Seeing Listening in the Eyes: Examining the Effects of Light Level and Fatigue on Pupil Dilation During a Speech-in-Noise Task

by

Jennifer Baldock

BBehavSc (Psych), BHlthSc (Hons)

Thesis
Submitted to Flinders University
for the degree of

Doctor of Philosophy

College of Nursing and Health Science
20 March 2022

TABLE OF CONTENTS

ABSTRACT

Listening effort can be defined as the cognitive processing that is required to attend to, and understand, an auditory message. Challenging listening conditions are encountered frequently throughout the average day and can increase the listening effort that is required of individuals, particularly for those with hearing loss. Excessive listening effort can lead to fatigue which may lead to communication breakdown and withdrawal, and can have negative consequences for occupational, social, and psychological wellbeing. To date, there is no clinically available measure of task-specific listening effort.

Pupil dilation is one physiological measure of listening effort that has received significant attention in research settings. This has resulted in a substantial body of literature describing task-evoked pupil response (TEPR) outcomes during auditory tasks, for individuals with and without hearing loss. It might be practical to use this measure in clinical audiology. This thesis provides an original contribution to knowledge regarding the use of TEPRs to measure listening effort and may have implications for continuing research in the area as well as potential clinical applications of TEPRs.

The empirical studies that make up this thesis were driven by gaps in the literature regarding two factors that may affect TEPRs during a speech-in-noise task: light level and fatigue. These factors were examined across three studies involving 36 participants.

Study 1 provides evidence that light level does affect TEPRs in a speech-in-noise task and that light level and task difficulty interact in their effect on TEPRs. A novel approach for using trial-level pupil data in the quantification of listening effort was also tested and described.

In Study 2, the potential mechanisms behind the relationship between light level and TEPRs were examined via mediation analyses. These analyses revealed that baseline pupil diameter partially mediated the relationship between light level and TEPRs by suppression (i.e., inconsistent mediation). This was the first study to provide evidence of this effect.

Study 3 provided preliminary evidence which suggested that there may have been an effect of time-on-task in the trial-level TEPRs across the test session and within each condition. This may be consistent with reduced physiological arousal because of fatigue

during the speech-in-noise task. This was supported by subjective fatigue and task-engagement reports but was not supported by performance measures.

TEPR-based measures of listening effort are a candidate for implementation in clinical audiology and may improve evaluation of clients' auditory difficulties and rehabilitative outcomes. While there are no current protocols for using TEPRs in clinical audiology, technological and statistical advances, and increased understanding of TEPRs may enable this implementation in future. The research reported in this thesis contributes to the growing body of work examining and enhancing understanding of TEPRs as a measure of listening effort.

DECLARATION

I certify that this thesis does not incorporate without acknowledgment any material previously submitted for a degree or diploma in any university; and that to the best of my knowledge and belief it does not contain any material previously published or written by another person except where due reference is made in the text.

Signed.. Jennifer Baldock

Date... 27/03/2022

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my primary supervisor, Dr Sarosh Kapadia. I am extremely grateful for the time that you have dedicated towards this research and me since I did Honours with you in 2015. There are countless ways in which you have helped my personal and professional development. I would not have been able to reach this stage without your supervisory support. Thank you for always being understanding and compassionate. Thank you for enabling me to explore multiple methods, avenues, and extracurricular activities. Thank you for always making time for discussions.

I would also like to express great thanks to my supervisory team, Dr Willem van Steenbrugge and Professor Jason McCarley who dedicated their time to provide support and valuable guidance.

I wish to acknowledge that this work was supported by an Australian Government Research Training Program Scholarship and the Ember Venning Post Graduate Research Scholarship in Speech Pathology and Audiology. Thank you to these funding sources who made this work possible.

Thank you to my supports in the College of Nursing and Health Science, including but not limited to: Brittany, Georgia, Shabs, Chelsea, Nicky, and Alison. Thank you to the other students from other colleges who have supported me, including but not limited to: Alycia, Marina, Cece, Lenore. Also, thanks to Dee, for swooping into the department during my final year and providing some much-needed friendship, laughs and guidance.

To Jess Young, my department PhD mate: Thank you so much for everything. You are a fantastic scholar and human.

Thanks also go to: The Office of Graduate Research Staff for being incredibly supportive and always having my back, the AIPDS interns for reminding me why we are all here, Leonie Randall for being such a massive help and a kind support (and for reminding me to go home when I was working too much), and Di Chamberlain and Ranjay Chakraborty.

I am incredibly appreciative of all the individuals who have helped me throughout this process. I have had people from different disciplines at my own University, to people from France, the Netherlands and America give me a portion of their time, just to help me understand something better. Some of these interactions have changed me for the better.

If, for some reason you are reading this, and you are someone who responded to me on Twitter, or via email when I asked a question – Thank you.

I wish to thank Dr Chelsea Parlett-Pelleriti for providing me with the occasional, judgement-free, statistical consults and for helping develop my knowledge of statistical inference and mixed-effects models. Additionally, I wish to thank Dr Matti Vuorre for guiding me through the application of Bayesian mediation analyses and providing me with helpful advice and direction.

I would also like to acknowledge that this work was completed on the lands of the Kaurna people and pay my respects to Elders; past, present, and emerging. Sovereignty was never ceded.

Finally, my biggest thanks go to my family and close support network.

To Jake, your excellence is unmatched! I am so fortunate to have had you by my side during this degree. Your support has been invaluable. Thank you for bringing me teas (or beers/wines) during those long days and nights on the computer. Thank you for always getting me to crack a smile and laugh when I was struggling. Thank you for reminding me that I was capable when I felt like I wasn't. Thank you for being you.

To Leila, thank you for also bringing me teas, being a wonderful friend, supporting me, and letting me whinge when I needed to.

Dad – where to even begin… I am not sure I would have finished this thesis without your calm and centred guidance. It has been an absolute privilege to have been able to learn from you in this way these past few years. Thank you for always being a phone call away when I needed a little sanity check and for always letting me use your computer to run my models -- I promise to get a computer with more cores now. Thank you for reading random things for me when I wanted to check if my interpretations were sound. Thank you for all the sage advice. I am a better scientist because of you.

Mum – also, where to even begin… You were (and are) a constant source of unwavering support which was so integral to getting this thesis finished. Thank you for making sure I took breaks and looked after myself. Thank you for always encouraging me and being my biggest cheerleader.

This thesis is dedicated to my Mum and Dad.

LIST OF PUBLICATIONS AND PRESENTATIONS

- Baldock, J., Kapadia, S., van Steenbrugge, W. & McCarley, J. (2019) The effect of light level of the task-evoked pupil response, Oral Presentation at Australian Society for Psychophysiology Conference, Wollongong, New South Wales, Australia.

- Baldock, J., Kapadia, S., van Steenbrugge, W. & McCarley, J. (2019) The effect of light level of the task-evoked pupil response during effortful listening. Poster presented at Australasian Cognitive Neuroscience Society Conference, Launceston, Tasmania, Australia.

- Baldock, J., Kapadia, S., van Steenbrugge, W. & McCarley, J. (2019) The effect of light level of the task-evoked pupil response during effortful listening. Poster presented at DOCFEST 2019, Flinders University, Adelaide, Australia.

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

AIC	Aikake Information Criterion
ANS	Autonomic Nervous System
BCI	Bayesian Credible Interval
BIC	Bayesian Information Criterion
BKB	Bamford-Kowal-Bench
CI	Confidence Interval
DTP	Dual Task Paradigms
EEG	Electroencephalogram
EMM	Estimated Marginal Means
fMRI	Functional Magnetic Resonance Imaging
FUEL	Framework for Understanding Effortful Listening
LC	Locus Coeruleus
LIV	Law of Initial Values
LL	Log Likelihood
LRT	Likeilhood Ratio Test
MAD	Median Absolute Deviation
MCAR	Missing Completely at Random Test
MEG	Magnetoencephalogram
MEM	Mixed-Effects Model
PCC	Participant/Condition Combination

PNS	Parasympathetic Nervous System
RANOVA	Repeated-Measures Analysis of Variance
SNR	Signal-to-Noise Ratio
SNS	Sympathetic Nervous System
TEPR	Task-Evoked Pupil Response

1 GENERAL INTRODUCTION

1.1 Statement of the Problem

People encounter sub-optimal listening conditions frequently throughout an average day. For example, at a dinner party, there will likely be various sources of noise besides an individual's immediate conversation of interest. This noise may comprise additional chatting, food-related sounds, music, laughing, and more. In situations like this, it can be more difficult to listen and may require more effort than usual to follow and contribute to a conversation. Situations like this are effortful and can lead to feelings of fatigue because individuals must allocate more cognitive resources, like attention processes, to the speaker of interest while allocating other cognitive resources, like active inhibition, to ignore the background noise from within a finite cognitive capacity (Pichora-Fuller et al., 2016).

Individuals with hearing loss typically need to expend even more effort to achieve their communication goals than individuals without hearing loss, especially in sub-optimal listening conditions. Even when sounds are understood accurately, the listening process is reported to be tiring and can result in diverse feelings of fatigue (Davis et al., 2021; Holman et al., 2019). Instances of excessive effort and fatigue during listening are frequently reported to audiologists (Pichora-Fuller et al., 2016). Despite this, there are currently no tools to objectively measure listening effort in clinical audiology.

Hearing aids may provide significant performance benefits through amplification and other technologies, for example, multiple-band wide dynamic range compression, directional microphones, and noise reduction schemes (Lunner et al., 2009, p. 395). Individualisation of hearing aid fitting is usually based on pure-tone audiometry. While pure-tone audiometry is valuable and important in the quantification of hearing loss, it cannot provide information about the effort an individual expends during listening.

Hearing aids will typically be fitted with similar settings for individuals who have the same audiogram, despite potential differences in underlying pathologies, and functional and cognitive abilities which could influence the benefits a user may gain (Lunner et al., 2009).

There is some evidence that hearing aids may reduce listening effort in their users (e.g., Fiedler et al., 2021); however, there are no clinical tools to evaluate their effectiveness in reducing listening effort in individual clients (Pichora-Fuller et al., 2016). A clinical measure of listening effort could help address the common complaint of excessive listening effort and fatigue, and thereby, could improve quality of life for individuals with hearing difficulties and hearing health care, in general.

Pupillometry, or measurement of the *task-evoked pupil response* (TEPR), has been used as a measure of mental effort since the 1960s (Hess & Polt, 1964). More recently, it has also been applied in hearing science as a measure of listening effort (Chapter 6). While pupillometry is a promising tool for the measurement of listening effort in research and potentially in clinics, there remains uncertainty about how it is affected by environmental and cognitive factors, such as light level and fatigue.

1.2 Thesis Aims and Significant Original Contribution to Knowledge

The central aim of this thesis is to examine how light level and fatigue affect TEPRs during a speech-in-noise task. The significant original contribution to knowledge emerges from the direct and comprehensive assessment of the effects of light level and time-on-task/fatigue on commonly measured TEPRs and in trial-level pupil data. This thesis provides the most comprehensive examination of the effects of light level and signal-to-noise ratio (SNR) on TEPRs during a speech-in-noise task to date. Evidence that light level affected TEPRs and that this effect was dependent on SNR was provided. Recent advances in statistical methods were employed to examine the novel use of trial-level pupil data in a listening effort paradigm. The findings reported support the use of trial-level pupil data in the current paradigm, provided appropriate statistical methods are employed. Additionally, the first evidence that baseline diameter mediates the relationship between light level and task-evoked pupil dilation was provided via novel application of mediation analyses. Relationships between performance and subjective fatigue and task-engagement were also examined and compared to the effect of time-on-task on trial-level TEPRs. Evidence of a time-on-task effect in trial-level pupil data was reported. These effects had not previously been established in trial-level pupil data. Furthermore, results suggested that baseline diameter may be most sensitive to time-on-task effects within-conditions compared to peak and mean dilation. Overall, the studies reported in this thesis contribute new knowledge to the study of TEPRs as a measure of listening effort.

1.3 Structure of Thesis

The first chapters describe the relevant background information pertaining to the process of hearing and hearing loss, followed by listening effort (Chapter 2) and listening fatigue (Chapter 3). The Framework for Understanding Effortful Listening (FUEL) is introduced here. Following this, a description of the various methods for measuring listening effort are described, ending with a detailed description of TEPRs (Chapter 4). The subsequent sections detail the physiological mechanisms that govern pupil size and pupil dilation (Chapter 5). The sections within this chapter include descriptions of the sympathetic and parasympathetic nervous system's contributions to TEPRs, the brain structures that may be involved in TEPR generation and introduces Adaptive Gain Theory – a theory which may be used to provide an explanation for the relationship between brain activity and TEPRs. The subsequent chapter describes the various aspects of pupillometry which may make it a viable option for a clinical measure of listening effort (Chapter 6). These sections lead into a detailed review of the literature regarding the effects of light level and fatigue on TEPRs and identify the gaps associated with these factors. The thesis aims, and general methods used are described in Chapter 7 and 8, respectively. A series of three studies based on the identified gaps follows (Chapters 9-14). After the study chapters, a comprehensive general discussion is provided (Chapter 15), which includes a discussion of the broad limitations of the thesis and future research directions that are warranted based on this work. This is followed by the general conclusion (Chapter 16).

2 HEARING AND LISTENING EFFORT

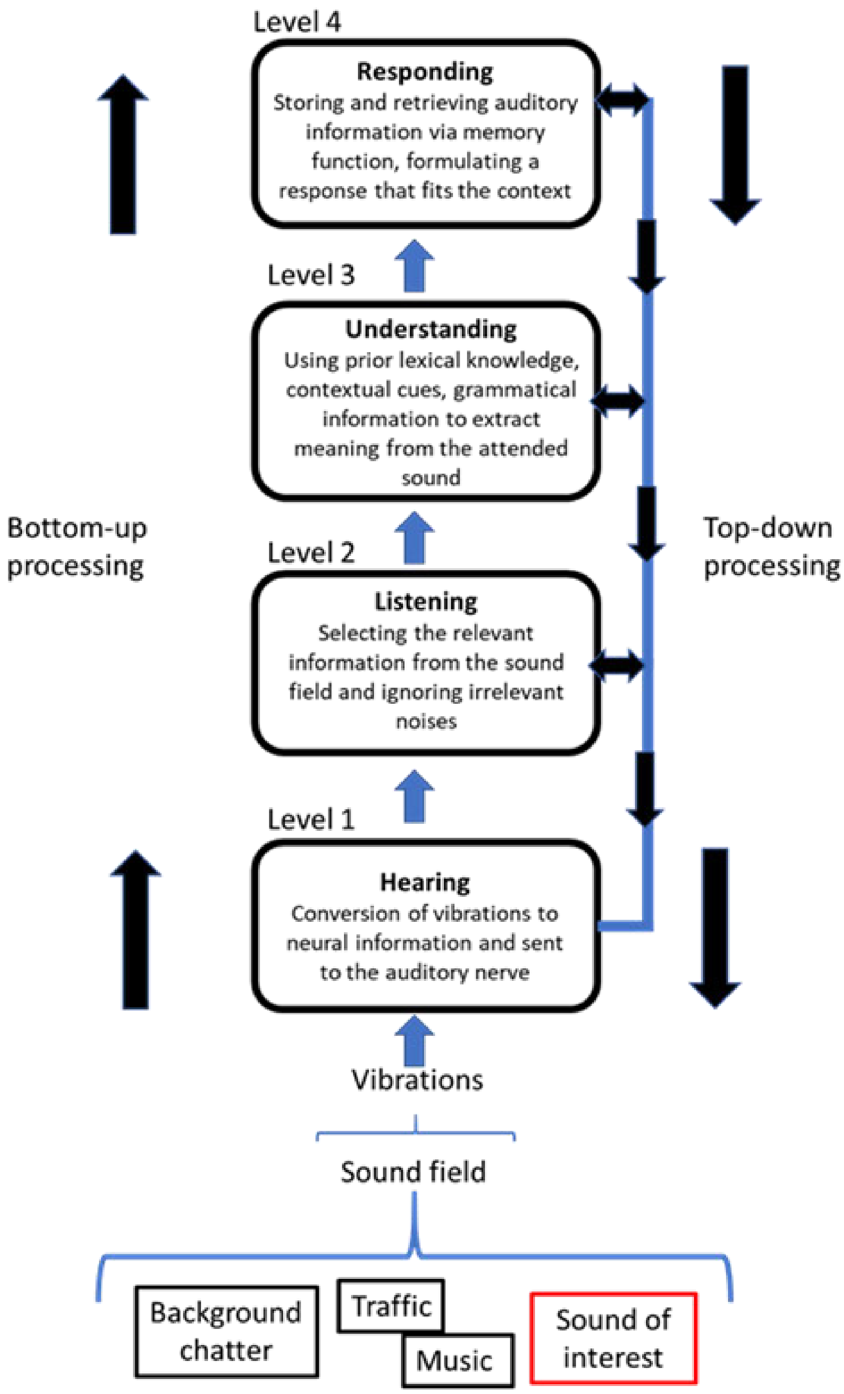
2.1 Hearing

Hearing, listening, understanding, and responding are interrelated functions that are necessary for successful, verbal communication. These functions are made possible by a combination of low-level processing in the cochlea to the cortex, as well as higher-level cognitive functioning such as attention, memory, and language (Edwards, 2007; Pichora-Fuller et al., 2016; Pichora-Fuller & Singh, 2006; Rönnberg et al., 2019). These low-level and higher-level functions are often referred to as bottom-up and top-down processing, respectively (Edwards, 2007; Stenfelt & Rönnberg, 2009).

Figure 1 depicts a model of a one-sided communication situation comprising a sound of interest, and background noise. At the lower-level (Level 1, Figure 1) (i.e., the starting point for bottom-up auditory processing), acoustic vibrations are received by the outer ear. Once these vibrations reach the inner ear, epithelial receptor cells equipped with mechanosensitive hair bundles are responsible for the conversion of the physical sound stimuli into electrical impulses via mechanoelectrical transduction (Hudspeth, 1997). These impulses move upwards in the auditory nerve and through multiple neural networks to the auditory cortex. Processing that moves in this upwards direction is often called “bottom-up” processing (Stenfelt & Rönnberg, 2009) (Figure 1). At this stage, the sound of interest and any audible background noise have been “translated” into neural information which is moving to the “listening” stage (Level 2, Figure 1) for selection of important information.

Figure 1

Depiction of the Interplay Between Bottom-Up and Top-Down Processing During Listening



Note. Adapted from "The signal-cognition interface: Interactions between degraded auditory signals and cognitive processes," by S. Stenfelt, and J. Rönnberg, 2009, *Scandinavian Journal of Psychology*, 50, p. 386. Copyright 2009 by the Scandinavian Psychological Associations.

In complex listening conditions (i.e., situations with more than one sound source, like the one depicted in Figure 1), the vibrations received by the ear represent an entire sound field, rather than individual soundwaves (Bregman, 1994). To selectively listen to one person’s voice in a sound field of many noises, the complex auditory scene must first be grouped into “streams”. This type of low-level processing was coined “auditory scene analysis” (Bregman, 1994). Auditory scene analysis involves segregating and/or integrating sound sources within mixtures of spectrally and temporally overlapping sounds to derive meaning.

Auditory stream segregation refers to the perceptual partitioning of multiple sounds into various streams (Paredes-Gallardo et al., 2018). Conversely, auditory stream integration refers to the perceptual combining of multiple sounds into a single stream. Initial processing by segregation or integration is based on the sound source’s physical similarities and temporal patterns. Segregation is likely when sounds are more distinct. Integration is likely when the sounds are more similar (Paredes-Gallardo et al., 2018). In addition to these low-level, stimulus-related influences, stream segregation/integration may be further facilitated by cognitive resources, like attention processes (Alain et al., 2001; Bregman, 1994; Paredes-Gallardo et al., 2018).

For example, Carlyon et al. (2001) found that when participants were instructed to selectively attend to a specific ear, they showed a reduced ability to separate distinct sounds over a period of time in the unattended ear (i.e., reduced “build-up” effect). In a recent review article, Sussman (2017) discussed multiple studies which showed that attention interacts with the automatic analysis of auditory scenes to facilitate task goals. Processing that moves in a downward direction and affects processing at lower levels is often called “top-down” processing (Stenfelt & Rönnberg, 2009) (Figure 1). Thus, there is a complex interplay between bottom-up and top-down processing, even at the lower-levels (Figure 1).

When operating optimally, precise information can be extracted from the various sounds present in a sound field by automatic analysis of the auditory scene, but with reference to specific communication goals. The use of communication goals represents a top-down cortical process which can influence the processing that occurs after (and during) auditory scene analysis by directing attention (Levels 3 and 4, Figure 1). Bottom-up and top-down processing of auditory information interact at different levels in the auditory system, depending on the listening task and the individual (Pichora-Fuller et al., 2016; Stenfelt & Rönnberg, 2009).

Individuals with hearing loss may experience difficulties with auditory scene analyses (Edwards, 2016; Middlebrooks & Simon, 2017). This is thought to be due to a reduced capacity to perceptually differentiate incoming auditory information, even when wearing hearing aids (Lesica, 2018). Additionally, individuals with hearing loss may need to deploy additional cognitive resources to achieve their communication goals, especially in complex listening situations (Pichora-Fuller et al., 2016). The following section will summarise hearing loss pathologies and associated consequences.

2.1.1 Hearing Loss

There are three types of hearing loss: (1) conductive hearing loss caused by the obstruction or difficulty with the passing of sound from the outer/middle ear (e.g., build-up of earwax, or otosclerosis), (2) sensorineural hearing loss caused by damage to the cochlea, hair cells or auditory nerve affecting the transmission of electrical information to brain, (3) mixed hearing loss, caused by co-occurring conductive and sensorineural hearing loss (American Speech-Language-Hearing Association, 2015). Individuals may also experience hearing difficulties without a hearing loss. Difficulty processing auditory information in the central auditory nervous system (auditory processing disorder) can result in listening difficulties even when there is no evidence of conductive or sensorineural hearing loss (Moore, 2015). Similarly, disruptions to and/or decline in an individual's cognitive ability may also result in difficulties with hearing (e.g., brain injuries, natural aging decline) (Pichora-Fuller et al., 2016).

According to the World Health Organization (2021) over 5% of the global population have a disabling hearing loss[1]. This prevalence estimate is expected to double by 2050 and does not include individuals with mild, non-disabling hearing loss even though these individuals may experience negative outcomes due to their hearing loss. One of the main impacts of hearing loss is on an individual's ability to communicate with others (World Health Organization, 2021). Hearing loss can directly affect one's speech recognition,

[1]Disabling hearing loss refers to hearing loss greater than 40 dB in the better hearing ear in adults (15 years or older) and greater than 30 dB in the better hearing ear in children (0 to 14 years)" (World Health Organisation, 2021).

communication, and language acquisition (Ohlenforst, Zekveld, Lunner, et al., 2017). It can also indirectly affect social, occupational, and psychological wellbeing (Pichora-Fuller et al., 2016).

It has been widely reported that listeners with hearing loss need to deploy more cognitive resources or *listening effort* to understand what is being said (McGarrigle et al., 2014). This is corroborated by the lived experiences of individuals with hearing loss, who frequently report that listening is tiring and excessively effortful to their audiologists (Pichora-Fuller et al., 2016). Due to the frequency of these reports, there has been a recent surge in interest among researchers and clinicians in the concept, definition, and measurement of listening effort. Understanding the challenges that individuals with hearing loss experience (e.g., excessive listening effort) is one way that hearing healthcare can be improved. Improvements in hearing health care are paramount for improving quality of life for individuals with hearing loss. The next section will introduce and define the concept of listening effort as set out by the Framework for Understanding Effortful Listening (FUEL) (Pichora-Fuller et al., 2016).

2.2 Listening Effort

Listening effort is a multidimensional phenomenon that can be defined as the cognitive processing required to attend to, and understand, an auditory message (McGarrigle et al., 2014). The FUEL defines listening effort as the “deliberate allocation of mental resources to overcome obstacles in goal pursuit when carrying out a [listening] task” (Pichora-Fuller et al., 2016, p. S10). This definition encapsulates the multidimensional nature of listening effort as it borrows from a variety of relevant theories and models, including: the Capacity Model of Attention (Kahneman, 1973), the Ease of Language Understanding Model (ELU) (Rönnberg et al., 2013; Rönnberg et al., 2008), Adaptive Gain Theory (Aston-Jones & Cohen, 2005), theories of fatigue (Hockey, 2013), and Motivational Intensity Theory (Brehm & Self, 1989). By combining these relevant and interrelated theories and models, the FUEL can provide a holistic framework from which to build an understanding of effortful listening.

Task-relevant and individual factors can influence the amount of listening effort that a person needs to expend in a listening task. Task-relevant factors that are external to the individual may include: background noise (Sarampalis et al., 2009), unrecognisable or degraded signals (e.g., cochlear implant transmission) (Wagner et al., 2016), accented speech (Van Engen & Peelle, 2014), and/or interference during sound transmission (e.g., reverberation) (Rennies et al., 2014). These factors may increase the cognitive demand of the task and therefore the amount of listening effort required to attend to and process auditory information. The FUEL also posits that individual factors such as hearing status (Ohlenforst, Zekveld, Lunner, et al., 2017), age (Gosselin & Gagné, 2011), motivation (Richter, 2016), level of fatigue (Hornsby, 2013), emotional state (Francis & Oliver, 2018), cognitive capacity (Rönnberg et al., 2011), and memory capacity (Rönnberg et al., 2014) can influence the amount of listening effort an individual expends in a task. One or a combination of these factors may represent "obstacles" that can influence what the individual's goal is and/or how much effort they are willing to expend to achieve this goal.

For example, a person may experience changes in their level of motivation to complete a task based on the amount of listening effort it requires. If the task is important to an individual (or has high reward), they may expend more listening effort to achieve their goal than an individual who decides that the task is not important and that the amount of listening effort needed for success is not "worth it". This notion was supported by Richter (2016) who examined the relationship between listening effort and task importance by measuring effort-related cardiovascular activity during an auditory discrimination task. Tasks with a higher reward (increased task importance) resulted in higher effort-related cardiovascular activity, indicating that more listening effort was expended in these tasks. More recently, Koelewijn et al. (2018) demonstrated that listening effort (as measured by effort-related pupil responses) was sensitive to monetary rewards, in a speech-in-noise task. Higher monetary reward resulted in larger pupil responses supporting the notion that motivation plays an important role in listening effort. However, these studies used participants without hearing loss (Koelewijn et al., 2018; Richter, 2016). Therefore, it is not yet known if this effect extends to individuals with hearing loss. An individual with hearing loss may have more obstacles to overcome and this may affect their motivation to complete listening tasks, even when rewards are available.

Listening has been shown to be disproportionately effortful for individuals with hearing loss and for those who show poorer auditory and cognitive processing (e.g., older adults, or individuals with brain injury) (Pichora-Fuller et al., 2016). The use of hearing aids or cochlear implants does not fully alleviate this (Alhanbali et al., 2017; Hornsby, 2013). In sub-optimal listening conditions, an individual who has hearing loss may abandon a task more quickly (i.e., withdraw effort) than someone without hearing loss. There is a greater cognitive burden placed on the individual with hearing loss to understand speech in noisy conditions and they may become fatigued more quickly (Bess & Hornsby, 2014; Edwards, 2007; Holman et al., 2019; Hornsby et al., 2016; Zekveld et al., 2011). This may lead to a reassessment of the importance of the task, a change in motivation and/or a decision to conserve cognitive energy, rather than to complete the task (Hockey, 2011, 2013; Richter, 2016). Such reassessments may lead to withdrawal and social isolation (Shukla et al., 2020).

There are many obstacles that can affect effort expenditure and hamper achieving a communication goal, for example, background noise, hearing loss, fatigue, and lack of motivation. Consistent effortful listening without adequate rest and recovery may exacerbate acute fatigue and may lead to a more chronic type of listening-related fatigue, which can have a variety of negative health and social consequences for individuals (McGarrigle et al., 2014; Pichora-Fuller et al., 2016; Wang, Naylor, et al., 2018). A clinical measure of listening effort may help mitigate the negative consequences associated with excessive listening effort and fatigue.

3 LISTENING-RELATED FATIGUE

Like listening effort, listening-related fatigue is a complex concept that is poorly understood (Hockey, 2013; Hornsby, 2013). Broadly, fatigue can be divided in two domains: physical and mental. Physical fatigue refers to tiredness of the body after heavy physical activity (van der Linden, 2011). Physical fatigue is beyond the scope of this thesis and thus, “fatigue” from here on refers to mental fatigue.

Mental fatigue refers to the effects of prolonged or intense periods of cognitive activity (Boksem & Tops, 2008). This type of fatigue is relatively common and may be experienced differently, depending on the person. It may be felt as tiredness, lethargy, low mood, exhaustion and/or unfocused and distractible mental states (Hockey, 2013; Hornsby et al., 2016). These feelings may lead to behavioural changes (like withdrawing effort) to avoid further fatigue and may manifest as performance decrements on tasks (Hornsby, 2013). Additionally, it could lead to reduced motivation during task performance.

Mental fatigue can be further categorised into two main types: chronic mental fatigue and acute mental fatigue. Chronic mental fatigue is characterised by persisting effects that may not be caused by cognitive activity but are often related to an underlying medical condition (e.g., chronic fatigue syndrome, depression, multiple sclerosis, or hearing loss) (van der Linden, 2011). Conversely, anyone may experience acute mental fatigue. This type of fatigue refers to transient reactions to periods of intense cognitive effort (van der Linden, 2011) such as, a challenging listening situation. Therefore, this type of fatigue may be characterised as task-induced fatigue.

As set out by the FUEL, to achieve communication goals (e.g., speech understanding) an individual must allocate cognitive resources to the task, which increases listening effort. While this can aid speech understanding, it may have consequences such as fatigue or exhaustion (Hornsby, 2013). The degree of interference mental fatigue imposes on an individual’s quality of life may vary with the severity and duration of the experienced fatigue, among other factors (e.g., hearing loss).

It is possible that consistent instances of intense listening effort can lead to subjective fatigue for individuals with hearing loss. When there is not adequate recovery, sustained and excessive listening effort may lead to a build-up of acute listening-related fatigue, which can lead to the more pervasive, chronic listening-related fatigue (Hornsby et al., 2016). This represents traditional fatigue theory, wherein work leads to energy depletion, as described in Rabinbach (1992).

Hockey (2013) argues that traditional fatigue theory is flawed. In traditional fatigue theory, it is suggested that more effort would lead to more fatigue (Rabinbach, 1992). However, evidence regarding a causal relationship between effort and fatigue is inconsistent. Because hearing loss may make listening tasks more effortful, one might expect that the degree of hearing loss would be related to the extent of fatigue reported. While individuals with hearing loss do typically report more fatigue than individuals without hearing loss, null findings regarding the relationship between the degree of hearing loss and fatigue have been reported (Alhanbali et al., 2017; Hornsby & Kipp, 2016). Hornsby and Kipp (2016) found that subjective fatigue and vigour were more strongly associated with *perceived* hearing handicap (i.e., psychosocial hearing difficulties) than with degree of hearing loss (measured by pure-tone audiometry). Correspondingly, Holman et al. (2019) reported that there was a link between fatigue and negative emotions related to having hearing loss. Therefore, there may be psychological factors besides effort expenditure that contribute to the development of listening-related fatigue.

The experience of fatigue may affect motivation in task performance. The Motivational Control Theory of Cognitive Fatigue (Hockey, 2011, 2013) posits that the experience of fatigue moderates an individual’s motivation to expend cognitive effort. Hockey (2011) and Hockey (2013) argued that subjective fatigue is an adaptive state and disagreed with the traditional view that fatigue is *just* a negative state. They argued that fatigue is functional and allows individuals to monitor their effort expenditure by evaluating constantly evolving effort-reward trade-offs. Motivation (and therefore, effort) may be reduced when the effort-reward trade off becomes unfavourable. When individuals continue to expend effort in the face of fatigue, they may experience negative consequences such as tiredness.

The effects of expending listening effort to the point of fatigue are far-reaching and significant. It may lead to withdrawal from communication tasks deemed to be “too difficult”, even when there are social and/or economic benefits. For example, participating in social gatherings may be excessively effortful for an individual with hearing loss due to the presence of multiple competing talkers and other environmental sounds. Situations such as this may be fatiguing and may result in embarrassment if the person cannot achieve their communication goals. Thus, they may decide to avoid these situations despite the social benefits. In turn, this avoidance can lead to feelings of isolation and loneliness (Shukla et al., 2020). Additionally, since hearing aids do not always completely resolve communication problems, individuals also may stop wearing prescribed hearing aids due to a lack of perceived benefits (Lesica, 2018; McCormack & Fortnum, 2013).

These short-term and long-term communication decisions (e.g., social avoidance and hearing aid rejection) which may be influenced by listening effort and listening-related fatigue can have adverse effects on psychosocial wellbeing (World Health Organization, 2021). For example, such decisions may lead to excessive stress, depression, chronic loneliness (Shukla et al., 2020) and/or anxiety (Brewster et al., 2018; Hetu et al., 1988; Jayakody et al., 2018; Nachtegaal, Kuik, et al., 2009; Nachtegaal, Smit, et al., 2009; Pronk et al., 2011; Saito et al., 2010). Furthermore, evidence suggests that daily-life functioning (Heyl & Wahl, 2012; Solheim et al., 2011), occupational wellbeing (Danermark & Gellerstedt, 2004; Kramer et al., 2006; Svinndal et al., 2018) and interpersonal/intimate relationships (Hetu et al., 1993) can also be affected by hearing loss.

Recent research has led to the development of a listening-specific fatigue scale (Vanderbilt Fatigue Scale for Adults) (Hornsby et al., 2021). This tool may be used to quantify listening-related fatigue in adults and may also be used to identify those at risk of severe listening related fatigue. However, there is a need for accurate measurement of listening effort in clinical audiology to address issues before listening-related fatigue negatively affects daily living. Measurement of listening effort in clinics could result in the adoption of strategies and/or technologies that may reduce listening effort and in turn, may reduce listening-related fatigue and mitigate associated effects. The following chapter describes multiple methods that can be used to measure listening effort.

4 MEASURING LISTENING EFFORT

Listening effort is usually assessed using: (1) subjective, (2) behavioural, or (3) physiological (and neuroimaging) methods (Pichora-Fuller et al., 2016). Figure 2 depicts a listening situation. Section A represents the auditory environment, and the relevant task and individual factors that may contribute to the cognitive demand of the task. Section B represents factors that influence the effort expended in listening. Section C shows how the effort expended in listening can be measured. The task and individual factors (Section A) as well as the cognitive factors (Section B) need to be considered when using the various measurement techniques (Section C). Techniques that are currently used to measure listening effort include:

1) Subjective measures: These allow researchers (and sometimes clinicians) to estimate a listener's self-perceived effort.

2) Behavioural or performance-based measures: These indicate how well a person can perform in certain situations, often using clinical tools.

3) Physiological measures: Physiological changes in the central and/or autonomic nervous systems can also be reflective of listening effort during task performance. These changes can be measured via neuroimaging methods, or by tracking physiological responses. These changes are often involuntary and are commonly measured concurrently with listening tasks. Therefore, they may provide an immediate picture of the processing that occurs during task performance, independent of behavioural and subjective outcomes.

These methods, their assumptions, and their application to listening effort research are discussed below.

Figure 2

Diagram of a Listening Situation

Figure removed due to copyright restriction

Figure is similar to original but contains "fatigue" and "emotional state" text boxes in the middle portion, an additional "subjective" measurement text box and fNIRs in the neuroimaging text box.

Note. This figure shows multiple environmental, individual, and cognitive factors that may affect listening effort and the outputs that can be used to measure listening effort. Adapted from “Listening effort: How the cognitive consequences of acoustic challenge are reflected in brain and behavior”, by J. E. Peelle, 2018, *Ear and Hearing*, 39, p. 205. Copyright 2017 The Authors.

4.1 Subjective Methods

Subjective measures have been used extensively to evaluate personal perceptions of listening effort (e.g., Alhanbali et al., 2017; Brons et al., 2014; Mackersie & Cones, 2011; McAuliffe et al., 2012; McGarrigle, Rakusen, et al., 2021; Nagle & Eadie, 2012; Panico & Healey, 2009; Picou et al., 2019; Picou et al., 2017; Picou & Ricketts, 2018; Picou et al., 2011; Rudner et al., 2012). These measures typically take the form of questionnaires or rating scales. An example of a questionnaire is “The Speech, Spatial and Qualities of Hearing Scale (SSQ)” (Gatehouse & Noble, 2004). This 49-item, multidimensional scale was designed to examine a variety of hearing difficulties in various real-world scenarios. It includes questions related to listening effort, such as: “Do you have to concentrate very much when listening to someone or something?”; “Can you easily ignore other sounds when trying to listen to something?”; and “Do you have to put in a lot of effort to hear what is being said in conversation with others?" (Gatehouse & Noble, 2004). Questions like these can provide valuable information about the personal experiences and feelings of an individual during listening, enabling the formation of individualised rehabilitation strategies based on their specific listening challenges. Additionally, these measures are usually quick and easy to administer as one-off measures.

The subjective nature of these methods may have limitations, though. Subjective measures are susceptible to self-report biases and interpersonal differences in effort thresholds and in the interpretation of what constitutes "effort" (e.g., what one individual deems effortful, may not be effortful to someone else) (McGarrigle et al., 2014). For example, Larsby et al. (2005) found that older individuals did not report more listening effort than younger individuals, even though they showed poorer performance when performing the same listening tasks. The finding that older individuals did not report more listening effort was unexpected as older people typically experience a decline in cognitive ability as they age and may need to expend more effort in listening (Pichora-Fuller & Souza, 2003). These factors may have been responsible for their poorer performance, but they did not lead to greater reports of subjective effort.

Moreover, subjective responses to questions about complex, multifaceted phenomena (e.g., effort) may be affected by heuristic response strategies. For example, as reported in Moore and Picou (2018), participants may substitute conceptually easier questions like "how well did I perform" for conceptually more difficult questions like "how much listening effort did I expend?". Therefore, participants may not directly answer the question under examination. However, this would not explain the results of Larsby et al. (2005), who found that despite different performance levels, older and younger people rated effort expenditure similarly. It is more likely that those results were due to differences in effort thresholds between the groups. Larsby et al. (2005) suggested their finding may reflect a reluctance to complain in the older group but this interpretation was not verified. Fundamentally, self-report biases may make subjective measures of listening effort unreliable.

When measuring listening effort in research settings, subjective measures are often used in combination with behavioural or physiological measures. However, results are rarely consistent across methodologies (Mackersie & Cones, 2011; Moore & Picou, 2018; Pichora-Fuller et al., 2016). This suggests that subjective measures of listening effort may not be reflecting the same underlying constructs as behavioural or physiological measures.

4.2 Behavioural and Performance-Based Methods

Difficulties associated with processing and understanding sounds and speech is a common aspect of hearing loss. Speech understanding is typically measured in clinical audiology by calculating the percentage of correctly identified material in a speech test. In a speech-in-noise task, the speech or background noise levels may be manipulated to examine if increased task difficulty leads to a performance decrement. These are valid tests for assessing the functional difficulties associated with hearing loss. However, performance scores do not reveal much about the allocation of cognitive resources in listening, nor do they necessarily reflect listening effort (Winn & Teece, 2021). In terms of listening effort, high performance could indicate that the task was relatively easy and required little effort. Alternatively, high performance could indicate that the task was difficult, but that the listener expended the effort needed to compensate for that difficulty.

Behavioural and performance-based methods of listening effort may aim to probe more than the number or percentage of stimuli a person understood. Based on Kahneman's (1973) theory that individuals have a limited cognitive capacity, behavioural performance on specific cognitive tasks can provide an indication of cognitive resource allocation during effortful listening. Typically, behavioural methods attempting to gauge listening effort will use tests that probe cognitive mechanisms like working memory, attention, or speed of processing (Pichora-Fuller et al., 2016).

Working memory involves the maintenance and manipulation of information and has a limited capacity (Baddeley, 2010). If part of a sentence is unclear, working memory resources may be allocated to derive meaning based on semantic context, while temporarily storing other essential elements of the sentence (Peelle, 2018). Working memory is regulated by attention (Gazzaley & Nobre, 2012; Sreenivasan & Jha, 2007). Attention is the allocation of resources to a single activity (selective attention) or multiple activities (divided attention). If the total (finite) cognitive resource capacity is exceeded, speed of processing slows down and/or errors are made (Pichora-Fuller et al., 2016). Testing these cognitive domains during listening tasks may be operationalised such that they can be used as measures of listening effort.

Measuring reaction/response times is one behavioural approach to quantifying listening effort (Gagné et al., 2017; Pals et al., 2015). Hecker et al. (1966) examined reaction time during a single speech intelligibility task. They found that as background noise became louder, making the speech less intelligible, reaction time increased, that is, speed of processing slowed. This finding has been consistently replicated in single task paradigms (Gustafson et al., 2014; Houben et al., 2013).

