opera	60	-0.38	0.07
Pool et al. 2000, E1	30	30	C	B	RC	music	TV music	60	-0.21	0.07
Pool et al. 2000, E2	48	24	C	B	RC	speech	TV soap opera	60	-0.57	0.06
Pool et al. 2000, E2	48	24	C	B	RC	music	TV music	60	-0.10	0.06
Dockrell & Shield 2006	52	52	C	B	RTS	noise	babble	65	-0.49	0.04
Dockrell & Shield 2006	52	52	C	B	RTS	noise	babble+environ.	65	0.58	0.04
Hyönä & Ekh. 2016, E1	42		A	W	RC	speech	native	82.5	-0.17	0.01
Hyönä & Ekh. 2016, E1	42		A	W	RC	speech	foreign	82.5	0.00	0.01
Hyönä & Ekh. 2016, E1	42		A	W	RS	speech	native	82.5	-0.02	0.01
Hyönä & Ekh. 2016, E1	42		A	W	RS	speech	foreign	82.5	0.06	0.01
Hyönä & Ekh. 2016, E2	36		A	W	RS	speech	scrambl.-differ.	82.5	-0.15	0.01
Hyönä & Ekh. 2016, E2	36		A	W	RS	speech	scrambl.-same	82.5	-0.18	0.01
Hyönä & Ekh. 2016, E3	35		A	W	RS	speech	native	82.5	-0.13	0.01
Hyönä & Ekh. 2016, E3	35		A	W	RS	speech	scrambled	82.5	-0.20	0.01
Hyönä & Ekh. 2016, E4	36		A	W	RS	speech	scrambled-sem.	82.5	-0.11	0.01
Hyönä & Ekh. 2016, E4	36		A	W	RS	speech	scrm-syn+sem	82.5	-0.14	0.01
Armstrong & Chng 2000	19	20	A	B	RC	speech	native	-	-0.09	0.10
Madsen 1987, E1	50	50	A	B	RC	music	various	75	-0.10	0.04
Sörqvist 2010, E1a	23		C	W	RC	noise	aircraft	57.5	-0.13	0.02
Sörqvist 2010, E1b	23		C	W	RC	speech	native	57.5	-0.51	0.03
Sörqvist et al. 2010, E1	24		A	W	RC	speech	native	65	-0.46	0.02
Sörqvist et al. 2010, E2	42		A	W	RC	speech	native	65	-0.30	0.01
Halin et al. 2014	32		A	W	RC	speech	native	65	-0.10	0.02
Halin et al. 2014, E1	31		A	W	PR[‡]	speech	native	65	-0.09	0.02
Halin et al. 2014, E1	31		A	W	PR[†]	speech	native	65	0.20	0.02
Halin et al. 2014, E2	29		A	W	PR[‡]	speech	native	65	-0.13	0.02
Halin et al. 2014, E2	29		A	W	PR[†]	speech	native	65	0.11	0.02
Haapakangas et al. 2011	54		A	W	PR[‡]	speech	native	48	-0.09	0.01
Haapakangas et al. 2011	54		A	W	PR[†]	speech	native	48	-0.11	0.01
Baker & Madell 1965	24		A	W	RC	speech	native	-	-0.70	0.03
Vasilev et al. 2017	40		A	W	RC	noise	speech-spectr.	60	-0.03	0.01

Study	N_C	N_E	Sample	Design	DV	Sound	Sound type	dB(A)	g	var
Vasilev et al. 2017	40		A	W	RC	speech	foreign	60	-0.01	0.01
Vasilev et al. 2017	40		A	W	RC	speech	native	60	-0.07	0.01
Vasilev et al. 2017	40		A	W	RS	noise	speech-spectr.	60	0.04	0.01
Vasilev et al. 2017	40		A	W	RS	speech	foreign	60	-0.06	0.01
Vasilev et al. 2017	40		A	W	RS	speech	native	60	-0.15	0.01
Falcon 2017, Sample 1	22	20	C	B	RC	music	classical	55	-0.26	0.09
Falcon 2017, Sample 2	25	28	C	B	RC	music	classical	55	1.32	0.09
Ahuja 2016	20		A	W	RC	music	liked	60	-0.71	0.04
Ahuja 2016	20		A	W	RC	music	disliked	60	-0.08	0.02
Kou et al. 2017	31	29	A	B	RC	music	pop (vocal)	65	0.37	0.07
Kou et al. 2017	31	32	A	B	RC	noise	office	65	-0.13	0.06
Sukowski et al. 2016	12		A	W	PR	speech	native	59.5	-0.62	0.05
Yan et al. 2017	42		A	W	RS	speech	native	62	-0.16	0.01
Yan et al. 2017	42		A	W	RS	speech	meaningless	62	0.06	0.01
Gillis 2016	24	47	A	B	RC	music	various	-	0.07	0.06

Table A1. A Summary of the studies that were included in the meta-analysis from Chapter 2 and their effect sizes. N_C: number of participants in the control (silence) condition. N_E: number of participants in the experimental (sound) condition. DV: dependent variable. RC: Reading comprehension. RS: reading speed. RTS: Reading test score. PR: Proofreading accuracy. g: Effect size in Hedges' g. var: effect size variance. A: adults. C: children. W: within-subject design. B: between-subject design.

† Non-contextual errors (proofreading accuracy)

‡Contextual errors (proofreading accuracy)

Appendix C

Supplementary analyses for the meta-analysis in Chapter 2.

Visualization of the Effect Sizes

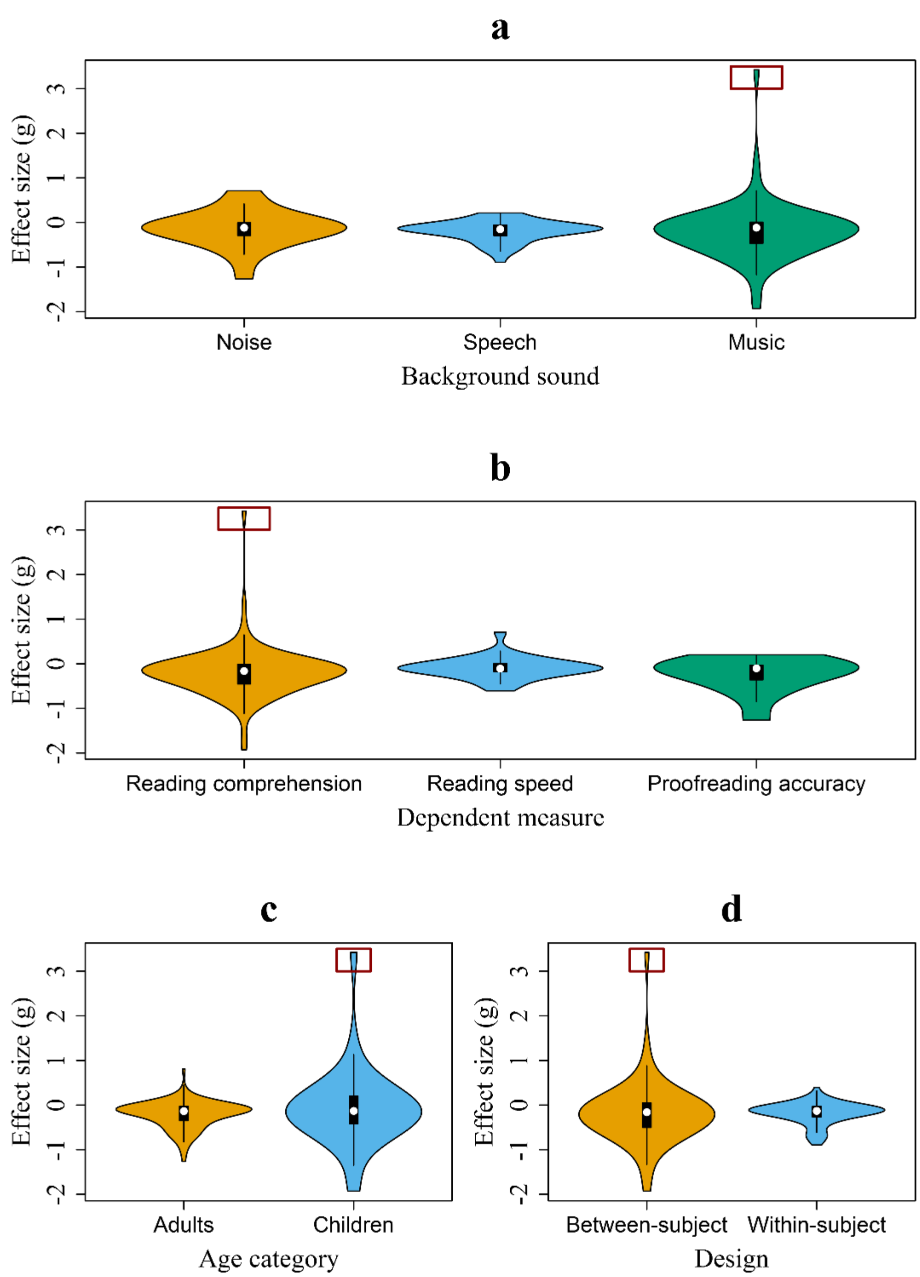

Figure A1. Box plots and probability densities of the effect sizes included in the meta-analysis. Breakdown shown by: background sound type (panel **a**), dependent measure (panel

b), age of participants (panel **c**), and study design (panel **d**; computed after transforming within-subject effect sizes with Morris & DeShon's, 2002, formula 11). Red rectangle shows one effect size that was excluded as an outlier.