Measuring listening effort (via reaction time or percentage correct) during a single task is common, however, dual task paradigms (DTP) can also be used to measure listening effort and aim to draw on multiple cognitive domains simultaneously. In DTPs, participants perform a primary task and secondary task simultaneously and are instructed to prioritise (focus on) the primary task. An increase in effort due to performing the primary task generally leads to a corresponding drop in the performance on the secondary task. When the primary task is listening, the magnitude of the performance decrease on the secondary task can be used as a measure of listening effort (Downs, 1982; Gagné et al., 2017; Pals et al., 2015).

Listening and speech processing often requires dual-tasking (or multi-tasking). For example, in a classroom, individuals are often required to listening to a teacher and to take notes. Therefore, the dual-task method has high ecological validity when assessing difficulties with performance in these situations (McGarrigle et al., 2019). However, using behavioural measures to assess listening effort relies on certain assumptions, as outlined by McGarrigle et al. (2014): (1) DTPs are sensitive to performance trade-offs between tasks only if the total workload in the two tasks exceeds the individual's entire (finite) cognitive capacity; (2) DTPs rely on the individual following instructions and prioritising the correct task over the other which is difficult to guarantee; and (3) there is also no way of measuring how much effort is allocated to one task over the other task. For example, Hicks and Tharpe (2002) found that children with hearing loss had slower reaction times in the secondary task than children without hearing loss, even though they showed similar performance scores in the primary task. While this was interpreted as evidence for increased listening effort, it is possible that the task was not effortful enough to consume a participant's entire cognitive resource capacity. Furthermore, the children with hearing loss may have prioritised accuracy over speed. The results of Hicks and Tharpe (2002) could reflect increased listening effort in the group with hearing loss, but this cannot be confirmed using DTPs.

DTPs may provide an indication of the performance costs of increased listening effort. However, it is unlikely that secondary task performance represents an objective index of listening effort, per se (McGarrigle et al., 2014). For instance, even when two individuals have the same level of performance in the primary and the secondary DTP task, it is still possible that one person was putting more effort into the task than the other person.

Fundamentally, listening can be more effortful for some individuals, even when this effort is not reflected in behavioural/performance measures (Gagné et al., 2017; Pichora-Fuller et al., 2016). Physiological methods of measuring listening effort may overcome the disadvantages associated with subjective and behavioural/performance methods by objectively and temporally tracking physiological states related to effortful listening when individuals are engaged in a listening task. Physiological measures may show differences in effort expenditure when performance measures and/or subjective measures do not.

4.3 Physiological Methods

Changes in central nervous system (CNS) and/or autonomic nervous system (ANS) activity during task performance can be tracked and used as physiological measures of listening effort (McGarrigle et al., 2014). The sections below detail physiological methods for measuring listening effort, including neuroimaging methods, skin conductance, cardiac responses, and the method used in the current thesis, pupil dilation.

4.3.1 Neuroimaging Methods

Neuroimaging techniques used to study listening effort include: magnetoencephalogram (MEG), electroencephalogram (EEG), function magnetic resonance imaging (fMRI), and functional near-infrared spectroscopy (fNIRS).

Electro- and magneto-encephalographic neural responses to acoustic stimuli provide precise markers of mental processing in the CNS and ANS during listening (Obleser et al., 2012; Weisz et al., 2011). EEG methods involve measurement of electrical activity in the brain, whereas MEG methods involve the measurement of magnetic sources in the brain.

In the context of listening effort, data derived from these methods can provide information about task-relevant neural activity (Lopes da Silva, 2013). Both EEG and MEG measurements are based on the same neurophysiological processes. However, due to the fissures in the cortex, localisation of specific brain activity using EEG can be difficult (Srinivasan & Nunez, 2012). Therefore, EEG is considered to have poor spatial resolution, but excellent temporal resolution given the number of samples measured per second (Burle et al., 2015). On the other hand, brain activity measured by MEG may be localised with greater accuracy (Cohen & Cuffin, 1983; Lopes da Silva, 2013) because magnetic signals are not distorted by the fissures (and/or other anatomy) (Burgess, 2019). Therefore, MEG may provide better source localisation than EEG, while still maintaining temporal resolution due to the high sampling rate (Gage & Baars, 2019).

Physiological CNS responses during effortful listening can also be measured using fMRI. This method involves tracking changes in blood-oxygen-level-dependent signals in the brain. Changes in blood oxygenation can reflect metabolic changes in neuronal activity during a listening task (Wild et al., 2012). The fMRI method has excellent spatial resolution, that is, it is valuable for the localisation of brain activity and functions (Mele et al., 2019).

The ability to localise brain regions related to excessive listening effort is imperative to some hearing research and could guide optimal rehabilitation strategies (Dimitrijevic et al., 2019; Rosemann & Thiel, 2019). For example, if clinicians could infer the specific sites of listening-related cognitive dysfunction in brain physiology, they could design specialised cognitive interventions to be used alongside appropriate hearing technologies (Dimitrijevic et al., 2019).

However, fMRI measures have considerably poorer temporal resolution when compared with EEG, MEG, and other peripheral physiological measures. For example, peak blood-oxygen-level-dependent responses typically occur 5-6 s after stimulus onset and last approximately 3 s, whereas EEG and MEG responses can be measured within milliseconds of stimulus onset and can capture the intricacies of processing fluctuations on the order of milliseconds (Glover, 2011; Mele et al., 2019). The information that MEG/EEG and fMRI techniques provide is highly distinct and complex, and separating listening effort from such rich brain activity data requires further research.

The use of neuroimaging techniques for measuring listening effort has several disadvantages, particularly when considering its use in clinical audiology. For fMRI, the definition of activation relies on an arbitrary measure of statistical significance that can be influenced by external factors (e.g., movements as little as a fraction of a millimetre can impact fMRI results) (Matthews et al., 2006). Additionally, even though the fMRI procedure is safe and non-invasive, subjects are required to remain motionless in a large cylindrical tube for up to two hours. For some individuals, this can result in experiences of claustrophobia or agitation which can skew results (Radiological Society of North America, 2020). Additionally, fMRI machines make noise when in operation (Peelle, 2014; Tomasi et al., 2005). Tomasi et al. (2005) found that fMRI noise affected blood-oxygen-level-dependent signals when performing a working memory task. fMRI noise is also likely to confound results when measuring listening effort. However, if appropriate care is taken, these effects may be mitigated to an extent (see Peelle, 2014).

An alternative neuroimaging method which can be used for the measurement of listening effort is fNIRS. Like fMRI, fNIRS measures blood oxygenation levels, however, it can only penetrate approximately 1.5 – 2 cm beneath the skull (Rovetti et al., 2019). Unlike fMRI, fNIRS can simultaneously measure the overabundance of oxygenated blood and the scarcity of deoxygenated blood (Scarapicchia et al., 2017) and is quiet when in operation. fNIRS represents a compromise between the benefits of fMRI and EEG/MEG as it offers competitive spatial and temporal resolution. The use of fNIRS to measure listening effort is gaining in popularity (e.g., Lawrence et al., 2018; Wijayasiri et al., 2017). However, it remains unclear whether fNIRS would be suitable for use in clinical audiology.

Fundamentally, EEG/MEG, fMRI, and fNIRS equipment is expensive, and greater expertise is needed to use it and to interpret the results when compared to peripheral physiological methods. Therefore, while these neuroimaging methods can provide important information about localisation and temporal processing in the brain for research purposes, they may not be a viable candidate for a clinical measure of listening effort, for the reasons outlined above. Peripheral physiological measures of listening effort may overcome some of these issues.

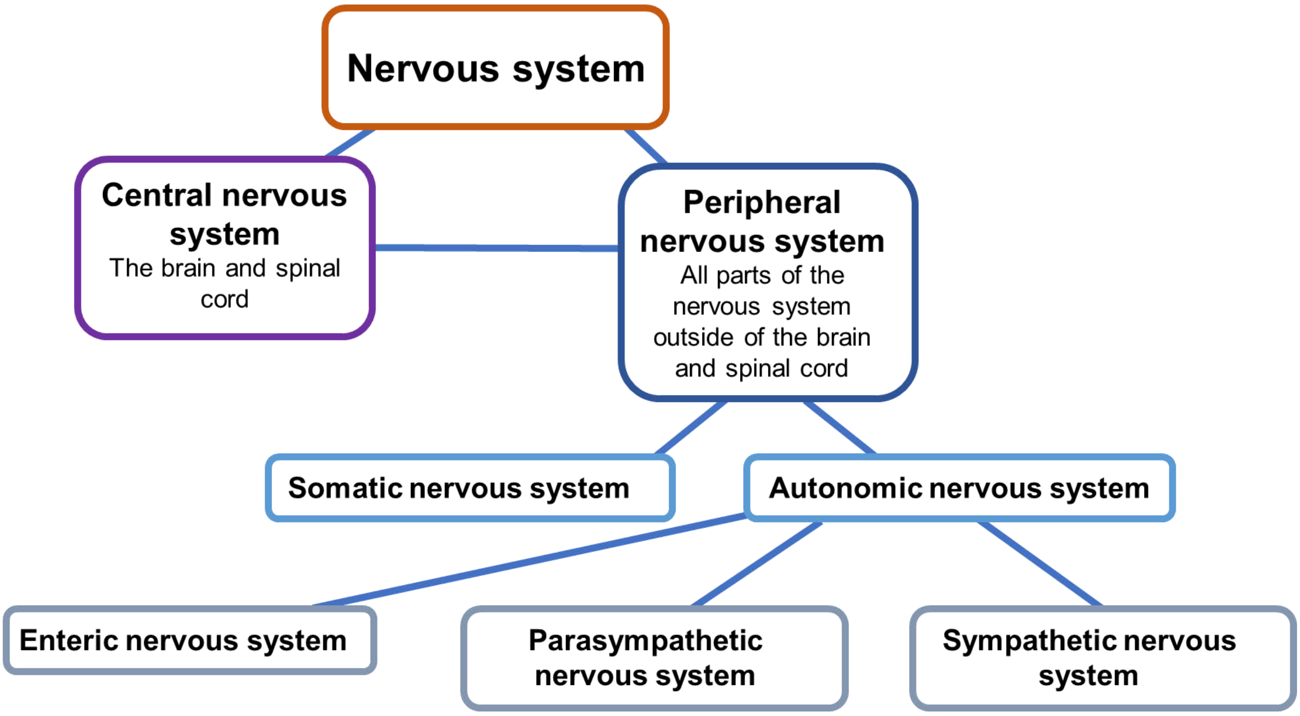
4.3.2 Peripheral Physiological Methods

There are a variety of ways to measure listening effort via activity changes in the ANS branch of the peripheral nervous system (specifically, via sympathetic and parasympathetic activity).

The ANS is divided into three branches: the sympathetic nervous system (SNS), the parasympathetic nervous system (PNS) and the enteric nervous system (Bear et al., 2016; Jänig, 2013) (Figure 3). The enteric nervous system is a local nervous system within the gastrointestinal tract and is beyond the scope of the current thesis. The SNS and the PNS are under the control of the peripheral nervous system which has afferent and efferent connections to the central nervous system (Figure 3). The SNS and PNS are directly relevant to the current thesis and will be discussed in more detail below.

Figure 3

The Branches of the Nervous System



Note. The connection between the central nervous system and the peripheral nervous system demonstrates the role of the central nervous system in interpreting threatening stimuli and regulating the stress response of the ANS (de Morree et al., 2013).

Involuntary and automatic bodily functions (e.g., heart rate, blood pressure, motility and secretion of the digestive system, bladder function, aspects of sexual function) are under the control of the ANS (Jänig, 2013). The main purpose of the ANS is to regulate the body's internal environment and the body's physiological responses to the external environment. Regulation of these responses aims to enable people to optimally respond to and navigate their internal and external environments. Cortical and subcortical central nervous system structures also contribute to how the ANS responds to stimuli (de Morree et al., 2013). The central autonomic network receives information from multiple sources (i.e., humoral, visceral, and environmental) and then integrates these inputs to generate autonomic, endocrine and behavioural outputs (Benarroch, 1993).

The SNS branch of the ANS is responsible for the body's response to stress and life-threatening situations, regulating metabolic resources, and the release and synthesis of associated hormones and neurotransmitters (e.g., norepinephrine). The SNS controls the "fight or flight" response. Increased arousal in the SNS leads to bodily changes that are associated with action preparation, for example, increased heart rate, blood pressure and electrodermal activity. Additionally, blood supply to major muscle groups is increased and the pupils of the eyes dilate (Francis & Oliver, 2018).

On the other hand, the PNS branch of the ANS is responsible for recovery and restoration of metabolic reserves and secretion of waste. In contrast to the "fight or flight" response, the PNS controls the "rest and digest" response (McCorry, 2007). It is associated with reducing heart rate and blood pressure and constriction of the gut, salivary glands, and pupils. The PNS does not operate to simply undo SNS responses. Like the SNS, the PNS can also operate protectively. An example of this is the parasympathetically controlled pupil light reflex (i.e., when the pupil constricts in response to bright light) (Francis & Oliver, 2018).

The measurement of peripheral physiological responses to listening tasks (e.g., cardiac responses, skin conductance, pupil dilation) can be used for non-invasive measurement of the ANS activity that occurs during effortful listening. These measures do not quantify a singular concept of "effort" but rather a combination of attention, engagement, anxiety, ability, cognitive processing, and arousal as reflected in ANS activity during task performance, and it is important to interpret these responses as such.

Furthermore, the effort that is exerted during task performance is not a direct measure of the effort demanded by the task (Winn et al., 2018). The effort that one exerts in a task is also affected by the individual's cognitive capacity, level of arousal, level of fatigue, motivation, goals, and perceived success. Therefore, responses that indicate less effort during task performance, do not necessarily mean the task was easier. They could reflect that the individual was not motivated to succeed in the task, or that the task was too difficult, and they disengaged. Peripheral physiological methods for measuring listening effort are discussed in more detail below.

4.3.2.1 Cardiac Responses

Changes in heart rate variability have been shown to be related to increases in background noise during speech-in-noise tasks, for individuals with hearing impairment (Mackersie et al., 2015) and for individuals with normal hearing (Seeman & Sims, 2015). Specifically, Mackersie et al. (2015) found reduced power in the PNS-driven high frequency band of heart rate variability in more adverse listening conditions. The findings of Mackersie et al. (2015) indicated suppression of parasympathetic activity in response to greater listening-related effort, but only for individuals with hearing loss. Seeman and Sims (2015) found that measures of heart rate variability in the time-domain were sensitive to task complexity and signal-to-noise ratio (SNR) in a sample of individuals without hearing loss. PNS-driven cardiac responses seem to be less sensitive measures of listening effort compared to SNS-driven cardiac responses (Slade et al., 2021).

The pre-ejection period is another way to quantify listening effort via cardiac responses. The pre-ejection period is the time between excitation of the left heart ventricle and the opening of the aortic valve and is associated with sympathetic activity (Richter, 2016). Richter (2016) provided the first evidence that the pre-ejection period could be a useful dependent variable in the measurement of listening effort. They found lower pre-ejection period reactivity in easy listening conditions and higher pre-ejection period reactivity in more demanding conditions, but only when the potential reward was high. This demonstrated that success importance plays a significant role in effort expenditure as reflected by the pre-ejection period. This also supports the notion that motivation is linked to listening effort. These findings were recently supported by Slade et al. (2021).

4.3.2.2 Skin Conductance

Measures of electrodermal activity (e.g., the skin conductance level and response) reflect the skin's capacity to conduct electricity. During stressful (or effortful) situations, the SNS increases activity in the eccrine sweat glands (Boucsein, 2012; Mackersie & Calderon-Moultrie, 2016). This increases the moisture on the skin's surface and leads to greater skin conductance level and response. Skin conductance level has been successfully used to study listening effort. However, findings have been inconsistent.

For instance, Mackersie and Calderon-Moultrie (2016) examined how speaking rate affected skin conductance level (sympathetic activity) and the high frequency band of heart rate variability (parasympathetic activity). While they did not find differences in performance between their fast and slow speaking rate conditions, they did find a significant increase in skin conductance level and a decrease in the high frequency band of heart rate variability in the fast-speaking rate condition. These findings correspond to greater sympathetic activity and greater withdrawal of parasympathetic activity in the fast-speaking rate condition. This suggested that participants may have expended more effort to sustain their performance in the fast-speaking rate condition.

Although skin conductance level was shown to be sensitive to speaker rate in Mackersie and Calderon-Moultrie (2016), multiple other studies have shown that skin conductance level is not sensitive to SNR (Holube et al., 2016; Mackersie et al., 2015; Seeman & Sims, 2015). This is an important disparity and may affect the usefulness of skin conductance level as a measure of listening effort. Individuals experiencing excessive listening effort often report that listening to speech-in-noise is particularly difficult. Therefore, a measure of listening effort should be sensitive to various SNR conditions. Skin conductance as a measure of listening effort appears to be less sensitive and less reliable to changes in listening effort than heart rate variability (Mackersie & Calderon-Moultrie, 2016; Mackersie et al., 2015) and/or pupil dilation (Giuliani et al., 2020).

4.3.2.3 Pupil Dilation

Measurement of pupil dilation or the TEPR is the physiological measure which is perhaps most commonly used in the measurement of listening effort (Winn et al., 2018). Variations in pupil dilation during task performance can provide an indication of the effort that individuals expend in specific listening situations.

Unlike skin conductance and specific cardiac responses, pupil dilation is controlled by a combination of SNS activity and PNS activity, rather than one or the other (Francis & Love, 2020). The SNS innervates the dilator muscles in the iris and the PNS innervates the constrictor (sphincter) muscles in the iris. Any change in pupil size can be the result of sympathetic excitation or withdrawal and/or parasympathetic excitation or withdrawal. This complicates attempts to separate the relative contributions of the SNS and PNS to pupil dilation[2]. Separation of SNS and PNS contributions to pupil dilation during effortful listening may lead to a greater understanding of the ANS mechanisms that underlie effortful listening (Slade et al., 2021). Nevertheless, TEPRs can be used to indirectly measure the combination of SNS and PNS activity that occurs during effortful listening, and this may be useful in clinical audiology.

The TEPR represents the phasic change from baseline diameter in response to an event (e.g., an auditory stimulus in a listening task) (Aston-Jones & Cohen, 2005). Like skin conductance, the slower, tonic pupil response can also be measured and is thought to reflect an individual's initial state of general arousal and is not due to phasic "bursts" of processing during task performance. Often, this tonic pupil size is used as the basal or baseline diameter (Beatty & Lucero-Wagoner, 2000). However, there have been a few studies carried out on the slow, tonic pupil response as a separate measure to the more static baseline diameter measure (e.g., Milne et al., 2021; Peysakhovich et al., 2017). Despite this, no standard parameter for tonic pupil size measurement has been established. There are multiple additional measures that can be extracted from the TEPR which are covered in more detail in Section 6.1.3.

Using pupillometry to measure listening effort has advantages over other peripheral physiological measures. For example, it appears to distinguish effort expenditure between various cognitive tasks across a wide range of domains, including those outside the auditory domain (Beatty, 1982b; Einhäuser, 2017; Winn et al., 2018). When compared with skin conductance, pupil dilation is more sensitive to changes in SNR (Giuliani et al., 2020). This is important for any measure of listening effort as manipulation of SNR conditions is often used to vary the amount of listening effort that is required in a task.

Evidence also suggests that the pupil response is more sensitive across a range of participant groups and listening conditions than PNS driven cardiac responses. For example, Mackersie et al. (2015) found that the high frequency band of heart rate variability only decreased in the most difficult listening conditions and only for individuals with hearing loss, whereas, multiple studies have shown that the pupil response is sensitive to a wide range of listening conditions for individuals with and without hearing loss (Ohlenforst, Zekveld, Lunner, et al., 2017; Zekveld et al., 2018; Zekveld & Kramer,

[2] If the goal is to separate the contributions of the SNS and PNS to pupil dilation during effortful listening, a method has recently been proposed by Wang, Kramer, et al. (2018) based on Steinhauer et al. (2004).

2014; Zekveld et al., 2010, 2011). This may indicate SNS contributions are important for a sensitive measure of listening effort as heart rate variability is typically mediated by the PNS. This notion was recently supported by Slade et al. (2021) who found an effect of listening demand on SNS-driven cardiac responses, but not on PNS-driven cardiac responses. It would be valuable to examine the relative sensitivities of SNS-driven cardiac responses and pupil dilation to listening demands.

It is possible that different physiological measures may be sensitive to qualitatively distinct kinds of cognitive challenges during effortful listening. For example, Francis et al. (2021) reported preliminary evidence which suggested that hearing acuity and personality factors may be more strongly associated with physiological responses to speech-in-noise perception, than to understanding non-native accented speech, which may be more dependent on working memory capacity. Furthermore, Francis et al. (2021) reported that skin conductance responses were more closely related to affective responses to noise-related interference (e.g., annoyance, anger, frustration) than cardiovascular responses. Conversely, cardiovascular responses may be more affected by working memory and lexical access demands.

Pupil dilation has also been shown to have better test-retest reliability than other measures (i.e., subjective measures, behavioural measures, EEG measures) (Alhanbali et al., 2019). Additionally, Giuliani et al. (2020) found pupil dilation to be significantly more reliable than skin conductance across test sessions. More research comparing the various peripheral physiological measures (including pupillometry) is required to establish which measures are most sensitive to the specific cognitive challenges of interest and to assess the implications of this for future research and clinical practice.

Despite the common use of pupillometry in the measurement of listening effort, factors like participant positioning (i.e., object nearness), accommodation reflexes, and environmental light level may affect TEPR generation (Joshi & Gold, 2020). In research settings, strategies that aim to avoid negative effects of these factors are often adopted. For example, standardisation of participant positions and gaze points, and standardisation or individualisation of light levels may be implemented (Winn et al., 2018).

However, it is not currently clear how TEPRs are affected by light level, and how comparable findings are between different laboratories or clinics where different light levels have been used. Furthermore, the multifaceted nature of listening effort and the physiological mechanisms that generate TEPRs are likely to be affected by a broad range of cognitive states. One factor that is relevant to the hearing domain is fatigue. Fatigue and light level require further examination in effortful listening tasks for continuing research in the area, and any future clinical applications.

As pointed out by Winn et al. (2018) individualised measures of listening-related pupil responses might be a goal for many audiologists and researchers (p. 24). Clinical assessments using these methods could be used, for example, to track clinical intervention progress and/or to compare the benefits of treatment approaches. However, there are currently no clinical protocols available and the use of pupillometry in clinical settings is not currently practical. Technological advancements and enhanced understanding of TEPRs may enable clinical use in future.

The primary aim of this thesis is to enhance understanding of factors that influence TEPRs during effortful listening to aid future research and clinical implementation. This is achieved by specifically addressing gaps in the literature related to the effects of light level and fatigue on commonly used pupil parameters in a clinically relevant, effortful listening task. The physiology of pupil size and the mechanisms involved with the pupil response to cognitive effort and task performance are detailed in the following chapter.

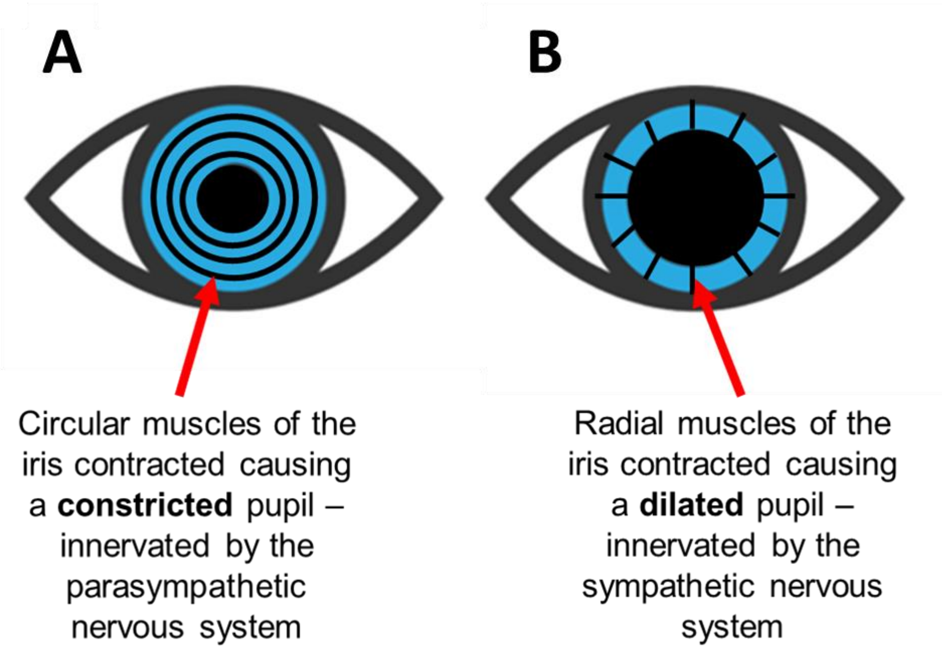
5 THE PHYSIOLOGY OF PUPIL SIZE

Changes in pupil size aid vision by controlling retinal illumination, reducing optical aberrations, and changing the depth of focus (Kardon, 2011). The adult human pupil diameter can vary in size from 2 mm to 8 mm under the control of the muscles in the iris (Spector, 1990). The size of the pupil is determined by the ANS, and more specifically, by interactions between the PNS and the SNS (Larsen & Waters, 2018; Lowenstein & Loewenfeld, 1962; Wilhelm, 2011).

The constrictor muscles in the iris are innervated by the PNS. They are responsible for constriction and limiting the amount of light that can enter the eye. Constrictor muscles are located around the perimeter of the pupil and are approximately 0.10-0.17 mm thick (Wilhelm, 2011). The parasympathetic pathway for constriction originates in the dorsal midbrain. The Edinger-Westphal (E-W) nucleus in the midbrain is stimulated by light falling on the retinas, which activates preganglionic parasympathetic neurons and leads to innervation of the ciliary ganglion, which then leads to constriction of the pupils (Figure 4 A) (Eckstein et al., 2017).

Figure 4

The Muscles and Innervations Involved with Constriction (A) and Dilation (B) of the Pupil



On the other hand, dilator muscles are small, myoepithelial cells (approximately 0.01 mm thick) and are spread throughout the iris (Figure 4 B) (Wilhelm, 2011). The dilator muscles in the iris are innervated by the SNS and are responsible for allowing more light to enter the eye, thus optimising vision.

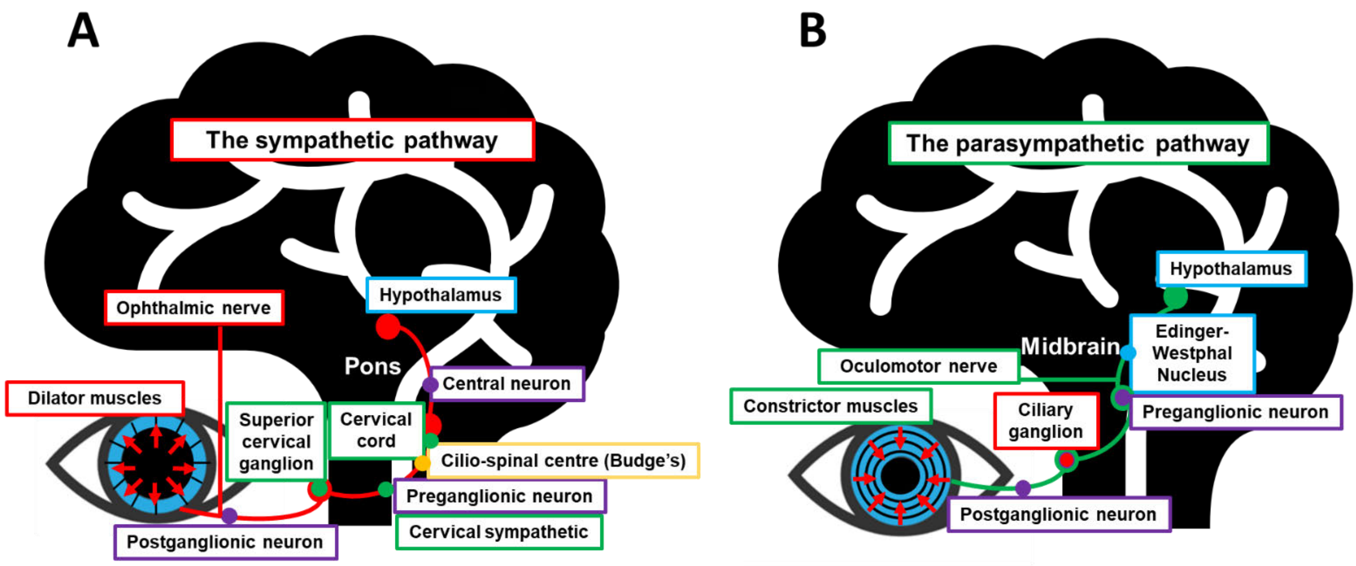
Constriction of the pupil in light and dilation of the pupil in dark form the "pupil light response" (Mathôt, 2018). The pupil can also dilate in response to psychosensory stimuli, for example, a loud noise or a cognitive task such as effortful listening. This type of dilation shares the mechanisms responsible for pupil dilation in darkness and was first referred to as the "pupillary reflex dilation" (Lowenstein & Loewenfeld, 1962). The pupillary reflex dilation differs from the pupil response to darkness in that it is a less uniform and less vigorous movement (Loewenfeld & Lowenstein, 1993). More recently, the pupillary reflex dilation has been referred to as the "psychosensory pupil response" (Mathôt, 2018). This term adequately indicates that this response can be driven by psychological stimuli (e.g., cognitive effort, affective reaction) and sensory stimuli (e.g., a loud, startling noise). Therefore, it is the psychosensory pupil response that generates the TEPR during effortful listening. This thesis is focussed on the psychosensory pupil response to effortful listening tasks, thus "TEPR" is used when referring to pupil responses to cognitive effort.

There is no single, specialised afferent pathway for the psychosensory pupil response (Loewenfeld & Lowenstein, 1993). Any afferent pathway and central connection that is involved in sensation or general arousal can initiate the psychosensory pupil response. Conversely, the efferent pathway for pupil dilation is specialised and complex.

Figure 5 (A) shows the sympathetic pathway which innervates the dilator muscles of the eyes (Joos & Melson, 2012; Ju, 2018; Loewenfeld & Lowenstein, 1993; Smith, 2009). Sympathetic innervation of dilator muscles begins at the hypothalamus. It then descends through the brain stem and spinal cord, to the ciliospinal centre (also known as Budge's centre) and exits through the superior cervical ganglion. Postganglionic fibres exit the superior cervical ganglion and run along the ophthalmic nerve by way of long ciliary fibres until they reach the dilator muscles of the eyes. Figure 5 (B) shows the parasympathetic pathway which innervates the constrictor muscles of the eyes (Joos & Melson, 2012; Ju, 2018; Loewenfeld & Lowenstein, 1993; Smith, 2009).

Figure 5

The Sympathetic Pupil Dilation Pathway (A), and the Parasympathetic Pupil Constriction Pathway (B)



Dilation occurs primarily due to direct sympathetic activation of the dilator muscles via the efferent pathway. However, in studies on cats, the pupil still displayed a dilatory response to psychosensory stimuli after the sympathetic pathway to the pupil was cut, albeit, a much smaller response was observed (Loewenfeld, 1958; Lowenstein & Loewenfeld, 1950). On the other hand, when the parasympathetic pathway was cut, there was only a minute reduction in dilation. This suggested that a portion of the observable dilation in response to psychosensory stimuli can be attributed to the PNS. More specifically, it can be attributed to parasympathetic inhibition of the pupillo-constrictor neurons of the Edinger-Westphal nucleus leading to relaxation of the constrictor muscles (Figure 5 B) (Lowenstein & Loewenfeld, 1962; Steinhauer et al., 2004; Wilhelm, 2011).

Steinhauer et al. (2004) measured pupil dilation in human participants while they performed two arithmetic tasks (a serial add-1 task and a serial subract-7 task). During the tasks, the activity of the PNS or SNS was systematically and pharmacologically blocked in the participants. Steinhauer et al. (2004) found no difference in dilation between two cognitive tasks (hard vs. easy) in human participants when the PNS was blocked, but they did find a difference between the two tasks when the SNS was blocked. Steinhauer et al. (2004) performed an additional experiment using light to separate PNS and SNS contributions to pupil dilation. They found a significant but relatively constant dilation between the two tasks when performed in darkness, when there is little parasympathetic tone. This suggested that inhibition of the parasympathetic pathway is an important component in the TEPR.

It has been proposed that pupil dilation related to inhibition of the parasympathetic pathway occurs via noradrenergic projections from the locus coeruleus (LC) nucleus to the Edinger-Westphal nucleus (Koss, 1986; Larsen & Waters, 2018). Evidence also suggests that noradrenergic projections from the LC also enhances sympathetic drive, resulting in further pupil dilation (Joshi & Gold, 2020).

5.1.1 The LC Norepinephrine System

Numerous studies have demonstrated evidence of a strong link between TEPRs and activity in the mammalian LC norepinephrine system (e.g. Alnæs et al., 2014; Aston-Jones

& Cohen, 2005; Costa & Rudebeck, 2016; Eckstein et al., 2017; Joshi et al., 2016; Murphy et al., 2014; Varazzani et al., 2015).

The LC is a neuromodulator nucleus in the dorsal pons of the mid brain that is involved in the regulation of arousal and autonomic function. The LC supplies numerous regions of the brain with norepinephrine; a neurotransmitter associated with emotions, arousal and executive function (Sara, 2009). Norepinephrine is predominately released in response to activation of nerve fibres of the sympathetic nervous system and is essential for the fight or flight response.

Due to the LC′s noradrenergic nature, it is excitatory when affecting sympathetic processes and inhibitory when affecting parasympathetic processes (Francis & Oliver, 2018). It has been suggested that the LC sends excitatory projections to the preganglionic sympathetic neurons in the intermediolateral cell column of the cervical cord (Figure 5 A) which directly promotes sympathetic drive (Joshi & Gold, 2020). The LC can also indirectly affect sympathetic drive via the reciprocal connectivity between the LC and the hypothalamus which also projects to the intermediolateral cell column of the cervical cord (Joshi & Gold, 2020). Additionally, the LC may also be involved in inhibiting parasympathetic activity via inhibitory projections to the Edinger-Westphal nucleus as discussed above (Zekveld et al., 2018). This results in additional pupil dilation due to relaxation of the constrictor muscles in the iris. Therefore, the LC may contribute to the mediation of the TEPR, via excitation in the sympathetic pathway and inhibition in the parasympathetic pathway (Einhäuser, 2017; Joshi & Gold, 2020; Koss, 1986; Larsen & Waters, 2018; Samuels & Szabadi, 2008).

With that said, a direct, causal link between LC activity and the autonomic nuclei that control pupil size has not been reliably established (Joshi & Gold, 2020; Nieuwenhuis et al., 2011). A multitude of brain areas might be co-activated during TEPRs, yet the LC is most commonly implicated (Joshi & Gold, 2020; Joshi et al., 2016; Larsen & Waters, 2018). Despite the uncertainty regarding the precise neural mechanisms that underlie TEPRs, it seems a robust, indirect proxy measure for the collective physiological brain changes that underlie cognitive effort and therefore, effortful listening.

Primarily based on animal studies, Aston-Jones and Cohen (2005) proposed “Adaptive Gain Theory” to explain the role of the LC in task engagement and performance in cognitive tasks. In line with more recent work (Joshi & Gold, 2020; Joshi et al., 2016; Larsen & Waters, 2018), Adaptive Gain Theory posits that the LC modulates cognition by regulating task engagement via coordination of neuronal activity and norepinephrine secretion. This coordination and activity may be reflected in pupil dynamics during task-engagement.

5.1.2 Adaptive Gain Theory

Aston-Jones and Cohen (2005) and Bouret and Sara (2005) identified two modes of LC activity: phasic and tonic. These modes correspond to stimulus-evoked and baseline release of norepinephrine, respectively. Phasic LC activity typically occurs in response to task-engagement, where the individual is "exploiting" a known source of reward (completing a task). It is characterised by low baseline neural firing and epochs of pronounced peaks in neural firing due to the release of norepinephrine from the LC. The norepinephrine released in the phasic mode of LC activity temporarily enhances the responsiveness of cortical areas, aiding the processing of task-relevant stimuli (therefore, aiding cognitive processing).

In contrast, tonic LC activity is characterised by a lack of phasic responses, higher tonic levels of norepinephrine release and high baseline firing rates that render individuals sensitive to a variety of environmental stimuli. Tonic activity typically occurs when an individual is not engaged in any task and instead "explores" the environment looking for a rewarding task. This is known as the "exploit/explore" trade-off (Gilzenrat et al., 2010; Jepma & Nieuwenhuis, 2011). This trade-off is reflected in the extent to which a system's control state favours behaviours associated with task-engagement versus disengagement and may also be reflected in pupil dynamics.

Gilzenrat et al. (2010) and Jepma and Nieuwenhuis (2011) found phasic and tonic activity in the LC was reflected in pupil dynamics. Baseline diameter (sometimes referred to as tonic pupil diameter) reflects the sustained or tonic component of the pupil response and may reflect an individual's current arousal levels (Unsworth & Robison, 2018) and intelligence/cognitive ability (Tsukahara & Engle, 2021). The task-relevant "spikes" related to cognitive processing represent the phasic component of the pupil response. This component is characterised by transient responses to stimuli and is usually expressed relative to the baseline diameter (Gilzenrat et al., 2010).

Gilzenrat et al. (2010) found that individuals with larger baseline diameters (and smaller TEPRs) performed worse than those with smaller baseline diameters (and larger TEPRs) in an auditory oddball task. These findings were consistent with predictions set out by Adaptive Gain Theory and the exploit-explore trade-off. Specifically, individuals with larger baseline diameters and smaller TEPRs performed worse which could indicate that they were not as engaged in the task (corresponding to exploration). Conversely, individuals with smaller baseline diameters and larger TEPRs performed better, which could indicate that they were more engaged in the task (corresponding to task exploitation). In the context of listening effort, the magnitude of the phasic pupil response can quantify the amount of listening effort that an individual expends in response to task-relevant stimuli and this magnitude may depend on how engaged in the task the individual is.

With reference to the FUEL, motivation and decision-making processes are significant components of the Adaptive Gain Theory (Nieuwenhuis et al., 2005). An individual may decide whether it is worth engaging in (i.e., exploiting) a task. In addition to receiving inputs from other brainstem nuclei (Berridge & Waterhouse, 2003), the LC receives input from brain areas that are known to be important in decision making, and representing task goals and affective value, for example, the prefrontal cortex, anterior cingulate cortex, and the orbitofrontal cortex (Nieuwenhuis et al., 2005). The idea that these brain areas are implicated in task-engagement compliments the FUEL by acknowledging that the magnitude of the phasic LC activity (and corresponding TEPR) may be influenced by the motivational significance of a listening task. Furthermore, the motivational significance of a listening task may be affected by the experience of fatigue, as predicted by the Motivational Control Theory of Cognitive Fatigue (Hockey, 2011, 2013).

Hopstaken et al. (2015b) found that TEPRs decreased as time-on-task increased in a visual n-back task with a duration of 2 hours. This finding corresponded to decreased subjective task-engagement and increased fatigue in their participants over the duration of the test session. It is possible that participants became fatigued during the test session and disengaged from the task when the costs of sustaining engagement were no longer worth the effort. As such, motivation for the task may have also decreased. When rewards were presented to participants (increasing motivation), TEPRs were restored to levels that were similar or higher to those at the beginning of the experiment. This may reflect re-engagement in the task for motivated participants.

TEPRs were recently shown to be sensitive to motivational manipulations (high or low monetary reward) in a listening effort paradigm. Koelewijn et al. (2018) observed larger TEPRs when participants were offered a high monetary reward compared to a low monetary reward during speech-in-noise tasks of equivalent difficulty. This effect did not replicate in Koelewijn et al. (2021). However, in post hoc time course analysis, Koelewijn et al. (2021) found evidence of larger pupil dilation in the high reward condition, relative to the low reward condition, only for the more difficult condition. This may indicate that TEPRs were only affected by the motivational manipulations in a more challenging condition when individuals may be more likely to experience task-induced fatigue.

Adaptive Gain Theory highlights the relationship between LC activity, task engagement, and the subsequent TEPRs and may aid explanations of the mechanisms involved in TEPR generation during effortful listening. This will be beneficial for ongoing research and potential clinical applications of the measure. The following chapter describes pupillometry as a method for measuring listening effort, and highlights gaps in the literature that need to be addressed.

6 PUPILLOMETRY

Based on the physiological properties of the pupil described in Chapter 5, TEPRs are accepted as reliable, objective measures of ANS activity induced by the dynamic central nervous system variations that underlie cognitive effort (Beatty & Lucero-Wagoner, 2000).