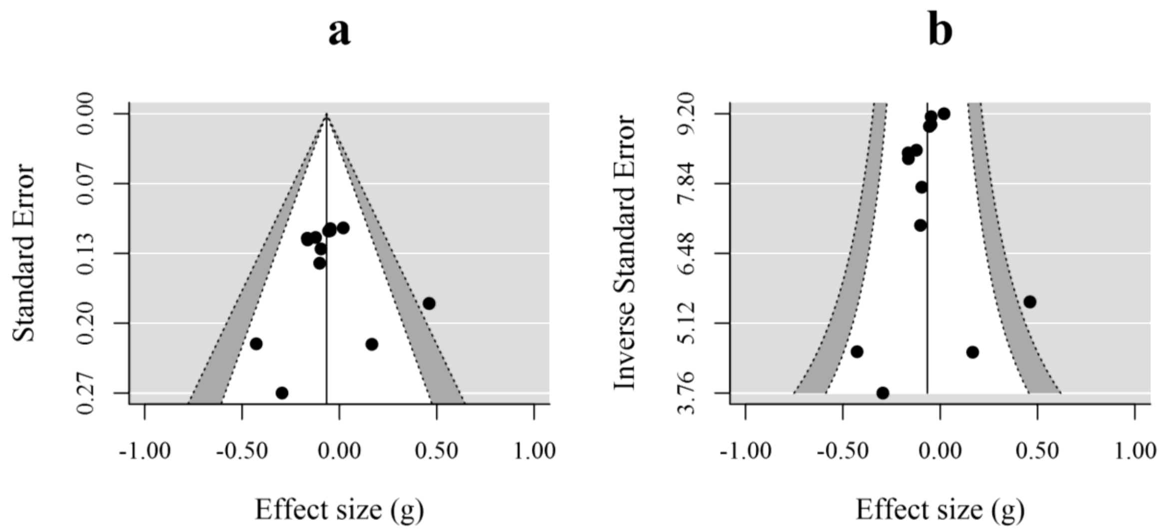

Figure A2. Funnel plot of reading speed effect sizes plotted against their standard error (**a**) and the inverse of their standard error (**b**).

Prior Sensitivity Analysis

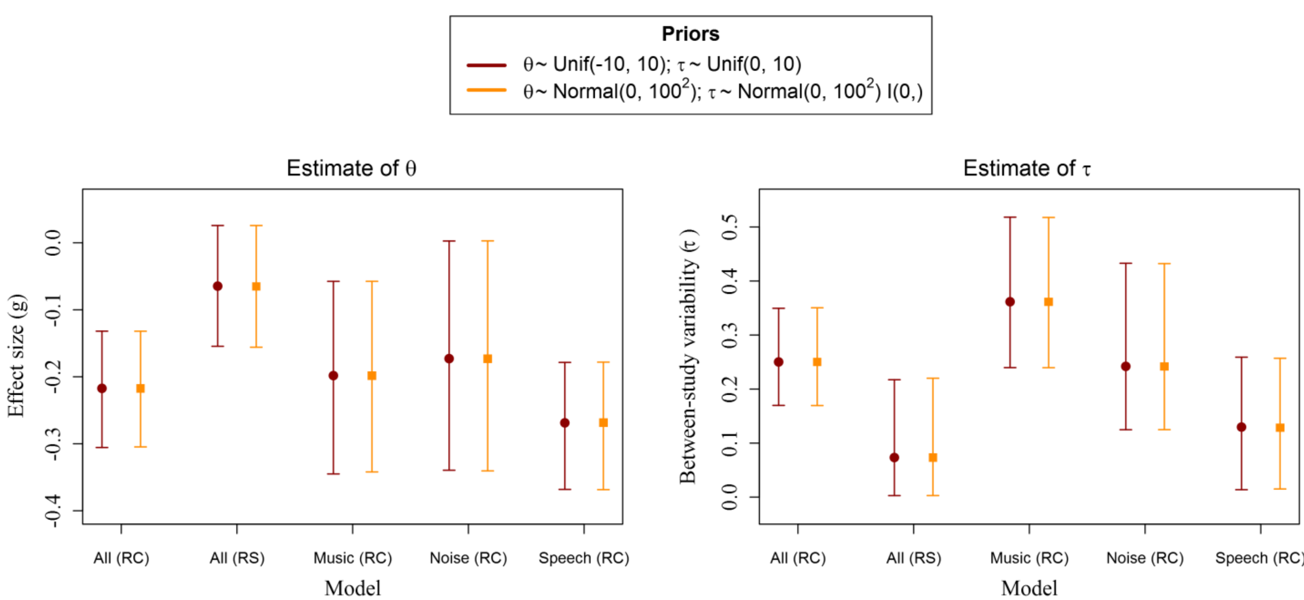

Figure A3. Sensitivity analysis with different priors on the θ and τ parameters for the main meta-analysis results. Uniform priors (dark red) were used in the analysis reported in

Chapter 2. The results show that using diffuse normal priors (orange) did not change the main results reported in the paper. All: all studies. RC: reading comprehension. RS: reading speed. Effective sample size of the MCMC chains for θ (from left to right): 91803, 96976, 20915, 25462, 93678, 100908, 92499, 98585, 47662, 54666. Effective sample size of the MCMC chains for τ (from left to right): 53392, 53451, 18985, 19517, 67441, 68050, 71910, 72202, 11786, 12392.

Robustness Check (Leave-one-out Method)

Robustness analyses were carried out by using the leave-one-out method (see Greenhouse & Iyengar, 2009) to ensure that individual studies did not have undue influence on the results. In this method, the meta-analysis is repeated by omitting one different study each time. The summary statistics of the results are reported in Table A2. Overall, the effect sizes changed little by omitting each one of the studies. The effect size range for proofreading accuracy was slightly bigger, but this was likely due to the small number of studies in this analysis (*N*= 7). This greater variability is not unusual for random-effects meta-analysis with few studies because there is more uncertainty in estimating the between-study variance in the model (see Welton et al., 2012).

Analysis	ES reported in the Thesis	Leave-one-out			
		Mean ES	SD of ES	Min ES	Max ES
Reading comprehension					
All sounds	-0.21	-0.21	0.006	-0.23	-0.19
Noise	-0.17	-0.17	0.02	-0.19	-0.11
Speech	-0.26	-0.26	0.01	-0.28	-0.24
Music	-0.19	-0.19	0.01	-0.22	-0.16
Reading speed	-0.06	-0.06	0.01	-0.09	-0.05
Proofreading accuracy	-0.14	-0.15	0.04	-0.19	-0.08

Table A2. Summary of the robustness analysis using the leave-one-out method.

Lyrical vs Non-Lyrical Music: Meta-regression Robustness Check

Some of the included studies had effect sizes for both lyrical and non-lyrical music. In order to avoid stochastical dependency among the effect sizes included in this meta-regression analysis, it was necessary to ensure that each study contributed one and only one effect size to either the “lyrical” or “non-lyrical” group. In Chapter 2, the effect sizes were divided into the two groups in a way that maximized the number of effect sizes per group. This is because meta-regressions with larger and more balanced number of observations per group would generally yield more informative results. However, to check for subjectivity in this decision, we did the opposite division of the effect size to compare the results (this will be referred to as the “alternative coding”). The resulting posterior distributions of the mean difference are plotted in Figure A4. As it can be seen, the estimated mean difference was slightly smaller. In the model reported in Chapter 2, there was 95% probability that lyrical music was more distracting than non-lyrical music. For the model with alternative coding, this probability was 83%. Therefore, even though there was slightly more uncertainty and the mean difference was slightly smaller with the alternative coding, the main conclusions remain unchanged.

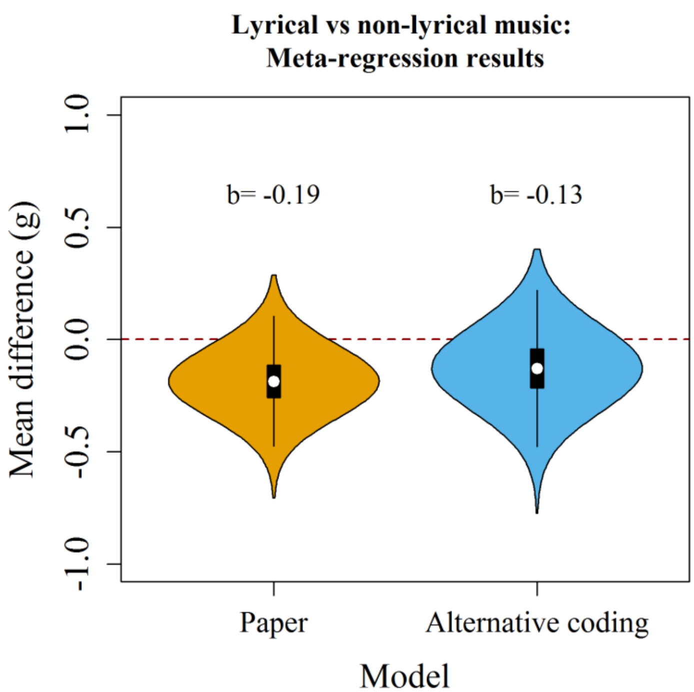

Figure A4. A plot of the posterior distributions of the estimated mean difference in effect sizes between lyrical and non-lyrical music. Plotted are the model reported in Chapter 2 (orange) and the model done with the alternative coding of the effect sizes (blue). The results indicate that the decision of which coding to use did not affect the conclusions in Chapter 2. Effective sample size of the MCMC chains for β: 11455 (model reported in Chapter 2), 11695 (model with alternative coding).

Unavailable Data

Due to that fact that four studies did not contain enough information to compute effect sizes and to include them in the meta-analysis, statistical simulations were carried out to explore the consequences of this. The relevant information about these studies is summarised in Table A3. For each study, a realistic interval was computed that should contain the effect size of interest given the available information. The simulations were done by taking 10 000 random draws from a Uniform distribution using the effect size bounds in Table A4. For the variance component, a random draw was also taken from a Uniform distribution with bounds corresponding to the range of variance values in the dataset. The random draws were taken from Uniform distributions to denote ignorance about where on the interval the real value may lie. Each randomly generated effect size was added to the dataset that was analysed in the paper and the meta-analysis was then repeated. The results from the simulations are presented in Table A4 and compared to the effect sizes reported in the main paper. As the simulations show, the results changed very little or not at all when the missing effect sizes were simulated and then added to the analyses. Therefore, the lack of access to the effect sizes of these fours studies did not bias the conclusions from the meta-analysis.

Study	N	Measure	Sound	Available information	Anticipated effect size
Hall (1952)	245	RC	Music	2.37% increase in reading score in the music condition	0< g <0.5
Gawron (1984)	32†	RC	Noise	Effect size known, but not the direction of the difference	|g|= 0.048
Slater (1968)	263	RC	Noise	No sign. differences and “no trends indicative of… [an] effect” (p. 242)	-0.2< g <0.2
Jones et al. (1990), E2	16	PR	Speech	F-value <1; effect size is negative based on the means in Table 2	-0.13< g <0 ‡

Table A3. Information about studies with unavailable data and their anticipated effect sizes. RC: reading comprehension accuracy. PR: proofreading accuracy. N: (combined) sample size. All effect sizes are with Morris and DeShon’s (2002) correction (where applicable).

† Only two schedules (2x16 participants) are relevant to the analysis

‡ -0.13 is the lowest possible bound since this would correspond to the effect size when the *F*-value is 1.

Analysis	ES (paper)	Results from 10 000 simulations			
		Mean ES	Range	Mean distribution	Variance distribution
PR	-0.14	-0.13 (0.01)	[-0.15, -0.10]	Uniform(-0.13, 0)	Uniform(0.01, 0.13)
RC: Music	-0.19	-0.19 (0.004)	[-0.20, -0.17]	Uniform(0, 0.5)	Uniform(0.01, 0.20)
RC: Noise	-0.17	-0.16 (0.01)	[-0.17, -0.13]	Uniform(-0.2, 0.2)†	Uniform(0.01, 0.08)

Table A4. Results from the statistical simulations with missing data (SDs in parenthesis). RC: reading comprehension. PR: proofreading.

† Used for Slater's (1968) study. For Gawron's (1984) study, the effect size was positive for half of the simulations (*g*= 0.048), and negative (*g*= -0.048) for the remaining half.

Appendix D

Sentence stimuli used in Chapter 3. The target words are formatted in bold. In each sentence, the first word in bold is the high frequency target word, and the second word in bold is the low frequency target word.

1. Mrs. Clark is a **social/ chatty** person who gets along with everybody.
2. Hannah enjoyed the **modern/ trendy** artworks at the museum of contemporary art.
3. The house was immediately recognisable by its **green/ beige** fence and big windows.
4. The building inspector examined the **large/ leaky** roof of the house in the morning.
5. Building a house requires a **massive/ sizable** amount of money that few people have.
6. The bike was **strong/ sturdy** enough to be used on the bumpy roads in the mountain.
7. In some schools, subjects such as art and music play a **special/ trivial** role in the curriculum.
8. The local schools were criticised for their **liberal/ lenient** policy towards pupil truancy.
9. To care for **soft/ pale** skin in the summer, the dermatologist recommended a special cream.
10. The company's updated logo featured a **blue/ cyan** star in the foreground.
11. Jim prepared a **nice/ neat** outfit for his best friend's wedding.
12. The crowd listened to the **famous/ solemn** hymn before the start of the main event.
13. The children all had a **lovely/ cheery** smile on their face in the photo taken at the zoo.
14. During the financial crisis, the **clever/ astute** merchant invested money in real estate.
15. The archaeologists suspected that the tool was **common/ unused** before the era of Neanderthals.
16. The TV guest explained how his **serious/ austere** attitude to life helped him stay disciplined.
17. The host family gave a **welcome/ sincere** gift to the noble couple that was visiting them.
18. Despite the high price, the hotel room had a **central/ frontal** view of the ocean.

19. There are many critics of the **classic/ archaic** prison system that is still in use today.
20. Sarah always tried to make a **careful/ prudent** use of her hard-earned money.
21. The children immediately liked the new cat with its **white/ furry** paws and playful behaviour.
22. Linda used the **short/ blunt** pencil to write a quick note on the fridge.
23. Some citizens were concerned with the **quick/ rapid** growth of tourism in the area.
24. The little town was known for its **local/ famed** brewery with 400 years of tradition.
25. People saw Samantha as a **happy/ nerdy** girl who did very well in her studies.
26. After winning the national competition, the **young/ agile** athlete had a promising career.
27. For many species, the **quiet/ dense** forest is an ideal natural habitat.
28. The taxi driver struggled to put the **heavy/ bulky** suitcase in the back of the car.
29. The small child was riding a bike while his **proud/ stern** father was keeping an eye on him.
30. The little girl liked to sleep with her **lucky/ plush** toy next to her.
31. David's friends were tired of listening to his **usual/ stale** jokes all the time.
32. The progress of the **bright/ docile** student made his teacher very proud.
33. Thanks to the **proper/ prompt** response of the mayor, the dispute was quickly resolved.
34. Mark tried to keep a **normal/ casual** tone after finding out about the secret.
35. Many spectators were fascinated by the **long/ wavy** hair of the lead actress.
36. Jane didn't like the taste of the **free/ iced** cake at her company's banquet party.
37. The hotel guests liked the **warm/ airy** reception area with its big windows and cosy sofas.
38. Alex is the **keen/ avid** type of golfer who plays every weekend regardless of the weather.
39. The mountaineers experienced **cold/ numb** feet due to the heavy snowfall.
40. For small children, even a **calm/ tame** dog can be a cause of fear.
41. In the big city, there was a **real/ dire** need to build more residential buildings.
42. John knew that the **next/ mock** exam requires little preparation.
43. Some people find the **natural/ calming** sound of rain drops conducive to sleep.