Hess and Polt (1964) were among the first to use pupillometry as a measure of cognitive effort. They demonstrated that difficult mental arithmetic problems (e.g., 16 x 23) produced larger TEPRs than simple arithmetic problems (e.g., 4 x 7). They also noted that dilation occurred after a stimulus was presented and during processing, while constriction and return to baseline diameter occurred only after responses had been given. Kahneman and Beatty (1966) demonstrated that TEPRs were sensitive to the cognitive load induced by short-term auditory memory tasks (strings of 3 - 7 digits presented for immediate recall). Furthermore, they observed that the TEPRs increased as a function of digit series length. Subsequently, Kahneman and Beatty (1967) examined pupil dilation in response to a pitch-discrimination task. They observed larger pupil sizes in more difficult tasks; when the tones were more similar and thus, more difficult to differentiate.

Beatty (1982) stated, “it [pupillometry] provides a reliable and sensitive indication of within-task variations in processing load. It generates a reasonable and orderly index of between task variations in processing load. It reflects differences in processing load between individuals who differ in psychometric ability when performing the same objective task. For these reasons, the TEPR provides a powerful analytic tool for the experimental study of processing load and the structure of processing resources” (p. 281). These characteristics conform to the criteria necessary for a physiological measure of effort as outlined by Kahneman (1973).

These seminal studies provided the foundation for the application of pupillometry as a measure of listening effort in hearing science (Zekveld et al., 2018). The next sections describe pupillometry methods and analysis techniques, followed by a discussion of the factors that are relevant for the use of pupillometry in measuring listening effort and the current thesis.

6.1 Pupillometry Methods and Analysis Techniques

TEPRs typically entail changes to pupil diameter that are small, usually less than 0.5 mm (Beatty & Lucero-Wagoner, 2000; Winn et al., 2018). With appropriately sensitive equipment (usually an infrared eye-tracker), the changes can be measured, and time-locked to a cognitive task. Raw pupil data must be inspected for artefacts and cleaned before being subject to statistical analyses. Artefacts in the pupil trace are likely to be due to eye blinks, movements during recording, or shifts in focus (e.g., accommodation responses). Small anomalies can be corrected using linear interpolation (Geller et al., 2020). However, if there are too many blinks and/or a large number of anomalies, the entire trial should be excluded from further analyses (Winn et al., 2018). Once these anomalies have been removed, the cleaned pupil data can be analysed.

6.1.1 Baseline Correction

An appropriate pre-stimulus baseline diameter should be determined and then corrected for, to examine the effects of experimental manipulations on the dilation response. There are two common baseline correction methods: (1) divisive baseline correction (corrected pupil size = pupil size/baseline) and (2) subtractive baseline correction (corrected pupil size = pupil size − baseline) (Mathôt et al., 2018). Both methods have been used in pupillometry research even though they may result in incongruent results. Recent attempts have been made to report best practice methods and standardise baseline correction in pupillometry research (Mathôt et al., 2018; Reilly et al., 2019). Mathôt et al. (2018) recommended that subtractive baseline correction be used as it is not as affected by artefacts as divisive baseline correction. However, subtractive baseline correction assumes that TEPRs are independent of the pupil’s initial baseline diameter. This is a common assertion that has been made in the literature (e.g., Beatty & Lucero-Wagoner, 2000), but conclusive evidence is lacking.

6.1.2 The Baseline-Response Relationship

There is conflicting evidence regarding the relationship between baseline diameter and subsequent responses. Adaptive Gain Theory posits that pronounced peaks and moderate baseline levels of LC activity (the phasic mode) are thought to reflect exploitation of a task and correspond to task-engagement. Conversely, larger baseline levels of LC activity (the tonic mode) attenuate the pronounced peaks and are thought to reflect exploration and task-disengagement (Aston-Jones & Cohen, 2005). As discussed in Section 5.1.2, Gilzenrat et al. (2010) and Jepma and Nieuwenhuis (2011) reported that this relationship was reflected in pupil dynamics, that is, larger baseline diameter corresponds to smaller TEPRs and smaller baseline diameters correspond to larger TEPRs. However, as pointed out by Gilzenrat et al. (2010) and Joshi and Gold (2020) another reason for this relationship could be related to the Law of Initial Values (LIV).

The LIV asserts that the magnitude of a physiological response is dependent on its initial baseline level (Lacey, 1956; Wilder, 1957, 1958). More specifically, larger baseline levels may constrain subsequent responses. In terms of the pupil, the LIV posits that large initial baseline diameters may constrain TEPRs due to the dynamic range of the pupil. Jin (1992) proposed a reconceptualization of this law whereby reactivity to a stimulus increases with higher initial values, until the initial value reaches the system's upper limit, for example, the upper limit of pupil size. However, these theorists did not examine the LIV in the pupil. Given the complexity of physiological systems, including the pupil, a blanket law like the LIV seems inadequate for informing specific predictions for systems in which the law's existence has not been explicitly verified.

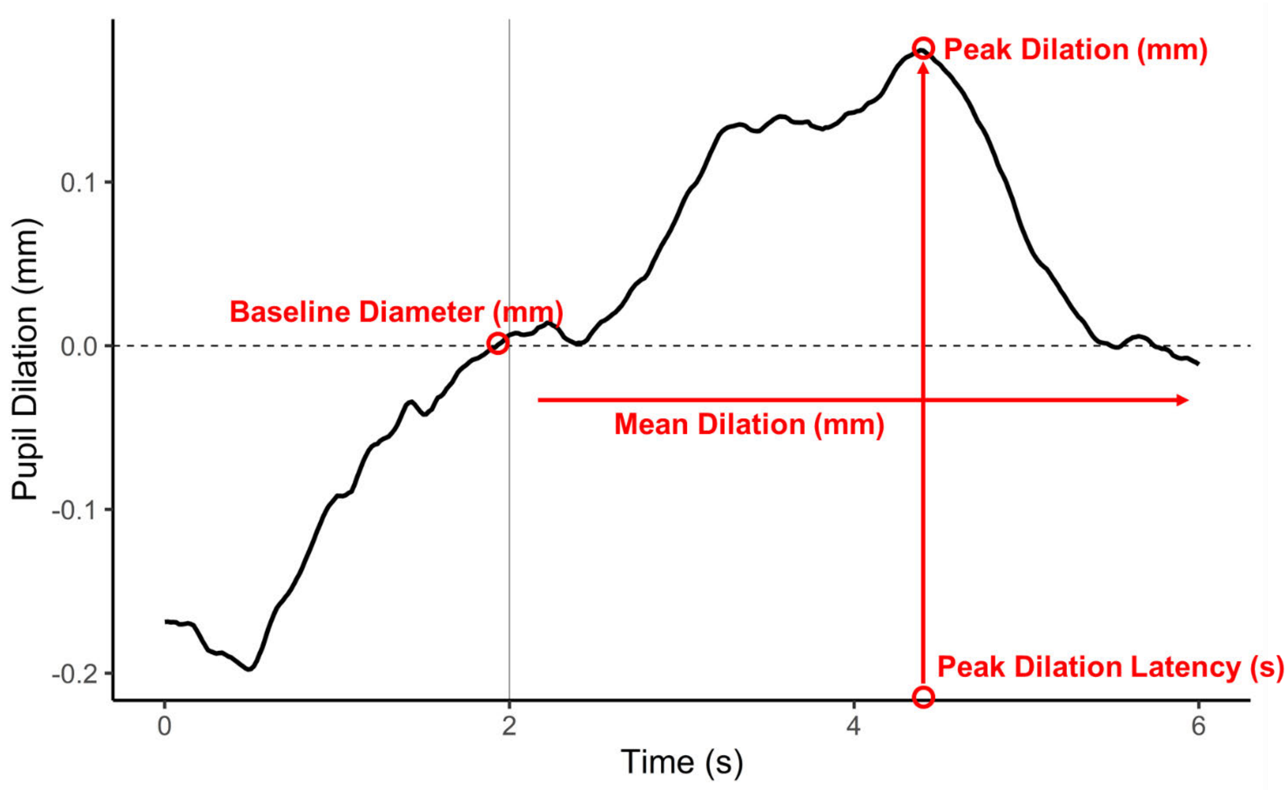
Gilzenrat et al. (2010) (Experiment 1B) and Reilly et al. (2019) examined if the LIV applied to TEPRs by manipulating light level to purposefully modulate baseline diameter. In response to equivalent stimuli, there were no differences in TEPRs, despite different light-induced baseline diameters. Therefore, they did not find evidence for the LIV in TEPRs. Gilzenrat et al. (2010) did find an inverse relationship between baseline diameter and TEPRs in Experiment 1A where light level was not manipulated and asserted that this was likely due to a systematic relationship between TEPRs and control state (as predicted by Adaptive Gain Theory), rather than the LIV. A similar relationship and conclusion was subsequently reported by Peysakhovich et al. (2017), who did not find an effect of light level on peak dilation, but did find an inverse relationship between tonic pupil diameter (mean pupil diameter during task performance) and peak dilation. Therefore, there is some evidence that a relationship between baseline diameter and subsequent pupil responses may exist, but more research is needed to elucidate the relationship and the mechanisms involved.

6.1.3 Pupil Parameters

In speech recognition experiments, the pupil is thought to start dilating approximately 0.5 s to 1.3 s after stimulus onset and the peak of the response may occur approximately 0.7 s to 1 s after stimulus offset (Winn et al., 2018). Typically, peak dilation and/or mean dilation values are measured. Figure 6 shows a typical TEPR waveform and TEPR measurements that are commonly extracted. Peak dilation is the largest dilation value during the pre-determined measurement epoch (e.g., 2 – 6 s in Figure 6) relative to a baseline diameter value measured before stimulus onset. Mean dilation is the average dilation during a pre-determined measurement epoch relative to a baseline diameter value measured before stimulus onset. Peak and mean dilation are used to quantify the amount of effort expended during task performance. Peak dilation latency is a measure of the time it takes for the pupil to reach peak dilation and is typically measured in seconds or milliseconds. Peak dilation latency may be a marker of cognitive processing speed (Hyönä et al., 1995; Zekveld et al., 2010). For the purposes of this thesis, peak dilation, mean dilation and peak dilation latency are the TEPR components of interest, and baseline diameter is an additional dependent variable.

Figure 6

A Typical Baseline-Corrected TEPR Waveform from the Current Dataset and the Commonly Measured Pupil Parameters



Note. This figure shows a typical TEPR waveform measured during a speech-in-noise task. The dashed horizontal line at 0.0 mm (on the y axis) indicates baseline-correction. The solid vertical line at 2 s on the x axis) indicates stimulus onset. Baseline diameter is typically measured just before stimulus onset.

Peak dilation may represent a more sensitive measure of listening effort than mean dilation, as dilation is not averaged over an extended period. Zekveld et al. (2010) observed more significant differences in pupil dilation between background noise levels when peak dilation measures were used rather than when mean dilation measures were used. Beatty and Lucero-Wagoner (2000) have argued that peak dilation is also independent of the number of data points in the measurement epoch. This is beneficial for tasks where trial durations vary between participants (e.g., in a maximum memory span task) (Beatty & Lucero-Wagoner, 2000). Conversely, peak dilation measures may be more subject to noise or random variations in pupil diameter measurements than mean dilation measures because it is notionally based on a single value as opposed to multiple values (Beatty & Lucero-Wagoner, 2000).

6.1.4 Signal-Averaged Versus Trial-level Pupil Data

Like other stimulus-evoked measurements (e.g., event-related potentials measured with EEG), TEPRs are typically measured over multiple trials, under the same conditions. For each participant, in each condition, TEPR trials are time-aligned and averaged together in a process called *signal-averaging*. The aim of signal-averaging is to average out components of the measured TEPR that are unrelated to the stimulus, so that the task-relevant pupil signal is clearer (Winn et al., 2018). Baseline diameter, peak dilation, mean dilation, and peak dilation latency values are usually measured from the averaged waveforms. This is the standard approach used in the analysis of TEPR data.

While signal-averaging has the benefit of reducing the noise in pupil signals, it will disregard potentially important information about the variance in pupil data that may be attributed to time-on-task, or the specific stimuli that were used, by condensing many observations per participant per condition, into one measurement per participant per condition.

Like event-related potentials in EEG research, creating signal-averaged waveforms relies on certain assumptions about the pupil response across trials; primarily, it assumes that task-relevant pupil activity is relatively constant across trials (Luck, 2014; Volpert-Esmond et al., 2018). This assumption may not be appropriate for TEPRs, as they are likely to change over the course of a test session due to learning effects, fatigue, and/or changes in motivation (Hopstaken et al., 2015a, 2015b; Pichora-Fuller et al., 2016; Winn et al., 2018).

In an experiment that uses sentence stimuli, different sentences should be used between trials to mitigate effects related to learning, habituation, and familiarisation, but the different sentences that are used should also be similar in syntactic and morphological complexity. When signal-averaging is used on pupil data, it is often assumed that stimuli were not different and did not contribute measurable variance to TEPRs. However, this assumption may not be appropriate (Volpert-Esmond et al., 2018). It is likely that responses to the same stimuli (e.g., sentences) between participants will be related. Statistical methods, like mixed-effects modelling, enable simultaneous modelling of complex fixed and random effects structures. These methods can account for the variance associated with factors like trial and stimuli (Baayen et al., 2008; Gelman & Hill, 2007), rather than ignoring and losing this information as a result of the signal-averaging approach.

Recently, researchers have been examining trial-level pupil data using techniques like mixed-effects modelling (Clewett et al., 2020; Clewett et al., 2018; Cohen Hoffing et al., 2020; Leuchs et al., 2017; Wetzel et al., 2020). For example, Cohen Hoffing et al. (2020) examined trial-level baseline diameter, peak dilation and peak dilation latency in a longitudinal study (8 biweekly sessions) using a classical mental arithmetic task. Their analyses showed that despite the noisiness of trial-level TEPRs, relatively simple parameters (like peak dilation) can be meaningfully related to behaviour on a trial-by-trial basis.

Additionally, Clewett et al. (2018) analysed fMRI data and trial-level mean dilation to examine LC activity during a memory encoding task. Monetary incentives and threats of punishment were used to motivate participants to remember target scenes. Trial-level mean dilation predicted memory for goal-relevant scene images when under threat. Based on the fMRI and trial-level pupil data, Clewett et al. (2018) found a positive correlation between greater threat-evoked mean dilation and greater activity related to scene encoding in the LC and parahippocampal cortex, a brain area which is specialised for scene processing. This demonstrated that trial-by-trial mean dilation may contain meaningful information in a memory encoding paradigm.

The use of trial-level pupil data has not been examined in a listening effort paradigm, but it warrants investigation through the use of mixed-effects modelling for several reasons: (1) the use of trial-level pupil data has been successfully applied in other cognitive domains, (2) gains in statistical power may be achieved, and (3) it provides a chance to account for additional sources of variance in pupil data and acknowledge the underlying variance structure of the data which may lead to a greater understanding of TEPRs.

In summary, task-relevant measurements can be extracted from averaged and trial-level pupil waveforms recorded during task performance. These measurements may provide valuable information about the listening effort that individuals expend in specific tasks.

6.2 Pupillometry and Listening Effort

Pupillometry has been widely used to measure listening effort in hearing research. The TEPR has been shown to be sensitive to SNR/speech intelligibility (Koelewijn, Zekveld, Festen, & Kramer, 2012; Koelewijn, Zekveld, et al., 2014; Koelewijn, Zekveld, Festen, Rönnberg, et al., 2012; Koelewijn et al., 2018, 2021; Wang, Naylor, et al., 2018; Wendt et al., 2016; Wendt et al., 2017; Zekveld & Kramer, 2014; Zekveld et al., 2010, 2011), masker type (Koelewijn, Zekveld, Festen, & Kramer, 2012; Ohlenforst et al., 2018; Rennies et al., 2019; Wendt et al., 2018; Zekveld et al., 2014), motivation in listening tasks based on monetary reward (Koelewijn et al., 2018), syntactic complexity (albeit in younger participants only) (Piquado et al., 2010), and selective and divided auditory attention (Baldock et al., 2019; Koelewijn, Shinn-Cunningham, et al., 2014). The following sections review the literature concerning SNR/speech intelligibility and masker type as these are of most relevance to the current thesis.

6.2.1 SNR and Speech Intelligibility

Speech intelligibility and listening to speech in noise is often implicated in anecdotal reports of excessive listening effort, that is, people find it most difficult and effortful to understand speech in situations with degraded signals (e.g., due to background noise). Thus, manipulation of task demands via speech intelligibility is common practice when measuring listening effort.

Speech intelligibility can be manipulated by varying the level of the background noise relative to the target signal (i.e., stimuli) (SNR). Louder background noise leads to reduced intelligibility of the signal. Manipulation of fixed SNR conditions controls the demand of the task but allows effort expenditure and performance to vary between participants.

The effects of speech intelligibility on TEPRs can also be examined using adaptive procedures. In these procedures the SNR is adaptively adjusted based on an individual’s performance until they reach a predetermined level of performance – the speech reception threshold (a percent correct score) (Smits & Houtgast, 2006). The TEPR is measured once the predetermined level of performance is achieved. The adaptive procedure allows researchers to control performance levels, allowing the demand of the task and effort expended to achieve the performance levels to vary between participants. Regardless of the intelligibility manipulation procedure, more adverse SNRs typically lead to larger pupil dilations for individuals with hearing loss (Ohlenforst, Zekveld, Lunner, et al., 2017) and without hearing loss (Zekveld et al., 2013), unless the task is too difficult.

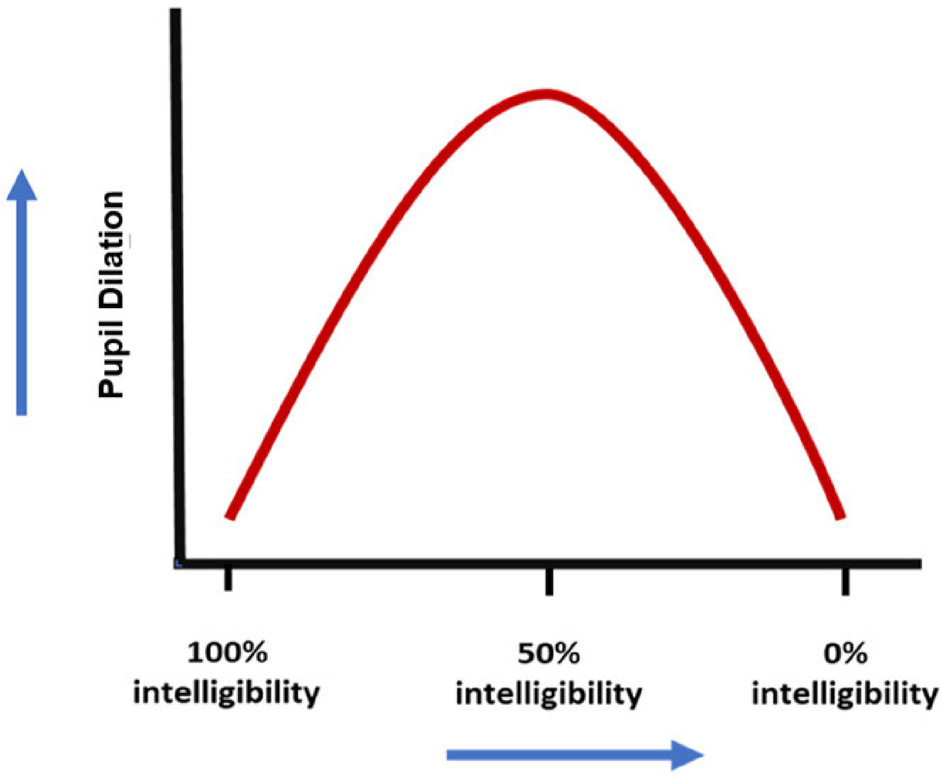
Kramer et al. (1997) provided the seminal evidence that pupillometry can be used to quantify hearing difficulty in individuals with and without hearing loss. Listening conditions were manipulated by variations to SNR using an adaptive procedure. Kramer et al. (1997) found smaller TEPRs when participants were listening in more favourable SNR conditions, which indicated that the task required less listening effort. TEPRs also indicated individuals with hearing loss benefitted less from easier listening conditions, that is, they showed larger TEPRs than individuals without hearing loss in easier listening conditions.

Zekveld et al. (2011) found similar results when they examined TEPRs in middle-aged adults with and without hearing loss during a speech reception threshold test targeting 50%, 71%, and 84% intelligibility. Individuals with hearing loss showed a smaller reduction in TEPRs in more intelligible conditions. This effect was described as the experience of “less release from effort”. Essentially, this means that individuals with hearing loss may need to expend more effort compared to individuals without hearing loss, even in easier listening situations.

Recent findings indicated that the relationship between speech intelligibility and peak dilation over a wide range of intelligibility levels (0 – 100% correct) followed an inverted-U shaped curve (Figure 7) with largest pupil dilation occurring when individuals were scoring approximately 50% correct (Ohlenforst, Zekveld, Lunner, et al., 2017). This suggests that TEPRs are non-linear; the easiest SNR conditions typically result in small TEPRs, the most adverse SNR conditions also typically result in small TEPRs, and intermediate SNR conditions typically result in the largest TEPRs (Ohlenforst, Zekveld, Lunner, et al., 2017; Wendt et al., 2018; Winn et al., 2018; Zekveld & Kramer, 2014).

Figure 7

Theoretical Depiction of the Inverted-U Shaped Response Curve of Pupil Dilation From 100% Intelligibility to 0% Intelligibility in a Speech-In-Noise Task



The inverted-U relationship has been observed in individuals with and without hearing loss (Ohlenforst, Zekveld, Lunner, et al., 2017). However, the curve was shifted by approximately 10 dB for the group with hearing loss. Individuals without hearing loss expended most effort in negative SNRs, whereas individuals with hearing loss expended effort across a wider range of SNR conditions (including in easier SNRs). These findings suggest that there are common effort mechanisms between groups, but there are differences in the conditions in which individuals with and without hearing loss expend the most effort.

The inverted-U relationship between pupil dilation and speech intelligibility resembles the classic Yerkes-Dodson relationship between arousal and performance (Yerkes & Dodson,

1908) and the relationship between LC function and performance proposed by Aston-Jones and Cohen (2005). In these models, intermediate arousal leads to optimal engagement and performance. The inverted-U relationship between speech intelligibility and pupil dilation may indicate that individuals are optimally aroused when they are performing at 50% correct, and that corresponds to the largest pupil dilation due to the involvement of the LC. When tasks become too difficult, individuals may withdraw effort, leading to a reduction in LC activity, arousal, and pupil dilation.

Research has shown that individuals with hearing loss display smaller TEPRs than individuals without hearing loss when a speech-in-noise task is set at 50% intelligibility (e.g., Kramer et al., 2016; Ohlenforst, Zekveld, Lunner, et al., 2017; Wang, Naylor, et al., 2018; Zekveld et al., 2011). This may counterintuitively suggest that individuals with hearing loss exert less effort than individuals without hearing loss at 50% intelligibility. However, because these studies used an adaptive testing procedure to target 50% intelligibility and individuals with hearing loss typically require better SNRs to achieve 50% intelligibility compared to individuals without hearing loss, there were differences in the SNRs used for the two groups in each of the studies. The group with hearing loss performed at 50% intelligibility in less adverse SNRs than the group without hearing loss. It is possible that the group with hearing loss may have exerted less effort to reach 50% intelligibility in the adaptive testing procedure.

Based on the FUEL (Pichora-Fuller et al., 2016) and Adaptive Gain Theory (Aston-Jones & Cohen, 2005), effort expenditure, and TEPRs, may have been modulated by motivation and/or fatigue. In the studies cited above (i.e.,Kramer et al., 2016; Ohlenforst, Zekveld, Lunner, et al., 2017; Wang, Naylor, et al., 2018; Zekveld et al., 2011), individuals with hearing loss may have experienced fatigue-induced changes to motivation (and individuals without hearing loss may not have) because the costs of effortful listening may be greater for the individuals with hearing loss. This may have led to less effort expended to reach 50% intelligibility (resulting in easier SNR conditions) and smaller TEPRs, if individuals with hearing loss were experiencing listening-related fatigue. Supporting this, Wang, Naylor, et al. (2018) reported a negative correlation between self-reported fatigue and peak dilation at 50% intelligibility.

Fundamentally, pupil dilation represents a sensitive measure of the effort that is expended in listening tasks. There are multiple factors that can affect the amount of effort an individual must expend in a listening task. One factor is the type of background noise that is present, or is purposefully manipulated, that is, masker type.

6.2.2 Masker Type

TEPRs are also sensitive to masker type (or type of background noise) in speech-in-noise tasks. Individuals might experience excessive listening effort, and therefore TEPR variations in noisy situations, because of energetic or informational masking. Energetic masking occurs when the audibility of an acoustic target is reduced by interference of acoustic noise in the periphery, that is, the acoustic signals of the target and noise are “blended” (Mattys et al., 2009). In contrast, informational masking is any masking that occurs after energetic masking has been accounted for in the early stages of auditory processing and typically interferes with intelligibility at a more cognitive level (Feng et al., 2018; Mattys et al., 2009). For example, informational masking may occur when speech of a known language interferes with the target. In this case, the masker speech may draw attention away from the target speech due to shared perceptual attributes (i.e., the linguistic characteristics of the masker may interfere with the target and affect performance) (Rennies et al., 2019).

Research has shown that at the same intelligibility levels, TEPRs vary based on the type of masker used (Koelewijn, Zekveld, Festen, & Kramer, 2012; Koelewijn, Zekveld, et al., 2014; Koelewijn, Zekveld, Festen, Rönnberg, et al., 2012; Wendt et al., 2016; Wendt et al., 2017). In speech-in-noise tasks targeting 50% and 84% intelligibility, Koelewijn, Zekveld, Festen and Kramer (2012) found that a one-talker masker condition resulted in larger TEPRs compared to a non-speech masker condition. This may be due to the lexical and semantic content of the one-talker masker, interfering with the target speech (i.e., informational masking) and may have created contextual overlap during cognitive processing (Kidd et al., 2008).

This conclusion was supported by Wendt et al. (2018) (Experiment 1). At 84% intelligibility, Wendt et al. (2018) found that TEPRs were largest for the 1-talker masker condition when compared with the fluctuating noise condition (no linguistic content) and the 4-talker masker condition (where linguistic information may not be as audible due to increased energetic masking). However, at 50% intelligibility there were no differences between masker types. This conflicts with Koelewijn, Zekveld, Festen and Kramer (2012) who found larger TEPRs for one-talker masker at 50% intelligibility.

In Koelewijn, Zekveld, Festen and Kramer (2012) and Experiment 1 of Wendt et al. (2018), different SNRs were required to reach 50% and 84% intelligibility between the different maskers. Due to the large effect that SNR has on peak dilation (Ohlenforst, Zekveld, Lunner, et al., 2017), the difference in TEPRs may be related to the SNR that was used as well as the masker type. Therefore, in Experiment 2 of Wendt et al. (2018), they examined effects of a 4-talker masker and a stationary noise masker on TEPRs in a speech-in-noise task at fixed SNRs (ranging from −20 dB and +8 dB). Interestingly, there was a negligible effect of masker type on TEPRs in corresponding SNRs. However, this may be due to slight methodological differences between Experiment 1 and 2. For example, in Experiment 2 the talkers in the 4-talker masker were spatially separated which might have made it easier to identify the target. Furthermore, the stationary noise masker did not involve temporal fluctuations which might have maximised the masking (Wendt et al., 2018).

TEPRs are also sensitive to other target/masker characteristics, like voice similarity. Zekveld et al. (2014) examined the effects of masker gender and spatial separation of masker and target in a speech-in-noise task targeting 50% intelligibility. TEPRs were largest for the condition with a masker of the same gender to the target (i.e., the voices were more similar), when compared to conditions where the target and the masker were of different genders (i.e., the voices were less similar) but were not affected by the spatial separation manipulation. Therefore, voice similarity increases listening effort more than decreasing the spatial separation of the target and the masker.

Fundamentally, TEPRs recorded in speech-in-noise tasks are sensitive to the properties of the maskers used. Decisions about masking properties are important for research and any future clinical application of pupillometry and will largely depend on measurement aims. For measuring listening effort in a clinical setting, multi-talker babble (speech maskers including more than one talker) may be more ecologically valid than other types of non-speech maskers (e.g., white noise, speech-shaped noise) as it simulates noise that people experience in most real-world scenarios (e.g., at a cocktail party, at a restaurant) (Silbert et al., 2014).

6.3 Clinical Use

Despite the reports of effortful listening and its disproportionate incidence in individuals with hearing loss, listening effort is not routinely assessed by audiologists as clinical measures are not readily available (McGarrigle et al., 2014; Miles et al., 2017; Pichora-Fuller et al., 2016). A reliable, objective, clinical measure of listening effort would likely result in a more comprehensive assessment of an individual's hearing difficulties by complementing current audiological assessment tools.

A clinical measure of listening effort could also be used to assess the extent to which an individual's hearing aids or hearing aid features may reduce effort during listening. This could then potentially better inform individualised intervention strategies (e.g., selection of specific signal-processing algorithms) and support hearing aid adoption.

Despite the high prevalence of hearing loss in the global population, there is a relatively low level of persistent hearing aid use (Gallagher & Woodside, 2018). Improved understanding of situations that are excessively effortful and/or increased knowledge of the intervention strategies that are most effective for reducing an individual's listening effort could also lead to more consistent and persistent use of hearing aids. Additionally, TEPRs could potentially be used to assess how an individual allocates cognitive resources in effortful listening situations (Bönitz et al., 2021) which may improve outcomes for individuals with hearing loss and aid the design of new hearing aid technologies (Rovetti et al., 2019).

The fundamental goal of hearing aids is to improve the perception and intelligibility of sounds and speech in quiet and noisy environments. Hearing aid features, such as amplification, may also reduce listening effort. However, Ohlenforst, Zekveld, Jansma, et al. (2017) conducted a systematic review and found no consistent results which indicated that hearing aid amplification lessens listening effort expenditure. The authors suggest that this finding could be due to a general lack of consistency and standardisation between studies, and a lack of statistical power in the design of the studies (Ohlenforst, Zekveld, Jansma, et al., 2017, p. 267).

In research settings, pupillometry has been used to measure the physiological benefit of advanced hearing aid features, such as signal processing algorithms including noise reduction schemes (Fiedler et al., 2021; Ohlenforst et al., 2018; Wendt et al., 2017). These studies have shown that noise reduction schemes can reduce listening effort as measured by TEPRs. This demonstrates the potential of TEPRs for clinical applications because the TEPR measurements provided information about how specific hearing aid features affected listening effort expenditure.

The potential clinical utility of pupillometry was addressed in this section. However, while pupillometry has the potential for use in clinical settings, it is important to note that changes in pupil diameter during task performance may not result in clear strategies for effort reduction (Winn et al., 2018). Furthermore, measurement of TEPRs may only be accurate under tightly controlled settings which may be difficult to achieve in clinical settings. For example, differences in effort expenditure between tasks and individuals may be obscured when different environmental light levels are involved. Different light levels may affect the sensitivity of TEPRs to detect differences in effort expenditure in different conditions (e.g., when using different hearing aid features). Additionally, TEPRs decrease in size due to effort overload and subsequent withdrawal of effort and may also decrease due to loss of motivation and/or fatigue as outlined in the FUEL. This may obscure differences in effort expenditure between tasks and individuals. Therefore, there are gaps in the literature pertaining to the effects of light level, and time-on-task and fatigue on the generation of TEPRs during effortful listening. The next sections will expand on these factors and review the literature that is associated with how light level, and time-on-task and fatigue affect TEPRs across multiple cognitive domains.

6.4 Light level

There is conflicting evidence regarding the effect of environmental light level on TEPRs during effortful tasks. Table 1 provides an overview of the research concerning the effects of light level on TEPRs in several cognitive domains, using various pupil parameters. There is an overall lack of consistency, standardisation, and detail reported in the studies regarding light level manipulations. This complicates any comparison and replication attempts. Furthermore, the pupil parameters reported across the studies are also inconsistent, often using absolute pupil diameter (uncorrected for baseline diameter) and in the case of Peysakhovich et al. (2017), using unjustified measures of tonic pupil diameter. Studies on the effect of light level on TEPRs are reviewed in the next section.

The studies in which absolute pupil diameter was used (Table 1) will not be reviewed further as the differences in absolute pupil diameter are likely to result from the effect of light level on baseline diameter, rather than an effect of light level on TEPRs.

Table 1. Overview of Research Examining Light Level Effects on TEPRs

Reference	Type	Purpose of study	Cognitive domain	Task used	Light level measurement	Performance	Pupil parameters reported	Evidence of an effect of light level
Bradshaw (1969)	Journal article	To examine whether TEPRs were independent of baseline levels via manipulation of illumination	Auditory	Reaction time task	Bright condition: 25 ft-L Dark condition: .56 ft-L.	Not reported	Peak amplitude and shape	No
Pomplun and Sunkara (2003)	Conference paper	To examine the effects of light level and cognitive workload on pupil size in human-compute r interaction	Visual	Reaction time task with 3 difficulty levels	White background: 82.4 cd/m2 Black background: < 1 cd/m2	Significant main effect of task difficulty	"Corrected" pupil size	Yes, no interaction between light level and task difficulty.
Steinhauer et al. (2004)	Journal article	To isolate contributions of the sympathetic and parasympathetic pathways to pupillary dilation during sustained processing	Mental arithmetic	Serial add 1 and serial subtract 7 tasks	Bright condition: 0.03 cd/m2 Dark condition: 0.59 cd/m2	Not reported	Baseline diameter (average pupil diameter over 5 s pre-task) Mean dilation (measured	Yes, interaction between light level and task difficulty

Reference	Type	Purpose of study	Cognitive domain	Task used	Light level measurement	Performance	Pupil parameters reported	Evidence of an effect of light level
							over the 60 s response period)	
Gilzenrat et al. (2010)	Journal article	Experiment 1B: To examine an if an earlier finding (experiment 1A) could be explained by the LIV	Auditory	Auditory oddball task/reaction time	Light levels individually set Bright condition pupil diameter: 3.7 mm Dark condition pupil diameter: 5.1 mm	No differences between conditions. Near ceiling.	Peak dilation	No
Xu et al. (2011)	Conference paper	To examine pupil size under different luminance conditions	Mental arithmetic/ visual	Level 1: sum binary digits (0,1) Level 2: sum 1 digit (1 to 9) Level 3: sum 2 digits (10 to 99) Level 4: sum 3 digits (100 to 999)	Adjusted by changing grayscale value of screen: L1: 32 L2: 96 L3: 160 L4: 224	Not reported	Average pupil diameter (uncorrected for baseline)	Yes

Reference	Type	Purpose of study	Cognitive domain	Task used	Light level measurement	Performance	Pupil parameters reported	Evidence of an effect of light level
Peysakhovich et al. (2015)	Journal article	To explore the interaction between luminance and memory load on pupil size and dilation	Memory/auditory	Short-term memory task (recall a paced sequence of digits)	Adjusted via computer screen: Grey: 24 cd/m2 White: 54.8 cd/m2	No differences between conditions	Baseline diameter Average pupil diameter during the retention pause (3 s following the stimuli) Peak dilation	Yes, interaction between light level and task difficulty
Pfleging et al. (2016)	Conference paper	To examine the effects of room illumination (part 1) and screen luminance (part 2) on the pupil size.	Auditory	Auditory Delayed Digit Recall Task (n-back task, 3 difficult levels)	Part 1: Level 1: M=133.50 lux, Level 2: M=247.55 lux, Level 3: M=387.14 lux Part 2: Level 1: M = 255.19 lux, Level 2: M= 308.10 lux, Level 3: M= 364.76 lux	Significant differences between performance for 0-back and 2-back and 1-back and 2 back.	Average pupil diameter (uncorrected for baseline)	Yes, interaction between light level and task difficulty

Reference	Type	Purpose of study	Cognitive domain	Task used	Light level measurement	Performance	Pupil parameters reported	Evidence of an effect of light level
Peysakhovich et al. (2017)	Journal article	To examine the effect of screen luminance on tonic and phasic pupil responses during cognitive processing	Memory/visual	The Toulouse n-back Task (Mandrick et al., 2016)	Illumination: 10 lux Computer screen Grey: ~ 11 cd/m2 White: ~ 28 cd/m2	Significant main effect of memory load on accuracy	Tonic pupil diameter: Mean absolute pupil diameter 1-3 s post-stimulus. Phasic pupil response: peak dilation (baseline-corrected maximum value 1-3 s post-stimulus)	Tonic: Yes, interaction between light level and task difficulty Phasic: No effect of light level, or interaction
Wang, Kramer, et al. (2018)	Journal article	To understand the contributions of the PNS and SNS on TEPRs during speech perception for individuals with	Auditory	Speech-in-noise task (adaptive SRT)	Bright condition: 360 cd/m2 Dark condition: 0.1 cd/m2	Equivalent performance as task targeted 50% correct	Baseline diameter Peak dilation	Yes. Difference in peak dilation between individuals with and without

Reference	Type	Purpose of study	Cognitive domain	Task used	Light level measurement	Performance	Pupil parameters reported	Evidence of an effect of light level
		and without hearing loss						hearing loss in bright and dark conditions
Reilly et al. (2019)	Journal article	To examine if TEPRs for transient events (e.g., target detection) scale linearly with baseline diameter	Auditory/visual	Auditory target detection (tones) and visual word monitoring	Bright condition: 753 lux Dark condition: 16 lux. Mid condition (word monitoring only): 350 lux	Tone detection: 99.91% correct Word monitoring: 99.28%	Baseline diameter Mean dilation Peak dilation	No
Książek et al. (2021)	Journal article	To evaluate several pupil parameters extracted from TEPRs. To examine the effects of SNR and luminance on TEPRs using from previous studies (Ohlenforst, Zekveld, Lunner, et al., 2017; Wang,	Auditory	Speech-in-noise task (adaptive SRT)	Bright condition: 360 cd/m2 Dark condition: 0.1 cd/m2	Not reported	Peak dilation Mean dilation Index of pupillary activity Growth curve analysis Principle components analysis	Yes, for all pupil measures. No interaction examined.

Reference	Type	Purpose of study	Cognitive domain	Task used	Light level measurement	Performance	Pupil parameters reported	Evidence of an effect of light level
		Naylor, et al., 2018)						

Beatty and Lucero-Wagoner (2000) conducted a comprehensive review of the pupillary system and its contribution to the field of psychophysiology. They stated that "all available evidence indicates that the extent of the pupillary dilation evoked by cognitive processing is independent of baseline pupillary diameter over a wide range of baseline values" (p. 148). This assertion was largely based on the results of Bradshaw (1969). More recent research suggests that this may not be accurate (Książek et al., 2021; Peysakhovich et al., 2015; Peysakhovich et al., 2017; Steinhauer et al., 2004; Wang, Kramer, et al., 2018). The next section is split into two parts. In the first part, the evidence for an effect of light level on TEPRs is reviewed and in the second part, evidence against an effect of light level on TEPRs is reviewed.

6.4.1 Evidence For the Effect of Light Level on TEPRs

Steinhauer et al. (2004) used a serial subtract-7 task (cognitively demanding), a serial add-1 task (less cognitively demanding) and two illumination conditions (light vs. dark) to investigate the contributions of the SNS and PNS to TEPRs (in this case, mean dilation measured over 60 s). They found that mean dilation increased at the onset of the task in both light and dark conditions. Additionally, they observed larger mean dilation when tasks were performed in light. This supports the notion that a component of TEPRs is related to inhibition of the parasympathetic pathway when tasks are performed in light.

Steinhauer et al. (2004) argued that parasympathetic activity has minimal impact on TEPRs when tasks are performed in darkness because the constrictor muscles in the iris receive minimal input from the parasympathetic pathway in darkness (Loewenfeld & Lowenstein, 1993; Steinhauer et al., 2004). When light level increases, constrictor muscles are stimulated, leading to pupil constriction. When activation related to cognitive processing reaches the Edinger-Westphal nucleus, it may result in relaxation of the constrictor muscles as a smaller, additional component of dilation via inhibition of the parasympathetic pathway (Steinhauer et al., 2004). Therefore, a TEPR elicited in darkness is thought to reflect a predominantly sympathetic response and a TEPR elicited in ambient light is thought to reflect both sympathetic and parasympathetic responses.

Furthermore, Steinhauer et al. (2004) observed a task difficulty-by-light level interaction where the average mean dilation was largest for the more cognitively demanding task when performed in light but there was no difference between tasks in dark. This may indicate an additional component of the task-evoked mean dilation can be attributed to parasympathetic inhibition, which is only present in tasks that are cognitively demanding. These results suggest that light levels and cognitive demand may interact in their effect on TEPRs, especially in a sustained processing task, when mean dilation is calculated over 60 s duration.

Wang, Kramer, et al. (2018) compared peak dilation between individuals with and without hearing loss in a speech-in-noise task targeting 50% intelligibility. They did not find a difference in peak dilation between individuals with and without hearing loss when the task was performed in the dark. Because TEPRs recorded in darkness are said to reflect a predominately sympathetic response (Steinhauer et al., 2004), this may indicate that the sympathetic activity during task performance is relatively similar between individuals with and without hearing loss. However, when compared to individuals with hearing loss, individuals without hearing loss showed significantly larger peak dilation when the task was performed in light. This could mean that the PNS is less active in individuals without hearing loss, which might lead to lesser constriction in the presence of light. These results provide further evidence of an effect of light level on TEPRs and show that these effects may be different for individuals with hearing loss.