44. The party included a number of **popular/ notable** guests from the capital.
45. Despite its long history, the **private/ baroque** castle was shrouded in mystery.
46. Many of the bus passengers became **nervous/ fearful** due to the bad road conditions.
47. Because of his **current/ chronic** medical condition, Jake could not travel abroad.
48. The farm consisted of large areas of **yellow/ fallow** soil that were intended for growing crops.
49. There were complaints despite the **double/ hourly** payment rate for the new project.
50. The general public was not happy about the **secret/ futile** actions of the defence ministry.
51. Situated at the outskirts of the city, the **public/ wooded** area was neglected for many years.
52. The construction company was fined for the **latest/ costly** delays in building the new stadium.
53. The start-up company was hoping to keep up the **average/ booming** rise in their sales.
54. The central bank imposed new regulations for **foreign/ virtual** currency in the country.
55. Katy remained positive despite the **sorry/ bleak** financial situation that she was in.
56. The scientists came up with a **likely/ viable** solution to the problem of growing food in space.
57. The landlord made the **final/ hasty** decision to sell his property and retire abroad.
58. The LED lamp was a **small/ faint** source of light that left most of the room dark.
59. The secretary was in **total/ utter** disbelief after she was accused of embezzlement.
60. The protagonist of the book relied on her **human/ feral** instincts to survive in the forest.
61. The news reporter caused controversy with his **unusual/ graphic** account of the events.
62. Doing exercise is believed to alleviate **certain/ cardiac** diseases and improve mental wellbeing.
63. The emergency services took **further/ drastic** measures to ensure the safety of the population.

64. The audience was **pleased/ stunned** after the fantastic performance of the theatre group.
65. The newly built **street/ avenue** was designed to reduce traffic in rush hours.
66. The camera recorded the **driver/ robber** while he was speeding down the highway.
67. The autobiography explained how the author's **brother/ sibling** helped him rise to fame.
68. Elizabeth likes to spend time with her **family/ spouse** whenever she comes home.
69. According to the commission, the **island/ lagoon** should be better protected from pollution.
70. For breakfast, Jason ate the **food/ stew** that was left from the previous night.
71. Many students could not understand the **idea/ gist** without reading the text twice.
72. Last night, Joseph stopped by the **wine/ pawn** shop on his way home.
73. During a family visit to the village, the child saw a **horse/ stork** behind the empty barn.
74. Jack visited the **market/ tavern** situated just outside of the city centre.
75. The old lady needed a **friend/ helper** who could assist her with cleaning the house.
76. Just like every other **person/ broker** working in the office, Lisa wanted to earn a lot of money.
77. After reports of **smell/ odour** coming from the basement, the gas company sent a response team.
78. During the summer, the **hotel/ motel** attracted many tourists with its low prices.
79. Mary sold the old **house/ shack** hidden in the shades of the nearby forest.
80. The traditional dish had the sweet **taste/ aroma** unique to the food in that region.
81. Karen read a book on how the concept of **faith/ karma** evolved throughout the centuries.
82. The documentary showed a series of **crime/ heist** cases that still remain unsolved.
83. For the upcoming performance, the **stage/ foyer** inside the theatre had to be decorated.
84. Not long after the **fight/ brawl** escalated in the bar, the police asked everybody to leave.
85. The teacher used a **glass/ prism** coated in black to demonstrate the properties of light.

86. The large production of **sugar/ maize** proved to be a big boost for the economy of the country.
87. The maintenance of the **water/ sewer** pipe made it necessary to close the whole road.
88. The manager asked the **staff/ clerk** handling the finances to be careful with the paperwork.
89. None of the trainees could work at the **speed/ tempo** set by their supervisor.
90. The lack of **rain/ haze** outside made it possible to conduct the field experiment.
91. Because of problems with the **line/ cord** last night, the telephone could not be used at all.
92. Megan prepared the **cheese/ lentil** soup that her grandmother taught her.
93. The budget included money for a **system/ server** upgrade that would solve the current issues.
94. The discovery of the missing **record/ folder** shed new light on the police investigation.
95. The professor explained that the **picture/ proverb** dates back to at least 500 years.
96. The new theory led to a **problem/ paradox** which was seemingly very difficult to solve.
97. The company developed a new type of hair **oil/ gel** intended for everyday use.
98. Because the man was sleeping, his **wife/ maid** answered the phone.
99. The instructor explained that the **mistake/ theorem** found in the textbook should be ignored.
100. Vanessa realised she had forgotten her **money/ purse** when she was in front of the supermarket.
101. In recent years, there are more students interested in **sport/ chess** events in the city.
102. The restaurant offered a **fruit/ mango** pudding as a desert of the day.
103. Last year, a law was enacted to help preserve the **fish/ swan** numbers in the wilderness.
104. The patient reported feeling a strange sensation in her **foot/ heel** prior to her admission.
105. The rules of the game are that if a **card/ dice** falls on the ground, the player skips a turn.

106. Historians believe that the ancient **city/ tomb** served as a tribute to the Roman gods.
107. The dog barked at the **bird/ crow** nesting in the branches of the tree.
108. The new book introduced readers to the **kitchen/ cuisine** typical for Mediterranean countries.
109. The investigation showed that the **light/ stove** started the fire in the kitchen.
110. The young family was looking to buy a new **table/ couch** suitable for their living room.
111. The businesswoman picked her **lunch/ scarf** lying on the table and headed towards the elevator.
112. The candidate wanted to appeal to every **media/ voter** following the election campaign.
113. Wearing floral clothes became the latest **style/ craze** quite quickly in some countries.
114. Margaret benefited a lot from the advice of her **father/ mentor** during her school years.
115. The television didn't believe that the new **show/ duet** idea would attract a lot of viewers.
116. The military leader examined the **weather/ terrain** map before he planned the operation.
117. The author's book lacked the **quality/ clarity** needed to get a good contract.
118. During the weekend, Jenifer removed all the **clothes/ clutter** scattered in her attic.
119. Only the fearless **captain/ swimmer** dared to go out in the sea when a storm was coming.
120. Sally bought flowers for the **bedroom/ doorway** upon moving into her new house.
121. In the library, the **history/ archive** section was closed due to the ongoing renovation works.
122. Some viewers didn't like the boring **message/ trailer** shown before the actual movie.
123. The new vaccine has wide implications for **society/ mankind** because of its high success rate.
124. After finishing his **meeting/ lecture** early in the morning, the student was free to go home.

125. Located next to the river, the historic **village/ mansion** offered a picturesque view of the area.
126. The maintenance of the impressive **garden/ facade** required a substantial amount of money.
127. To ensure the quality of bulbs, the company tests every **piece/ batch** produced in the factory.
128. After the house was fully cleaned, the **floor/ porch** shined in the sunlight.

Appendix E

Supplementary analyses from Chapter 3.

Chapter 3: Post-hoc Analysis of All Words in the Sentence

One possible explanation for the lack of a significant interaction between background sound and lexical frequency in Chapter 3 could be a lack of statistical power. In order to test this possibility, frequency norms were obtained for all words in the sentence and these were entered into a model for which we used the entire eye movement data set. This analysis provided greater statistical power because all of the words in the sentence are analysed, not just the target word. The breakdown of the whole data by fixation duration measure is presented in Figure A5 and the results from the LMMs are shown in Table A5. Consistent with the target word analysis, lexical frequency failed to interact with the contrast between English speech and the remaining background sound conditions. Additionally, TVT was significantly longer in English speech compared to all other sound conditions (Silence: d= -0.15; Noise: d= -0.17; Mandarin: d= -0.10), which also replicates the results from the target word analysis. However, similar to the target word analysis, there were no effects of English speech on first-pass reading measures (FFD or GD). There were also no statistically significant differences between the Mandarin and Noise condition for any of the measures (all ps $\geq$ .07).