Using data from Wang, Kramer, et al. (2018), Książek et al. (2021) examined how light level affected several TEPR measures recorded during a speech-in-noise task. The pupil measures included peak dilation, mean dilation, index of pupillary activity. Analysis methods included growth curve analysis and principal component analysis. The index of pupillary activity measures sudden changes in pupil dilation (dilation reflexes) over a specified period of time (Książek et al., 2021, p. 3). Growth curve analysis and principal component analysis were used to examine task-relevant changes in pupil size over time. Książek et al. (2021) found larger peak and mean dilation in the condition with light. Furthermore, all other pupil measures were significantly affected by light level. Because performance was set at 50% intelligibility, examination of a possible interaction between light level and SNR was not possible. However, Książek et al. (2021) provided evidence that light level affects TEPRs measured during a speech-in-noise task.

Peysakhovich et al. (2015) investigated how different light levels affected TEPRs (peak dilation) in a short-term memory task. Participants were required to recall digit sequences at three difficulty levels (sequence lengths of 5, 7, and 9) under two light levels (white and grey computer screens). Half of the participants were required to recall and repeat back the digit sequences (load-on-memory condition), and the other half were required to passively listen to the sequences (control condition). They found that peak dilation depended on light levels. They also found a significant task difficulty-by-light level interaction. Peak dilation for the load-on-memory condition was larger when performed in darker conditions (grey screen). These findings suggest an effect of light level in the opposite direction to the findings of Steinhauer et al. (2004), Wang, Kramer, et al. (2018) and Książek et al. (2021) as the darker condition produced larger peak dilation, rather than the brighter condition.

Peysakhovich et al. (2015) suggested that the discrepancy between their results and those by Steinhauer et al. (2004) could be related to task differences (arithmetic/sustain processing vs. a short-term memory task) as well as how the TEPR was quantified (e.g., mean dilation over 60 s vs. peak dilation). The methodological differences make comparing these findings difficult. Despite this, the results in Peysakhovich et al. (2015) suggest that peak dilation is affected by light level in a digit recall task.

Peysakhovich et al. (2017) examined the effect of screen luminance on TEPRs during the Toulouse n-back Task (Mandrick et al., 2016). Participants performed transient arithmetic tasks while simultaneously sustaining their attention and remembering their answers to the arithmetic questions from one or two steps back to perform the n-back component (1-back or 2-back) under two light levels (grey and white computer screens). Specifically, they were interested in the effects of screen luminance on tonic pupil diameter and phasic pupil response (peak dilation). They defined tonic pupil diameter as mean pupil diameter (uncorrected for baseline diameter), over the 1 – 3 s period post-stimulus. However, it is not clear that this operationalisation measured what they purported to measure, especially because it was measured over an epoch post-stimulus and was confounded by the task-relevant dilation.

There is a lack of consistency in the literature regarding measurement of tonic pupil diameter. For example, Beatty (1982a) defined tonic pupil diameter as the pupil diameter measured at stimulus onset, Mandrick et al. (2016) defined tonic pupil diameter as the median pupil diameter for each block of 12 trials, and Gilzenrat et al. (2010) used (pre-stimulus) baseline diameter as a proxy to examine tonic LC activity.

Nevertheless, as predicted, Peysakhovich et al. (2017) found a significant task difficulty-by-light level interaction for their tonic pupil diameter measure whereby the difference between tonic pupil diameter in the 1-back condition and the 2-back condition was larger in dimmer conditions. Peak dilation was not affected by light level and was modulated only by task difficulty. If the authors had reported a baseline diameter that was independent of the task, the effect of luminance on tonic pupil diameter may have been clearer. Furthermore, their use of absolute pupil diameter makes drawing conclusions difficult. Despite drawbacks, Peysakhovich et al. (2017) reported evidence which suggests that TEPRs (post stimulus pupil diameter) may be affected by light level in memory paradigms.

The results of Steinhauer et al. (2004), Wang, Kramer, et al. (2018) and Książek et al. (2021) all showed larger pupil dilation in light conditions, compared to dark conditions[3]. Peysakhovich et al. (2015) and Peysakhovich et al. (2017) reported the opposite, finding larger TEPRs in dimmer conditions. The discrepancy between the findings might be due to the differences in how light level was manipulated. Steinhauer et al. (2004), Wang, Kramer, et al. (2018) and Książek et al. (2021) adjusted the illumination in the room, whereas Peysakhovich et al. (2015) and Peysakhovich et al. (2017) manipulated light levels by adjusting the brightness of a computer screen. Furthermore, the darkest condition in Peysakhovich et al. (2015) and Peysakhovich et al. (2017) was not as dark as in Steinhauer et al. (2004), Wang, Kramer, et al. (2018) and Książek et al. (2021). Therefore, there may have been more parasympathetic activity innervating the constrictor muscles in the darker condition. In Peysakhovich et al. (2015) and Peysakhovich et al. (2017), TEPRs in the dimmer condition may have been affected by stimulation of the dilator muscles and parasympathetic inhibition, and TEPRs in the brighter condition may have been constrained by constriction leading to smaller TEPRs. Despite these contrasting findings, they still provide evidence suggesting that light level can affect TEPRs in a variety of cognitive domains and this may also depend on the type of task under examination (e.g., sustained vs. transient tasks). Furthermore, light level may affect TEPRs recorded during effortful listening tasks. Therefore, this effect should be examined in more detail.

[3] Though, Wang, Kramer, et al. (2018) and Książek et al. (2021) used the same dataset.

6.4.2 Evidence Against the Effect of Light Level on TEPRs

The absence of an effect of light level on TEPRs was reported in three of the studies included in Table 1 (Bradshaw, 1969; Gilzenrat et al., 2010; Reilly et al., 2019). Bradshaw (1969) was the first to investigate the effect of light on TEPRs by examining how the pupil responded to an auditory target detection task under dark versus bright conditions. The dark condition increased baseline diameter by 33% but there was negligible evidence of changes in TEPRs. Bradshaw's study had many issues that may have affected the accuracy of their findings. For example, they used a small sample (N = 7), the equipment had a low sample rate (2.7 Hz) and they used visual inspection of response functions to determine their results.

Gilzenrat et al. (2010) found that larger baseline diameter corresponded to smaller TEPRs and vice versa. They subsequently examined the relationship between light-induced baseline diameter and peak dilation (experiment 1B) to verify if those findings (experiment 1A) could be explained by the LIV. They adjusted light levels so that baseline diameter was approximately 3.7 mm in the light condition and 5.1 mm in the dark condition, ensuring that there was always 1.4 mm between conditions (to align with natural baseline diameter differences in experiment 1A). Therefore, light levels varied between individuals. Participants performed the task at near ceiling levels (96.9% hit rate). Peak dilation did not differ between light and dark conditions. Thus, the findings showed that peak dilation during an auditory oddball task was not affected by light level.

More recently, Reilly et al. (2019) examined the effects of light level induced baseline diameter on peak dilation during two tasks: an auditory signal detection task (experiment 1) and a visual monitoring of words task (responding when a target word is presented) (experiment 2). Participants responded with high accuracy across both tasks (experiment 1: 99.91% correct and experiment 2: 99.28% correct) suggesting that participants may not have found the tasks especially demanding. The data showed that peak dilation was independent of light-induced baseline diameter for both light levels in both tasks. This showed that light level did not affect the peak dilation, in line with the findings of Gilzenrat et al. (2010) and Bradshaw (1969)

6.4.3 Summary

In summary, the literature provides conflicting results regarding the effect of light levels on TEPRs. It is possible that the studies in which an effect of light level was demonstrated used tasks that were more cognitively demanding than those that did not demonstrate an effect. For instance, the Toulouse n-back task used in Peysakhovich et al. (2017) requires significantly more cognitive effort than a simple reaction time task (Bradshaw, 1969) or an auditory oddball task (Gilzenrat et al., 2010), where no effect of light level was found. Furthermore, the studies showing no effect of light level reported ceiling performance levels (Table 1). The results of Steinhauer et al. (2004) also support this since the subtract-7 task showed an effect of light level whereas the simpler, add-1 task did not. This suggests TEPRs may be more susceptible to the effects of light levels in more cognitively effortful tasks. Therefore, it is important to examine the effects of light level on TEPRs in a listening task, as TEPRs for different types of tasks and amounts of listening effort may be differentially affected by light levels, which may lead to erroneous conclusions.

Overall, there was a lack of standardisation in how light levels were manipulated, measured, and reported between the studies reviewed. Currently, there are no standard practices for objectively measuring light levels in pupillometry studies. Recently, Tsukahara and Engle (2021) reported that excessively bright light levels may restrict the range of pupil size values to such a degree that results will be biased towards a null relationship (at least in studies examining differences in baseline diameter). Thus, they made two recommendations regarding light levels for pupillometry studies (Tsukahara & Engle, 2021, p. 15):

- Report lighting levels and luminance values from at least two sources: overall room lighting and screen lighting (by placing the light meter directly on the monitor).
- Be very explicit about how luminance values were obtained i.e., what instrument and unit was used, and where was it positioned.

The studies in the auditory domain reviewed here have only used two light levels representing the two extremes, dark and light. For pupillometry to be used in clinical audiology, it would be beneficial to examine light level effects on TEPRs across a range of less extreme, more ecologically valid light levels.

Many research groups investigating listening effort have employed a protocol that involves setting the ambient lighting level during the test session to correspond to the mid-point of the pupil's dynamic range for each participant (Winn et al., 2018). This results in different light levels between participants. The rationale for this protocol is to avoid the possibility of pupil dilation reaching ceiling level, and consequently, the pupil being unable to dilate to its full extent (Chapman et al., 1999). This protocol has also been used in data normalisation, for instance, to correct for potential age differences in the pupil's dynamic range (Piquado et al., 2010). If TEPRs are not independent of baseline diameter and are affected by light level, the use of individualised light level adjustment may weaken previous research findings and the comparability of findings between experimental set ups may be hindered. This may be particularly detrimental when the effect of light level on TEPRs also interacts with the amount of listening effort expended.

In experimental and clinical settings, the effects of light level on TEPRs measured during effortful listening is an important consideration. This may be especially important when measuring listening effort in individuals with a hearing loss or older adults showing cognitive decline who may have to expend more effort in listening than younger individuals without hearing loss (Pichora-Fuller et al., 2016). Degree of hearing loss could lead to larger effects of light level on TEPRs due to the additional listening effort that may be required from individuals with hearing loss. Given the physiological mechanisms that control pupil size and the additional listening effort that individuals with hearing loss may have to expend during listening, it is possible that listening effort and the associated SNS and PNS responses could interact and modulate pupil responses, obscuring the effort-induced response.

Overall, there is a gap in the literature regarding the effects of light level on TEPRs during effortful listening. Wang, Kramer, et al. (2018) and Książek et al. (2021) provided some initial evidence that an effect exists, but a more thorough examination of the effect across multiple light levels and a larger range of task difficulty (i.e., SNRs), and the potential interaction between light level and task difficulty is warranted.

6.5 Fatigue

There is substantial, converging evidence that significant cognitive resources are required for listening, especially in adverse listening conditions (Pichora-Fuller et al., 2016). The study of effort in listening is entangled with the study of fatigue, as excessive listening effort can often lead to reports of fatigue. After a period of cognitive effort expenditure, individuals may experience a variety of psychological fatigue effects, including subjective feelings of tiredness and compromised task performance. More research is needed to elucidate the relationship between effortful listening and fatigue and, how fatigue manifests in any potential clinical measure of effortful listening. There could be short-term and long-term effects of fatigue due to effortful listening. Likewise, there may be short-term effects (e.g., time-on-task) and long-term effects (e.g., daily-life) of fatigue on physiological measures of listening effort, like the TEPR. Listening-related fatigue may lead to differences in ANS functioning and this may affect TEPRs.

6.5.1 Daily-Life Fatigue and TEPRs

In a series of studies (Wang, Kramer, et al., 2018; Wang, Naylor, et al., 2018; Wang, Zekveld, et al., 2018), Wang and colleagues reported cumulative evidence that: (1) subjective, daily-life fatigue was associated with greater PNS activation, (2) greater subjective, daily-life fatigue was associated with smaller peak dilation in a speech-in-noise task, and (3) individuals with better hearing acuity have larger peak dilation, likely because there is a larger effect of parasympathetic inhibition (due to less parasympathetic activation). The heightened PNS activity that individuals with hearing loss may experience might block the inhibitory effect of cognitive processing (and related dilation) at the parasympathetic pathway, which could mean that the constrictor muscles in the iris do not relax as much as they do in an individual without hearing loss, and/or an individual who is not fatigued.

The PNS is thought to be involved in fatigue as it facilitates a “rest and digest” state in the human body and it has been suggested that PNS activity could be protective against the stress associated with hearing loss (Hasson et al., 2009). Therefore, it was suggested that differences in peak dilation for individuals with and without hearing loss may also reflect differences in PNS activation, and fundamentally fatigue, rather than just listening effort (Wang, Kramer, et al., 2018; Wang, Naylor, et al., 2018). These findings may suggest that TEPRs could also be affected by other types of fatigue, including the more transient, task-induced fatigue.

6.5.2 Task-Induced Fatigue, Time-On-Task and TEPRs

In a recent review article, Zekveld et al. (2018) collated 19 studies that reported larger pupil sizes at the beginning of a test session, compared to the end. It is generally accepted that TEPRs decrease over the course of a test session (Winn et al., 2018). However, Winn et al. (2018) proposed that TEPRs measured in tasks of 1 – 1.5 hr typically do not show a significant time-on-task effect and that fatigue is avoidable in most listeners when tasks are less than 2 hours (p.12). A recent systematic review found evidence that contradicts this proposition (Bafna & Hansen, 2021). Bafna and Hansen (2021) reviewed articles that examined the effects of fatigue/time-on-task on pupil responses recorded during a variety of tasks (e.g., neurocognitive tasks, simulated driving, aircraft flying, attention tasks, etc.). They found that cognitive tasks ranging from 30 minutes to 1.5 hours were sufficiently long to reduce peak and mean dilation as a function of mental fatigue. Additionally, they found that cognitive tasks ranging from 1 – 2.5 hr were sufficiently long to reduce baseline diameter as a function of mental fatigue.

Fatigue and time-on-task effects may occur within trials as well. McGarrigle et al. (2017b) used growth curve analysis to examine average within-trial fluctuations in the pupil response during a sustained listening task in a sample of individuals without hearing loss. After initial peak dilation, McGarrigle et al. (2017b) found a steeper decline in within-trial pupil size in the latter half of the test session, which was steeper still during the more difficult task. This response is consistent with reduced physiological arousal which is consistent with fatigue. However, it is unclear whether fatigue effects are apparent in common TEPRs measurements (peak and mean dilation) and baseline diameter. The literature concerning this effect is reviewed in the following sections.

Table 2 provides a chronologically organised overview of studies in which the effects of task-induced fatigue and/or time-on-task in TEPRs were examined. Studies which have examined these effects in children (e.g., Brännström et al., 2021; McGarrigle et al., 2017a) were not included in the table as examination of this sample group is beyond the scope of the current thesis.

Table 2. Overview of Research Examining Task-Induced Fatigue and/or Time-On-Task Effects on TEPRs

Reference	**Purpose of study**	**Cognitive domain**	**Task used**	**Task length**	**Subjective fatigue measured**	**Performance measured**	**Time-on-task effect measured**	**TEPRs reported**	**Evidence for fatigue/time-on-task effect on TEPRs**
Beatty (1982a)	To clarify the type of activation process that underlies sustained attention using TEPRs	Auditory	Vigilance/tone detection	48 min	No	Yes. Performance decreased as a function of time-on-task	Yes	Baseline diameter Peak dilation	No, for baseline diameter Yes, for peak dilation (decreased)
Zekveld et al. (2010)	To examine the effect of sentence intelligibility on TEPRs	Auditory	Speech-in-noise task	1.5 hr	No	Yes – but performance was fixed at three levels during the task	Yes	Baseline diameter Mean dilation	Baseline diameter and mean dilation decreased during the first SRT test Baseline diameter decreased during the third SRT test
Murphy et al. (2011)	To examine if baseline diameter can be used to index task engagement and if EPRs	Auditory	Auditory oddball task	37 min	No	RT and accuracy– no time-on-task effect	Yes	Baseline diameter Pupil dilation defined as maximum dilation	Yes, for baseline diameter (increased) Yes, for pupil dilation (decreased)

Reference	Purpose of study	Cognitive domain	Task used	Task length	Subjective fatigue measured	Performance measured	Time-on-task effect measured	TEPRs reported	Evidence for fatigue/time-on-task effect on TEPRs
	are similarly sensitive							0.4 - 2 s post stimulus minus minimum pupil diameter 0–0.4 s post stimulus	
Hopstaken et al. (2015a)	To examine the relationships between TEPRs task-engagement and mental fatigue	Visual	n-back	2 h	Yes	Yes – performance declined over the session	Yes	Baseline diameter Peak dilation	No, for baseline diameter Yes, for peak dilation (decreased)
Hopstaken et al. (2015b)	To examine the link between mental fatigue and task-engagement using subjective, behavioural, and psychophysiol	Visual	n-back	2 h	Yes	Yes – performance declined over the session	Yes	Baseline diameter	Yes, for baseline diameter (decreased)

Reference	Purpose of study	Cognitive domain	Task used	Task length	Subjective fatigue measured	Performance measured	Time-on-task effect measured	TEPRs reported	Evidence for fatigue/time-on-task effect on TEPRs
	ogical measures								
van den Brink et al. (2016)	To examine the relationship between pupil diameter and performance during a sustained attention task .	Visual	gradual continuous performance task	40 min	No	Yes - performance declined over the session	Yes	Baseline diameter and baseline diameter derivative	Yes, for baseline diameter (decreased)
Sibley et al. (2019)	To examine within-task learning using pupil size	Visual	Spatial orientation aptitude test	Not reported	No	Yes – accuracy increased, and reaction time decreased over the session	Yes	Maximum pupil size, baseline diameter and peak dilation	Yes, decrease across all TEPRs
Alhanbali et al. (2020)	To examine the relationships between performance, subjective fatigue and effort, and baseline	Auditory	Digits-in-noise task	Not reported	Yes	Yes	No	Baseline diameter Peak dilation	Yes, for baseline diameter (decreased) No, for peak dilation

Reference	Purpose of study	Cognitive domain	Task used	Task length	Subjective fatigue measured	Performance measured	Time-on-task effect measured	TEPRs reported	Evidence for fatigue/time-on-task effect on TEPRs
	diameter and peak dilation.								
McGarrigle, Rakusen, et al. (2021)	To examine the covariance of effort, tiredness from listening and TEPRs at multiple time points	Auditory	Speech-in-noise	1 h	Yes	Yes – performance improved over the session	Yes	Peak dilation	Yes (decreased)
McGarrigle, Knight, et al. (2021)	To examine differences in effortful listening between young and old adults using TEPRs	Auditory	Speech-in-noise	1.1 -1.2 hr	Yes	Yes – performance improved over the session	Yes	Normalised TEPR	Yes (decreased)

In Table 2, an effect of time-on-task in TEPRs for tasks as short as 37 min duration was reported (Murphy et al., 2011) and this effect has been replicated across tasks and domains. It is possible that time-on-task effects in TEPRs do not reflect fatigue but rather, a learning effect (Foroughi et al., 2017; Sibley et al., 2019) or a reaction to a monotonous task (Zekveld et al., 2010). Zekveld et al. (2010) found evidence for a time-on-task effect that resulted in a decrease in mean dilation and baseline diameter in a speech-in-noise task, however, they did not measure subjective fatigue or performance as a function of time-on-task. Therefore, it is unclear if their findings were the result of performing a monotonous task, or if the findings were related to fatigue.

For the purposes of this thesis, only studies which have examined the effect of task-induced fatigue and/or time-on-task on TEPRs and included measures of subjective fatigue and/or performance are discussed further. The following section is split into two parts. In the first part, the literature associated with the effect of task-induced fatigue and/or time-on-task on baseline diameter is reviewed and in the second part, the literature associated with the effect of task-induced fatigue and/or time-on-task on peak dilation is reviewed.

6.5.2.1 Fatigue and Time-On-Task in Baseline Diameter

Baseline diameter is largely determined by environmental light levels, but it is also modulated by the LC norepinephrine system via excitatory effects on the sympathetic dilator muscles and inhibitory effects on the parasympathetic constrictor muscles (Aminihajibashi et al., 2019). Activation of the LC norepinephrine system corresponds to increased arousal. A fatigued individual may experience a decrease in LC activity and arousal (McGarrigle et al., 2017b) and therefore, these effects may be apparent in baseline diameter.

Hopstaken et al. (2015a) simultaneously measured subjective, behavioural, and physiological responses during a visual n-back task of varying difficulty (1-back, 2-back, and 3-back). Consistent with fatigue and their predictions, they found that baseline diameter decreased as a function of time-on-task. Additionally, baseline diameter was significantly associated with subjective fatigue and task-engagement measured after each time block and these associations were stronger in more difficult tasks. However, the difficulty levels were not counterbalanced; the most difficult task always came last, and participants may have already been substantially fatigued when faced with the most difficult task. Therefore, the main effect of task-difficulty that was reported should be interpreted with caution.

In another study by Hopstaken et al. (2015b) (using only the visual 2-back task) baseline diameter did not significantly decrease by time block. They reasoned that the discrepancy in the findings may have been due to the experimental environment being less arousing than it was in Hopstaken et al. (2015a) (no EEG equipment). Because physiological arousal contributes to baseline diameter, a less arousing environment may have led to “floor effects” in initial baseline diameter and may have resulted in a limited range for decrease. However, if this was the case, another interpretation of their baseline diameter results could be that participants habituated to the experimental set up over time in Hopstaken et al. (2015a), that is, they became less aroused by the environment over time. Additionally, it would be interesting to know if the light level in the experimental set up was consistent between studies. If it was not, this may also explain some variation in baseline diameter between studies. Tsukahara and Engle (2021) recently showed that there is less variability in baseline diameter when light levels are too bright. Again, it is recommended that researchers comprehensively measure and report environmental lighting conditions for their experimental set ups. Despite this, the results are inconsistent and highlight the need to further examine the relationship between fatigue and baseline diameter.

Alhanbali et al. (2020) examined whether performance and subjective, task-related fatigue and effort predicted baseline diameter during an auditory digits-in-noise task. They found a relationship between smaller baseline diameter and higher scores on a distinct dimension of the “Visual Analogue Scale to Evaluate Fatigue Severity” scale (Shahid et al., 2012) related to tiredness and drowsiness. Although they did not report time-on-task effects, these findings align with those of Hopstaken et al. (2015a) and provide support for the existence of a relationship between baseline diameter and subjective fatigue in an auditory task.

However, because peak and mean dilation capture the effort associated with discrete tasks, these measures are likely more useful for clinical audiology than baseline diameter. The following section will assess evidence relating to the presence of a fatigue effect in dilation responses.

6.5.2.2 Fatigue and Time-On-Task Effects in Pupil Responses

It is possible that fatigue will affect peak and mean dilation, similar to how it affects baseline diameter. This is because the LC norepinephrine system has an influence on pupil size and it is also involved in arousal, attention, and behavioural/performance regulation (Section 5.1.1) (Aminihajibashi et al., 2019; Aston-Jones & Cohen, 2005; Sara, 2009).

Although Hopstaken et al. (2015b) (described in the section above) did not find an effect of time-on-task on baseline diameter, they did find that increasing time-on-task led to smaller peak dilation, which may also be an indication of task-related fatigue. These time-on-task effects were significantly related to increased subjective fatigue, decreased subjective task-engagement, and decreased performance as a function of time block. Therefore, in a visual n-back task of 2 hours duration, peak dilation appeared to be affected by fatigue. These effects were reversed when rewards were offered in the last block (peak dilation and performance returned to their initial magnitude). This highlighted the motivational component of fatigue in task performance and effort expenditure.

However, just because peak dilation and performance were restored when rewards were presented, does not mean that individuals were no longer fatigued. If fatigue is a protective mechanism that leads to a reassessment of goals (as the Motivational Control Theory of Fatigue posits), then introducing rewards in the last block of the task may indeed have led to a reassessment of costs/benefits, which then resulted in re-engagement and effort expenditure, despite fatigue. It is possible (though, speculative) that the fatigue may return post-task and may take longer to recover from. Individuals with hearing loss may experience this post-task fatigue after effortful listening, even though they may be significantly motivated to participate in conversation and can do so without error. Daily-life fatigue (or fatigue that is hard to recover from) is commonly reported to audiologists.

In the auditory domain, McGarrigle, Rakusen, et al. (2021) examined the relationship between TEPRs, subjective effort and subjective “tiredness from listening” across six blocks, during a speech-in-noise task in a sample of individuals without hearing loss. Consistent with results reported in Hopstaken et al. (2015b), McGarrigle, Rakusen, et al. (2021) found that mean TEPRs decreased as a function of time block. Furthermore, they reported a correlation between smaller peak dilation and increased subjective “tiredness from listening” ratings in two experiments. They did not find an association between peak dilation and listening effort. This may suggest that tiredness from listening is more closely associated with peak dilation over task duration, than listening effort. These findings also suggest that subjective experiences of tiredness during listening may be reflected in peak dilation in individuals without hearing loss.

In a similar study, McGarrigle, Knight, et al. (2021) replicated the finding that tiredness from listening increases over the course of a speech-in-noise task but did not find differences between older and younger listeners. In this study, they also examined the change in mean TEPRs over the course of a speech-in-noise task using growth curve analysis. Averaged across participants, they found that mean TEPRs decreased as a function of trial number and that the decrease was more pronounced for younger adults. McGarrigle, Knight, et al. (2021) suggested that this may reflect a neural compensatory process in older listeners during a speech-in-noise task which enables prolonged cognitive effort expenditure when faced with reduced sensory and/or cognitive acuity. This could mean that older listeners exert more effort to sustain their performance in speech-in-noise tasks (even if they are feeling fatigued) when compared to younger listeners.

When Hopstaken and colleagues found a relationship between subjective fatigue and baseline diameter (Hopstaken et al., 2015a) and a relationship between subjective fatigue and peak dilation (Hopstaken et al., 2015b), they measured fatigue by explicitly asking participants “how tired do you feel?” throughout the test session, which was similar to the scale used in McGarrigle, Rakusen, et al. (2021) and McGarrigle, Knight, et al. (2021). Similarly, in Alhanbali et al. (2020), smaller baseline diameter was only significantly related to an increase in a distinct dimension of the fatigue scale related to “tiredness and drowsiness”. Therefore, there is mounting evidence that the perceptual experience of tiredness during task performance may be reflected in both baseline diameter and peak dilation.

6.5.3 Summary

In summary, there is evidence that time-on-task can affect TEPRs in listening tasks and that this effect may be associated with the subjective experience of fatigue and/or task-disengagement. Based on the literature reviewed above, smaller TEPRs might be indicative of a fatigue effect, rather than less effort. However, there is a gap in the literature pertaining to the effect of time-on-task on peak dilation, mean dilation, and baseline diameter during a speech-in-noise task at the trial-level. Furthermore, it is unknown if the physiological manifestation of fatigue in TEPRs at the trial-level coincides with subjective feelings of fatigue and performance decrement over the course of a speech-in-noise task. It is important to address these gaps as knowledge gains in these areas could facilitate implementation of standards in terms of appropriate test durations and could aid interpretation of TEPRs when examining listening effort.

7 THESIS AIMS

The studies reported in this thesis aim to contribute to the growing body of work examining TEPRs as a measure of listening effort. This is achieved by examining the effects of light level and fatigue on TEPRs during a speech-in-noise task.

The primary aim of Study 1 was to address the gap in the literature pertaining to the effects of light level on TEPRs during a speech-in-noise task and to assess the interaction between light level and SNR regarding how they affect TEPRs. Previous findings regarding the effect of light level on TEPRs are conflicting. Further, it has not been established whether light level and SNR interact during a speech-in-noise task. A secondary aim of Study 1 was to explore the use of trial-level pupil data in data analysis. Trial-level pupil data are not typically used in pupillometry analyses, but the use of more advanced statistical methods makes this possible and potentially beneficial. The research reported in Study 1 is explicitly focussed on the relationship between light level, SNR and TEPRs during a speech-in-noise task using signal-averaged pupil data (Parts A and B) and trial-level pupil data (Part C).

The aim of Study 2 was to ascertain the mechanisms that contributed to the effect of light level on pupil dilation. To achieve this, Study 2 examines the mediation effect of baseline diameter in the relationship between light level and pupil dilation using a Bayesian regression approach.

The primary aim of Study 3 was to address the gap in the literature pertaining to the effect of time-on-task on peak dilation, mean dilation and baseline diameter at the trial-level. The effect of time-on-task on TEPRs at the trial-level has not yet been established. Furthermore, it is unknown if time-on-task effects at the trial-level are consistent with subjective fatigue and task-engagement and performance during a speech-in-noise task. Study 3 (split into 2 Parts) examines the effect of time block on subjective fatigue and task-engagement, the effect of time block on performance, and the relationships between these measures (Part A) and the effect of time-on-task on peak dilation, mean dilation, and baseline diameter at the trial-level (Part B).

8 GENERAL METHODS

Participants, data collection methods, and some pre-processing methods were the same for each study reported in this thesis. Therefore, this chapter details the methods that apply to all subsequent study chapters. Methods and analyses sections specific to each study (1 – 3) are described in each individual chapter.

8.1 Participants

A priori power analyses were performed for all dependent variables (baseline diameter, peak dilation latency, peak dilation and mean dilation) using the Power Analysis and Sample Size System (PASS) (Hintze, 2015) based on pilot data collected. The power analyses indicated that the peak dilation latency parameter required the largest number of participants (31 participants) to achieve at least .8 power, at an alpha value of .05.

Convenience sampling was used to recruit participants via advertisements (Appendix 1) displayed throughout Flinders University and Medical Centre, and across Adelaide, South Australia, for one 2 hr appointment at Flinders Medical Centre. Individuals were screened prior to participation via a pre-task questionnaire (Appendix 2) and pure-tone audiometry to ensure they fit the inclusion criteria and did not meet any exclusion criteria.

Inclusion criteria:

- Aged between 18 – 40 years old
- English as first language
- No reported listening, speech, language, or cognitive difficulties
- Normal pure-tone audiometry thresholds (i.e., thresholds less than or equal to 20 dB HL at 250, 500, 1000, 2000, 4000, and 8000 Hz)

Exclusion criteria:

- Self-reported pupil abnormalities

- Self-reported issues that may affect pupil recordings

The final sample comprised 36 healthy adults (20 female) aged between 18-37 (M = 27.2, SD = 3.6) who fit the inclusion criteria. Participants received an honorarium of $20 for their participation. This research was approved by the Southern Adelaide Clinical Human Research Ethics Committee of the Southern Adelaide Local Health Network (Appendix 3).

8.2 Materials

8.2.1 BKB sentence test

The Bamford-Kowal-Bench (BKB) sentence test has been widely used in clinical audiology (Bench et al., 1979). The National Acoustic Laboratories' CDs of speech and noise for hearing aid evaluation include a recorded version of the BKB sentence test spoken with Australian English accents (Keidser et al., 2002). With permission from the National Acoustics Laboratories, the Australian version was used for the current studies.

The sentences and background noise in the Australian version of the BKB sentence test were filtered to match the international long-term average speech spectrum (Keidser et al., 2002). The background noise used in this test is a mixture of four female talkers and four male talkers (multi-talker babble). Other types of background noise that are used in speech-in-noise tasks are not as ecologically valid as multi-talker babble (e.g., white noise and speech-shaped noise) (Silbert et al., 2014). Multi-talker babble replicates noise that listener’s encounter regularly and have acoustic properties that are similar to the signal of interest (e.g., the BKB sentences) (Silbert et al., 2014). Both energetic and informational masking was anticipated to affect intelligibility of target sentences in this task. The BKB sentence test is made up of 21 lists (16 unique sentences per list). Sixteen lists were used in the current study resulting in 256 unique sentences (Appendix 4). Lists were excluded if they resulted in significantly different performance in the norms reported in Keidser et al. (2002). Participants were given the following instructions prior to the BKB sentence test:

> *“We are going to do some speech-in-noise testing while we measure changes to your pupil diameter. This can give us an indication of the amount of listening*

effort that the task requires so it is important that you try as hard as you can. We are also going to be changing the brightness of this screen to see how it affects what your pupil is doing during the task.

During the task there will be a man talking and several other talkers in the background. After 2 s of background noise, the target man will say a sentence. After the background noise stops, repeat the sentence the man says. The sentence that the man says may be easy to hear because the background noise is quieter than the man's voice or it may be difficult to hear because the background noise is louder.

Regardless of whether you think you heard the whole sentence or not, please try to guess and repeat as much of each sentence as possible as you are scored based on the number of target words that you get correct. Also, try to limit movement and blinking as much as possible, whilst keeping your eyes looking at the black dot in the middle of the screen.

If you are struggling not to blink, the best time to do so is after the response prompt, before the next trial starts.

Any questions?"

The BKB sentence test was scored according to the guidelines in the BKB sentence test manual (Keidser et al., 2002). In each list (of 16 sentences), there were 50 target words. Participants received one point for every target word they got correct.

Audacity® 2.2.2 audio editor and recorder software (Audacity Team, 2019) was used to edit the full-length BKB sentence list tracks into 16 individual, 6 s trials with the target sentence beginning 2 s[4] after the background babble noise. Additionally, a prompt tone to indicate the beginning of the target sentence was also removed because the sentences always started 2 s after the background noise.

The 16 BKB sentence lists and the corresponding background noise were mixed off-line and presented at four SNR levels (3 dB, 0 dB, -3 dB, -6 dB). This was achieved by varying the level of background noise (as was done in Keidser et al., 2002) using Audacity. The content of the BKB sentence test was otherwise unchanged. This range of SNR levels

[4]The precision with which the timing of stimuli could be specified within Audacity was estimated as +/- 0.08s.

provides a wide spectrum of difficulty and performance, based on the norms reported in Keidser et al. (2002). The speech-in-noise stimuli were presented binaurally via Sennheiser HD215 headphones at 65 dB HL.

8.2.2 Subjective fatigue measures

Subjective fatigue and task engagement were measured four times throughout the speech in noise task using visual analogue scales and asking participants to rate: “How fatigued do you feel right now?” and “How engaged in the task are you right now?”. More information on these measures and the associated methods is provided in Chapter 13.

8.3 Procedure

First, participants read a letter of introduction that outlined the purpose of the study and the procedure and provided informed, written consent. After granting consent, participants were asked a series of demographic and general health questions on the pre-task questionnaire (Appendix 2). Participation was terminated if the responses to the pre-task questionnaire revealed that the participant did not meet the inclusion criteria. Participant’s hearing was then screened to ensure they had hearing within normal limits (see Section 8.1). Subsequently, participants received instructions on the BKB sentence test and were provided with two practice trials to ensure they understood the task.

Participants were seated in front of a computer monitor (Dell E2210 22-inch Widescreen Flat Panel Monitor). There were four light level conditions. Lighting was manipulated by changing the background colour of the screen during testing (dark grey to white) (see Table 3 for colour codes in HSV). To control the approximate amount of light that reached participant’s eyes, head position was stabilised using a commercially available chin rest, positioned 60 cm away from the computer screen for all participants. Light level measurements were taken at the beginning of each test day from a standard mark on the chin rest at approximate eye level using a Digitech light meter (QM1587). Measurements were taken with the light meter's sensor facing the computer screen, 60 cm away from the computer screen (see Table 3 for light level measurements).

Table 3. Light level Measurements Measured in Lux and Associated HSV Codes

Light level	Average lux (SD)	HSV colour code (computer screen colour)
L1	21.98 (1.29)	0, 0, 0.1
L2	41.07 (2.71)	0, 0, 0.4
L3	65.31 (3.17)	0, 0, 0.6
L4	95.14 (0.71)	0, 0, 1

Speech-in-noise stimuli and light levels were presented using custom software written in PsychoPy (Peirce et al., 2019). The ambient illumination in the room stayed constant (i.e., the computer screen was the only light source used to manipulate light levels). The computer screen was a solid colour (depending on the light level condition), apart from a black dot of approximately 2 mm diameter at its centre which remained throughout the testing. Participants were instructed to keep their gaze on the dot during pupil recordings and to limit blinking during the sentence trials. Each condition began with a 10 s accommodation period.

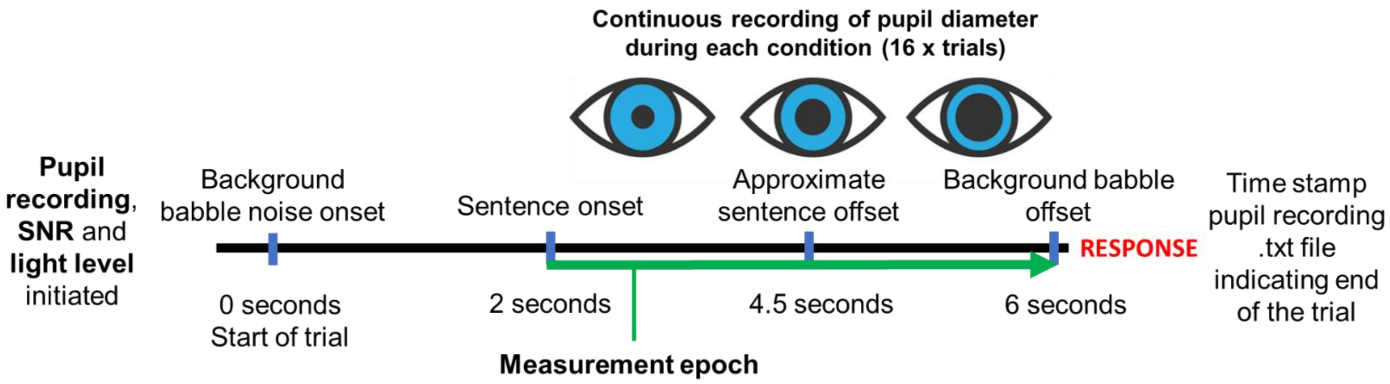
The BKB sentence test was divided into four blocks based on the four different light levels (L1, L2, L3, L4). Every participant received the same 16 lists from the BKB sentence test. Each light level block included four BKB sentence lists presented at each SNR. Sentence list presentation was counterbalanced across participants using a Latin square design (Appendix 5). This ensured that participants did not receive the sentence lists in the same order. The order of light level and order of SNR presentation within each light level were also counterbalanced using a Latin square design (64 trials per block) (Appendix 6). This ensured that participants did not receive light levels or SNR conditions in the same order. After each block of four sentence lists, participants rated their level of fatigue and task-engagement using a visual analogue scale (details of this portion of the research are presented in Chapter 13).

Figure 8 depicts a single trial of the speech-in-noise task. Once the light level and SNR had been selected, the pupil recording began, marking the beginning of each condition.

Within each trial, the background babble noise commenced first and continued for the duration of the trial (6 s). At the 2 s point, the unique sentence stimuli were presented. Each sentence lasted approximately 2.5 s. Therefore, the background noise continued for approximately 1.5 s after sentence offset. Participants were required to respond only once the background noise stopped, that is, the cessation of the background noise was used as the prompt to respond. A key press by the experimenter indicated the end of the trial and a time stamp was automatically recorded in the pupil recording .txt file. There was a slight pause (approximately 3 s) from one trial to the next to allow the pupil to return to its baseline diameter. These trials were repeated 16 times (using unique BKB sentence stimuli) for each of the 16 light level by SNR conditions. The speech-in-noise testing lasted approximately 1 hr.

Figure 8

Example of One Trial in the Speech-in-Noise Task



Note. Trials were repeated 16 times per condition. Each participant completed 256 (using unique sentences) trials over 16 (light level x SNR) conditions.

8.3.1 Pupillometry

All raw pupil data are available at https://osf.io/am6uv/.

Pupil diameters of left and right eyes were recorded using the SMI RED-m (SensoMotoric Instruments, Berlin, Germany) eye tracking system, with a sampling rate of 60 Hz per second. Pupil diameter was recorded continuously for each condition. Only the pupil data recorded between noise onset and noise offset was retained for processing (Figure 8).

This resulted in a total measurement epoch of 6 s per trial. Because the eye-tracker's sample rate was set at 60 Hz, pupil diameter was measured 360 times per trial.

Participants were told that they could blink during the verbal response period, as pupil diameter recorded during this time was not included in the measurement epoch. The experimenter visually monitored the participants and the quality of the pupil capture during the testing. If participants were noticeably moving and/or excessively blinking, participants were reminded to limit movement and blinking as much as possible during the recording.

Data are presented for the left pupil only.

8.4 Pre-Processing of Pupil Data

Pupil data were pre-processed in R open source software (version 4.0.1) (R Core Team, 2020) and R Studio (version 1.3.1073) (RStudio Team, 2020). The package *GazeR* was used (Geller et al., 2020). Prior to pre-processing, adjustments to data configuration were made to facilitate the use of *GazeR's* functions. These adjustments were completed using Visual Basic in Excel. The adjustments involved allocating each pupil measurement sample a subject ID (a unique identifier indicating the participant/condition combination [PCC])[5] and additional identifying information. A copy of the Visual Basic macro and the R code for pre-processing are available at https://osf.io/am6uv/. Modifications to the following functions within *GazeR* were made: extend_blinks, speed_pupil, smooth_interpolate_pupil, peak_pupil_dilation, count_missing_pupil. The modified functions can also be found found https://osf.io/am6uv/. Geller et al. (2020) was used for guidance in the pre-processing strategy, however, there were adjustments to the order in which some steps were applied. These adjustments are detailed below.

8.4.1 Initial Missing Data Deletion

[5] The PCC label format: Participant number_light level_SNR. For example, P24_ L4_SNR0 means participant 24, in L4, in SNR 0 dB.

One PCC was missing from the raw dataset due to an experimenter error during the testing which resulted in no pupil recording for that PCC (P24_ L4_SNR0). Additionally, the SMI eye-tracking system malfunctioned on some occasions during the testing. This resulted in no pupil measurements being recorded for some PCCs. PCCs that recorded no pupil measurements over the 16 trials of that condition were removed from the full data set. This resulted in the removal of 10 PCCs[6]. Data was visually inspected and one more PCC (P17_L3_SNR3) had to be removed due to experimenter error (wrong light setting applied). The baseline diameter for this PCC was too large to have been measured in L3. Trial 2 was removed from P13_L4_SNR3 as the eye tracker did not record any pupil measurements over the 6 s trial period. In these initial processing stages, 177 trials were removed from the dataset (though most of these did not contain data), leaving 98.07% of raw pupillometric data remaining.

8.4.2 De-Blinking

Eye blinks can introduce significant artefacts to raw pupil data (Geller et al., 2020). During eye blinks, the eyelids obstruct the pupil as they close. As pupil diameter becomes increasingly more obstructed by the eyelids, the eye tracker’s pupil size measurement will rapidly decrease until it is 0 mm (indicating loss of tracking, maybe due to a fully closed eye during a blink). As the eyelids open again, the eye tracker’s pupil diameter measurement will rapidly increase until it has full “capture” of the pupil diameter again. Because of this, all blinks will be accompanied by erroneous values on either side of the eye closure as eyelid obstruction increases and decreases during blinks. These aberrations can be accounted for by “de-blinking” the pupil diameter data. De-blinking involves identifying the blinks, removing data during the blink, removing data on either side of the blink, and interpolating the data across the removed data. The *GazeR* function “extend_blinks” was edited and “mod_extend_blinks” was made to suit the specific needs of this dataset. This function identifies blinks by the presence of NAs in the raw data. As recommended, segments of traces that were identified as blinks were extended by 100 ms

[6] Missing PCCs: P01_L1_SNR3, P03_L4_SNR-6, P04_L1_SNR-6, P12_L4_SNR-6, P16_L2_SNR3, P25_L3_SNR3, P28_L3_SNR-6, P30_L1_SNR-6, P30_L3_SNR0, P32_L1_SNR3.

before and after the blink to replace the erroneous pupil diameter measurements caused by partial obstruction of the pupil during blinks (Geller et al., 2020).

The next steps in pupil pre-processing are usually to smooth and interpolate across the extended blinks. *GazeR* combines these steps in a function called "smooth_interpolate_pupil". This function was edited to suit the current study and was called "mod_smooth_interpolate_pupil". The name of the resulting vector was changed, and a "group_by" statement was added to ensure that interpolation did not occur across trials.

Pupil data are typically noisy. Smoothing raw pupil data is a common way of reducing some of the noise in the pupil signal. A 5-point moving average to smooth the raw pupil data was used. In the function, 5-points represents the total number of points used in the averaging (5 = point of interest plus two points either side). Segments of the smoothed pupil trace that contained the extended blinks were reconstructed with linear interpolation (Geller et al., 2020). Linear interpolation uses linear polynomials to reconstruct data within a range of known data points.

8.4.3 Outlier and Artefact Rejection

8.4.3.1 Outliers

In the next stage of pre-processing, pupil data that were outside of the feasible range of the pupil were removed (Mathôt et al., 2018). Following Mathôt et al. (2018), a histogram of the entire dataset was visually inspected to determine unlikely pupil diameter values. Pupil diameter data outside the range of 2 mm and 8 mm were changed into NAs. This is slightly more conservative than the 1.5 mm – 9 mm range suggested in Kret and Sjak-Shie (2019). However, in line with the range identified in the visual inspection of these data, the healthy adult pupil should range from 2 mm to 8 mm in diameter (Spector, 1990). Using this range, 263 outliers (0.008% of data) were removed from the dataset. The outlier data were examined to ensure that the outliers did not occur exclusively in the brightest light level (potentially indicating measurements of extremely constricted pupils due to the light manipulation). There was no apparent relationship between light level and data points deemed outliers. The majority of outliers were well beyond the likely pupil diameter range. Linear interpolation was used to reconstruct the NAs that were introduced by removing outliers. The function “na_interpolation” from the package *imputeTS* (Moritz & Bartz-Beielstein, 2017) was used for this interpolation as the data did not need to be smoothed again, and the *GazeR* function combines these steps.

8.4.3.2 Median Absolute Deviation Violations

Sometimes the change in pupil diameter can happen too rapidly to be considered an accurate pupil measurement. Median Absolute Deviation (MAD) can be used to detect these artefacts (Geller et al., 2020). The *GazeR* function “speed_pupil” calculates the normalised dilation speed for each time point (absolute change in pupil size between samples divided by the temporal separation between them). The dilation speed variable is multiplied by a constant to detect MAD violations. The constant informs the sensitivity threshold[7]. Larger constants require more extreme values to be marked as a MAD violation. The *GazeR* function “calc_mad” adds the MAD values to the median dilation speed variable and converts values that exceed the threshold to NAs. A modified version of the “speed_pupil” function was used for the current data, “mod_speed_pupil”. In this modified version, the maximum of the absolute values of the backwards and forwards pupil change was measured, rather than just the maximum. By selecting the maximum, the slower of the two speeds would have been picked when the pupil is constricting and negative[8]. 3497 MAD violations (0.1%) were identified in the data and were converted to NAs. Linear interpolation (na_interpolation, from the *imputeTS* package) was used to reconstruct the NAs that were introduced by removing MAD violations.

8.4.3.3 Data Removal

The next step in the pre-processing strategy was to count all the NAs introduced in the above steps (extended blinks, outliers, MAD violations) and remove any trials or PCCs which had more than 20% missing data (Winn et al., 2018). Excessive missing data (due to removal in previous stages) can indicate poor data quality. Therefore, it is recommended that trials and PCCs with excessive missing data are removed (Geller et al., 2020; Winn et al., 2018). To count missing data and remove threshold breaches, a modified version (“mod_count_missing_pupil”) of *GazeR’s* “count_missing_pupil” function was used. Only the print messages were modified, to be reflective of the unique conditions in the dataset. The actual function code did not change. Four PCCs (P33_L3_SNR0, P33_L3_SNR3, P33_L4_SNR0, P33_L4_SNR3) breached the PCC threshold and were

[7] The constant used in this case was 16, in line with Geller et al (2020).
[8] This issue was raised with Geller et al (2020) in personal communications and was rectified in a later version of GazeR.

removed and 244 individual trials breached the individual trial threshold and were removed for having too many missing values. This resulted in 3.1% of data being discarded.

9 STUDY 1A - THE EFFECTS OF SNR AND LIGHT LEVEL ON TEPRS DURING A SPEECH-IN-NOISE TASK

9.1 Background and Aims

The aim of Study 1A was to investigate the effects of light level and SNR on TEPRs during the BKB sentence test. Based on established effects, it was expected that baseline diameter would decrease in brighter light level. In accordance with previous findings (reviewed in Zekveld et al., 2018), it was also expected that peak and mean dilation would be larger and peak dilation latency would be greater in the more adverse SNR conditions. Furthermore, it was expected that the performance would decline systematically as a function of SNR.

The background literature regarding the effect of light level on TEPRs is conflicting (Section 6.4). The interaction effect of light level and task difficulty on TEPRs has not yet been examined in a listening effort paradigm. Therefore, this study aimed to fill that gap. Due to the similarities in methods used to manipulate light level in the current thesis and in Peysakhovich et al. (2015) and Peysakhovich et al. (2017), it is possible that the findings reported here will align with those. However, the tasks that were used in Peysakhovich et al. (2015) and Peysakhovich et al. (2017) were different to the speech-in-noise task used here. The task used in the current study was more similar to the task used in Wang, Kramer, et al. (2018). Due to the variability in methods and outcomes in the literature, it was not possible to confidently predict effects for peak and mean dilation. The effects of light level, SNR and the interaction were tested for peak and mean dilation, peak dilation latency, and baseline diameter.

9.2 Significance

Understanding the possible effect of light level on TEPRs during a clinically relevant speech-in-noise task will provide valuable information for the potential use of pupillometry as a clinical measure of listening effort and for continuing research in the area. A deeper understanding of factors that affect pupil dilation in listening tasks will shed light on the suitability of the measure for clinical use which, ultimately, may promote advancements in the treatment of hearing difficulties.

9.3 Methods

9.3.1 Participants

Information related to the participants can be found in Chapter 8 - General Methods, Section 8.1.

9.3.2 Materials

Information related to the materials can be found in Chapter 8 - General Methods, Section 8.2.

9.3.3 Procedure

Information related to the procedure can be found in Chapter 8 - General Methods, Section 8.3.

9.3.4 Pupil Data Pre-Processing

Information related to the pre-processing can be found in Chapter 8 - General Methods, Section 8.4. R code for the additional pre-processing used in Study 1A can be found https://osf.io/am6uv/.

9.3.4.1 Signal-Averaging

After the initial pre-processing procedures described in the General Methods section (Chapter 8), pupil data were signal-averaged; that is, time-locked pupil measurements within trials for a given PCC were averaged across the 16 trials to give a single average pupil trace for each condition (light level by SNR).

9.3.4.2 Baseline Correction

After signal-averaging, pupil traces were baseline-corrected. Baseline-corrected pupil data allows the reporting of the change in pupil size from a specific point in the pupil trace (in this case, sentence onset). The subtractive baseline correction method was employed (Mathôt et al., 2018). Baseline diameter was defined as the median[9] pupil size between 1.5 s and 2 s[10] of the 6 s measurement epoch. This baseline diameter was then subtracted from each pupil measurement resulting in the baseline-corrected pupil measurements. Because the effects of light level were being tested and light level is known to affect baseline diameter, a baseline diameter variable was also created.

9.3.4.3 Pupil Parameters

Mean dilation, peak dilation and peak dilation latency parameters were obtained from the same measurement epoch: 2 s to 6 s (sentence onset to noise offset/response prompt). This measurement epoch excludes participant's verbal responses to the stimuli, avoiding potential confounds related to pupillary motor responses. Within the baseline-corrected, averaged trace, the pupil parameters were defined as follows: (1) mean dilation was the

[9] The definition of baseline diameter was based upon recommendations in Mathôt (2018) and Geller et al (2020).

[10] Sentence onset occurred at 2 seconds in the measurement epoch. Any change in pupil size due to the 2 s of background noise was accounted for in the baseline correction.

average dilation in the measurement epoch, (2) peak dilation was the largest dilation value in the measurement epoch, and (3) peak dilation latency the time (in seconds) it took for the pupil to reach peak dilation after sentence onset.

9.3.4.4 Missing Data

After pre-processing, pupil values from 16 PCCs were missing. To rule out the presence of any influential patterns in the missingness of the data, Little's Missing Completely at Random Test (MCAR) was run on mean dilation, peak dilation, peak dilation latency, and baseline diameter (Little, 1988) using an R function written by Stemmler (2020). Little's MCAR test was non-significant ($p = > .05$). This indicated that the missing data were missing completely at random, and imputation could be used to generated values for the missing data to create a balanced dataset.

To avoid list-wise deletion in the analyses and retain as much of the data as possible, the "ImputeEM" from the package *mvdalab* 1.4 (Afanador et al., 2021) was used for imputation of missing values. This function uses an Expectation Maximisation algorithm to impute the most likely values where data are missing.

9.3.5 Data Analyses

Data analyses and visualisations were completed in R open source software (version 4.0.1) (R Core Team, 2020) and R Studio (version 1.3.1073) (RStudio Team, 2020). Analysis and visualisation codes are available at https://osf.io/am6uv/.

Data visualisations were completed using *ggplot2* (version 3.3.3) (Wickham, 2016). Assumption checks were completed using the *rstatix* package (version 0.7.0) (Kassambara, 2021). To investigate if light levels affected TEPRs during the BKB sentence test with varying SNR conditions, five 4x4 repeated-measures analyses of variance (RANOVAs) were performed on performance and each pupil parameter: baseline diameter, peak dilation latency, peak dilation, and mean dilation, using the *ez* Package (version 4.4-0) (Lawrence, 2016). The Greenhouse-Geisser correction was used to correct for violations of sphericity. Post hoc analyses were performed where appropriate to examine potential significant differences between levels of both independent variables.

9.4 Results

9.4.1 Assumption Checks

Examination of studentised residuals identified a few outliers in the data and the Shapiro-Wilks Test of the studentised residuals revealed that some data showed non-normal distributions. However, as the design was balanced, used a sample size larger than n = 12, and scores were roughly symmetrical, a p value of < .025 was used to correct for outliers and non-normality (Keppel, 1991).

9.4.2 Performance

Means and standard deviations for performance on the BKB sentence test are presented in Table 4. RANOVA results are presented in Table 5. There was no main effect of light level on performance. There was a significant main effect of SNR on performance. As expected, more adverse SNRs resulted in significantly worse performance on the BKB sentence test. There was also a significant interaction between light level and SNR in their effects on performance (Table 5).

Table 4. Means and Standard Deviations for Performance as a Function of a 4(light level) X 4(SNR) Design

	SNR (dB)							
	-6		-3		0		3	
Light level	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
L1	4.28	2.44	17.75	5.56	37.25	4.65	46.97	1.95
L2	3.44	1.84	16.50	6.65	37.81	4.87	46.92	1.90
L3	4.36	3.04	19.97	6.32	36.78	5.96	46.14	2.51
L4	5.19	3.47	19.64	4.60	36.69	5.38	45.81	2.15

Note. M and *SD* represent mean and standard deviation, respectively. Scores represent average words correct out of 50.

Table 5. RANOVA Results for BKB Performance as a Function of Light Level and SNR (dB)

Predictor	df_{Num}	df_{Den}	*Epsilon*	*F*	*p*	η^2_g
Light level	2.38	83.34	0.79	0.55	.608	.00
SNR	2.43	85.19	0.81	3198.74	<.001	.94
Light level x SNR	5.67	198.31	0.63	3.36	.004	.03

Note. Df_{Num} indicates degrees of freedom numerator. Df_{Den} indicates degrees of freedom denominator. Epsilon indicates Greenhouse-Geisser multiplier for degrees of freedom, *p value*s and degrees of freedom in the Table incorporate this correction. H^2_g indicates generalized eta squared.

9.4.2.1 Post Hoc Analyses

Post hoc simple effects analyses were conducted to examine performance differences between SNRs within each light level and light levels within each SNR level. After applying the Bonferroni correction for 48 multiple comparisons (alpha is significant at $p = .001$), there were significant differences in performance between all SNRs within light levels ($p < .001$) and there were no significant differences in performance in any of the comparisons of light level within SNR levels ($p > .001$).

9.4.3 Baseline Diameter

Means and standard deviations for baseline diameter are presented in Table 6 and Figure 9. RANOVA results are presented in Table 7. There were significant main effects of light level and SNR on baseline diameter. Brighter light levels resulted in smaller baseline diameter and baseline diameter was larger in more adverse SNR conditions. However, there was also a significant interaction between light level and SNR in their effects on baseline diameter.

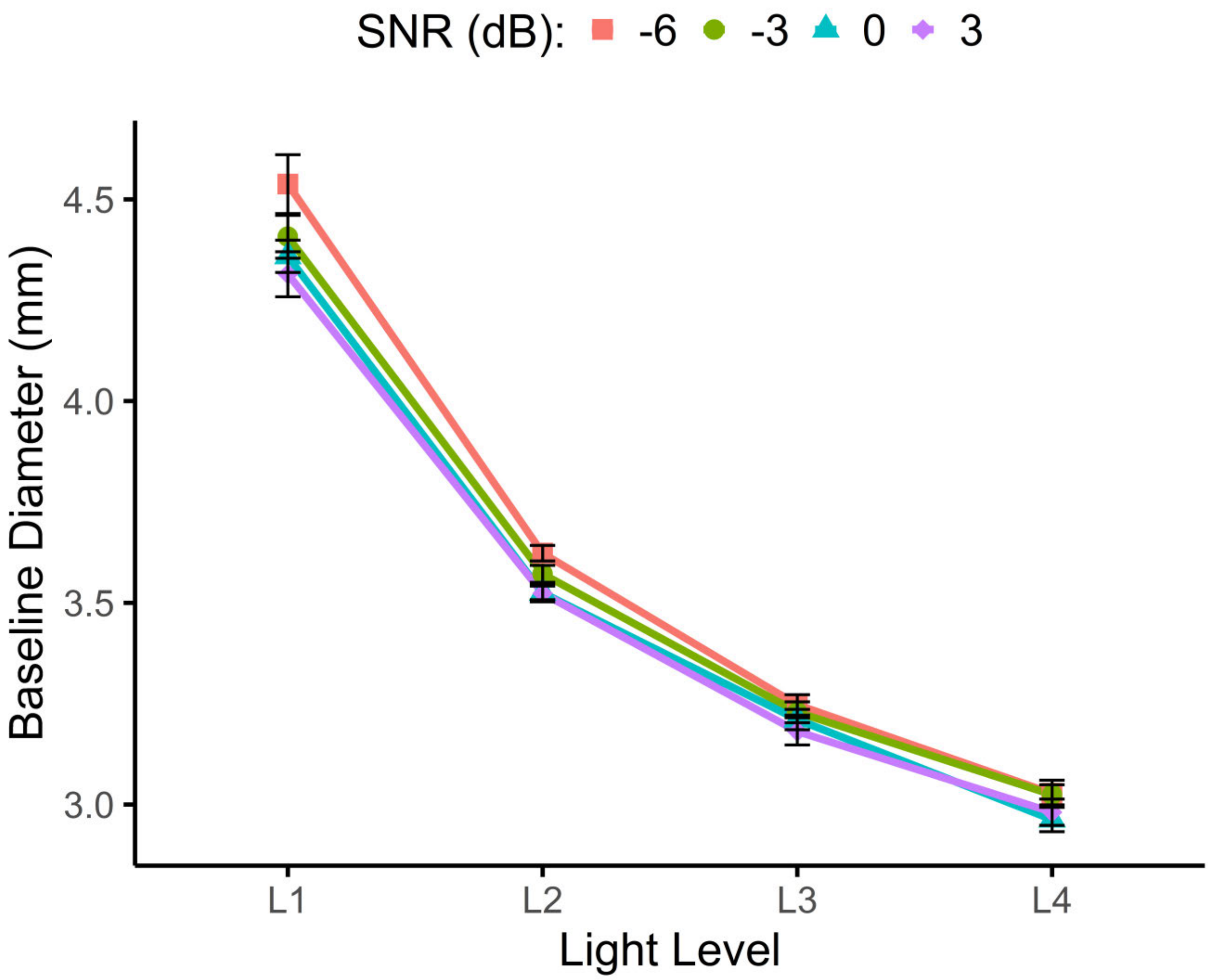
Table 6. Means and Standard Deviations for Baseline Diameter as a Function of a 4(light level) X 4(SNR) Design

	SNR (dB)							
	-6		-3		0		3	
Light level	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
L1	4.54	0.69	4.41	0.66	4.36	0.57	4.31	0.63
L2	3.62	0.41	3.57	0.43	3.52	0.39	3.52	0.36
L3	3.25	0.35	3.23	0.33	3.21	0.35	3.18	0.29
L4	3.03	0.34	3.02	0.30	2.96	0.33	2.98	0.32

Note. M and *SD* represent mean and standard deviation, respectively.

Figure 9

Mean Baseline Diameter



Note. This figure shows mean baseline diameter (within-subjects standard error bars) across all participants in the 16 light level by SNR conditions. Brightness successively increases from L1 (dimmest) to L4 (brightest).

Table 7. RANOVA Results for Baseline Diameter as a Function of Light Level and SNR (dB)

Predictor	df_{Num}	df_{Den}	*Epsilon*	*F*	*p*	η^2_g
Light level	1.28	44.82	0.43	351.69	<.001	.60
SNR	2.23	78.07	0.74	12.82	<.001	.01
Light level x SNR	4.31	150.76	0.48	2.45	.044	.00

Note. df_{Num} indicates degrees of freedom numerator. df_{Den} indicates degrees of freedom denominator. Epsilon indicates Greenhouse-Geisser multiplier for degrees of freedom, *p value*s and degrees of freedom in the Table incorporate this correction. η^2_g indicates generalized eta squared.

9.4.3.1 Post Hoc Analyses

Due to the well-known effect of light level on baseline diameter, post hoc simple effects analyses were conducted to examine baseline diameter differences between SNR levels within each light level. After applying the Bonferroni correction for 24 multiple comparisons (alpha is significant at $p = .002$), baseline diameter in SNR -6 dB significantly differed from SNR 3 dB and SNR 0 dB at L2. SNR -6 dB and SNR 3 dB also significantly differed at L1. All other comparisons were not significant (Figure 9).

9.4.4 Peak Dilation Latency

Means and standard deviations for peak dilation latency are presented in Table 8 and Figure 10. RANOVA results are presented in Table 9. There were significant main effects of light level and SNR on peak dilation latency. The interaction term was not significant.

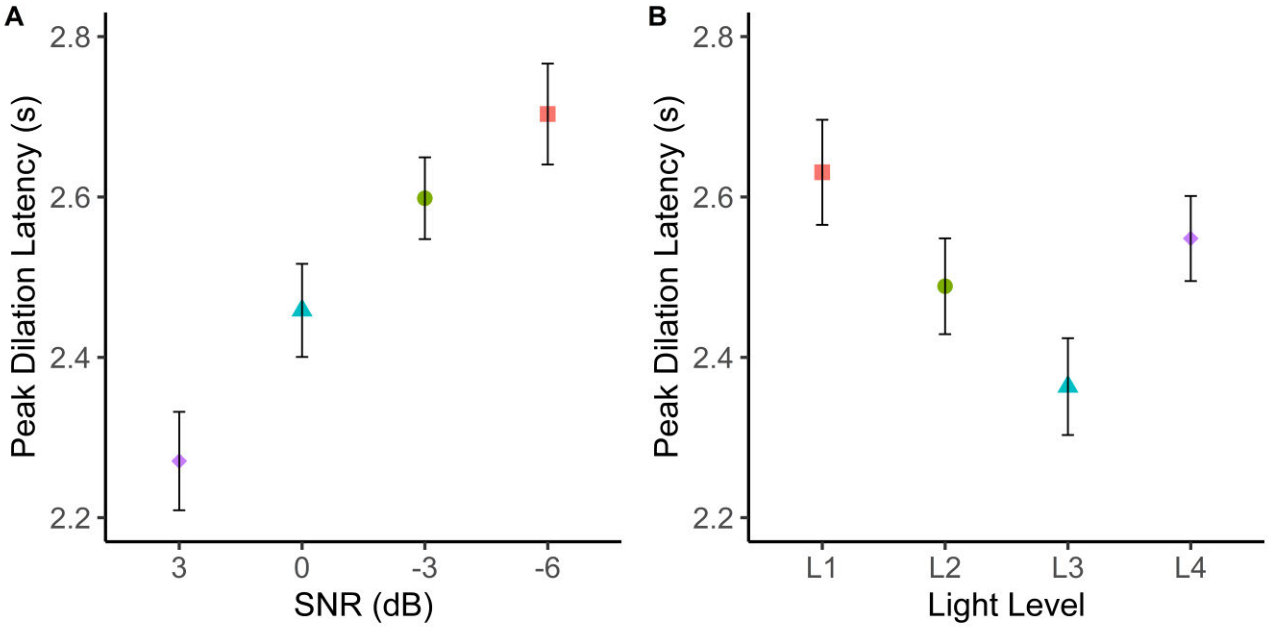
Table 8. Means and Standard Deviations for Peak Dilation Latency (s) as a Function of a 4(light level) X 4(SNR) design

	SNR (dB)							
	-6		-3		0		3	
Light level	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
L1	2.92	0.83	2.79	0.64	2.57	0.71	2.25	0.73
L2	2.66	0.82	2.53	0.63	2.46	0.69	2.31	0.68
L3	2.58	0.77	2.48	0.57	2.26	0.62	2.13	0.76
L4	2.65	0.65	2.60	0.55	2.55	0.56	2.39	0.63

Note. M and *SD* represent mean and standard deviation, respectively.

Figure 10

Mean Peak Dilation Latency



Note. Figure 10 A shows mean peak dilation latency (s) (within-subjects standard error bars) across all participants for SNR (dB). Figure 10 B Mean peak dilation latency (s) (within-subjects standard error bars) across all participants for light level. Brightness successively increases from L1 (dimmest) to L4 (brightest).

Table 9. RANOVA Results for Peak Dilation Latency as a Function of Light Level and SNR (dB)

Predictor	df_{Num}	df_{Den}	*Epsilon*	*F*	*p*	η^2_g
Light level	2.70	94.65	0.90	3.85	.015	.02
SNR	2.70	94.50	0.90	10.09	<.001	.05
Light level x SNR	6.21	217.23	0.69	0.81	.565	.01

Note. df_{Num} indicates degrees of freedom numerator. df_{Den} indicates degrees of freedom denominator. Epsilon indicates Greenhouse-Geisser multiplier for degrees of freedom, *p values* and degrees of freedom in the Table incorporate this correction. η^2_g indicates generalized eta squared.

9.4.4.1 Post Hoc Analyses

Post hoc main effects analyses were conducted to examine significant differences in peak dilation latency between SNR levels (Table 10), and between light levels (Table 11). Bonferroni correction for 12 comparisons (alpha is significant at $p = .005$) was applied.

Main effects analyses showed that peak dilation latency increased as a function of SNR. More adverse SNR conditions resulted in longer peak dilation latency values (Figure 10 A). All comparisons were significantly different, except the comparisons SNR -6 dB and SNR -3 dB, and SNR -3 dB and SNR 0 dB.

Main effects analyses also showed that peak dilation latency was affected by light level (Figure 10 B). Only comparisons between L1 and L3, and L3 and L4 showed significant differences. L1 (the dimmest condition) showed the largest peak dilation latency and L4 (the brightest condition) showed the next largest. L3 resulted in the shortest peak dilation latency.

Table 10. Post Hoc Main Effects Analyses for Peak Dilation Latency by SNR (dB)

Comparison (SNR)		Mean difference (s)	95% CI – low	95% CI – high	Pairwise *t* statistic	*p*	*Sig.*
-6	-3	0.1	-0.03	0.24	1.51	.13	
-6	0	0.24	0.10	0.39	3.23	.002	*
-6	3	0.43	0.28	0.58	5.78	<.001	*
-3	0	0.14	0.01	0.27	2.05	.04	
-3	3	0.33	0.18	0.47	4.49	<.001	*
0	3	0.19	0.06	0.32	2.9	.004	*

Note. p is not Bonferroni corrected. * in Sig. column indicates significant differences after correcting for multiple comparisons (e.g., *p* is significant at 0.005). Degrees of freedom for all comparisons = 143. CI = confidence interval.

Table 11. Post Hoc Main Effects Analyses for Peak Dilation Latency by Light Level

Comparison (light level)		Mean difference (s)	95% CI – low	95% CI – high	Pairwise *t* statistic	*p*	*Sig.*
1	2	0.14	-0.01	0.29	1.88	.06	
1	3	0.27	0.13	0.41	3.72	<.001	*
1	4	0.08	-0.07	0.23	1.1	.27	
2	3	0.13	-0.01	0.26	1.8	.07	
2	4	-0.06	-0.19	0.07	-0.88	.38	
3	4	-0.18	-0.31	-0.06	-2.93	.004	*

Note. p is not Bonferroni corrected. * in Sig. column indicates significant differences after correcting for multiple comparisons (e.g., *p* is significant at 0.005). Degrees of freedom for all comparisons = 143. CI = confidence interval

9.4.5 Peak Dilation

Means and standard deviations for peak dilation are presented in Table 12 and Figure 11. RANOVA results are presented in Table 13. As expected, there were significant main effects of light level and SNR on peak dilation. Peak dilation was smaller in brighter light levels and larger in more adverse SNR conditions. However, there was also a significant interaction between light level and SNR in their effects on peak dilation.

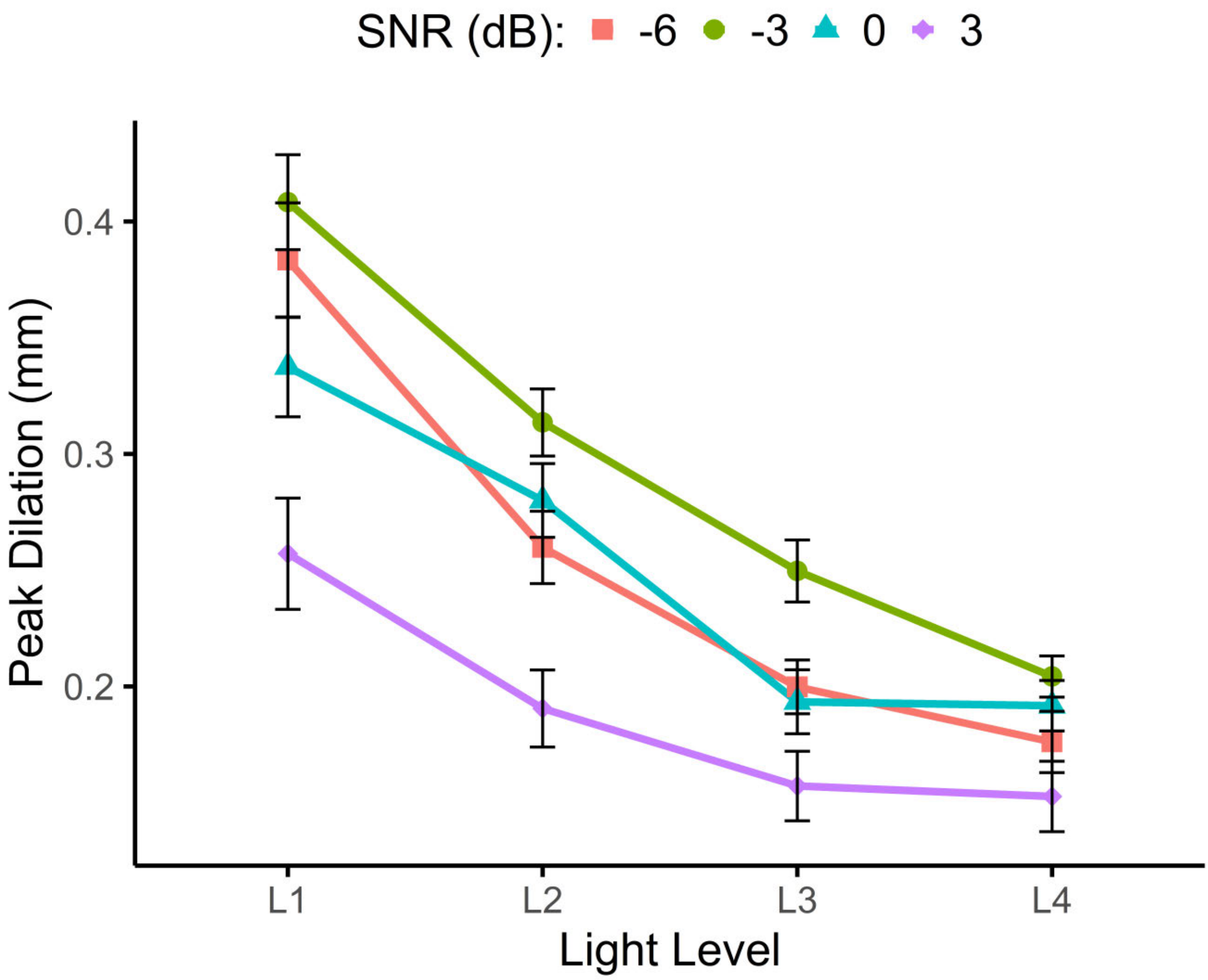
Table 12. Means and Standard Deviations for Peak Dilation (mm) as a Function of a 4(Light Level) X 4(SNR) Design

	SNR (dB)							
	-6		-3		0		3	
Light level	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
1	0.38	0.20	0.41	0.19	0.34	0.18	0.26	0.16
2	0.26	0.14	0.31	0.15	0.28	0.14	0.19	0.10
3	0.20	0.11	0.25	0.12	0.19	0.10	0.16	0.11
4	0.18	0.10	0.20	0.09	0.19	0.09	0.15	0.10

Note. M and *SD* represent mean and standard deviation, respectively.

Figure 11

Mean Peak Dilation



Note. This figure shows mean peak dilation (within-subjects standard error bars) across all participants in the 16 light level by SNR (dB) conditions. Brightness successively increases from L1 (dimmest) to L4 (brightest).

Table 13. RANOVA Results for Peak Dilation as a Function of light level and SNR (dB)

Predictor	df_{Num}	df_{Den}	*Epsilon*	*F*	*p*	η^2_g
light level	2.01	70.46	0.67	36.68	<.001	.19
SNR	1.72	60.24	0.57	22.72	<.001	.07
light level x SNR	5.89	206.19	0.65	3.71	.002	.02

Note. Df_{Num} indicates degrees of freedom numerator. Df_{Den} indicates degrees of freedom denominator. Epsilon indicates Greenhouse-Geisser multiplier for degrees of freedom, *p value*s and degrees of freedom in the Table incorporate this correction. H^2_g indicates generalized eta squared.

9.4.5.1 Post Hoc Analyses

Post hoc simple effects analyses were conducted to examine significant differences in peak dilation between SNR levels within each light level (Table 14), and between each light level within each SNR level (Table 15). The Bonferroni correction for 48 comparisons (alpha is significant at $p = .001$) was applied.

Brighter light levels typically resulted in smaller peak dilation and more adverse SNR conditions typically resulted in larger peak dilation (Figure 10). There were more significant differences in peak dilation between SNR levels in the two dimmest light levels (1 and 2) than there were in more bright light levels (3 and 4) (Table 14).

Additionally, there were more significant differences in peak dilation between light levels in SNR levels that demanded more listening effort (SNR -6 dB and SNR -3 dB) than there were in SNR levels that demanded less listening effort (SNR 0 dB). There was only 1 significant difference in peak dilation in the least demanding SNR level (SNR 3 dB) (Table 15).

Table 14. Post Hoc Simple Effects Analyses for Peak Dilation – Light Level as Grouping Variable

Light Level	Comparison (SNR dB)		Mean difference (mm)	95% CI – low	95% CI – high	Pairwise *t* statistic	*p*	*Sig.*
L1	-6	-3	-0.02	-0.07	0.02	-1.16	.25	
L1	-6	0	0.05	-0.01	0.10	1.73	.09	
L1	-6	3	0.13	0.06	0.19	3.98	<.001	*
L1	-3	0	0.07	0.02	0.12	2.69	.01	
L1	-3	3	0.15	0.09	0.22	4.73	<.001	*
L1	0	3	0.08	0.04	0.12	3.83	<.001	*
L2	-6	-3	-0.05	-0.08	-0.03	-3.90	<.001	*
L2	-6	0	-0.02	-0.06	0.02	-0.92	.36	
L2	-6	3	0.07	0.02	0.12	2.86	.007	
L2	-3	0	0.03	0.00	0.07	2.06	.05	
L2	-3	3	0.12	0.08	0.17	5.36	<.001	*
L2	0	3	0.09	0.05	0.12	5.22	<.001	*
L3	-6	-3	-0.05	-0.08	-0.02	-3.43	.002	
L3	-6	0	0.01	-0.03	0.04	0.37	.71	
L3	-6	3	0.04	0.01	0.08	2.42	.02	
L3	-3	0	0.06	0.03	0.09	3.70	<.001	*

Light Level	Comparison (SNR dB)		Mean difference (mm)	95% CI – low	95% CI – high	Pairwise *t* statistic	*p*	*Sig.*
L3	-3	3	0.09	0.06	0.12	6.61	<.001	*
L3	0	3	0.04	0.01	0.06	2.88	.007	
L4	-6	-3	-0.03	-0.05	0.00	-2.15	.04	
L4	-6	0	-0.02	-0.04	0.01	-1.11	.28	
L4	-6	3	0.02	-0.01	0.05	1.55	.13	
L4	-3	0	0.01	-0.01	0.03	1.16	.25	
L4	-3	3	0.05	0.03	0.08	4.22	<.001	*
L4	0	3	0.04	0.01	0.06	3.12	.004	

Note. p is not Bonferroni corrected. * in *Sig.* column indicates significant differences after correcting for multiple comparisons (e.g., *p* is significant at 0.001). Degrees of freedom for all comparisons = 35.

Table 15. Post Hoc Simple Effects Analyses for Peak Dilation – SNR as Grouping Variable

SNR (dB)	Comparison (light level)		Mean difference (mm)	95% CI – low	95% CI – high	Pairwise *t* statistic	*p*	*Sig.*
-6	L1	L2	0.12	0.07	0.17	4.89	<.001	*
-6	L1	L3	0.18	0.13	0.24	6.51	<.001	*
-6	L1	L4	0.21	0.15	0.26	7.30	<.001	*
-6	L2	L3	0.06	0.02	0.09	3.48	.001	*
-6	L2	L4	0.08	0.05	0.12	4.54	<.001	*
-6	L3	L4	0.02	0.00	0.05	1.86	.07	
-3	L1	L2	0.09	0.05	0.14	4.52	<.001	*
-3	L1	L3	0.16	0.10	0.21	5.98	<.001	*
-3	L1	L4	0.20	0.16	0.25	8.70	<.001	*
-3	L2	L3	0.06	0.03	0.10	3.55	.001	*
-3	L2	L4	0.11	0.07	0.15	5.82	<.001	*
-3	L3	L4	0.05	0.02	0.07	3.39	.002	
0	L1	L2	0.06	0.00	0.11	2.03	.05	
0	L1	L3	0.14	0.09	0.20	5.21	<.001	*
0	L1	L4	0.15	0.09	0.20	5.65	<.001	*
0	L2	L3	0.09	0.05	0.12	4.76	<.001	*

SNR (dB)	Comparison (light level)		Mean difference (mm)	95% CI – low	95% CI – high	Pairwise *t* statistic	*p*	*Sig.*
0	L2	L4	0.09	0.05	0.13	4.34	<.001	*
0	L3	L4	0.00	-0.03	0.03	0.11	.91	
3	L1	L2	0.07	0.01	0.12	2.39	.02	
3	L1	L3	0.10	0.04	0.16	3.41	.002	
3	L1	L4	0.10	0.04	0.17	3.47	.001	*
3	L2	L3	0.03	0.00	0.07	1.90	.07	
3	L2	L4	0.04	0.00	0.08	1.83	.08	
3	L3	L4	0.00	-0.03	0.04	0.25	.8	

Note. p is not Bonferroni corrected. * in *Sig.* column indicates significant differences after correcting for multiple comparisons (e.g., p is significant at 0.001). Degrees of freedom for all comparisons = 35.

9.4.6 Mean Dilation

Means and standard deviations for mean dilation are presented in Table 16 and Figure 12. RANOVA results are presented in Table 17. As expected, there were significant main effects of light level and SNR on mean dilation. Mean dilation was smaller in brighter light levels and larger in more adverse SNR conditions. However, there was also a significant interaction between light level and SNR in their effects on mean dilation.