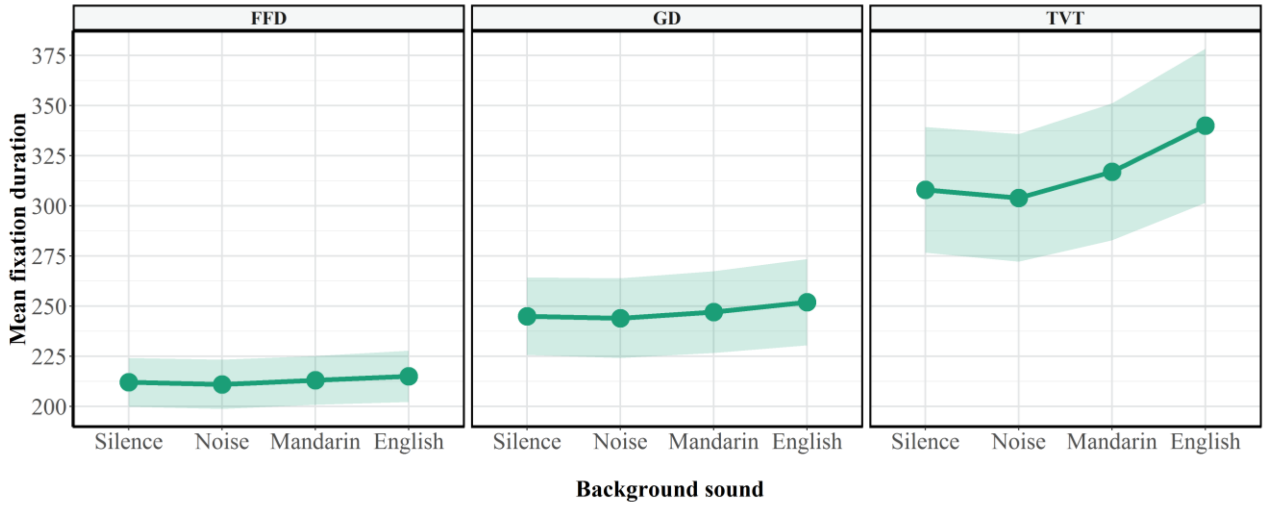

Figure A5. Mean fixation time measures on all words in the sentence in Chapter 3 for each of the background sound conditions. FFD: first fixation duration. GD: gaze duration. TVT: total viewing time. Shading shows standard error.

Fixed effect	FFD				GD				TVT[1]			
	b	SE	t	p	b	SE	t	p	b	SE	t	p
Intercept	5.3	.02	324.7	**<.001**	5.4	.02	240.2	**<.001**	5.6	.03	167.2	**<.001**
Freq	-.04	<.01	-13.14	**<.001**	-.09	<.01	-21.19	**<.001**	-.15	<.01	-29.70	**<.001**
Eng vs Slc	-.01	.01	-1.58	.24	-.02	.01	-1.98	.11	-.06	.02	-3.71	**.001**
Eng vs Noise	-.01	.01	-1.59	.24	-.02	.01	-2.15	.08	-.09	.02	-5.74	**<.001**
Eng vs Mnd	<-.01	.01	-.71	.70	-.01	.01	-1.40	.34	-.05	.01	-3.76	**.001**
Freq: Eng vs Slc	-.01	<.01	-1.69	.18	-.01	.01	-1.07	.57	.01	.01	1.75	.16
Freq: Eng vs Noise	-.01	<.01	-1.15	.50	<-.01	.01	-.17	1	.01	.01	1.23	.44
Freq: Eng vs Mnd	<-.01	<.01	-.82	.82	<-.01	.01	-.63	1	.01	.01	1.11	.54

Table A5. LMM results from the post-hoc analyses with word frequency in Chapter 3. Freq: Lexical frequency. Eng: English. Slc: Silence. Mnd: Mandarin. FFD: first fixation duration. GD: gaze duration. TVT: total viewing time. Statistically significant *p*-values are formatted in bold.

[1] Background sound was removed as a random slope for items due to convergence issues (intercept was retained).

Chapter 3: Sensitivity Analysis of the Bayes Factor T-tests on Comprehension Accuracy

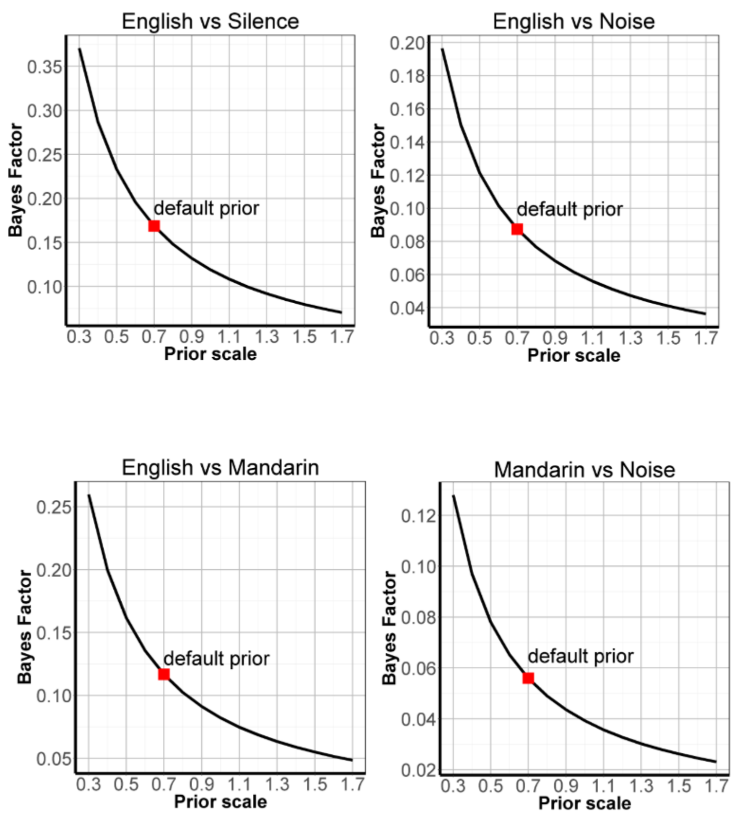

Figure A6. Sensitivity analysis of the computed Bayes Factor from Chapter 3 with a range of realistic priors. Bayes Factors greater than 1 indicate evidence for the alternative hypothesis, while Bayes Factors smaller than 1 indicate evidence in support for the null hypothesis. Red square shows the default prior that was used in the analysis. The figures show that the results are robust and were not influenced by the default width of the Cauchy prior distribution that was chosen ($r = \sqrt{2}/2$).

Appendix F

Supplementary analyses from Chapter 4.

Chapter 4: Sensitivity Analysis of the Bayes Factors on Comprehension Accuracy

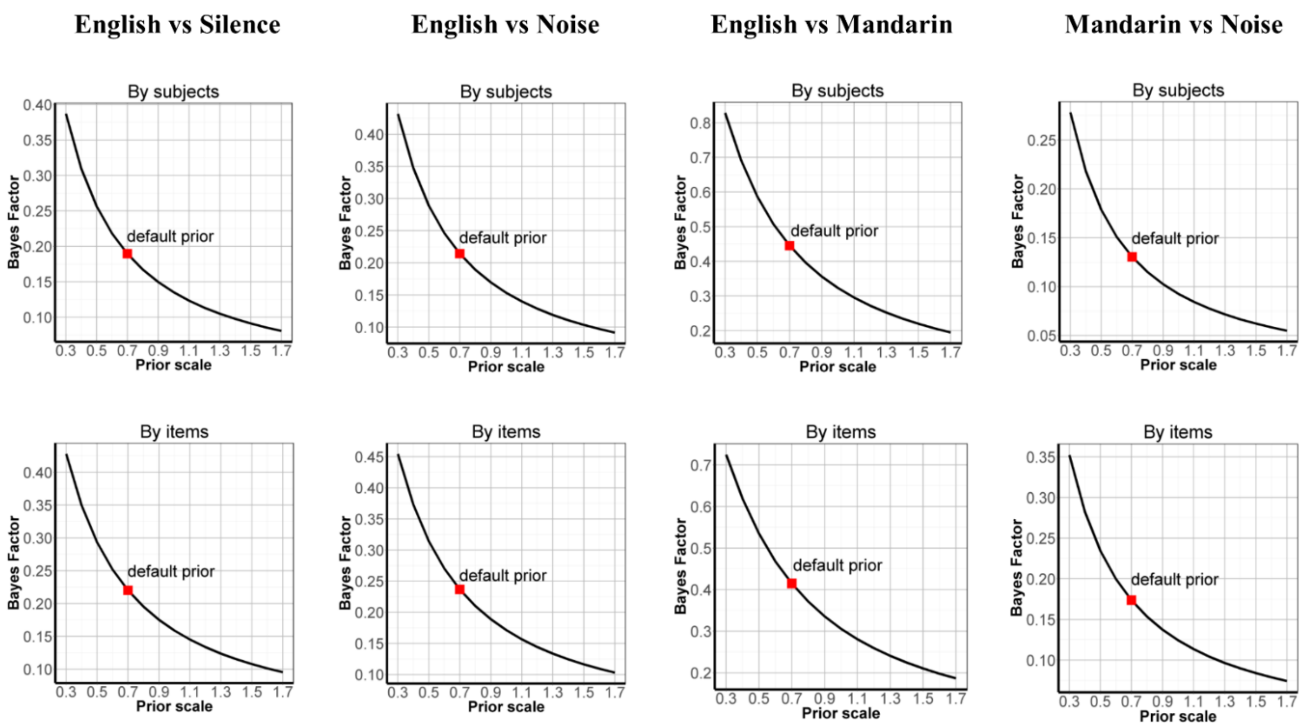

Figure A7. Sensitivity analysis of the computed Bayes Factor from Chapter 4 with a range of realistic priors. Bayes Factors greater than 1 indicate evidence for the alternative hypothesis, while Bayes Factors smaller than 1 indicate evidence in support for the null hypothesis. Red square shows the default prior that was used in the analysis. The figures show that the results are robust and were not influenced by the default width of the Cauchy prior distribution that was chosen ($r = \sqrt{2}/2$).

Appendix G

Supplementary analyses from Chapter 5.

Chapter 5: Sensitivity Analysis of Bayes Factor Tests of Comprehension Accuracy

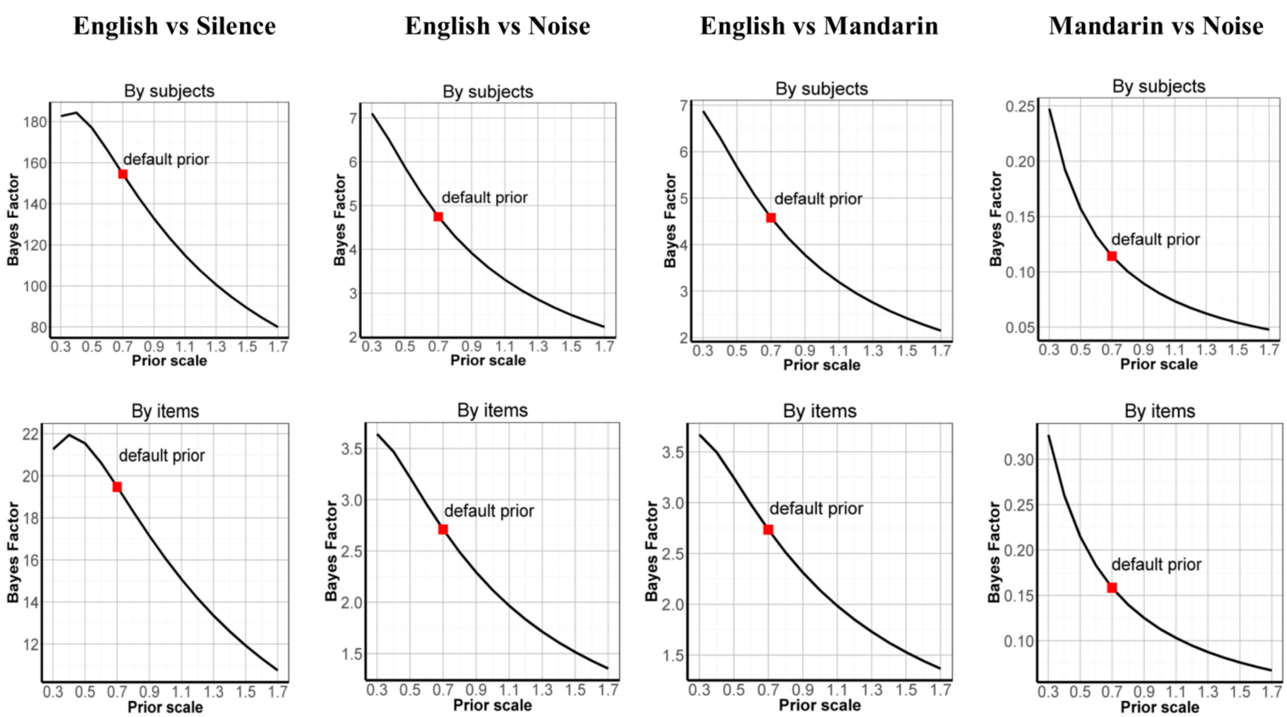

Figure A8. Sensitivity analysis of the Bayes Factor regression analyses comparing comprehension accuracy across the sound conditions in Chapter 5, Experiment 1. The results show that using a range of realistic prior distributions does not change the results from the analysis. Therefore, the conclusions are not influenced by the width of the chosen Cauchy prior distribution ($r = \sqrt{2}/2$).

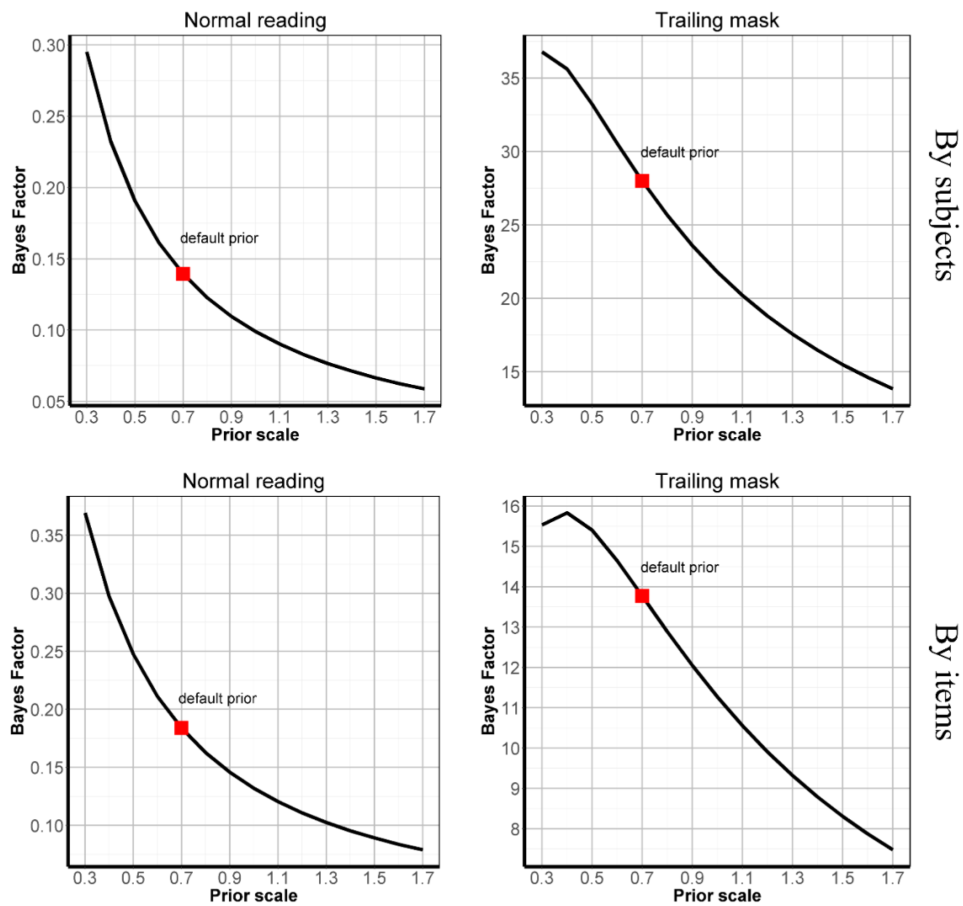

Figure A9. Sensitivity analysis of the Bayes Factor model comparing comprehension accuracy between English speech and Silence as a function of reading condition (normal vs trailing mask) in Chapter 5, Experiment 2. The results show that using a range of realistic prior distributions does not change the results from the analysis. Therefore, the conclusions are not influenced by the width of the chosen Cauchy prior distribution ($r = \sqrt{2}/2$).

Appendix H

Stimuli sentences used in Chapter 6. The words in bold denote the target words on which the sounds were played in the sentence.