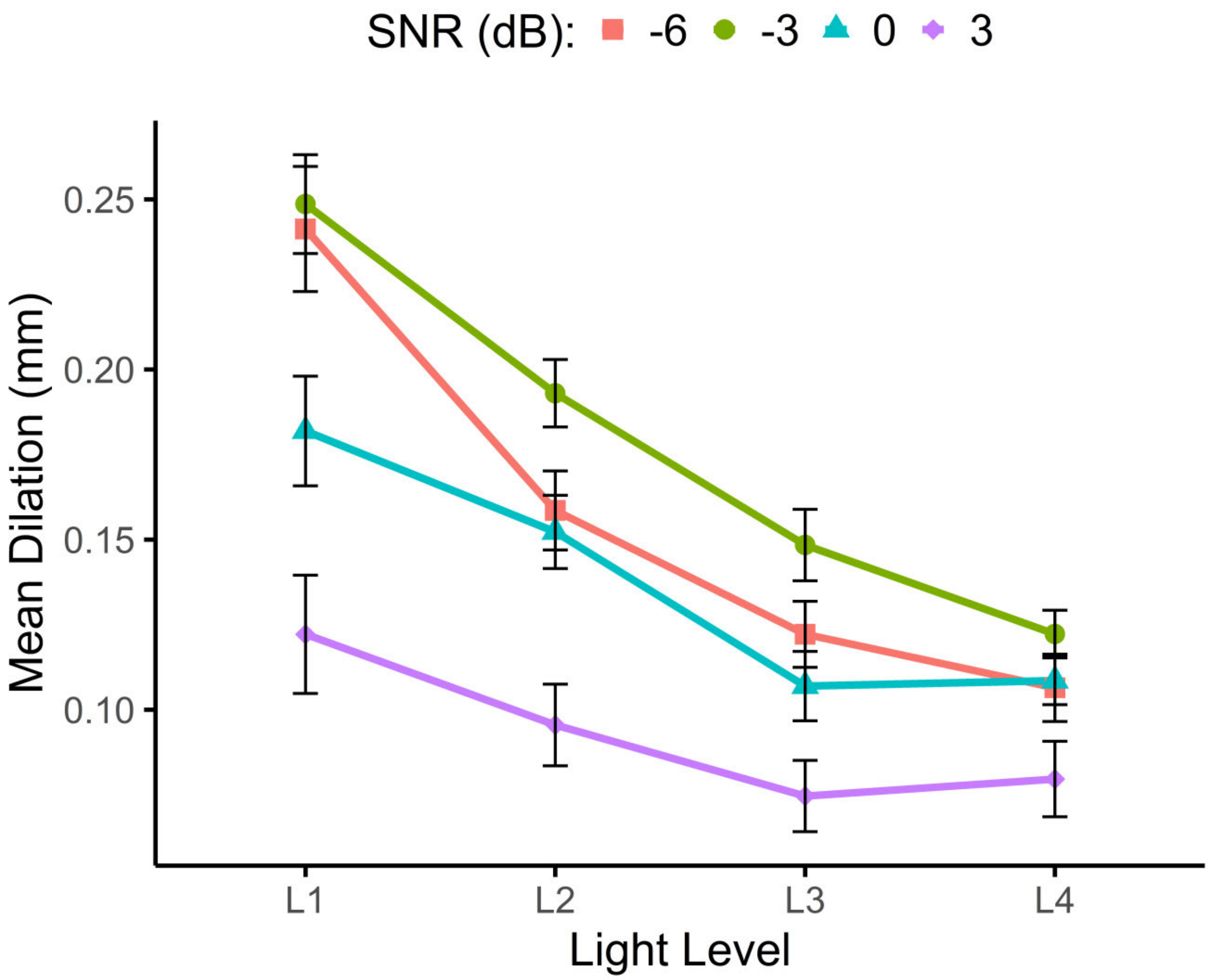
Table 16. Means and Standard Deviations for Mean Dilation as a Function of a 4(light level) X 4(SNR) design

	SNR (dB)							
	-6		-3		0		3	
Light level	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
1	0.24	0.15	0.25	0.13	0.18	0.12	0.12	0.12
2	0.16	0.11	0.19	0.10	0.15	0.09	0.10	0.07
3	0.12	0.08	0.15	0.09	0.11	0.07	0.07	0.07
4	0.11	0.08	0.12	0.06	0.11	0.06	0.08	0.06

Note. M and *SD* represent mean and standard deviation, respectively.

Figure 12

Average Mean Dilation



Note. This figure shows average mean dilation (within-subjects standard error bars) across all participants in the 16 light level by SNR (dB) conditions. Brightness successively increases from L1 (dimmest) to L4 (brightest).

Table 17. RANOVA Results for Mean Dilation as a Function of Light Level and SNR (dB)

Predictor	df_{Num}	df_{Den}	*Epsilon*	*F*	*p*	η^2_g
light level	1.87	65.59	0.62	22.43	<.001	.14
SNR	1.84	64.53	0.61	30.69	<.001	.10
light level x SNR	6.41	224.18	0.71	4.79	<.001	.02

Note. df_{Num} indicates degrees of freedom numerator. df_{Den} indicates degrees of freedom denominator. Epsilon indicates Greenhouse-Geisser multiplier for degrees of freedom, *p values* and degrees of freedom in the Table incorporate this correction. η^2_g indicates generalized eta squared.

9.4.6.1 Post Hoc Analyses

Post hoc simple effects analyses were conducted to examine significant differences in mean dilation between SNR levels within each light level (Table 18), and between each light level within each SNR level (Table 19). The Bonferroni correction for 48 comparisons (alpha is significant at p = .001) was applied.

There were more significant differences in mean dilation between SNR levels in dimmer light levels (Table 18). Only the difference between SNR -3 dB and SNR 3 dB was significant at light level 4.

Additionally, there were more significant differences in mean dilation between light levels in SNR levels that demanded more listening effort (SNR -6 dB and SNR -3 dB) than there were in SNR levels that demanded less listening effort (SNR 0 dB). There were no significant differences in mean dilation between any light levels in the least demanding SNR level (SNR 3 dB) (Table 19).

Table 18. Post Hoc Simple Effects Analyses for Mean Dilation – Light Level as Grouping Variable

Light Level	Comparison (SNR dB)		Mean difference (mm)	95% CI – low	95% CI – high	*Pairwise t statistic*	*p*	*Sig.*
L1	-6	-3	-0.01	-0.04	0.03	-0.43	.67	
L1	-6	0	0.06	0.02	0.10	3.23	.003	
L1	-6	3	0.12	0.08	0.16	5.62	<.001	*
L1	-3	0	0.07	0.03	0.11	3.35	.002	
L1	-3	3	0.13	0.08	0.17	5.68	<.001	*
L1	0	3	0.06	0.03	0.09	3.97	<.001	*
L2	-6	-3	-0.03	-0.06	-0.01	-2.97	.005	
L2	-6	0	0.01	-0.02	0.04	0.44	.67	
L2	-6	3	0.06	0.03	0.10	3.67	<.001	*
L2	-3	0	0.04	0.02	0.06	3.92	<.001	*
L2	-3	3	0.10	0.06	0.13	5.89	<.001	*
L2	0	3	0.06	0.03	0.08	4.53	<.001	*
L3	-6	-3	-0.03	-0.05	0.00	-2.16	.04	
L3	-6	0	0.02	-0.01	0.04	1.07	.29	
L3	-6	3	0.05	0.02	0.08	3.49	.001	*
L3	-3	0	0.04	0.02	0.07	3.49	.001	*

Light Level	Comparison (SNR dB)		Mean difference (mm)	95% CI – low	95% CI – high	*Pairwise t statistic*	*p*	*Sig.*
L3	-3	3	0.07	0.05	0.10	6.65	<.001	*
L3	0	3	0.03	0.01	0.05	3.60	<.001	*
L4	-6	-3	-0.02	-0.04	0.01	-1.48	.15	
L4	-6	0	0.00	-0.02	0.02	-0.21	.83	
L4	-6	3	0.03	0.00	0.05	2.31	.03	
L4	-3	0	0.01	0.00	0.03	1.64	.11	
L4	-3	3	0.04	0.02	0.06	4.59	<.001	*
L4	0	3	0.03	0.01	0.05	3.10	.004	

Note. p is not Bonferroni corrected. * in *Sig.* column indicates significant differences after correcting for multiple comparisons (e.g., *p* is significant at 0.001). Degrees of freedom for all comparisons = 35.

Table 19. Post Hoc Simple Effects Analyses for Mean Dilation – SNR as Grouping Variable

SNR (dB)	Comparison (light level)		Mean difference (mm)	95% CI – low	95% CI – high	Pairwise *t* statistic	*p*	*Sig.*
-6	L1	L2	0.08	0.04	0.12	4.24	<.001	*
-6	L1	L3	0.12	0.08	0.16	5.56	<.001	*
-6	L1	L4	0.13	0.09	0.18	6.25	<.001	*
-6	L2	L3	0.04	0.01	0.06	2.69	.01	
-6	L2	L4	0.05	0.03	0.08	3.93	<.001	*
-6	L3	L4	0.02	-0.01	0.04	1.48	.15	
-3	L1	L2	0.06	0.03	0.09	3.70	<.001	*
-3	L1	L3	0.10	0.06	0.14	5.16	<.001	*
-3	L1	L4	0.13	0.09	0.16	7.65	<.001	*
-3	L2	L3	0.04	0.02	0.07	3.40	.002	
-3	L2	L4	0.07	0.04	0.10	5.34	<.001	*
-3	L3	L4	0.03	0.00	0.05	2.35	.03	
0	L1	L2	0.03	-0.01	0.07	1.39	.17	
0	L1	L3	0.07	0.03	0.12	3.69	<.001	*
0	L1	L4	0.07	0.04	0.11	3.95	<.001	*
0	L2	L3	0.05	0.02	0.07	3.74	<.001	*

SNR (dB)	Comparison (light level)		Mean difference (mm)	95% CI – low	95% CI – high	Pairwise *t* statistic	*p*	*Sig.*
0	L2	L4	0.04	0.02	0.07	3.20	.003	
0	L3	L4	0.00	-0.02	0.02	-0.14	.89	
3	L1	L2	0.03	-0.02	0.07	1.28	.21	
3	L1	L3	0.05	0.00	0.09	2.26	.03	
3	L1	L4	0.04	0.00	0.09	1.87	.07	
3	L2	L3	0.02	-0.01	0.05	1.62	.12	
3	L2	L4	0.02	-0.01	0.05	1.08	.29	
3	L3	L4	0.00	-0.03	0.02	-0.41	.69	

Note. *p* is not Bonferroni corrected. * in *Sig.* column indicates significant differences after correcting for multiple comparisons (e.g., *p* is significant at 0.001). Degrees of freedom for all comparisons = 35.

9.5 Discussion

9.5.1 Summary of Results

The aim of Study 1 was to examine the effects of four light levels on TEPRs during the BKB sentence test with four SNR conditions in a sample of young adults without hearing loss. Pupil dilation increased from sentence onset in all conditions. This indicated task engagement and effort expenditure during all SNR by light level conditions.

As expected, task performance decreased as a function of SNR. Participants performed near ceiling[11] in SNR 3 dB, and this declined progressively in more adverse SNRs (i.e., SNR 0 dB, SNR -3 dB and SNR -6 dB).

Additionally, baseline diameter was larger in dimmer light levels. In L1, baseline diameter was largest and decreased in size progressively in brighter light levels (i.e., L2, L3, and L4). There was an effect of SNR on baseline diameter. SNR was manipulated by adjusting the levels of background noise. Therefore, the effect of SNR on baseline diameter was likely due to the presence of differing levels of background noise when baseline diameter was measured and may reflect a preparation effect depending on how loud the background was in a condition. For example, participants were likely able to anticipate how hard the forthcoming trial would be based on the loudness of the background noise prior to sentence presentation. Baseline correction in peak and mean dilation accounted for this effect but only if dilations measured were independent of baseline diameter.

There was also a significant interaction effect between light level and SNR on baseline diameter. Post hoc analyses revealed that significant differences in baseline diameter between SNR conditions occurred in the two dimmest light levels (i.e., L1 and L2). This may indicate that preparation effects in baseline diameter were dependent on the light level and that baseline diameter was more sensitive to preparation effects in dimmer light levels.

There was a significant main effect of SNR on peak dilation latency. In line with previous findings (e.g., Zekveld et al., 2010), more adverse SNR conditions resulted in longer peak

[11] The highest possible score for a given condition of the BKB sentence test in the present study was 50/50 target words.

dilation latency. This may indicate that more time was required for the cognitive processing involved in understanding speech-in-noise in more adverse SNR conditions and this may also reflect increased listening effort in more adverse SNR conditions.

Peak dilation latency was also significantly affected by light level. For example, peak dilation latency was largest in L1 (dimmest) and L4 (brightest) compared to the intermediate levels (i.e., L2 and L3). Peak dilation latency only significantly differed for the comparison between L1 and L3. Nevertheless, these results may suggest that light level can affect peak dilation latency in a speech-in-noise task. Additional research is required to examine the underlying mechanisms of this relationship.

The original contribution of Study 1 concerned the effects of light level and SNR on peak and mean dilation. Peak and mean dilation were larger in more adverse SNR conditions (except for SNR -6 dB) and in dimmer light levels. Furthermore, light level and SNR interacted in their effects on peak and mean dilation such that there were more significant differences in peak and mean dilation between light levels in more adverse SNRs and there were fewer differences in peak and mean dilation between SNRs in brighter light levels.

The following discussion is structured such that the effects of light level and SNR on peak dilation and peak dilation latency are discussed, followed by a discussion of the interaction effects. Because peak and mean dilation generally followed the same pattern, only peak dilation is discussed further. Peak dilation may represent the more sensitive measure of listening effort and may be the most likely choice for use in clinical audiology.

9.5.2 Effect of SNR on Peak Dilation and Peak Dilation Latency

SNR -3 dB resulted in the largest peak dilation of all SNRs, across all light levels. This may suggest that participants expended most effort during SNR -3 dB. This finding is somewhat counterintuitive as SNR -6 dB was a more adverse listening condition than SNR -3 dB. One potential explanation for this finding is that SNR -6 dB may have been too difficult for the participants. Therefore, the smaller peak dilation in SNR -6 dB may have reflected cognitive overload and withdrawal of effort due to a lack of success and loss of motivation.

This interpretation is supported by predictions set out in Motivational Intensity Theory (Brehm & Self, 1989) and the performance scores in SNR -6 dB and SNR -3 dB. Motivational Intensity Theory posits that effort is a function of task difficulty, only when successful performance is possible, and the importance of the task justifies the effort that is required (Richter et al., 2016, p. 151). For SNR -6 dB, participants scored an average of 4.31 words correct out of 50 (8.62% correct). Therefore, participants may have deemed this task "too hard" (i.e., lack of adequate success) and may not have been as motivated to expend as much effort in this condition. In SNR -3 dB, participants scored an average of 18.46 words correct out of 50 (36.92% correct). Therefore, participants may have been more willing and motivated to expend more effort because their effort expenditure was "rewarded" with better task performance (Richter et al., 2016), hence the larger peak dilation.

Participants still scored an average of 8.62% words correct in SNR -6 dB. This indicated that they were still engaged in the task to an extent. However, it is possible that they expended more effort in the beginning of the condition but withdrew effort towards the end, once the lack of success was clear.

These findings partially support previous findings that the relationship between peak dilation and speech intelligibility follows an inverted-U shaped curve, with the smallest peak dilations occurring when intelligibility is at 100% correct (easy) and 0% correct (impossible) (Figure 7) (Ohlenforst, Zekveld, Lunner, et al., 2017; Wendt et al., 2018; Winn et al., 2018; Zekveld & Kramer, 2014). However, peak dilation in SNR -6 dB was larger than in SNR 0 dB despite the low average intelligibility score. Therefore, these findings demonstrate that peak dilation is not a measure of "task difficulty" per se but a measure of how much effort one expends in a task. Furthermore, the amount of effort that is expended in a task may be influenced by several individual factors, like perceived success and/or motivation. In this study, participants may have been sufficiently motivated to expend effort in SNR -6 dB such that poor performance did not lead to as large a reduction in pupil size as may be predicted based on the typical inverted-U shaped curve relationship between peak dilation and speech intelligibility.

In line with previous findings (e.g., Zekveld et al., 2010), peak dilation latency systematically increased in more adverse SNRs. This indicated that SNR affects the time course of effortful cognitive processing in speech-in-noise tasks. However, peak dilation latency as a measure of listening effort, did not follow the typical inverted-U shaped curve that peak dilation followed. This suggests that peak dilation latency might be more reflective of task demand rather than effort expenditure as measured by peak and mean pupil dilation. Participants may have exerted effort for a greater amount of time to try to understand the sentence that was presented in SNR -6 dB, but they may not have expended as much effort as they did in SNR -3 dB due to a lack of success and/or motivation due to the difficult of the task, as outlined above.

Alternatively, it is possible that participants used a different strategy in SNR -6 dB. For example, they may have engaged less with semantic processing of whole sentences in SNR -6 and instead were only listening for words or word parts due to the difficult nature of this condition and an inability to perceive the entire sentence. This may explain the longer peak dilation latency in SNR -6 dB.

Peak dilation latency as a measure of listening effort may provide insight into the processing of speech-in-noise that is different to that of peak and mean dilation. The current results also suggest that peak dilation was affected by light level. Therefore, the timing of pupil responses and how this is affected by light level should be examined in more detail in future.

9.5.3 Effect of Light Level on Peak Dilation

Peak dilation decreased in brighter light levels. This result conflicts with expectations proposed in the LIV, where larger baseline levels of a physiological system may hinder subsequent reactivity due to the dynamic range of the pupil (Lacey, 1956; Wilder, 1957, 1958). In this study, baseline diameter never reached the upper limits of the pupil’s range. Based on the physiology of the eye, the standard adult pupil can range in size from 2 mm to 8 mm and the largest mean baseline diameter value in this study was 4.54 mm. Furthermore, peak dilations are typically on the order of 0.5 mm above baseline diameter. Therefore, the dynamic range of the pupil was unlikely to have been a limiting factor. However, these findings may align with the reconceptualization of the LIV proposed by Jin (1992), that is, responses to stimuli will be larger when there are larger initial values, until the initial value reaches the upper limits of the system.

These results conflict with several previous findings regarding TEPRs (e.g., Beatty & Lucero-Wagoner, 2000; Bradshaw, 1969; Gilzenrat et al., 2010; Reilly et al., 2019) which have indicated that peak dilation was not affected by light level. These results indicated that peak dilation can be affected by light level and may not be independent of baseline diameter. The findings of the current study also conflict with the findings of Steinhauer et al. (2004), Wang, Kramer, et al. (2018) and Książek et al. (2021) who reported effects of light level but showed larger pupil dilation in bright conditions, compared to dark conditions. This may be due to differences in how light levels were manipulated (e.g., illumination vs. screen luminance) and the light levels (bright and dark vs. four mid-range levels) that were used.

The current findings were more consistent with previous work that used similar methods to manipulate light level. Peysakhovich et al. (2015) also found larger peak dilation in their dimmer condition. They also manipulated light level by changing the background colour of a computer screen which resulted in mid-range light levels. Peysakhovich et al. (2015) used a vastly different task to the one used in the current study and in Wang, Kramer, et al. (2018) and Książek et al. (2021) (i.e., digit recall vs. speech-in-noise).This may suggest that methods used to manipulate light level and the degree of light level can significantly alter TEPRs, regardless of the cognitive task under examination. Wang, Kramer, et al. (2018) and Książek et al. (2021) did not examine interaction effects and therefore, will not be discussed further.

9.5.4 The Light Level by SNR Interaction

9.5.4.1 The Effect of SNR on Peak Dilation in Different Light Levels

The sensitivity of peak dilation to detect differences in listening effort between the SNR conditions was reduced under brighter light levels. In the brightest light level (L4), peak dilation significantly differed when comparing SNR 3 dB and SNR -3 dB. All other comparisons were not significant. There were more significant differences in peak dilation between SNR conditions in dimmer light levels (L2, L3, L4) than there were in the brightest condition (L4). This could be because the brightest light level caused extreme pupil constriction that limited the amount of dilation that could occur.

The possibility of this effect existing was recently raised by Winn et al. (2018). The authors posited that bright background screen colours (e.g., white) may cause extreme pupil constriction and that TEPRs elicited by listening tasks might not be robust enough to appear. The results of the current study supports this suggestion by showing that the effect of SNR on TEPRs does not override the pupil constriction that occurs in response to bright light from a computer screen because there were more significant differences between SNR conditions in dimmer light levels.

Although this was not explicitly studied, it is possible that when responding to bright light, the PNS's innervation of the iris constrictor muscles was too strong to be inhibited at the Edinger-Westphal nucleus by projections from the LC (and other brain regions) during effortful listening. Additionally, the response to bright light may also have been too strong to allow the sympathetically driven dilator muscles in the iris to fully contract, which may have hindered dilation ability.

Based on the current findings, researchers and clinicians should avoid the use of white backgrounds when using computer screens and excessively bright illumination conditions when assessing listening effort via pupil dilation. These conditions may negatively affect the sensitivity of TEPRs when measuring listening effort. Dimmer, grey backgrounds and dimmer illumination conditions may be preferable.

9.5.4.2 The Effect of Light Level on Peak Dilation in Different SNRs

The results of the current study also indicated that the effect of light level on peak dilation was larger at lower SNRs. For instance, at an SNR of +3 dB, peak dilation only significantly differed between the L1 and L4. However, differences between light levels arose in more effortful SNR conditions. At SNR -3 dB, peak dilation significantly differed between all light levels (except for the comparison between L2 and L3). Therefore, in addition to brighter light levels diminishing the sensitivity of peak dilation to detect changes in listening effort, these results also suggest that SNR influenced the effect of changing light levels on peak dilation.

No differences in peak dilation were found between light levels at SNR 3 dB. Reilly et al. (2019) found that peak dilation was not affected by light level in a perceptual discrimination task (Experiment 1). They suggested that this could be because the task was too easy and did not induce enough cognitive load to elicit large enough peak dilation. They performed a follow up experiment (Experiment 2) to further test this finding using a more difficult task (a visual word-monitoring task). They found similar results (e.g., no effect of light level) and concluded that the peak dilation was not affected by the light levels used, nor did peak dilation show that it was dependent on initial baseline diameter. However, it is not clear that Experiment 2 rectified the issues in Experiment 1. For example, response accuracy was still very high in Experiment 2 (99.28%) indicating that this task also may not have induced enough cognitive load to result in large enough peak dilations.

While performance and effort are different phenomena, performance at ceiling indicated the tasks used by Reilly et al. (2019) were relatively easy for the participants. In the current study, this could be why an effect of light level was not found in the easiest condition (SNR 3 dB). For example, at SNR 3 dB, the task may not have been difficult enough to elicit large enough peak dilations to reveal differences between light levels. More adverse SNR conditions may have been difficult enough to elicit large enough peak dilations to reveal differences between light levels.

These results are supported by Peysakhovich et al. (2015). They found a significant interaction between the effects of light level and task difficulty on peak dilation. They found the effect of light level on peak dilation was only significant in the more cognitively demanding “load on memory” condition. The methodological similarities between Peysakhovich et al. (2015) and the current study regarding light level manipulation and the use of peak dilation provide support for the findings reported here. However, Peysakhovich et al. (2017) did not find this relationship between peak dilation and light level. Instead, they found a relationship similar to that reported here and in Peysakhovich et al. (2015) for mean pupil diameter (referred to as tonic pupil diameter in their study) and light level when participants performed the Toulouse n-back Task. This may indicate that there are differences in how light levels affect TEPRs depending on the types of tasks that are performed during measurement.

While Steinhauer et al. (2004) also found an interaction between light level and task difficulty and how these affected mean dilation, they found larger mean dilation in the brighter, most difficult condition. Steinhauer et al. (2004) reported that their results reflected a greater amount of central processing activity reaching the Edinger-Westphal nucleus (caused by the greater cognitive load induced by the more difficult subtract-7 task) (see Section 6.4.1) resulting in an additional component dilation via inhibition of the parasympathetic pathway leading to relaxation of the sphincter muscles. In the light condition, there was more parasympathetic activity innervating the constrictor muscles. This may mean that there was more parasympathetic activity available to inhibit via central activity related to cognitive processing. This may have led to the larger TEPRs via a combination of the sympathetic dilation and more parasympathetic inhibition than was possible when the task was performed in darkness. None of the tasks in the current study or in Peysakhovich et al. (2015) were performed in darkness like they were in Steinhauer et al. (2004). Thus, inevitably, there was still a degree of parasympathetic activity innervating the constrictor muscles and causing constriction in the dimmest condition of the current study (more so than there would be in complete darkness). Consequently, there may have been more parasympathetic activity innervating the constrictor muscles that was able to be inhibited which provided an additional component of dilation in the dimmest conditions.

In the most difficult and dimmest condition in the current study, pupil dilation may have been caused by stimulation of the dilator muscles due to sympathetic activity and parasympathetic inhibition via projections from the LC. Parasympathetic activity may have been able to be inhibited to a larger degree than would have been possible in darkness. This may have led to greater relaxation of the constrictor muscles, and therefore, more dilation. This explanation is supported by Joshi and Gold (2020).

In the brightest condition of the current study, the parasympathetic activity reaching the constrictor muscles (induced by the bright light) may have been "too strong" to allow the sympathetically innervated dilator muscles to dilate the pupil to its full extent, and may have led to relaxation of the dilator muscles (Joshi & Gold, 2020). Moreover, central activity (related to the expended listening effort) reaching the Edinger-Westphal nucleus may not have been strong enough to inhibit the parasympathetic activity (which causes the constriction), thus impeding the ability of effort-related central activity to contribute to TEPRs via the parasympathetic pathway.

The discrepancy in results between the current study and Steinhauer et al. (2004) may also be explained by methodological differences. They manipulated ambient illumination. In the current study, light level was manipulated via a computer screen. Furthermore, Steinhauer et al. (2004) reported mean dilation over 1 minute of a sustained processing task. Peak and mean dilation over multiple, transient cognitive tasks (4 s duration) was reported in the current study. Therefore, the ability to compare the current findings to Steinhauer et al. (2004) is hindered.

In summary, there are complex systems that control pupil size when responding to effortful listening tasks under different light levels. While explanations of the physiological mechanisms which may have led to the specific results reported here have been provided, they are speculative. As suggested by Joshi and Gold (2020), the strength of activity and dynamics of the contributions of the PNS and SNS are difficult to ascertain based solely on observations of the pupil (pg. 20). Moreover, the discrepancy in results between the current study and Steinhauer et al. (2004) may be the result of methodological differences in how light level was manipulated (ambient illumination vs. computer screen colour), the light conditions used (light and dark vs. four levels of intermediate light) and how pupil responses were measured (peak dilation over 4 s vs. mean dilation over 60 s).

These findings also call into question the method of individually setting illumination levels to induced baseline diameters to be in the middle of each participant's dynamic pupil range. While this ensures that pupils have a large enough range to allow full dilation where appropriate, the light levels may have an unintended effect on peak and mean dilation, which may not be consistent across participants/clients or laboratories/clinics.

Instead of individually setting illumination levels, It may be more appropriate to employ fixed illumination conditions across participants/clients for measuring listening effort via TEPRs. However, any standard conditions that are established must ensure that full dilation is achievable for young and older individuals because the dynamic range of the pupil typically differs between these groups (Piquado et al., 2010). Furthermore, the conditions must also be replicable across laboratories and clinics. The standardisation of measurement conditions requires further research.

9.5.5 Limitations

The purpose of Study 1A was to examine the effects of light level and SNR on TEPRs during a speech-in-noise task. Illumination (luminous flux per unit area, i.e., lux) was measured from a standard point on a chin rest, 60 cm away from the computer screen. By measuring light level in this way, the total amount of luminous flux per unit area falling on a standard surface near eye level was measured. This means that both the additional illumination that was present in the audiology booth and the light emitted from the computer screen were measured. Due to differences in head shape and eye position, the intensity of light reaching participants' eyes likely varied to a degree. Therefore, a potential limitation of the current study was the absence of screen luminance (cd/m^2) measurements for each light level. This would have resulted in a more comprehensive report of the light levels used in the current study and may aid any replication attempts. To account for this, the HSV colour codes used to manipulate the light levels were provided (Table 3) in the Chapter 8. A detailed description of the procedures related to light manipulations and measurement was also provided.

Another potential limitation that should be addressed is the possible effect of participant's subjective responses to the light levels. Participants may have found the brightest light level (L4) uncomfortable. This may have affected their engagement in the auditory task and may have partly contributed to the lack of differences between SNR conditions in L4. However, participants did not appear uncomfortable in any light level in the pilot testing or the experimental sessions. In future, it could be beneficial to administer a questionnaire asking participants to rate their comfort levels under various light levels to rule out this possibility.

9.5.6 Recommendations

Due to the results of the current study and the apparent sensitivity of the pupil response to environmental light levels, it is recommended that future research using pupillometry report the procedures employed to measure light conditions in similar amounts of detail for replicability, transparency, and comparability purposes. Detailed reports regarding the positioning of the light meter at the time of light measurements are particularly important as pointed out by Tsukahara and Engle (2021). Researchers can use this information to perform accurate replications of studies and to make judgements on the validity of findings.

The importance of accurate and detailed descriptions of environmental light levels for accurate study replication was recently demonstrated by Tsukahara and Engle (2021). Tsukahara et al. (2016) found a robust relationship between baseline diameter and measures of fluid intelligence and working memory capacity. However, other researchers (Unsworth et al., 2020; Unsworth et al., 2019) have not been able to replicate those findings. Tsukahara and Engle (2021) reanalysed data from Tsukahara et al. (2016) and found less variability in baseline diameter was more likely to result in weak, non-signficant correlations between working memory capacity and baseline diameter. In follow up studies, they demonstrated that a bright computer screen monitor could reduce variability in baseline diameter[12] and that this was the likely cause of the replication failures in Unsworth et al. (2020) and Unsworth et al. (2019). They reaffirmed their earlier finding of a correlation between baseline diameter and measures of fluid intelligence and working memory capacity, but only when the screen was not too bright (which may lead to excessive constriction of sphincter muscles). Thus, seemingly innocuous decisions about screen brightness and environmental illumation can have unfavourable consequences on conclusions that may be drawn from pupil measurements. Based on the current findings, it can also be recommended that researchers and clinicians use dimmer light levels when using pupillometry to measure listening effort to ensure sensitivity of the measure.

9.5.7 Conclusion

These results demonstrate that light level affects TEPRs in a speech-in-noise task. Peak dilation measured in brighter light levels was not as sensitive to variations in SNR conditions. These findings differ from some past studies. Evidence that light levels and SNR interact in their effects on TEPRs measured during a speech-in-noise task was also presented. These findings may suggest that effect of light level on peak dilation may be dependent on the lighting environment, and the type and difficulty of the cognitive task being undertaken. Furthermore, these effects may be due to complex interactions between the SNS and PNS related to how they the control pupil size.

Measurement of TEPRs may provide a unique opportunity to quantify listening effort in clinical audiology. These results may have implications for the use of TEPRs to measure listening effort in research and clinical settings as TEPRs may not be independent of light levels. Careful consideration and detailed reporting of light levels should be standard for any future research and clinical applications.

[12] Smaller standard deviations in the brighter conditions compared to dimmer conditions for baseline diameter were also found in the current dataset

10 STUDY 1B - COMPARING TRADITIONAL REPEATED MEASURES ANOVA AND MIXED-EFFECTS MODELLING

10.1 Background and Aims

The primary aims of this chapter were to: (1) to examine the use of mixed-effects modelling for the analysis of signal-averaged pupil data, and (2) provide a conceptual bridge between Study 1A and Study 1C. There are significant limitations to RANOVA that can be overcome using mixed-effects modelling. Furthermore, mixed-effects modelling may enable the use of trial-level pupil data (Study 1C) (Volpert-Esmond et al., 2018). The results of the traditional RANOVA (Study 1A) and a mixed-effects model (MEM) are compared. In the next section, the characteristics of repeated-measures designs, and the traditional analysis techniques are described, followed by a description of mixed-effects modelling.

10.1.1 Repeated-Measures Designs

A study design is deemed “repeated-measures” when all participants experience all treatment conditions in a study (unlike a between-subjects design when two or more groups experience different treatment conditions). In repeated-measures designs, it is expected that data from the same participant will be related across conditions.

Traditionally, repeated-measures data are analysed with a RANOVA (as in Chapter 9). However, there are limitations associated with the use of RANOVA. These limitations are described in the sections below which are based on the recent paper by Brown (2021, p. 1).

In RANOVA, the differences between condition means are assessed, while accounting for the relatedness of data measured from the same participants. If researchers wanted to assess the differences between condition means, while accounting for the relatedness of data measured using the same stimuli[13], separate analyses (F1 and F2) would be performed for participants and stimuli and a quasi-F statistic would be computed. Fundamentally, participant-level variability and stimuli-level variability cannot be included simultaneously in RANOVA. Therefore, observations within a condition must be aggregated either across participants or stimuli. When data are aggregated in this way, information pertinent to the variability either within participants or stimuli is averaged out and statistical power may be reduced, which can affect the ability to detect a significant effect, if one exists.

Missing data are also not handled efficiently by RANOVA. Missing data are a common occurrence in large datasets and may be due to, for example: experimenter error, equipment malfunctions, or data rejection. If a dataset has missing values, RANOVA handles this via "listwise deletion" . This means that when a single data point is missing in a dataset, all data for that participant are typically excluded from further analysis[14]. Loss of data via listwise deletion reduces sample size and may lead to inflated standard error estimates and a concomitant reduction in statistical power (Brown, 2021). This is another limitation of RANOVA.

When using RANOVA, data manipulation may need to be carried out because RANOVAs are not particularly flexible in terms of the data that can be used. For example, the dependent variable must be continuous, and the independent variables must be categorical. Thus, continuous predictors (e.g., time) need to be "binned" and treated as categorical variables which can also reduce statistical power.

10.1.2 Mixed-Effects Modelling

The limitations of RANOVA can be circumvented by applying mixed-effects modelling where appropriate. Researchers have been reporting the benefits of mixed-effects modelling for decades (Bagiella et al., 2000; Judd et al., 2012; McCulloch, 2005). There

[13] For example, syllables, words, or sentences may be used as stimuli. It is expected that responses to the same stimuli (i.e., a sentence in the current study) will be related. Data measured from the same stimuli are not independent.

[14] List-wise deletion can be avoided if data meet specific criteria for missingness and values are imputed to replace the missing cases. Appropriate imputation methods may involve expectation maximisation algorithms which replace missing values with the most likely value based on the data. This method was used in Study 1A. On the other hand, mean imputation is an easier, more straightforward method of imputation, however, this method is less acceptable as it has undesirable effects (see Schafer & Graham, 2002).

has been a recent surge in the applications of mixed-effects modelling (also referred to as multilevel models, mixed models, hierarchical models) in psychological research; especially for data that have hierarchical structures or for data gathered in repeated-measures designs (Meteyard & Davies, 2020). Uncertainty in decision making regarding how MEMs should be applied and reported have been cited as reasons for the relatively slow uptake of mixed-effects modelling for analysis of psychological data (Meteyard & Davies, 2020).

Mixed-effects modelling can be used in the analysis of data gathered in repeated-measured designs where participant responses are related, that is, samples are not independent. MEMs can model both fixed effects and random effects (hence, "mixed-effects"). By including random effects, MEMs estimate the effects of a treatment condition while simultaneously accounting for non-independence in the data due to participants, stimuli, and/or any other variable that is justified in the experimental design (e.g., condition/trial order). For example, by including a random intercept effect of participant, MEMs estimates a unique intercept for each participant (rather than one intercept across participants) (Volpert-Esmond et al., 2021). Additionally, the slope for a fixed effect (independent variable) by participant can be estimated separately in MEMs (Volpert-Esmond et al., 2021). Including random slopes in MEMs is beneficial as it is unlikely that all participants will respond to the experimental conditions in the same way.

In the current study in which sentences were used as stimuli[15], if "participant" and "sentence" were included in the model as random effects, the model partitions the variance in the response variable that is associated with specific participants and sentences from the error term (i.e., more variance is accounted for). The inclusion of these random effects acknowledges the non-independent samples in the data, enhances statistical power and increases the ability to detect effects of independent variables (if they exist) (Gelman & Hill, 2007).

Unlike RANOVA, MEMs are robust to missing data. MEMs deal with missing data and/or extreme values by partial pooling (also called shrinkage, regularisation and "borrowing strength") (Gelman & Hill, 2007; Gelman et al., 2012; Mahr, 2017). For example, when a participant record has missing data, partial pooling enables MEMs to compute estimates informed by other participants who have complete data. This means that extreme values

[15] 256 unique sentences were used in the current study and the same sentence battery was used for every participant. The conditions in which each sentence was presented, and the order of sentence presentation was varied.

are pulled towards the average. Values which are already close to that average are not pulled as much. Participants and/or items with missing data have less of an impact on parameter estimates than those without missing data. Therefore, missing data does not result in a required removal of data and statistical power is maintained when MEMs are used.

In summary, mixed-effects modelling has many benefits that make it more suitable for the analysis of data with repeated-measures than the traditional RANOVA. The use of repeated-measures designs for listening effort research using pupillometry is common. Mixed-effects modelling (and affiliates, e.g., growth curve analysis) are beginning to be applied to pupillometric data in hearing sciences (e.g., McGarrigle et al., 2017b; Ohlenforst, Zekveld, Lunner, et al., 2017; Wagner et al., 2019). Mixed-effects modelling results in estimates of coefficients which provide information about the relationship between each level of predictor variable and the dependent variable, while controlling for other respective predictor variables in the model. By using mixed-effects modelling to analyse the data collected for this thesis, the relationships between light level, SNR and TEPRs can be estimated, while accounting for the variance in TEPRs that can be attributed to participants, and simultaneously, other random effects.

It was expected that the results of the RANOVA in Chapter 9 and the MEM would be similar, but that the MEM would have more power to detect significant differences. Potentially, this would lead to the identification of a greater number of significant differences between conditions. Following the results of the MEM, the relative advantages and disadvantages of using mixed-effects modelling over RANOVA in the analysis of the data in the current study and the prospect of using trial-level pupil data are discussed.

10.2 Significance

Mixed-effects modelling represents a powerful way to analyse repeated-measures data. In this chapter, an MEM and the traditional RANOVA reported in Chapter 9 are compared to investigate the benefits of mixed-effects modelling. Improvements in the way repeated-measures data are analysed may have significant benefits for future research through improved inferences.

10.3 Method

This chapter comprises a re-analysis of the data presented in Chapter 9 using mixed-effects modelling. In Chapter 9, RANOVA results for peak and mean dilation followed very similar patterns, with peak dilation appearing slightly more sensitive to light level and SNR manipulations. Therefore, only peak dilation will be reported in this chapter, as it was assumed that mean dilation would follow a similar pattern.

10.3.1 Participants

Information related to the participants can be found in Chapter 8 - General Methods, Section 8.1.

10.3.2 Materials

Information related to the materials can be found in Chapter 8 - General Methods, Section 8.2.

10.3.3 Procedure

Information related to the procedure can be found in Chapter 8 - General Methods, Section 8.3.

10.3.4 Pupil Data Pre-Processing

Information related to the pre-processing can be found in Chapter 8 - General Methods, Section 8.4.

The peak dilation data reported in this chapter were identical to those used in the RANOVA in Chapter 9, without imputation of missing values. Missing data are common in pupillometry datasets, due to measurement capture loss of eye-tracking equipment (often due to gaze shifts). A strength of MEMs is that they are robust to missing data, thus, no imputation or listwise deletion was necessary.

10.3.5 Data Analysis

Data analyses and visualisations were completed in R open source software (version 4.0.1) (R Core Team, 2020) and R Studio (version 1.3.1073) (RStudio Team, 2020). Analysis and visualisation codes are available at https://osf.io/am6uv/n.

The model building process and results are reported based on recommendations set out in the recent paper: "Best practice guidance for linear mixed-effects models in psychological science" (Meteyard & Davies, 2020).

Data visualisations were completed using *ggplot2* (version 3.3.3) (Wickham, 2016). Data were analysed using the *lmer* function of the *lme4* R package (version 3.5.2) (Bates et al., 2015) and *p* values from summary outputs for mixed-effects models were obtained via the *lmerTest* R package (version 3.1.0) (Kuznetsova et al., 2017). The default variance-covariance structure specified by *lmer* models is unstructured, meaning that the model specifies no pattern in the covariance matrix, that is, no observations are "equally correlated" and there is no "structure" between neighbouring values (Barnett et al., 2010). This specification typically results in the best model fit but may come at the cost of using up degrees of freedom, because each variance and covariance is estimated separately from the rest of the data (that is, many parameters need to be estimated). A *lmer* model with the default variance-covariance structure was used because *lmer* models are particularly suited for fully-crossed designs (rather than nested designs[16]). Furthermore, there was no evidence that the use of a different variance-covariance structure would be more suitable for the current data.

[16] In nested designs, experimental units at 1 level are "nested" within another level i.e., students nested within schools. Students are not exposed to all levels of the "nesting factor" (schools). In crossed designs, responses in all combinations of factors are measured i.e., like in the current design, where all participant responses are measured for all possible combinations of light level (4 levels) and SNR (4 levels), thus they were exposed to all 16 combinations.

10.3.5.1 Fixed Effects

The fixed effects were light level, SNR, and the interaction between light level and SNR. By including these fixed effects, four variance components are estimated in the MEM (Schielzeth & Nakagawa, 2013):

1. Main effect (marginal) variance explained by light level: This is the variance in peak dilation explained by the light level averaged across SNR levels.
2. Main effect (marginal) variance explained by SNR: This is the variance in peak dilation explained by SNR averaged across light levels.
3. Interaction variance explained by light level x SNR: This is the variance in peak dilation explained by the specific combinations of light level × SNR after considering the average effect of the light level across all SNR levels and the average effect of SNR levels across light levels.
4. Residual variance: This is the variance in peak dilation that is unexplained by light level, SNR, and the interaction between light level and SNR. It is the variance in peak dilation that remains after accounting for the mean peak dilation for each combination of factor levels (light level and SNR).