1) The couple **bought** another **carpet** for **their** living **room** before **returning** home from shopping.
2) Theodore eventually **found** the **right** road **after** getting **lost** twice **driving** his car in the area.
3) The teenagers **used** the **private** area **behind** the **apartments** to **practise** playing basketball.
4) Susan contemplated **buying** the **antique** painting **when** she **visited** the **auction** house in the capital.
5) Cathy was **nervous** about **giving** the **short** speech **before** the **panel** of judges.
6) The midwife **admired** the **serene** baby **following** the **difficult** birth **earlier** in the day.
7) The cook **ordered** the **fresh** organic **fruits** to **support** the **farmers** who **produce** them locally.
8) The mayor **stepped** down **after** his **insensitive** comments **created** a **social** media scandal.
9) Yesterday, Nicole **made** a **brief** comment **concerning** her **manager** that **caused** a scandal.
10) Without her **glasses**, Margaret **struggled** to **read** the **road** signs **during** her driving test.
11) The workers **replaced** the **faulty** bathtub **which** had **caused** the **entire** first **floor** to be flooded.
12) The clever **merchant** invested **money** in **real** estate **during** the **financial** crisis that shook the market.
13) Katy remained **positive,** despite **facing** financial **problems** that **threatened** to **ruin** her business.
14) Carmen couldn't **find** her **yellow** shirt **when** she **unpacked** her **clothes** from the trip.
15) His mother **purchased** the **crimson** cottage **after** going **through** many **online** ads for a summer home.
16) When the **issue** was **discovered**, the **employee** unwisely **took** the **blame** for the incident.
17) The officer **started** his **investigation** by **questioning** any **witnesses** who **might** have seen the perpetrator.
18) For the **position**, only **competent** and **experienced** candidates **would** be **invited** for an interview.
19) The young **author** didn’t **realise** that **writing** her **article** would **cause** so much controversy.
20) The new **hotel** had **impressive** dining **halls** that **featured** expensive **pieces** of art.

21) The loud **construction** noise **annoyed** many **residents,** who **preferred** to **spend** a quiet afternoon at home.
22) The lawyers **criticised** the **latest** report **because** it **provided** no **useful** evidence for the case.
23) The forecast **warned** about **strong** winds, **accompanied** by **occasional** rain, **throughout** most of the day.
24) They debated **whether** using **better** surveillance **equipment** would **deter** potential **thieves** and improve security.
25) The child **pestered** the **yellow** fish **that** was **hiding** behind **pondweed** leaves in the aquarium.
26) The witness **remembered** very **clearly** the **brown** leather **jacket** that **belonged** to the suspect.
27) The detective **noted** the **evasive** statement **that** Sam **gave** to **explain** his suspicious behaviour.
28) The waitress **assured** the **customer** that **every** food **item** was **prepared** from fresh ingredients.
29) The players **remained** focused **despite** initially **losing** the **first** three **rounds** of the game.
30) The student **lost** points **because** she **used** faulty **reasoning** to **explain** the concept.
31) The increasing **number** of **cyber** crime **offences** has **prompted** discussions **about** tougher laws.
32) The residents’ **peace** was **disturbed** by **thousands** of **visitors** who **flocked** to the town centre.
33) The technician **provided** expert **knowledge** that **helped** locate **suspects** in **connection** with the robbery.
34) The new **street** was **designed** to **reduce** traffic **congestion** and **improve** safety for all road users.
35) The old **lady** needed **somebody** to **help** her **clean** the **spacious** three-bedroom flat.
36) The small **hotel** was **known** for **providing** cheap **accommodation** and **good** quality service.
37) The candidate **wanted** to **appeal** to **every** media **following** the **election** campaign this year.
38) The general **examined** the **evidence** of **military** drills **conducted** in **secret** by the enemy.
39) The central **bank** published **revised** regulations **regarding** foreign **currency** and **money** transfers from abroad.
40) The landlord **made** the **decision** to **sell** his **property** and **retire** to the countryside.
41) The brave **little** boy **tried** to **rescue** the **dog** which **couldn’t** get out of the river.
42) Surprisingly, the **trip** was **cancelled** after **reports** of **unrest** caused **concern** among the organisers.
43) Some local **farmers** received **money** from **government** programmes **because** they **wanted** to buy more land.
44) The enormous **cave** is **largely** unexplored **because** its **remote** location **makes** it difficult to reach.
45) Melissa decided **against** going **home** after **work** and, **instead,** took **salsa** lessons with her friends.

46) To Jane's **surprise**, the **flower** somehow **managed** to **survive** the **blazing** heat with very little water.
47) The taxi **driver** had **trouble** finding **customers** and **decided** to **call** it a day.
48) The doctor **examined** the **patient** who **complained** of **constant** headaches **during** the past week.
49) The scientists **investigated** the **bright** energy **bursts** that **appeared** to **come** from outer space.
50) Because the **road** looked **steep** and **difficult** to **climb**, the **cyclists** pushed their bikes.
51) The president **wants** to **stay** in **politics** after **completing** his **term** later this year.
52) Kristin was **surprised** to **receive** an **honourable** mention **during** the **prize** award ceremony.
53) The car **engine** problems **started** because **someone** had **damaged** a **valve** during the maintenance.
54) Hannah enjoyed **visiting** the **Museum** of **Modern** Art **during** her **holiday** in the USA.
55) The beautiful **island** was **unknown** to **most** people **visiting** the **tropical** country due to its small size.
56) The start-up **company** scored **record** high **sales** soon **after** it **opened** for business.
57) The reclusive **monastery** had **hardly** any **contact** with **outsiders** and **maintained** a simple way of life.
58) The students **learned** how **computer** programs **could** soon **assist** medical **staff** with making diagnoses.
59) The water **supply** was **interrupted** while **works** were **underway** to **replace** an old pipe.
60) The documentary **showed** a **series** of **crime** cases **which** remain **unsolved** until today.
61) The mountain **trail** was **steep** and **difficult** to **navigate** without **proper** climbing gear.
62) John always **admired** the **courage** his **brother** showed **when** facing **difficult** situations in life.
63) The admiral **ordered** his **troops** to **investigate** the **strange** signal **detected** from the seabed.
64) Many endangered **species** will **receive** much **needed** protection **once** the **natural** reserve is open.
65) The flight **attendant** was **offered** her **position** after **completing** a **long** training programme.
66) The client **asked** for **another** evening **dress** that **matched** her **velvet** boutique shoes.
67) The government **tried** to **support** the **interest** in **solar** power **technology** in the private sector.
68) The buffet **offered** many **different** food **options,** which **included** both **continental** and British meals.
69) Because the **weather** was **cold** and **windy**, the **children** weren't **allowed** to play outside.
70) The landlady **asked** the **tenants** to **keep** the **house** clean **during** their stay there.

71) The couple **celebrated** their **tenth** wedding **anniversary** by **visiting** the **place** where they first met.
72) The new **computer** factory **promised** to **create** more **jobs** and **attract** specialists from nearby towns.
73) The charity **tried** to **raise** more **money** for **improving** healthcare **facilities** in developing countries.
74) The supermarket **chain** had **ambitions** to **expand** their **operation** to **nearby** countries in the future.
75) The club **aimed** to **foster** cultural **awareness** and **friendship** among **people** of different nations.
76) The strong **currency** attracted **investments** from **foreign** companies **interested** in **expanding** their business.
77) The young **mother** was **tired** of **seeing** her **children** play **pranks** on each other all the time.
78) The campers **realised** that **forested** areas **offer** better **protection** from **wild** animals and mosquitoes.
79) The severe **drought** affected **many** farmers **despite** the **water** irrigation **systems** that were installed.
80) The heavy **rain** storm **forced** some **businesses** in **town** to **remain** closed after the streets were flooded.
81) The poet **spent** the **whole** summer **writing** short **stories** for **another** book project.
82) The mission **failed** to **uncover** the **enormous** pirate **treasure,** which **some** believe is buried in the ocean.
83) The journalist **finished** her **assignment** and **prepared** several **good** questions **before** the interview.
84) The forest **fire** spread **quickly** and **gave** the **firefighters** little **time** to bring it under control.
85) The reporters **covering** the **case** were **barred** from **entering** the **courtroom** during the first hearing.
86) The old **mine** was **finally** shut **down** after **decades** of **digging** had completely exhausted it.
87) The mathematical **proof** was **clever,** but **difficult** to **understand** for **most** of the scientists.
88) The school **children** enjoyed **visiting** the **Museum** of **Natural** History **during** their trip to the capital.
89) The couple **ordered** a **double** cheese **pizza** after **deciding** to **spend** the night at home.
90) Mary always **devoted** some **time** to **helping** her **parents** with **household** chores during the weekend.
91) The paper **factory** was **temporarily** closed **following** several **health** and **safety** violations this year.
92) During the **latest** press **conference,** the **mayor** revealed **plans** to **build** a cinema complex.
93) The seismic **activity** worried **government** officials **because** it **signalled** a **possible** volcanic eruption.

94) The new **church** was **built** with **donations** from **companies** and **members** of the public.
95) The green **park** attracted **many** city **dwellers,** who **wanted** to **enjoy** the summer days outside.
96) Plans to **build** the **final** underground **station** were **scrapped** after **historical** ruins were unearthed.
97) The instructor **demonstrated** the **technique** for **passing** the **ball** to **another** player on the field.
98) The art **gallery** was **recently** renovated **thanks** to **funding** from **wealthy** benefactors from the region.
99) The old **gravel** road **outside** of **town** was **perfect** for **mountain** bike competitions.
100) The nurse **conducted** the **required** tests **diligently** and **informed** the **doctor** about the results.
101) The children **practised** their **lines** regularly **before** the **school** play **which** had been organised for their parents.
102) The Arctic **outpost** was **impossible** to **reach** without **helicopters** that **could** fly there.
103) The deep **canyon** was **formed** over **hundreds** of **thousands** of **years** due to erosion of the rocks.
104) The peaceful **protest** quickly **turned** violent **after** two **groups** of **demonstrators** started fighting.
105) The beautiful **lake** was **popular** among **hikers,** who **would** usually **camp** there in the evening.
106) The music **festival** was **overpriced** but **many** fans **still** bought **tickets** to attend the event.
107) The multinational **company** was **convicted** of **insurance** fraud **after** secret **documents** were leaked to the press.
108) The light **breeze** outside **made** it **easier** to **bear** the **summer** heat during the day.
109) The local **florist** had **decades** of **experience** in **making** quick **deliveries** for different celebrations.
110) The new **book** was **published** in **electronic** format **because** most **readers** use mobile devices.
111) The principal **changed** his **clothes** after **large** coffee **stains** completely **ruined** his outfit.
112) The old **library** was **moved** to **another** building **because** space **restrictions** made it hard to store new books.
113) The annual **swimming** competition **became** a **major** success **when** one **thousand** people took part.
114) The ancient **forest** was **closely** protected **because** it **contained** the **oldest** trees on Earth.
115) Surprisingly, the **movie** sequel **received** better **reviews** from **film** critics **compared** to the original.
116) The small **archipelago** had **long** produced **various** exotic **spices** and **exported** them abroad.
117) The precious **stone** was **found** by **children** playing **near** the **rapid** mountain stream.

118) The remote **village** was **long** abandoned **because** all **residents** had **moved** to nearby towns.
119) The tourist **town** was **mostly** deserted **during** the **long** winter **months** when businesses closed down.
120) Many citizens **opposed** the **plans** to **invest** in **nuclear** power **plants** due to their environmental impact.

Appendix I

Supplementary analyses from Chapter 6.

Fixation Durations on the Next Word

One interesting question is whether the deviant sound affected fixation durations not only on the currently fixated word, but also on the next word in the sentence. For example, Inhoff, Connie, Eiter, Radach, and Heller (2004) reported that a phonologically similar spoken word presented at the start of fixation on a target word also led to longer gaze durations on the post-target word compared to both a phonologically unrelated word and a spoken version of the target word itself. Therefore, to determine whether the effect of deviant sounds was constrained only to the target word or also affected the post-target word, we analysed fixation durations on the next word in the sentence. The descriptive statistics are presented in Table A6 below. There were no significant differences between the standard and the deviant sound (all ps ≥ 0.75) or between the standard sound and the silence baseline (all ps ≥ 0.19) on any of the measures. Therefore, the results suggest that the effect of the deviant sound did not spill over to the next word in the sentence. This finding is consistent with the view that the deviant sound caused saccadic inhibition since this effect was constrained only to the word where the sound was first heard.

Sound type	FFD	SFD	GD	TVT
Silence	239 (92)	242 (92)	281 (132)	319 (184)
Standard	245 (93)	246 (94)	288 (166)	317 (203)
Deviant	241 (90)	241 (90)	284 (151)	315 (187)

Table A6. Mean fixation durations on the next word in the sentence after playing the sound in Chapter 6 (SDs in parenthesis).

Comparison between Slow and Fast Readers

To determine whether the magnitude of the sound deviance effect was modulated by the reading speed of participants, a separate analysis was carried out in which participants were divided into two equal groups of slow and fast readers. This was done by first calculating the average reading speed for each participant, and then using a median split (equal to a reading speed of 213.5 words per minute) to divide participants into "fast" and "slow" readers. The descriptive statistics are shown in Figure A10 and the LMM analyses are shown in Table A7. As expected, slow readers had longer fixation durations than fast readers, and this difference was statistically significant in all measures. However, the deviant sound effect failed to interact with reader type for any of the measures. This suggests that the magnitude of deviance distraction was not modulated by whether participants were fast or slow readers.

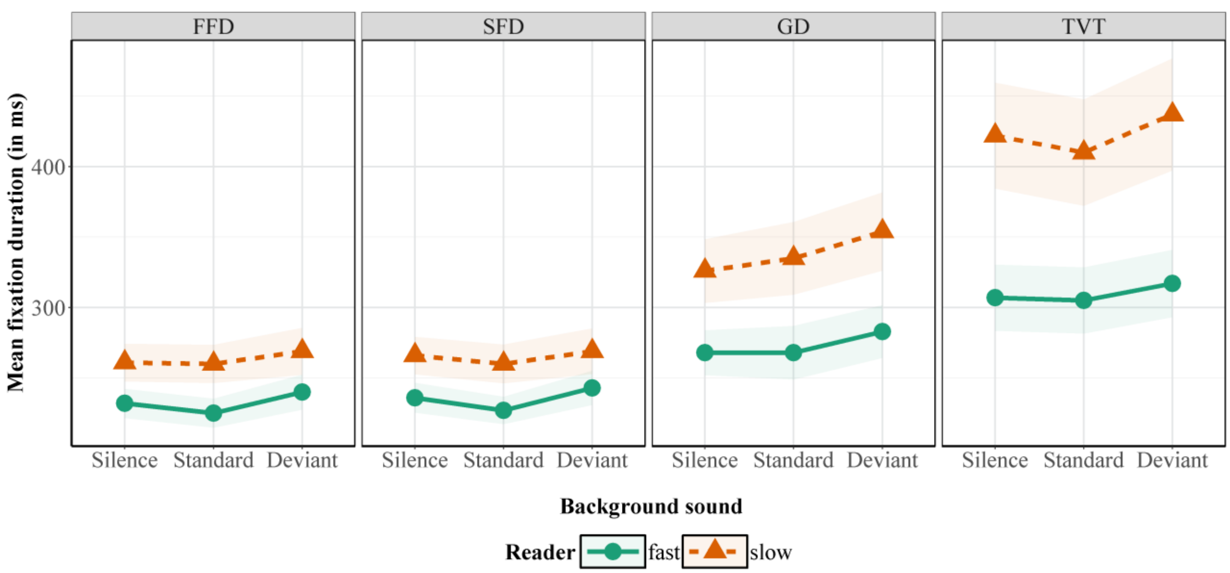

Figure A10. Mean fixation durations on the target words for the three sound conditions in Chapter 6, broken down by reader type (fast vs slow). Participants were divided into fast and slow readers based on a median split of their reading speed (median= 213.5 words per minute). Shading indicates the standard error.

Effect	FFD				SFD			
	b	SE	t	p	b	SE	t	p
Intercept	5.4	.02	281.1	**<.001**	5.4	.02	270.9	**<.001**
Reader	.07	.02	3.40	**.001**	.06	.02	3.00	**.004**
Deviant	.03	.02	2.02	**.05**	.04	.02	2.37	**.02**
Standard	.02	.01	1.34	.18	.03	.02	1.85	.07
Reader: Deviant	-.02	.02	-1.48	.14	-.02	.02	-.98	.33
Reader: Standard	-.01	.01	-.41	.68	<.01	.02	.06	.95

Effect	GD				TVT			
	b	SE	t	p	b	SE	t	p
Intercept	5.6	.03	218.1	**<.001**	5.7	.03	199.1	**<.001**
Reader	.10	.02	4.04	**<.001**	.13	.03	4.65	**<.001**
Deviant	.05	.02	2.76	**.008**	.04	.02	2.25	**.03**
Standard	<.01	.02	.20	.84	.01	.02	.58	.56
Reader: Deviant	<.01	.02	-.16	.87	.01	.02	.52	.60
Reader: Standard	-.01	.02	-.36	.71	.01	.02	.56	.57

Table A7. Interactions between sound condition and reader type (fast vs slow) in fixation durations on the target words in Chapter 6. Statistically significant *p*-values are formatted in bold.