10.3.5.2 Random Effects

The maximal random effects structure justified by the design was used. Random effects that did not improve the model, or resulted in singular fit[17] were removed (Barr et al., 2013). The random effects structure in the model allows for specification of grouping variables in the data that lead to non-independence of observations (Volpert-Esmond et al., 2021).

Two random intercept effects were included in the final model: Participant and condition order.

1. Participant was included as a random effect to account for the fact that responses from the same participant will be related. For example, the use of the same participants for multiple observations introduced non-independence to the data. Differences between participants that are related to genetic, environmental, social, and developmental factors can be modelled via this method (Baayen et al., 2008). The addition of this random effect to the MEM replicates the structure of the RANOVA reported in Section 9.4.5.

[17] "Singular fit" means that the some of the dimensions of the variance-covariance matrices have been estimated to be zero.

2. Condition order was included as a random effect to account for the relationship between responses measured at the same time over the test session. Condition order is a "time-dependent", ordered variable that indicated when a condition was completed by a participant. For example, "1" for the condition order variable indicated that the measurement was recorded during the first condition for that participant, and "16" for the condition order variable indicated that the measurement was recorded during the last condition. Condition order also introduces non-independence in the data. For example, responses recorded during conditions which are presented last may be related to each other. Multiple studies have demonstrated an effect of fatigue, or time-on-task on the pupil – TEPRs typically decrease with increasing time-on-task (see Section 6.5.2). This addition demonstrates the flexibility of MEMs when compared to RANOVA. The experimental design used in the current study justifies the inclusion of condition order, and its simultaneous inclusion was only possible using mixed-effects modelling.

10.3.5.3 Model Building

Details regarding the random effects structure for each model and the model comparison results can be found in Table 20.

Model comparisons were conducted using likelihood ratio tests. All random slopes that were tested resulted in a singular fit and were removed.

Table 20. Study 1B Model Comparison and the Model Building/Selection Process for the Signal-Averaged Peak Dilation MEM

Sampling Units		N total observations = 560, N Participants = 36, N Condition order = 1-16										
Model specification	**Model name**	**Simpler Model**	**Fixed Effects added**		**Random Effects**		**Model fit**				**LRT Test against nested**	
					Participant	**Condition order**	**AIC**	**BIC**	**LL**	**df**	**df**	**X^2**
Random effect only	MEM_null	-	-		Intercept		-659.5	-646.5	332.7	4	-	-
Both random effects only	MEM_null2	MEM_null	-		Intercept	intercept	-685.5	-668.2	346.8	4	-	-
Fixed Effect - main effects	MEM_fixedeff_sn r	MEM_null2	SNR		Intercept	Intercept	-739.35	-709.05	376.67	7	3	59.822 ***
Fixed Effect- main effects x 2	MEM_fixedeff_ll	MEM_fixedeff_sn r	SNR+ light level		Intercept	Intercept	-937.89	-894.61	478.95	10	3	204.55 ***
Two-way interaction	MEM_int	MEM_fixedeff_ll	SNRx light level		Intercept	Intercept	-942.42	-860.19	490.21	19	9	22.526 **

Note. AIC = Aikake Information Criterion, BIC = Bayesian Information Criterion, LL = LogLikelihood, LRT – Likeilhood Ratio Test.* = p <.05, ** = p <.01, *** = p <.001.

10.4 Results

10.4.1 Assumption Checks

Residual and QQ plots indicated homoscedasticity, linearity, and normality of residuals, thus the MEM assumptions were met.

10.4.2 Signal-Averaged Peak Dilation MEM: Final Model Results

The results of the final model are presented in Table 21. The estimated beta values reflect the estimated change in peak dilation from the base levels. The base levels were as follows: L1 for light level, SNR -3 dB for SNR, and the combination of L1 SNR -3 dB for the interaction terms.

Across SNRs, peak dilation decreased from L1 (except in SNR -6 dB). Across light levels, peak dilation decreased from SNR -3 dB. When L3 and SNR 3 dB co-occur, the predicted absolute value of peak dilation is higher than expected based on the main effects of L1 and SNR -3 dB alone. When L4 and SNR 0 dB co-occur, the predicted absolute value of peak dilation is higher than expected based on the main effects of L1 and SNR -3 dB alone. When L4 and SNR 3 dB co-occur, the predicted absolute value of peak dilation is higher than expected based on the main effects of L1 dB and SNR -3 dB alone.

Table 21. Final Signal-Averaged Peak Dilation MEM Output Table for Study 1B

Fixed Effects					
	Estimated Beta	SE	95% CI	*t* (df)	*p*
Intercept	0.41	0.02	0.36 - 0.45	17.22 (116.38)	<.001 ***
SNR -6 dB	-0.03	0.02	-0.07 – 0.01	-1.39 (509.72)	.16
SNR 0 dB	-0.07	0.02	-0.11 – -0.03	-3.39 (509.32)	<.001 ***
SNR 3 dB	-0.16	0.02	-0.2 – -0.12	-7.37 (509.36)	<.001 ***
L2	-0.09	0.02	-0.14 – -0.05	-4.46 (509.32)	<.001 ***
L3	-0.15	0.02	-0.2 – -0.12	-7.44 (509.33)	<.001 ***
L4	-0.2	0.02	-0.24 – -0.16	-9.51 (509.33)	<.001 ***
SNR -6 dB:L2	-0.02	0.03	-0.08 – 0.04	-0.78 (509.57)	.44
SNR 0 dB:L2	0.04	0.03	-0.02 – 0.1	1.25 (509.44)	.21
SNR 3 dB:L2	0.03	0.03	-0.03 – 0.09	1.04 (509.48)	.3
SNR -6 dB:L3	-0.02	0.03	-0.08 – 0.04	-0.68 (509.75)	.5
SNR 0 dB:L3	0.01	0.03	-0.04 – 0.07	0.49 (509.69)	.63
SNR 3 dB:L3	0.07	0.03	0.01 – 0.13	2.26 (509.4)	.02 *
SNR -6 dB:L4	0.01	0.03	-0.05 – 0.06	0.17 (509.59)	.86
SNR 0 dB:L4	0.06	0.03	0.004 – 0.12	2.11 (509.52)	.04 *
SNR 3 dB:L4	0.11	0.03	0.05 – 0.17	3.63 (509.7)	<.001 ***

Random Effects			
	Variance	S.D.	Correlation
Participant (Intercept)	0.01	0.09	NA
Condition Order (Intercept)	0.001	0.04	NA
Residual	0.008	0.1	NA

Model fit

R^2	Marginal (fixed effects)	Conditional (fixed and random effects)
	0.23	0.66

Key: *p* values for fixed effects calculated using Satterthwaites approximations, * <.05, ** <.01, *** <.001.
Confidence Intervals have been calculated using the Wald method via the *confint.merMod* function in *lme4*. R^2 values were computed via the *r2* function in the *Performance* package (Lüdecke et al., 2021) based the calculation by Johnson (2014).
Model equation: peakdilation ~ snr * lightlevel + (1 | participant) + (1 | cond_order)
Base levels for beta estimates are SNR -3 dB, L1. All effects are estimated with respect to the base levels (i.e., treatment encoding).

10.4.3 Signal-Averaged Peak Dilation MEM: ANOVA Results

The *anova* function from the *lmerTest* R package (version 3.1-3) (Kuznetsova et al., 2017) was used to compute scaled F-values and degrees of freedom using Satterthwaites' method to compare the earlier findings reported in Chapter 9 to those of the current analysis. For the MEM reported here, there was a significant main effect of SNR on peak dilation, $F(3, 509.2) = 33.03$, $p < .001$ and a significant main effect of light level on peak dilation, $F(3, 509.4) = 87.04$, $p < .001$. The interaction between light level and SNR on peak dilation was also significant, $F(9, 509.53) = 2.56$, $p = .006$. The results of the MEM were similar to the results of the RANOVA in Chapter 9.

10.4.4 Signal-Averaged Peak Dilation MEM: Post Hoc Analyses

Post hoc comparisons were completed using the *emmeans* R package (Lenth, 2021). There were a few differences between the results of the RANOVA post hoc analyses (Table 14, Table 15) and the signal-averaged peak dilation MEM post hoc analyses (Table 22, Table 23).

When using light level as the grouping variable (Table 22), the comparison between SNR -3 dB and 0 was significant at L1. Additionally, the comparison between SNR-6 and SNR 3 dB at L2 was also significant. Conversely, the comparison between

SNR -6 dB and SNR -3 dB at L2, the comparison between SNR -3 dB and SNR 0 dB at L3, and the comparison between SNR -3 dB and SNR 3 dB at L4 all became non-significant.

When using SNR as the grouping variable (Table 23), the comparison of peak dilations between L2 and L3 at SNR -6 dB and the comparison of peak dilation between L2 and L3 at SNR -3 dB became non-significant. Conversely, the comparison of peak dilation between L1 and L3 at SNR 3 dB became significant.

Table 22. Signal-Averaged Peak Dilation MEM Post Hoc Simple Effects Analyses for Peak Dilation – Light Level as Grouping Variable

Light Level	Comparison (SNR dB)		Mean difference (mm)	95%CI – low	95% CI – high	df	Pairwise *t* statistic	*p*	*Sig.*	*Sig. in 9.4.5.1*
L1	-6	-3	-0.03	-0.07	0.01	525.21	-1.37	.17		
L1	-6	0	0.04	0	0.08	524.86	1.92	.06		
L1	-6	3	0.13	0.08	0.17	525.22	5.79	<.001	*	*
L1	-3	0	0.07	0.03	0.11	524.8	3.34	<.001	*	
L1	-3	3	0.16	0.12	0.2	524.83	7.26	<.001	*	*
L1	0	3	0.09	0.04	0.13	524.92	3.96	<.001	*	*
L2	-6	-3	-0.05	-0.1	-0.01	524.79	-2.49	.01		*
L2	-6	0	-0.02	-0.06	0.02	524.58	-0.88	.34		
L2	-6	3	0.07	0.03	0.12	524.93	3.39	<.001	*	
L2	-3	0	0.03	-0.01	0.08	524.8	1.61	.11		
L2	-3	3	0.13	0.08	0.17	524.7	5.87	<.001	*	*
L2	0	3	0.09	0.05	0.13	524.94	4.27	<.001	*	*
L3	-6	-3	-0.05	-0.09	-0.01	524.8	2.34	.02		
L3	-6	0	0.01	-0.04	0.05	524.92	0.3	.77		

Light Level	Comparison (SNR dB)		Mean difference (mm)	95%CI – low	95% CI – high	df	Pairwise *t* statistic	*p*	*Sig.*	*Sig. in 9.4.5.1*
L3	-6	3	0.04	0	0.08	525.12	1.74	.08		
L3	-3	0	0.06	0.01	0.1	524.98	2.62	.01		*
L3	-3	3	0.09	0.05	0.13	525.11	4.06	<.001	*	*
L3	0	3	0.03	-0.01	0.08	525.12	1.44	.15		
L4	-6	-3	-0.02	-0.07	0.02	524.79	-1.13	.26		
L4	-6	0	-0.02	-0.06	0.03	524.96	-0.74	.46		
L4	-6	3	0.02	-0.02	0.07	524.86	1.08	.28		
L4	-3	0	0.01	-0.03	0.05	524.97	0.38	.71		
L4	-3	3	0.05	0.01	0.09	525	2.23	.03		*
L4	0	3	0.04	0	0.08	525.03	1.83	.07		

Note. p is not Bonferroni corrected. * in *Sig.* column indicates significant differences after using the Bonferroni correction for multiple comparisons (e.g., *p* is significant at 0.001). Df = degrees of freedom (method: Kenward-roger). Df are fractional because whole-plot and subplot variations are combined when standard errors are estimated. CI= Confidence interval.

Table 23. Signal-Averaged Peak Dilation MEM Post Hoc Simple Effects Analyses for Peak Dilation – SNR as Grouping Variable

SNR (dB)	Comparison (light level)		Mean difference (mm)	95% CI – low	95% CI – high	df	Pairwise *t* statistic	*p*	*Sig.*	*Sig. in 9.4.5.1*
-6	L1	L2	0.12	0.07	0.16	525.07	5.40	<.001	*	*
-6	L1	L3	0.18	0.13	0.22	525.20	8.10	<.001	*	*
-6	L1	L4	0.20	0.15	0.24	525.10	8.85	<.001	*	*
-6	L2	L3	0.06	0.02	0.10	524.80	2.78	.006		*
-6	L2	L4	0.08	0.04	0.12	524.78	3.58	<.001	*	*
-6	L3	L4	0.02	-0.03	0.06	524.81	0.81	.42		
-3	L1	L2	0.09	0.05	0.14	524.80	4.39	<.001	*	*
-3	L1	L3	0.16	0.11	0.20	524.80	7.33	<.001	*	*
-3	L1	L4	0.20	0.16	0.24	524.81	9.37	<.001	*	*
-3	L2	L3	0.06	0.02	0.10	524.79	2.94	.003		*
-3	L2	L4	0.11	0.06	0.15	524.80	4.98	<.001	*	*
-3	L3	L4	0.04	0.00	0.09	524.80	2.04	.04		
0	L1	L2	0.06	0.01	0.10	524.80	2.66	.008		
0	L1	L3	0.14	0.10	0.19	525.11	6.54	<.001	*	*
0	L1	L4	0.14	0.09	0.18	525.04	6.31	<.001	*	*

SNR (dB)	Comparison (light level)		Mean difference (mm)	95% CI – low	95% CI – high	df	Pairwise *t* statistic	*p*	*Sig.*	*Sig. in 9.4.5.1*
0	L2	L3	0.09	0.04	0.13	524.97	3.93	<.001	*	*
0	L2	L4	0.08	0.04	0.12	524.93	3.70	<.001	*	*
0	L3	L4	-0.01	-0.05	0.04	524.97	-0.23	.82		
3	L1	L2	0.06	0.02	0.11	525.22	2.87	.004		
3	L1	L3	0.09	0.04	0.13	525.28	3.96	<.001	*	
3	L1	L4	0.09	0.05	0.13	525.38	4.15	<.001	*	*
3	L2	L3	0.03	-0.02	0.07	525.24	1.15	.25		
3	L2	L4	0.03	-0.01	0.07	524.96	1.30	.19		
3	L3	L4	0.00	-0.04	0.05	524.96	0.13	.89		

Note. p is not Bonferroni corrected. * in *Sig.* column indicates significant differences after using the Bonferroni correction for multiple comparisons (e.g., p is significant at 0.001). Df = degrees of freedom (method: Kenward-roger). Df are fractional because whole-plot and subplot variations are combined when standard errors are estimated. CI = confidence interval.

10.5 Discussion

10.5.1 Summary of Results

The aim of Study 1B was to examine the effects of light level and SNR on signal-averaged peak dilation using mixed-effects modelling. Specifically, the benefits of using mixed-effects modelling relative to RANOVA in the analysis of pupillometric data was investigated. The MEM provided information about the magnitude and direction of the effects for each level of each predictor from the base levels (L1, SNR -3 dB). To compare the earlier findings reported in Chapter 9 to those of the current analysis, *F* values and post hoc analyses were computed. As expected, significant main effects were found for light level and SNR and there was also a significant interaction.

The MEM post hoc analyses aligned with and strengthened the findings reported in 9.4.5.1 in that, peak dilation is more sensitive to changes in SNR conditions, in dimmer light levels. A greater number of significant differences were found in peak dilation between SNR conditions in the dimmer light levels (L1 and L2) and fewer significant differences were found in peak dilation between SNR conditions in the brighter light levels (L3 and L4) in the MEM. Additionally, there were no significant differences in peak dilation between SNR conditions in the brightest light level (L4) in the MEM.

In Section 9.4.5.1, evidence that more adverse SNR conditions (SNR -6 dB and SNR -3 dB) led to a greater number of significant differences in peak dilation between light levels was reported. These differences were reduced in the MEM when compared to the RANOVA. However, there was still a greater number of significant differences in peak dilation between light levels in the more adverse SNR conditions (SNR -6 dB, SNR -3 dB and SNR 0 dB) in the MEM, when compared to the easiest SNR condition (SNR 3 dB). Two comparisons were significant in SNR 3 dB (i.e., L1 vs. L3 and L1 vs. L4).

The MEM reported here and the RANOVA reported in Section 9.4.5 were not nested models, therefore, the differences in results between the two analysis methods cannot be tested statistically. However, the differences in the patterns of the significant results can be described. The next section compares the significant effects computed in MEM reported here and the RANOVA reported in Section 9.4.5.

10.5.2 The Effects of SNR and Light Level on Peak Dilation Using Mixed-Effects Modelling

Consistent with findings from the RANOVA, the MEM post hoc analyses confirmed that peak dilation is more sensitive to the SNR in dimmer light levels. The pattern of significant results was more pronounced in the MEM. There were more significant differences in peak dilation between SNR conditions in dimmer light levels and there were less significant differences in peak dilation between SNR conditions in brighter light levels. This reaffirmed that dimmer light levels should be used to achieve maximum sensitivity of peak dilation when measuring listening effort using a speech-in-noise task.

Further support was also reported for the finding that more difficult SNR conditions resulted in more differences in peak dilation at different light levels. However, there was a less definitive pattern that more difficult SNR conditions resulted in more differences between light levels. Four significant differences between light levels in SNR -6 dB, SNR -3 dB, and SNR 0 dB were found in the MEM. Additionally, there were two significant differences in SNR 3 dB. Unlike the results when using the RANOVA, the comparison between SNR -6 dB and SNR -3 dB was non-significant in the MEM. On the other hand, one difference in SNR 3 dB was significant in the MEM when compared to the RANOVA.

It is possible that the variation in significant differences between the RANOVA results and the MEM results was due to the additional statistical power that can be achieved using MEMs, by way of their structure and ability to account for additional sources of variance (i.e., condition order) (Quené & van den Bergh, 2004).

10.5.3 MEM Versus RANOVA

Due to the overall congruence between the results reported here and those reported in Chapter 9, the effects of light level and SNR and the interaction will not be discussed again here. A discussion regarding these results is provided in Section 9.5. The rest of this discussion focuses on the merits of mixed-effects modelling for the analysis of the repeated-measures data collected in this study.

There are multiple advantages to using MEMs to analyse repeated-measures data that were outlined in Section 10.1.2. An additional advantage involves the maximum likelihood methods that are inherent to MEMs. These methods make use of all the information available in the data. This can lead to more precise effect estimates and avoids the information loss that occurs when data are subject to averaging to analyse differences in a set of means (Boisgontier & Cheval, 2016; Cairns, 1986; Detry & Ma, 2016; Judd et al., 2012). Furthermore, compared with RANOVA, MEMs may have enhanced statistical power when analysing effects, mainly due to more accurate modelling of the variance-covariance matrix (or matrices) at each level of sampling (e.g., condition order and participants in the current chapter) (Quené & van den Bergh, 2004).

There are many sources of variation that contribute to pupil size and dilation (Tryon, 1975; Winn et al., 2018). Due to the flexibility of MEMs, participant and condition order were simultaneously accounted for in the final model (10.4.2). Condition order is an important random effect to control for as time-on-task and fatigue are known sources of variation in TEPRs (see Section 6.5). Condition order represents a time dependent measure (e.g., 1 = first condition, 2 = second condition, etc). Because all participants completed all 16 conditions of the speech-in-noise task, it was likely that responses measured in same condition (e.g., the first condition) were related. Therefore, it would be remiss to not account for this relatedness in the model. The addition of this random effect significantly improved the model fit to the data. Condition order accounted for 6.4% of the total variance in the final model reported in Section 10.4.2. The model partitioned variance associated with condition order from the error term. Thereby, the power to detect the fixed effects increased.

This chapter aimed to highlight the benefits of using mixed-effects modelling in the analysis of pupil data and to provide a conceptual bridge to the next chapter which examines the use of trial-level pupil data in MEMs. One of the main reasons that MEMs are gaining in popularity in psychology research is in their ability to account for participant and stimuli (e.g., sentence) variability, simultaneously (Baayen et al., 2008). Due to the traditional "signal-averaging" approach employed in most pupillometry studies, it is not possible to include a random effect like stimulus (i.e., sentence) and account for the variability in TEPRs that may be attributed to the unique sentences or trials when data are signal-averaged by participant.

10.5.4 Future Directions

Due to advancements in the application of mixed-effects modelling in psychological research, the ability of researchers to make well-informed inferences about data, with acknowledgement of and accountability to experimental design has improved. The utility of MEMs in being able to account for multiple sources of variation inherent in the experimental design was demonstrated in this chapter. Due to the signal-averaging approach used, the variance associated with individual trials or sentences could not be accounted for. This possibility is explored in the next chapter.

Pupil signals are inherently noisy which is why pupil data are typically signal-averaged over multiple trials. However, this method relies on certain assumptions. Namely, that the pupil response associated with the event of interest is consistent over trials/sentences and that it does not change with time. This is likely not the case for pupil responses. Furthermore, when pupil data are signal-averaged over multiple trials a large and rich dataset is condensed to one data point per condition, per participant. Volpert-Esmond et al. (2018) recently demonstrated the utility of analysing (non-signal-averaged) trial-level EEG data via mixed-effects modelling. Furthermore, some pupillometry researchers have already demonstrated the feasibility of using trial-level pupil data (see Section 6.1.4).

The power and flexibility of MEMs enables the use of trial-level pupil data and may lead to fundamental improvements in understanding of the factors that contribute to pupil signals at the trial-level. By using the trial-level pupil data, random effects of sentence, trial number, and participant can be included in the model. By partitioning a greater number of sources of variance from the error term, the power to detect fixed effects may be increased without first having to signal-average. By signal-averaging, potentially important information about the variance that exists within pupil signals may be lost. At worst, this may lead to erroneous conclusions about TEPRs. Due to the random noise that is characteristic of pupil signals, it is possible that the use of trial-level data may not provide any benefits to statistical power or the conclusions that can be drawn. The next chapter explores the use of trial-level pupil data with MEMs.

10.5.5 Conclusion

This chapter has investigated and demonstrated the various potential benefits of using mixed-effects modelling in place of traditional RANOVAs to analyse pupil data with repeated-measures. The MEM did not result in any major differences in statistical interpretation of the current dataset, however, by accounting for additional sources of variation, the power to detect fixed effects in the model may have increased. Furthermore, no imputation or list-wise deletion to account for missing data was needed when mixed-effects modelling was used. The flexible nature of MEMs allows researchers to have tighter control over analysis methods.

11 STUDY 1C - USING TRIAL-LEVEL PUPIL DATA AND MIXED-EFFECTS MODELLING TO EXAMINE THE EFFECTS OF SNR AND LIGHT LEVEL ON TEPRS

11.1 Background and Aims

The primary aim of this chapter was to examine the feasibility of using trial-level pupil data to analyse the effects of light level and SNR on TEPRs using mixed-effects modelling. In doing so, random effects like sentence, trial number, and participant can be simultaneously included in the model. The inclusion of these random effects offers the potential to further improve the power to detect the effects of light level and SNR and the interaction between these variables. Examination of trial-level pupil data may enhance understanding of the factors that contribute to pupil dilation during effortful listening.

Researchers may have avoided using trial-level pupil data in TEPR analyses due to the assumption that such data are too noisy to show task-evoked responses. Typically, signal-averaging is employed to reduce the noise in trial-level pupil data. The averaged response is assumed to reflect the true response to an experimental condition and the difference between averaged responses is assumed to reflect the true effect of experimental manipulations. However, signal-averaging ignores potentially meaningful variance inherent to trial-level pupil data, including how a response may change, trial by trial and/or sentence by sentence.

Recent applications of mixed-effects modelling in psychological research have demonstrated that trial-level pupil data can be analysed without signal averaging (Clewett et al., 2020; Clewett et al., 2018; Cohen Hoffing et al., 2020; Leuchs et al., 2017; Volpert-Esmond et al., 2018; Volpert-Esmond et al., 2021; Wetzel et al., 2020) and therefore, will not be susceptible to the drawbacks of signal-averaging (detailed in Section 11.1.1). It is possible that some of the variance in trial-level TEPRs can be partitioned from the error term by the inclusion of random effects like sentence and trial number when using mixed-effects modelling. This is not possible in analyses that use signal-averaged data. In signal-averaging, trial-level pupil data are averaged across trials which attenuates potentially meaningful variance. The following section describes typical pupil data and the signal-averaging method for pupil data, as well as the advantages and disadvantages of this method.

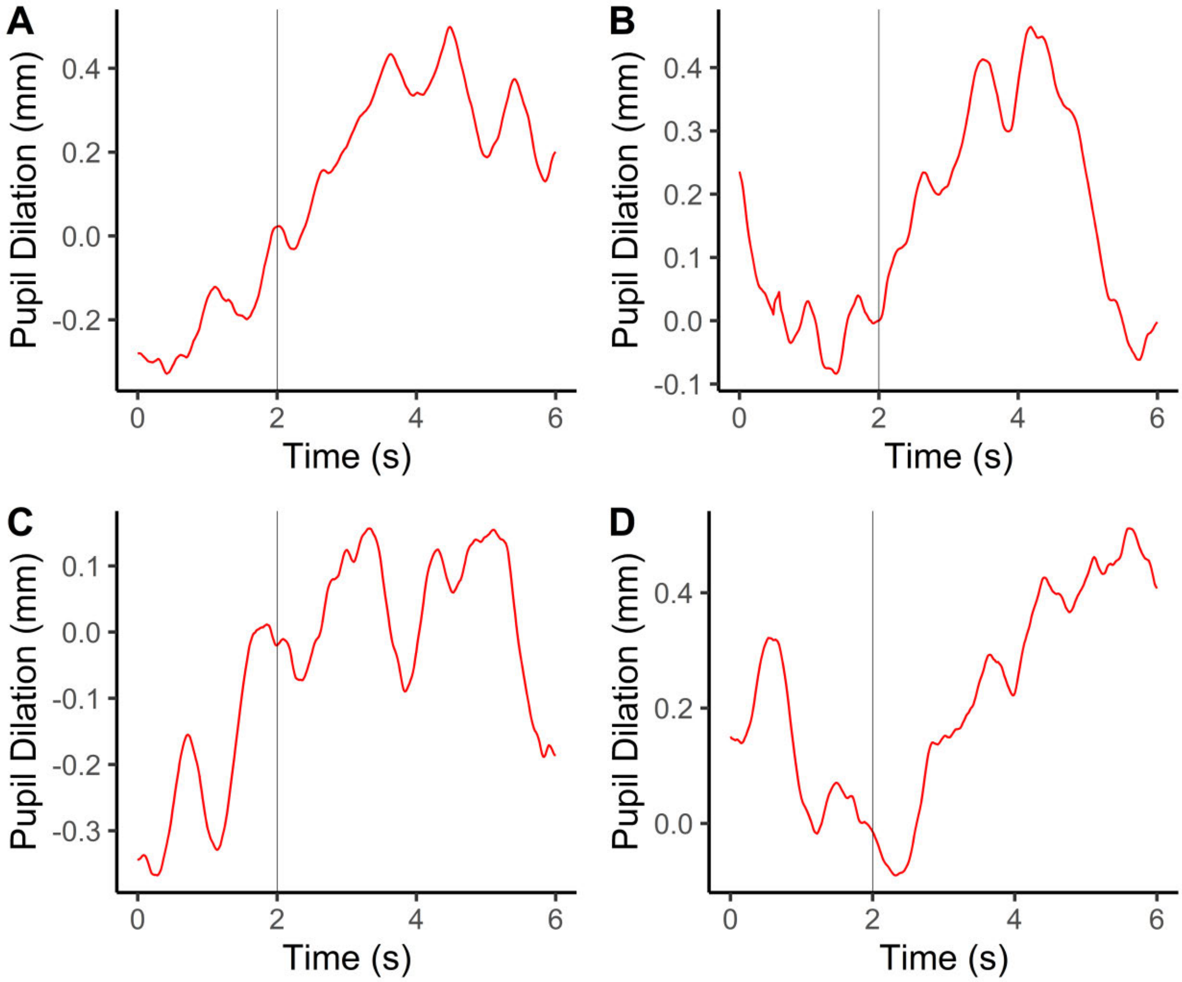
11.1.1 Signal-Averaging of TEPRs

Pupil data are typically noisy, even after baseline correction and initial smoothing, because there are many factors that can affect pupil size and dilation that are independent of the task/condition. These include factors that are internal, such as, arousal or control state and external, such as, light conditions and eye/head movements (Joshi & Gold, 2020).

Variability and fluctuations in the pupil response between trials can be seen in Figure 13. Figure 13 displays four separate smoothed, baseline-corrected pupil trials measured under the same conditions in the same participant and taken from the current dataset. It can be seen in Figure 13 D that peak dilation does not occur until near the end of the trial, unlike in Figure 13 A, B, C. Furthermore, Figure 13 A, B, C appear to show two prominent peaks, rather than one peak. This variability may be due to factors which were not related to the event of interest (the experimental stimuli). Therefore, task-related pupil signals are typically buried amongst seemingly random noise which researchers may not be able to control (Winn et al., 2018). This can make the task-related pupil signals difficult to detect in trial-level data.

Figure 13

Four Examples of Single, Baseline-Corrected, Smooth Pupil Traces from the Current Dataset



Note. This figure shows four examples of four single baseline-corrected and smoothed trials from current dataset (see Chapter 8 for general methods). Pupil responses measured from the same participant in L4, SNR -3 dB. The y axis indicates pupil dilation (mm) and the x axis indicates time (s). The vertical line at 2 s indicates sentence onset. A: Trial 1, B: Trial 2, C: Trial 3, D: Trial 4.

To reduce the effects of noise in the signal, pupil data are traditionally analysed in a series of steps. First, several replicate raw pupil diameter waveform measurements are made during a task assumed to be of consistent difficulty. Second, raw pupil diameter waveform data are pre-processed to remove blinks and other artefacts (see

Section 8.4 for details on how this was done in the current study). Third, individual pupil diameter waveforms are aligned with a time-locking event, usually at the onset of a stimulus. The waveforms of all trials are subsequently averaged (see Section 9.3.4.1 for details on how this was done for Study 1A and B). This method attenuates the noise in the pupil signal, so that the event-related pupil signal is more distinct. Theoretically, signal-averaging improves the SNR by a ratio of $\sqrt{N}$, where N is the number of waveforms averaged. In the present study that means the SNR in the averaged signal would be four times larger than that in a single trial, provided the task-relevant pupil response followed the same temporal pattern after onset of the stimulus.

The fourth step is to measure pupil parameters such as baseline diameter, peak dilation, mean dilation, and peak dilation latency in the averaged waveform at the participant-level. Typically, these parameters are then used as dependent variables in traditional statistical tests, like ANOVAs. Within ANOVAs, the participant-level pupil measures are averaged again, by experimental condition, to determine statistical differences between condition means

Much of the current discussion and analysis has been inspired by methodological development discussions in the ERP domain (Volpert-Esmond et al., 2018; Volpert-Esmond et al., 2021; Vossen et al., 2011) and the recent use of trial-level pupil data in cognitive domains (other than auditory) (Clewett et al., 2020; Clewett et al., 2018; Cohen Hoffing et al., 2020; Leuchs et al., 2017; Wetzel et al., 2020).

Like ERP data, signal-averaging of pupil data relies on several assumptions (Luck, 2014). For one, it assumes that the pupil activity related to the time-locked events is the same for every trial. Pupil activity related to an event/task/stimulus is likely to vary from trial to trial. Thus, this assumption is usually violated in pupil data (and ERP data). This is not necessarily an issue, if researchers are specifically interested in using the mean of a set of responses as a measure of central tendency. However, the mean is not always a good measure of central tendency, especially for skewed, non-normal data (Luck, 2014).

There may be negative effects for analyses and inference when the variations in pupil activity (usually deemed “noise”) are not random, that is, there are systematic variations in pupil responses across trials (e.g., due to boredom, fatigue, learning, inattention, habituation, etc.) (Vossen et al., 2011). For example, if pupil responses are larger in the first half of the condition but very small in the second half of the condition (as may happen with fatigue, for example), the averaged waveform will be an intermediate set of pupil measures that may not reflect any pupil trace measured (Luck, 2014). Fatigue effects have been demonstrated in TEPRs (see Section 6.5.2) but have not been examined in trial-level TEPRs. Similarly, the pupil response may be consistently small for most of the trials and large for a few trials, leading to a skewed distribution of response values. These variations may systematically differ as a function of time or experimental condition (e.g., light level and SNR in the current study).

A similar issue relates to variations in response latency (Luck, 2014). When using pupillometry to measure cognitive effort, more effortful tasks generally show longer response latency (as was shown in Study 1A in Section 9.4.4), often interpreted as reflecting processing load (Koelewijn et al., 2015; Van Der Meer et al., 2010; Zekveld et al., 2011). Processing speed can be affected by fatigue (DeLuca, 2005). Therefore, it is possible that there will be consistent and systematic variations in peak dilation latency at the trial-level between conditions, but there could also be consistent and systematic variations in peak dilation latency at the trial-level within-conditions based on whether they are easy and/or hard or if the participant is fatigued.

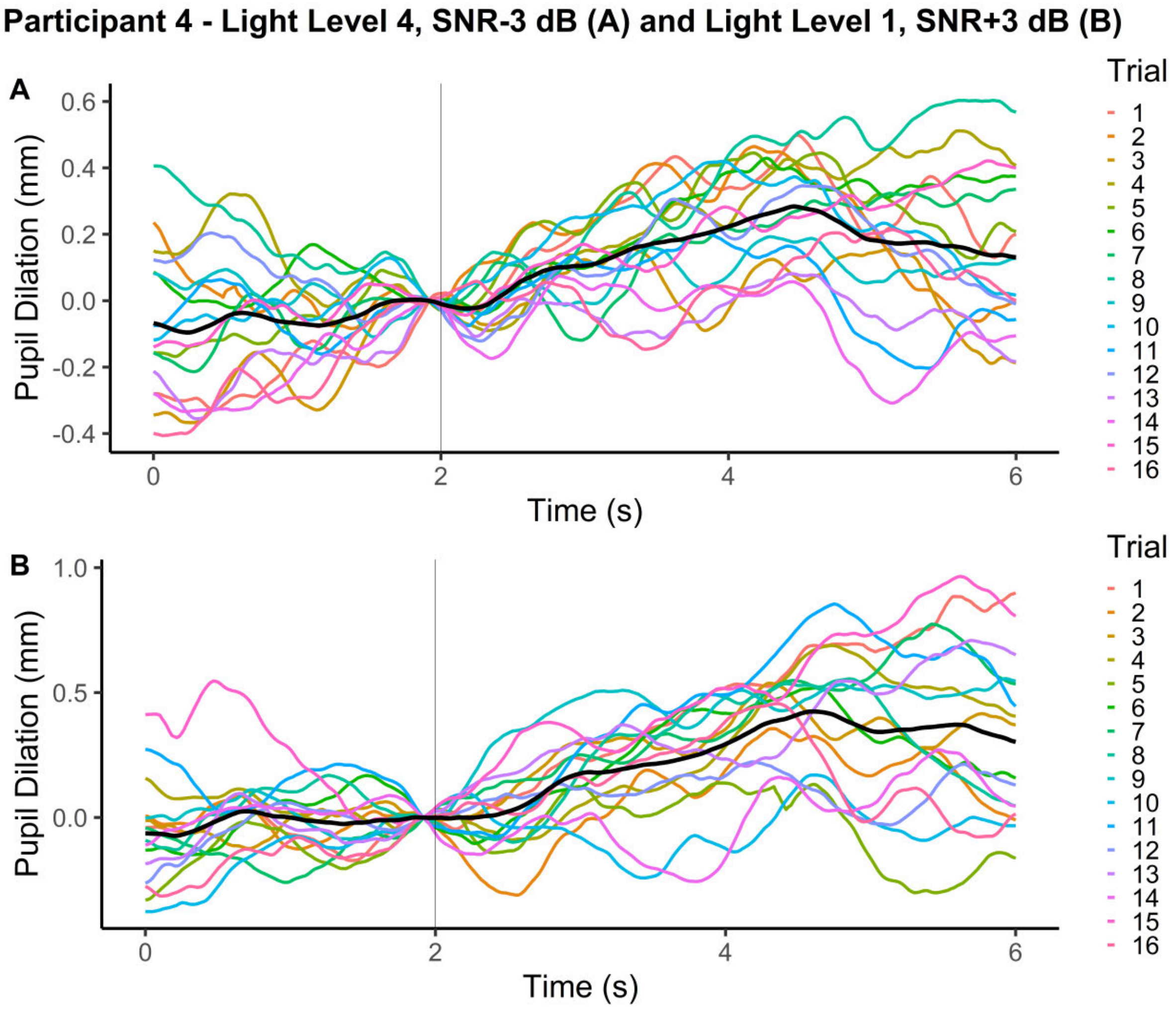
In pupillometry studies, it is possible that peak dilation latency is more varied across trials (within each condition) in more adverse SNR conditions. If this is the case, signal-averaging will attenuate peak dilation to a greater extent in the averaged waveform in the conditions with high variability in peak dilation latency than in conditions with less variability. This may happen, even when the amplitudes of those responses are the same.

Furthermore, if the pupil responses of different trials measured under the same condition contain both positive and negative sections (as is often the case in pupil data) and when these positive and negative sections occur at the same time, this could lead to “cancellation” of the effect, when signal-averaged. This has the potential to lead to spurious conclusions about differences in response amplitude between conditions in signal-averaged data. The partial pooling method inherent in mixed-effects modelling acknowledges all information in the data and accounts for extreme cases without attenuating signals (Mahr, 2017).

As seen in Figure 14, there is substantial variation within and between trials and conditions. Figure 14 A also shows that the pupil waveforms for trials 13, 14, 15, and 16 are generally smaller in amplitude than other trials. This pattern is less consistent in Figure 14 B. This may be indicative of a fatigue effect in the more adverse SNR -3 dB condition, when performed in L4 (Figure 14 A) which is not apparent in the easier SNR 3 dB, when performed in L1 (Figure 14 B).

Figure 14

16 Baseline-Corrected, Smoothed Pupil Traces from Two Conditions in Current Dataset



Note. Two series of 16 pupil waveforms representing 16 trials from two conditions from the same participant. The y axis indicates pupil dilation (mm) and the x axis indicates time (s). The vertical line at the 2 s point indicates sentence onset. The black line is the signal-averaged trace. Series A comes from the adverse SNR -3 dB performed in L4. Series B comes from the less adverse SNR 3 dB performed in L1.

In Figure 14, signal-averaging (represented by the black line) has resulted in some trial characteristics, like amplitude, being attenuated. If the fluctuations in the trials were due to random noise, this would not be an issue. However, it is possible that measurable factors contributed to this variance which have been ignored in signal-averaging. Accounting for these factors may strengthen analyses used to detect effects by partitioning sources of variance from the error term without attenuating signals. It also enables the acknowledgement of the underlying variance structure of the raw data which may lead to a better understanding of the factors that contribute to pupil responses.

In summary, while signal-averaging has the benefit of reducing the noise and increasing the ability to detect a signal in pupil data; information about the variance in the signal regarding how a response may change between trials and over time is lost. Furthermore, large and information-rich datasets are condensed into a single datum per participant and/or per condition (Volpert-Esmond et al., 2018). Because signal-averaging is typically applied when examining TEPRs, examination of trial-by-trial variations in TEPRs has often been neglected in the past. Some of the noise in unaveraged TEPRs, may be attributable to time-on-task or fatigue effects, habituation, stimuli (e.g., sentence), or another measurable phenomenon.

To date, trial-by-trial variations in TEPRs have not been examined via the use of trial-level pupil data. This gap is likely due to the noisy nature of raw pupil responses and the assumption that signal-averaging must be done (Winn et al., 2018). It may also be partly due to statistical constraints inherent to traditional analysis methods like ANOVA. For example, it is not possible to include multiple effects which might lead to variance in trial-level pupil data while simultaneously examining the effects of the experimental variables in ANOVAs. Alternative analysis methods such as mixed-effects modelling allow for these effects to be simultaneously included in models.

For these reasons and the potential drawbacks of signal-averaging explained above, examination of the use of mixed-effects modelling with pupil data that has not been subject to signal-averaging is warranted.

11.1.2 Analysing Trial-Level TEPRs Using Mixed-Effects Modelling

Chapter 10 presented mixed-effects modelling as a flexible method for the analysis of data with repeated measures (Meteyard & Davies, 2020). Due to the signal-averaging of pupil data that was applied in the previous chapters, it was not possible to take full advantage of the flexibility of mixed-effects modelling methods. For example, it was not possible to include random effects like trial number or sentence in the MEM reported in Chapter 10. It is possible that some of the variability in the trial-level pupil data has a meaningful and measurable origin which led to systematic variation in the data. In addition to retaining potentially important information about the variance in the data, using the trial-level data and including variables like trial number and sentence as random effects may increase the power of the analysis to detect the fixed effects.

However, the potential benefits of analysing trial-level pupil data come at the cost of working with noisier data. As mentioned in Section 10.1.2, MEMs deal with extreme values by partial pooling (i.e., the same way they handle missing data). Therefore, extreme values (which could be due to noise in the signal) are pulled towards an average (Mahr, 2017). Values which are already close to that average are not pulled as much. Participants and/or items with extreme values have less of an impact on parameter estimates than those without. Therefore, MEMs may handle the noisiness of trial-level pupil data more efficiently.

It is possible that trial-level pupil data collected for this study is too noisy to provide valid estimates of how light level and SNR affect TEPRs. If this is the case, the effect pattern reported in Chapters 9 and 10 would not be apparent in the results of analyses using the trial-level pupil data. It is also possible that the inclusion of sentence and trial number as random effects will not partition enough variance from the error term in trial-level peak and mean dilation to be worth working with the additional noise inherent in trial-level pupil data. If this is the case, the model fit would not improve with the inclusion of these random effects in the model building process. The variance explained by the random effects would also not increase. The trade-off between the benefits and costs of using trial-level versus signal-averaged data (which will be influenced by the nature and aims of the study) has not been explored in TEPRs in a listening effort paradigm. This chapter seeks to address that gap. The relative advantages and disadvantages of using trial-level pupil data over signal-averaged pupil data for the analysis of data in the current experimental paradigm are discussed.

11.2 Significance

Trial-level pupil data has not been examined in a listening effort paradigm. This chapter aimed to enhance the understanding of the factors that contribute to TEPRs at the trial-level which may have implications for data processing and methods of analysis.

11.3 Method

This chapter comprises an analysis of trial-level pupil data using mixed-effects modelling. As mentioned in Section 9.5.1, peak dilation appears more sensitive to light level and SNR manipulations. Trial-level peak dilation was recently shown to be meaningfully related to behaviour a mental arithmetic task (Cohen Hoffing et al., 2020). However, because peak dilation is based on a single value, it is more likely to be affected by random noise than mean dilation. Therefore, both peak and mean dilation are reported in the following section, but it was expected that peak and mean dilation would show similar patterns, provided the peak dilation values were not a product of noise in the trial-level pupil data.

11.3.1 Participants

Information related to the participants can be found in Chapter 8 - General Methods, Section 8.1.

11.3.2 Materials

Information related to the materials can be found in Chapter 8 - General Methods, Section 8.2.

11.3.3 Procedure

Information related to the procedure can be found in Chapter 8 - General Methods, Section 8.3.

11.3.4 Pupil Data Pre-Processing

Information related to the procedure can be found in Chapter 8 - General Methods, Section 8.4 for details regarding the pre-processing of the pupil data. Analysis code for the additional pre-processing used for Study 1C are available at https://osf.io/am6uv/.

11.3.4.1 Baseline Correction

The same subtractive baseline correction method as reported in Section 9.3.4.2 was used for the trial-level data.

11.3.4.2 Pupil Parameters

Within the baseline-corrected trace, peak and mean dilation were measured for each pupil waveform using the same procedure as reported in Section 9.3.4.3.

11.3.5 Data Analyses

Data analyses and visualisations were completed in R open source software (version 4.0.1) (R Core Team, 2020) and R Studio (version 1.3.1073) (RStudio Team, 2020). Analysis and visualisation codes are available at https://osf.io/am6uv/.

As in Chapter 10, the model building process and results are reported based on recommendations set out in the recent paper: “Best practice guidance for linear mixed-effects models in psychological science” (Meteyard & Davies, 2020).

Data visualisations were completed using *ggplot2* (version 3.3.3) (Wickham, 2016). Data were analysed using the *lmer* function of the *lme4* R (version 3.5.2) (Bates et al., 2015) and *p* values from summary outputs for MEMs were obtained via the *lmerTest* package (version 3.1.0) (Kuznetsova et al., 2017). Model comparisons were conducted using likelihood ratio tests (Table 24 and Table 25). As in Chapter 10, the default variance-covariance structure specified by *lmer* models (unstructured) was used. This means that the model specifies no pattern in the covariance matrix, that is, no observations are “equally correlated” and there is no “structure” between neighbouring values (Barnett et al., 2010)

11.3.5.1 Fixed Effects

As in Chapter 10, the fixed effects were light level, SNR, and the interaction between light level and SNR.

11.3.5.2 Random Effects

The maximal random effects structure justified by the design was used. Random effects that did not improve the model, or resulted in singular fit were removed (Barr et al., 2013). The random effects structure in the model allows us to specify grouping variables in the data that lead to non-independence of observations (Volpert-Esmond et al., 2021).

Three random intercept effects were included in the model: Participant, trial number, and sentence.

1. Participant was included as a random effect to account for the fact that responses from the same participant will be related.
2. By-trial number random intercepts were included to account for the relationship between responses measured at each trial over the test session (like condition order in Chapter 10). Trial number also introduced non-independence to the data and thus, can be modelled similarly. Trial number is a variable that is somewhat time-dependent and reflects the order in which the specific trial occurred. For example, 1 in the trial number variable indicated that trial was the first trial completed over the entire test session. Trial number 256 in the trial number variable indicated the last trial across the entire test session. Trials that occurred at the beginning or end of a test session may be related across participants. While it is not a “time” variable per

se, its inclusion accounted for changes in the pupil response that may occur over the trials in a test session. By including by-trial number random intercepts, the model estimates each trial number's deviation from the fixed intercept, which reflects that some trial numbers may elicit larger or smaller pupil response than others. The effects of SNR, light level and their interaction refer to the average trial without incorporating the uncertainty associated with how the effects change over trials. Trial number is examined as a fixed effect in Chapter 14

3. By-sentence random intercepts were included to account for the fact that responses to the same sentence stimuli cannot be classed as "independent". For the analyses using signal-averaged data (Chapters 9 and 10), it was assumed that all the sentences were equivalent in terms of the TEPR evoked. Thus, it was also assumed that when presented under the same conditions, different sentences would elicit the same (or very similar) pupil responses. While the sentences were designed to be equivalent and lists that showed performance differences were not included in this study, the unique sentences may still elicit differences in pupil responses. By including by-sentence random intercepts, the model estimates each sentence's deviation from the fixed intercept, which reflects that some sentences elicit larger pupil response than others. The effects of SNR, light level and their interaction refer to the average sentence without incorporating the uncertainty associated with how the effects change over sentences.

In the model building process, all possible random slopes justified by the design were tested for peak and mean dilation MEMs. In the final model for peak dilation, random slopes for light level by participant were included. This random slope was the only slope that allowed for a non-singular fit. Thus, the model also estimated by-participant adjustments to the slope of the light level fixed effect. The inclusion of random slopes for light level by participant acknowledges that all participants do not respond to light level in the same way. The significant improvement in model fit (Table 24) indicated that this was an important component that should be included in the final model. All random slopes for mean dilation resulted in a singular fit and therefore, were not included in the final model.

11.3.5.3 Model Building

Details regarding the random effects structure for each *lmer* model and the model comparison results can be found in Table 24 and Table 25.

Table 24. Study 1C, Model Comparison and the Model Building/Selection Process for Peak Dilation

	Sampling Units			N total observations = 8749, N Participants = 36, N Trial Number = 256, N Sentence = 256										
	Model specification	**Model name**	**Simpler Model**	**Fixed Effects added**		**Random Effects**			**Model fit**				**LRT Test against nested**	
						Participant	**Trial**	**Sentence**	**AIC**	**BIC**	**LL**	**df**	**df**	**X^2**
1	Participant random effect	MEM_null	-	-		intercept			-191.92	-170.69	98.96	3	-	-
2	Participant and trial number random effects	MEM_null 2	MEM_null	-		intercept	intercept		-317.45	-289.14	162.73	4	1	127.53 ***
3	Participant, trial number and sentence random effects	MEM_null 3	MEM_null2	-		intercept	intercept	intercept	-444.48	-409.10	227.24	5	1	129.03 ***
4	Participant, trial number and sentence random effects	MEM_null 3_Part1	MEM_null3	-		Intercept and random slope by light level	intercept	intercept	-1817.42	-1718.3	922.71	14	9	1390.9 ****
	Note: All additional random slopes resulted in singular fit thus were not taken forward in modelling (Barr, Levy, Scheepers, & Tily, 2013).													
4	Fixed effect x 1	MEM_fixedeff_snr	MEM_null3	SNR		Intercept and random slope by light level	intercept	intercept	-2051.4	-1931.1	1042.69	17	3	239.96 ***
5	Fixed effect x 2	MEM_fixedeff_ll	MEM_fixedeff_snr	SNR + light level		Intercept and random slope by light level	intercept	intercepts	-2103.0	-1961.4	1071.5	20	3	57.58 ***
6	Two-way interaction	MEM_int_TL	MEM_fixedeff_ll	SNR x light level		Intercept and random	intercept	intercept	-2139.1	-1933.9	1098.6	29	9	54.17 ***

						slope by light level								

Note. AIC = Aikake Information Criterion, BIC = Bayesian Information Criterion, LL = LogLikelihood, LRT – Likeilhood Ratio Test, * = p <.05, ** = p <.01, *** = p <.001.

Table 25. Study 1C, Model Comparison and the Model Building/Selection Process for Mean Dilation

	Model specification	Model name	Simpler Model	Fixed Effects added		Random Effects			Model fit				LRT Test against nested	
	Sampling Units			N total observations = 8749, N Participants = 36, N Trial Number = 256, N Sentence = 256										
						Participant	Trial	Sentence	AIC	BIC	LL	df	df	X^2
1	Participant random effect	MEM_null	-	-		intercept			-4021.5	-4000.3	2013.8	3	-	-
2	Participant and trial number random effects	MEM_null 2	MEM_null	-		intercept	Intercept		-4132.7	-4104.4	2070.3	4	1	113. 13***
3	Participant, trial number and sentence random effects	MEM_null 3	MEM_null2	-		intercept	Intercept	intercept	-4176.7	-4141.3	2093.3	5	1	45.9 9 ***
	Note: All random slopes resulted in singular fit thus were not taken forward in modelling (Barr, Levy, Scheepers, & Tily, 2013).													
4	Fixed effect x 1	MEM_fixe deff_snr	MEM_null3	SNR		Intercept	Intercept	Intercept	-4421.4	-4364.7	2218.7	8	3	250. 69***
5	Fixed effect x 2	MEM_fixe deff_ll	MEM_fixed eff_snr	SNR + light level		Intercept	intercept	Intercept	-4706.1	-4628.3	2364.1	11	3	290. 77***
6	Two-way interaction	MEM_int_ TL	MEM_fixed eff_ll	SNR x light level		Intercept	intercept	intercept	-4748.6	-4607.1	2394.3	20	9	60.4 7***

Note. AIC = Aikake Information Criterion, BIC = Bayesian Information Criterion, LL = LogLikelihood, LRT – Likeilhood Ratio Test, * = *p* <.05, ** = *p* <.01, *** = *p* <.001.

11.3.5.4 Bayesian Regression Modelling in Stan

The “keeping it maximal” advice was followed in the *lmer* models reported (Barr et al., 2013). Only terms that did not improve the model fit and/or resulted in a singular fit were removed. In *lmer* analyses, if a random effect structure is too complex to be supported by the data, the model returns a warning stating “singular fit”. This means that some dimensions of the variance-covariance matrix have been estimated as exactly zero (Bates et al., 2021, p. 49).

It is acknowledged that the inclusion of by-participant and by-trial number slopes for the SNR, light level and the interaction between light level and SNR makes theoretical sense, that is, different participants may respond to the independent variables differently and different trial numbers may affect responses to the independent variables differently. Thus, the deletion of these random effects in the *lmer* model may have resulted in distorted estimates and SEs.

To confirm that deletion of these terms did not distort *lmer* model estimates and SEs, a Bayesian regression analysis was conducted for peak dilation using the *brms* R package (Bürkner, 2018). This analysis used the full maximal random-effects structure: random intercepts for each participant, sentence, and trial number; and by-participant and by-trial random slopes for SNR, light level and SNR by light level interaction. Bayesian analyses provide valid estimates for the maximal random effects structure, regardless of singular fits calculated in *lmer.*

11.4 Results

11.4.1 Trial-Level Peak Dilation MEM

11.4.1.1 Model Assumptions

Inspection of residual plots indicated linearity. However, they also indicated some heteroskedasticity (fanning pattern). It is possible that this heteroskedasticity is due to an unmodeled fixed or random effect. While MEMs assume homoskedasticity, model estimates are usually robust to violations of this assumption (Schielzeth et al., 2020). Log-transforming the response variable did not lessen the heteroskedasticity. Thus, it should be noted that the final model that is reported in the following section showed some heteroskedasticity in the residuals, but this is not anticipated to negatively impact the model estimates. Additionally, QQ plots indicated slight non-normality of residuals. However, MEMs are also robust to this type of violation (Gelman & Hill, 2007; Knief & Forstmeier, 2020).

11.4.1.2 Model Convergence

The final *lmer* model did not converge on the first run. To address this, the control parameters were adjusted with an optimiser as suggested by Brown (2021). The “all_fit” function from the *afex* R package (Singmann et al., 2021) was used to assess which optimiser was suitable. The “bobyqa”[18] optimiser led to model convergence.

11.4.1.3 Descriptive Statistics

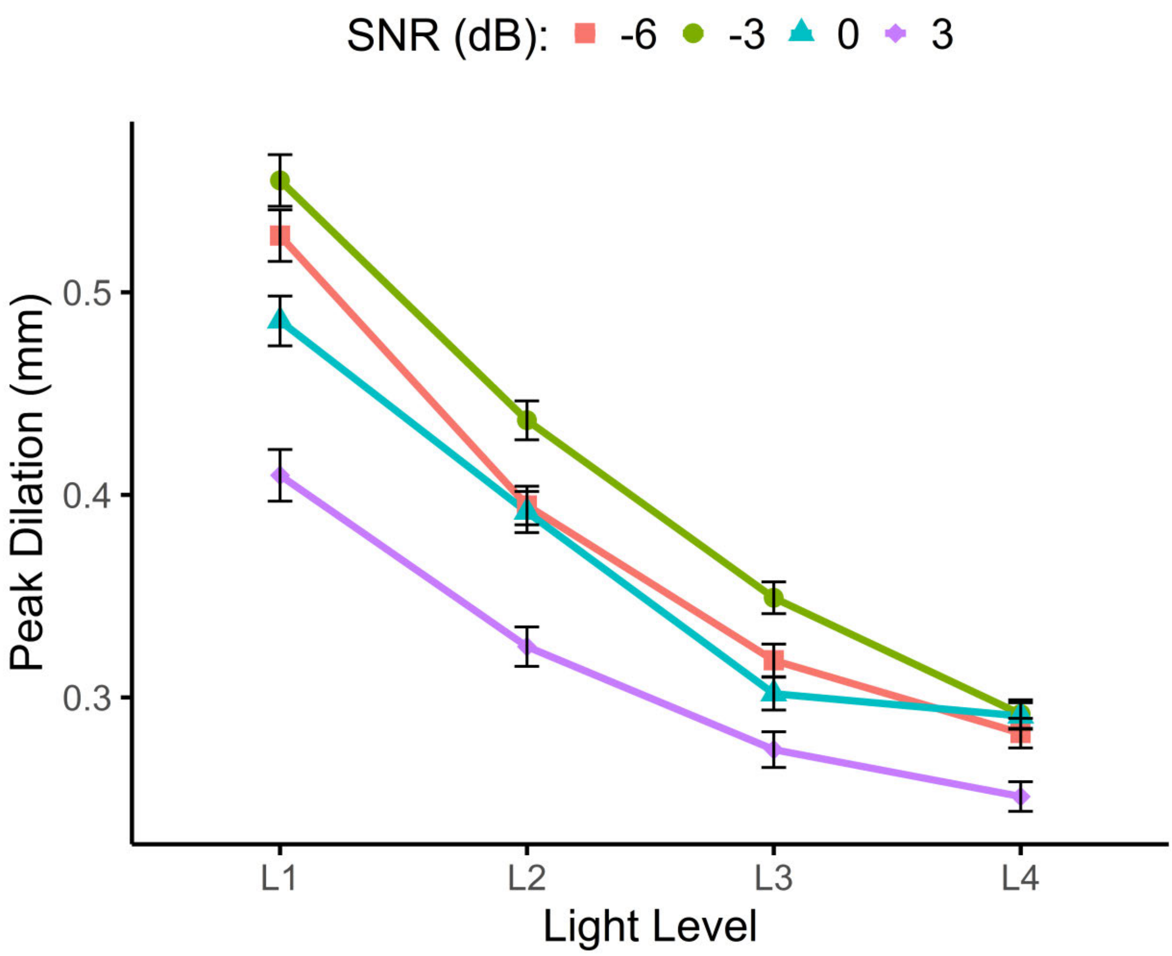
Descriptive statistics for trial-level peak dilation data are presented in Figure 15. This initial visualisation of the trial-level peak dilation data showed a similar pattern of results to the signal-averaged peak dilation data (Figure 11). Due to the increase in *N* achieved through using the trial-level peak dilation data, Figure 15 also showed smaller within-subject standard error bars than Figure 11. This verified that the current exploration of using trial-level pupil data may provide similar estimates to

[18] Based on the results of all_fit(), this argument was added to the final model: control = lmerControl(optimizer = "bobyqa"

signal-averaged data, and by using mixed-effects modelling with this data, variance associated with trial number and sentence may be partitioned from the error term.

Figure 15

Mean Trial-Level Peak Dilation



Note. This figure shows mean trial-level peak dilation (within-subjects standard error bars) across all participants in the 16 light*SNR conditions

11.4.1.4 Trial-Level Peak Dilation MEM: Final Model Results

The final model results are presented in Table 26. The estimated beta values reflect the estimated change in peak dilation from the base levels. The base levels were as follows: L1 for light level, SNR -3 dB for SNR, and the combination of L1 SNR -3 dB for the interaction terms.

Significance levels for the estimated beta values were consistent between the signal-averaged peak dilation MEM (Table 21) and the trial-level peak dilation MEM regarding the fixed effects light level and SNR. However, there are a greater number of significant differences in the estimated beta values in Table 26. Across SNRs, peak dilation decreased from L1. Across light levels, peak dilation decreased from SNR -3 dB (except in SNR -6 dB). When L2 and SNR 3 dB co-occur, the predicted absolute value of peak dilation is higher than expected based on the main effects of L1 and SNR -3 dB alone. When L3 and SNR 3 dB co-occur, the predicted absolute value of peak dilation is higher than expected based on the main effects of L1 and SNR -3 dB alone. When L4 and SNR 0 dB co-occur, the predicted absolute value of peak dilation is higher than expected based on the main effects of L1 and SNR -3 dB alone. When L4 and SNR 3 dB co-occur, the predicted absolute value of peak dilation is higher than expected based on the main effects of L1 and SNR -3 dB alone.

Table 26. Final Peak Dilation MEM Output Table for Study 1C

Fixed Effects					
	Beta estimate	SE	95% CI	t (df)	*p*
Intercept	0.55	0.031	0.496 - 0.615	18.17 (41.62)	<.001 ***
SNR -6 dB	-0.02	0.012	-0.047 - 0.003	-1.73 (8411.65)	.08
SNR 0 dB	-0.07	0.013	-0.097 - -0.048	-5.79 (8371.49)	<.001 ***
SNR 3 dB	-0.15	0.022	-0.174 - -0.124	-11.69 (8048.67)	<.001 ***

L2	-0.12	0.022	-0.163 - -0.075	-5.34 (61.72)	<.001 ***
L3	-0.21	0.018	-0.257 - -0.159	-8.31 (53.59)	<.001 ***
L4	-0.26	0.018	-0.305 - -0.220	-12.02 (63.14)	<.001 ***

SNR-6 dB:L2	-0.02	0.018	-0.056 - 0.014	-1.18 (8400.86)	.27
SNR 0 dB:L2	0.03	0.018	-0.007 - 0.062	1.58 (8364.47)	.11
SNR 3 dB:L2	0.04	0.018	0.002 - 0.072	2.06 (8199.72)	.04 *
SNR -6 dB:L3	-0.002	0.018	-0.037 - 0.032	-0.14 (8327.38)	.89
SNR 0 dB:L3	0.02	0.018	-0.012 - 0.058	1.3 (7956.49)	.19
SNR 3 dB:L3	0.07	0.018	0.038 - 0.110	4.03 (6611.81)	<.001 ***
SNR -6 dB:L4	0.01	0.031	-0.024 - 0.045	0.59 (8418.41)	.55
SNR 0 dB:L4	0.07	0.013	0.034 - 0.104	3.88 (8412.44)	<.001 ***
SNR 3 dB:L4	0.1	0.012	0.070 - 0.140	5.86 (8239.66)	<.001 ***

Random Effects

		Variance	SD	Correlation		
Light level | Participant (Intercept)		0.03	0.17	NA		
	L2 (slope)	0.01	0.11	-.63		
	L3 (slope)	0.02	0.13	-.76	0.85	
	L4 (slope)	0.01	0.11	-.92	0.70	0.90
Sentence (Intercept)		0.001	0.03	NA		
Trial Number (Intercept)		0.001	0.03	NA		
Residual		0.04	0.2	NA		

Model fit

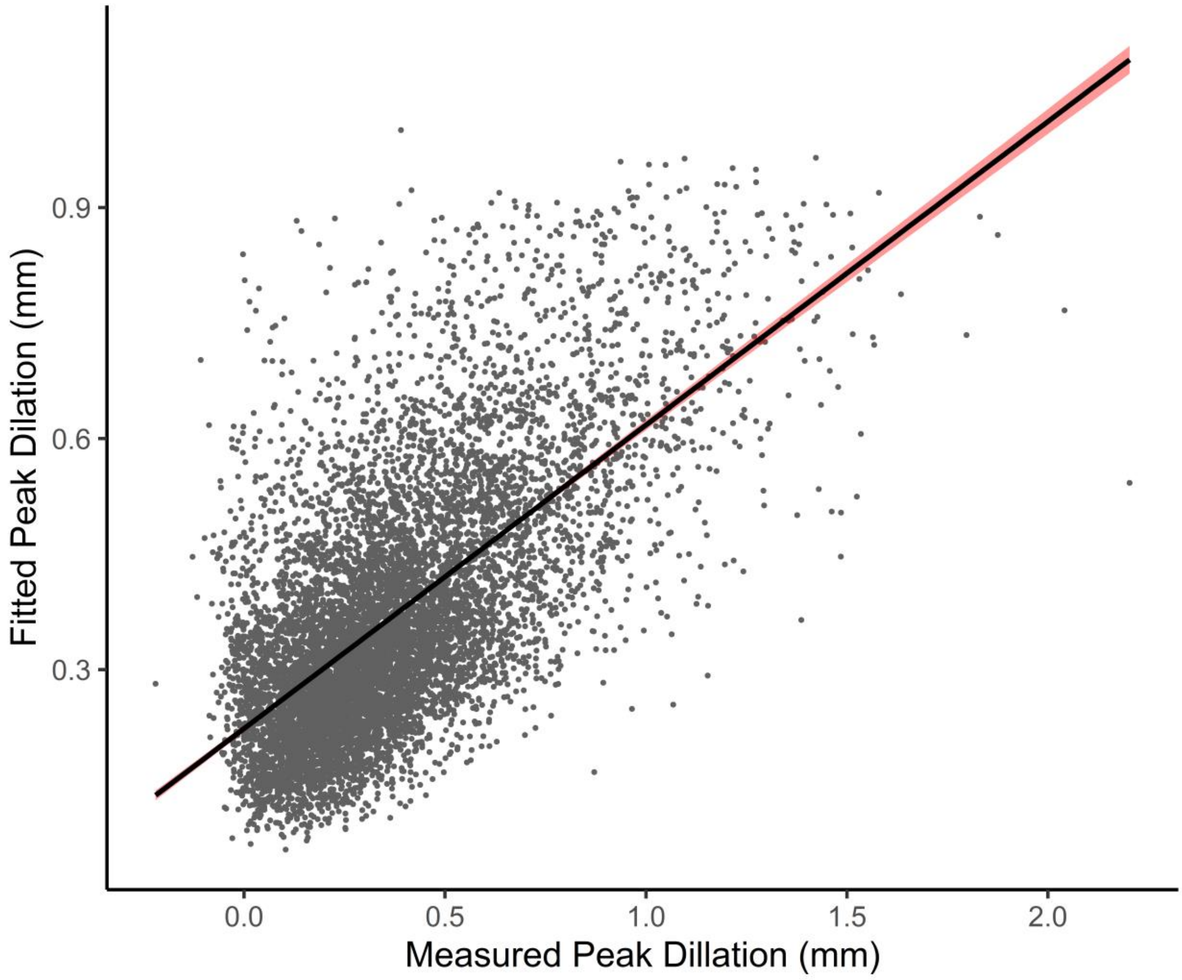
R^2	Marginal (fixed effects)	Conditional (fixed and random effects)
	0.12	0.39

Key: *p* values for fixed effects were calculated using Satterthwaite's approximations, * <.05, ** <.01, *** <.001.
Confidence Intervals have been calculated using the Wald method via the *confint.merMod* function in *lme4*.
R^2 values were computed via the *r2* function in the *Performance* package (Lüdecke et al., 2021) based the calculation by Johnson (2014).
Model equation: peakdilation ~ snr * lightlevel + (1 +lightlevel | participant) + (1 | trial_exp) + (1 | sentence)
Base levels for beta estimates are SNR -3 dB, L1. All effects are estimated with respect to the base levels.
SE = Standard Error, CI = confidence interval, SD = standard deviation

Figure 16 shows the linear trend line and confidence interval (in red) in a scatter plot of the peak dilation values estimated in the MEM and the measured peak dilation values. The scattering around the trendline aligns with the R^2, that is, there is unexplained variance in the measured peak dilation values which could not be used to estimate the fitted values. Figure 17 shows the same relationship but is split by condition. There is more scattering around the trendline in L1 compared to the other light levels. This indicates that there was more variance in measured peak dilation values in these conditions.

Figure 16

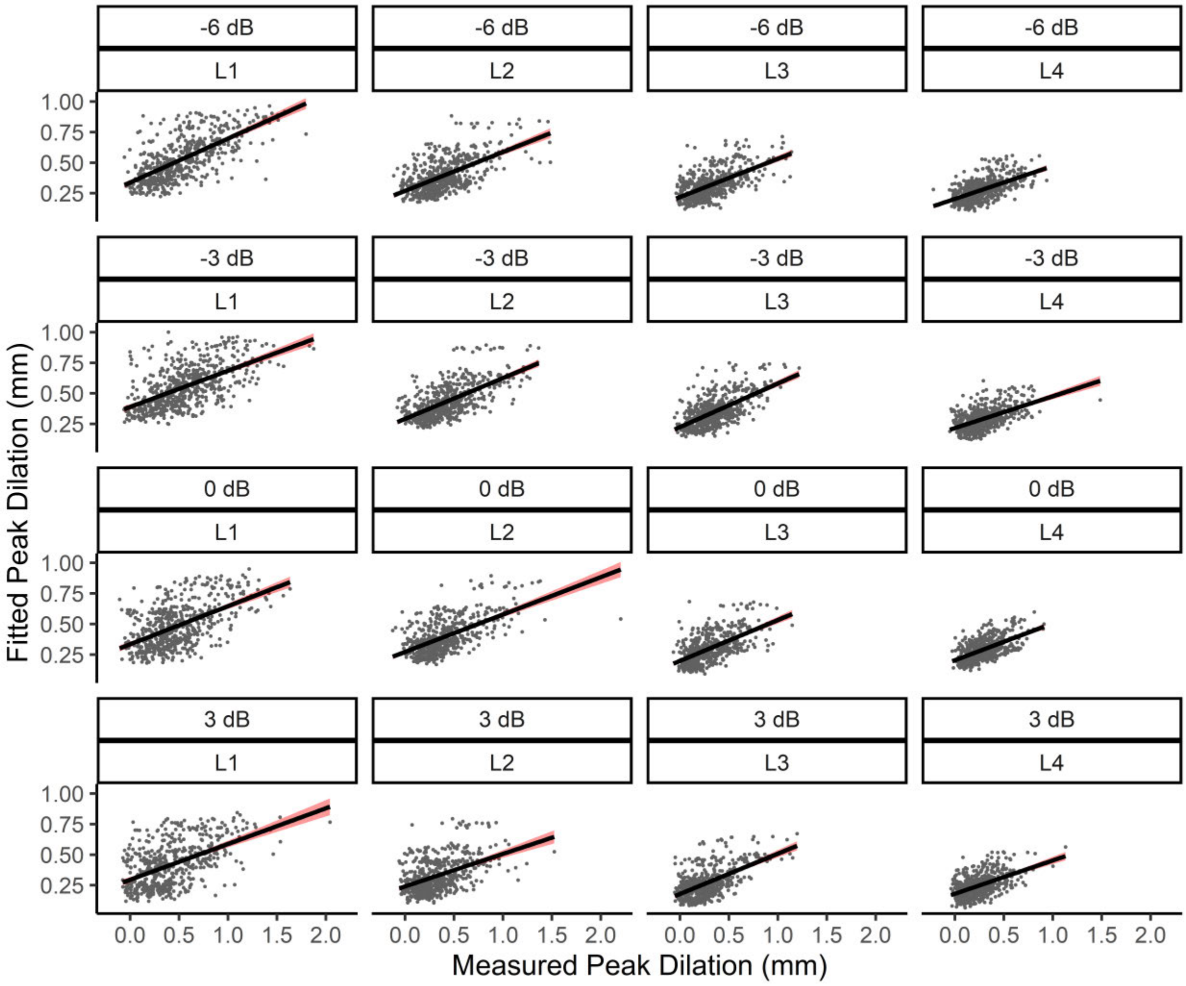
Scatterplot of Fitted Peak Dilation Values by the Measured Peak Dilation Values



Note. This figure shows a scatter plot of peak dilation values fitted by the "Trial-Level Peak Dilation MEM: Final Model" (y axis) and the measured peak dilation values (x axis). The trendline and confidence interval (in red) represents the linear relationship between all fitted and measured peak dilation values.

Figure 17

Scatterplot of Fitted Peak Dilation Values by the Measured Peak Dilation Values Split by Condition



Note. This figure shows a scatter plot of peak dilation values fitted by the "Trial-Level Peak Dilation MEM: Final Model" (y axis) and the measured peak dilation values (x axis) split by condition. The trendline and confidence interval (in red) represents the linear relationship between all fitted and measured peak dilation values.

11.4.1.5 Trial-Level Peak Dilation MEM: ANOVA Results

The *anova* function from the *lmerTest* R package (version 3.1-3) (Kuznetsova et al., 2017) was used to compute scaled F-values and denominator df using Satterthwaites' method. In line with the RANOVA in Chapter 9 and signal-averaged peak dilation MEM post hoc analyses in Chapter 10, when using trial-Level peak dilation in an MEM, there was a significant main effect of SNR on peak dilation, $F(3, 8308) = 81.67$, p <.001 and a significant main effect of light level on peak dilation, $F(3, 36.1) = 46.69$, p <.001. The interaction between light level and SNR on peak dilation was also significant, $F(9, 8171) = 6.04$, p <.001.

11.4.1.6 Trial-Level Peak Dilation MEM: Post Hoc Analyses

Post hoc comparisons were completed using the *emmeans* R package (Lenth, 2021). There were a few differences between the results of the signal-averaged peak dilation MEM post hoc analyses (Table 22, Table 23) and the trial-level peak dilation MEM post hoc analyses (Table 27, Table 28).

In the trial-level peak dilation MEM, when using light level as the grouping variable (Table 27), the comparison between SNR -6 dB and 0 was significant at L1. Additionally, all comparisons at L2 became significant (except the comparison between SNR -3 dB and SNR 0 dB). At L3, the comparisons between SNR -6 dB and SNR 0 dB and SNR -6 dB and SNR 3 dB became significant. At L4, the comparisons between SNR -6 dB and SNR 3 dB, and SNR 0 dB and SNR 3 dB became significant.

In the trial-level peak dilation MEM, when using SNR as the grouping variable (Table 28), the comparison between L2 and L3 was significant at SNR -6 dB and SNR -3 dB. This aligns with the findings of the RANOVA results reported in Chapter 9. In contrast to both the RANOVA and the signal-averaged peak dilation MEM (Table 23), the comparison between L3 and L4 was significant at SNR -3 dB and the comparison between L1 and L2 was significant at SNR 0 dB in the trial-level peak dilation MEM. Comparisons between L1 and L2, and L2 and L4 at SNR 3 dB were also significant.

Table 27. Trial-Level Peak Dilation MEM, Post Hoc Simple Effects Analyses for Peak Dilation – Light Level as Grouping Variable

Light Level	Comparison (SNR dB)		Mean difference (mm)	95% CI – low	95% CI – high	df	Pairwise *t* statistic	*p*	*Sig.*	*Sig. in 10.4.4*
L1	-6	-3	0.02	-0.003	0.05	8395.27	1.73	.08		
L1	-6	0	0.07	0.05	0.10	8410.98	5.79	<.001	*	
L1	-6	3	0.15	0.12	0.17	8315.51	11.67	<.001	*	*
L1	-3	0	0.05	0.03	0.08	8396.45	3.96	<.001	*	*
L1	-3	3	0.13	0.10	0.15	8455.38	9.85	<.001	*	*
L1	0	3	0.08	0.05	0.10	8434.79	6.05	<.001	*	*
L2	-6	-3	0.04	0.02	0.07	8399.77	3.45	<.001	*	
L2	-6	0	0.04	0.02	0.07	8399.18	3.57	<.001	*	
L2	-6	3	0.11	0.09	0.14	8389.91	8.88	<.001	*	*
L2	-3	0	0.002	-0.02	0.03	8358.41	0.13	.9		
L2	-3	3	0.07	0.04	0.09	8412.91	5.53	<.001	*	*
L2	0	3	0.07	0.04	0.09	8351.84	5.41	<.001	*	*
L3	-6	-3	0.02	-0.0002	0.05	8418.67	1.94	.05		
L3	-6	0	0.05	0.02	0.07	8437.76	3.86	<.001	*	
L3	-6	3	0.08	0.05	0.10	8332.19	5.87	<.001	*	

L3	-3	0	0.02	-0.0006	0.05	8333.21	1.92	.06		
L3	-3	3	0.05	0.03	0.08	8446.48	3.97	<.001	*	*
L3	0	3	0.03	0.001	0.05	8414.56	2.05	.04		
L4	-6	-3	0.01	-0.01	0.04	8403.04	0.89	.37		
L4	-6	0	0.003	-0.02	0.03	8382.48	0.26	.79		
L4	-6	3	0.04	0.02	0.07	8449.47	3.51	<.001	*	
L4	-3	0	-0.01	-0.03	0.02	8356.60	-0.62	.54		
L4	-3	3	0.03	0.01	0.06	8414.89	2.57	.01		
L4	0	3	0.04	0.02	0.07	8434.38	3.19	.001	*	

Note. p is not Bonferroni corrected. * in *Sig.* column indicates significant differences after correcting for multiple comparisons (e.g., *p* is significant at 0.001). df = degrees of freedom (method: Kenward-roger). df are fractional because whole-plot and subplot variations are combined when standard errors are estimated.

Table 28. Trial-Level Peak Dilation MEM, Post Hoc Simple Effects Analyses for Peak Dilation – SNR as Grouping Variable

SNR (dB)	Comparison (light level)		Mean difference (mm)	95% CI – low	95% CI – high	df	*Pairwise t statistic*	*p*	*Sig.*	*Sig. in 10.4.4*
-6	L1	L2	0.14	0.09	0.19	65.88	6.18	<.001	*	*
-6	L1	L3	0.21	0.16	0.26	58.23	8.26	<.001	*	*
-6	L1	L4	0.25	0.21	0.3	68.71	11.26	<.001	*	*
-6	L2	L3	0.07	0.04	0.11	107.81	4.09	<.001	*	
-6	L2	L4	0.11	0.07	0.15	87.64	5.81	<.001	*	*
-6	L3	L4	0.04	0.01	0.07	141.47	2.54	.01		
-3	L1	L2	0.12	0.07	0.16	64.66	5.28	<.001	*	*
-3	L1	L3	0.21	0.16	0.26	57.05	8.2	<.001	*	*
-3	L1	L4	0.26	0.22	0.31	65.54	11.88	<.001	*	*
-3	L2	L3	0.09	0.05	0.12	106.97	5.17	<.001	*	
-3	L2	L4	0.14	0.11	0.18	84.62	7.51	<.001	*	*
-3	L3	L4	0.05	0.02	0.09	133.05	3.4	<.001	*	
0	L1	L2	0.09	0.05	0.14	64.4	4.05	<.001	*	
0	L1	L3	0.18	0.13	0.24	57.42	7.27	<.001	*	*
0	L1	L4	0.19	0.15	0.24	67.45	8.7	<.001	*	*

0	L2	L3	0.09	0.06	0.13	108.94	5.4	<.001	*	*
0	L2	L4	0.1	0.06	0.14	85.8	5.34	<.001	*	*
0	L3	L4	0.01	-0.02	0.04	142.51	0.53	0.6		
3	L1	L2	0.08	0.04	0.13	66.78	3.61	<.001	*	
3	L1	L3	0.13	0.08	0.19	58.99	5.25	<.001	*	*
3	L1	L4	0.16	0.11	0.2	68.17	7.06	<.001	*	*
3	L2	L3	0.05	0.02	0.09	114.11	2.99	.003		
3	L2	L4	0.08	0.04	0.11	86.28	3.93	<.001	*	
3	L3	L4	0.02	-0.01	0.06	142.53	1.41	.16		

Note. p is not Bonferroni corrected. * in *Sig.* column indicates significant differences after correcting for multiple comparisons (e.g., *p* is significant at 0.001). df = degrees of freedom (method: Kenward-roger). df are fractional because whole-plot and subplot variations are combined when standard errors are estimated.

11.4.2 Trial-Level Mean Dilation MEM

11.4.2.1 Model Assumptions

Inspection of residual plots indicated linearity. However, they also indicated slight heteroskedasticity. While MEMs assume homoskedasticity, model estimates are usually robust to violations of this assumption (Schielzeth et al., 2020). Thus, it should be noted that the final model that is reported in the following section showed some heteroskedasticity in the residuals, but this is not anticipated to negatively impact the model estimates. Additionally, QQ plots indicated slight non-normality of residuals. However, MEMs are also robust to this type of violation (Gelman & Hill, 2007; Knief & Forstmeier, 2020; Schielzeth et al., 2020).

11.4.2.2 Model Convergence

All *lmer* models for mean dilation converged.

11.4.2.3 Trial-level Mean Dilation MEM: Final Model Results

The final model results are presented in Table 29. The estimated beta values reflect the estimated change in mean dilation from the base levels. The base levels were as follows: L1 for light level, SNR -3 dB for SNR, and the combination of L1 SNR -3 dB for the interaction terms. Significance levels for the estimated beta values were consistent between the trial-level peak dilation MEM (Table 26) and trial-level mean dilation MEM for the fixed effects light level and SNR.

Table 29. Final Mean Dilation MEM Output Table for Study 1C

Fixed Effects					
	Beta estimate	SE	95% CI	t (df)	*p*
Intercept	0.25	0.013	0.22 - 0.27	18.81 (86.97)	<.001 ***
SNR -6 dB	-0.012	0.011	-0.03 - 0.01	-1.08 (8466.94)	.28
SNR 0 dB	-0.07	0.011	-0.09 - -0.05	-6.34 (8471.35)	<.001 ***
SNR 3 dB	-0.13	0.011	-0.15 - -0.11	-11.98 (8267.93)	<.001 ***
L2	-0.06	0.011	-0.08 - -0.03	-5.05 (7065.24)	<.001 ***
L3	-0.1	0.011	-0.12 - -0.08	-9.02 (6365.37)	<.001 ***
L4	-0.13	0.011	-0.15 - -0.10	-11.48 (7021.95)	<.001 ***
SNR-6 dB:L2	-0.02	0.015	-0.05 - 0.01	-1.52 (8470.3)	.12
SNR 0 dB:L2	0.03	0.015	-0.003 - 0.06	1.78 (8470.68)	.08
SNR 3 dB:L2	0.03	0.016	0.0004 - 0.06	2 (8366.44)	.05*
SNR -6 dB:L3	-0.02	0.015	-0.05 - 0.01	-1.13 (8443.68)	.26
SNR 0 dB:L3	0.03	0.016	-0.004 - 0.06	1.71 (8217.74)	.08
SNR 3 dB:L3	0.06	0.016	0.03 - 0.09	3.95 (7074.03)	<.001 ***
SNR -6 dB:L4	-0.003	0.016	-0.03 - 0.03	-0.22 (8472.45)	.83
SNR 0 dB:L4	0.06	0.015	0.03 - 0.09	3.75 (8476.28)	<.001 ***
SNR 3 dB:L4	0.09	0.016	0.06 - 0.12	5.82 (8396.81)	<.001 ***

Random Effects			
	Variance	SD	Correlation
Participant (Intercept)	0.004	0.06	NA
Sentence (Intercept)	0.0004	0.03	NA
Trial Number (Intercept)	0.001	0.03	NA
Residual	0.03	0.17	NA

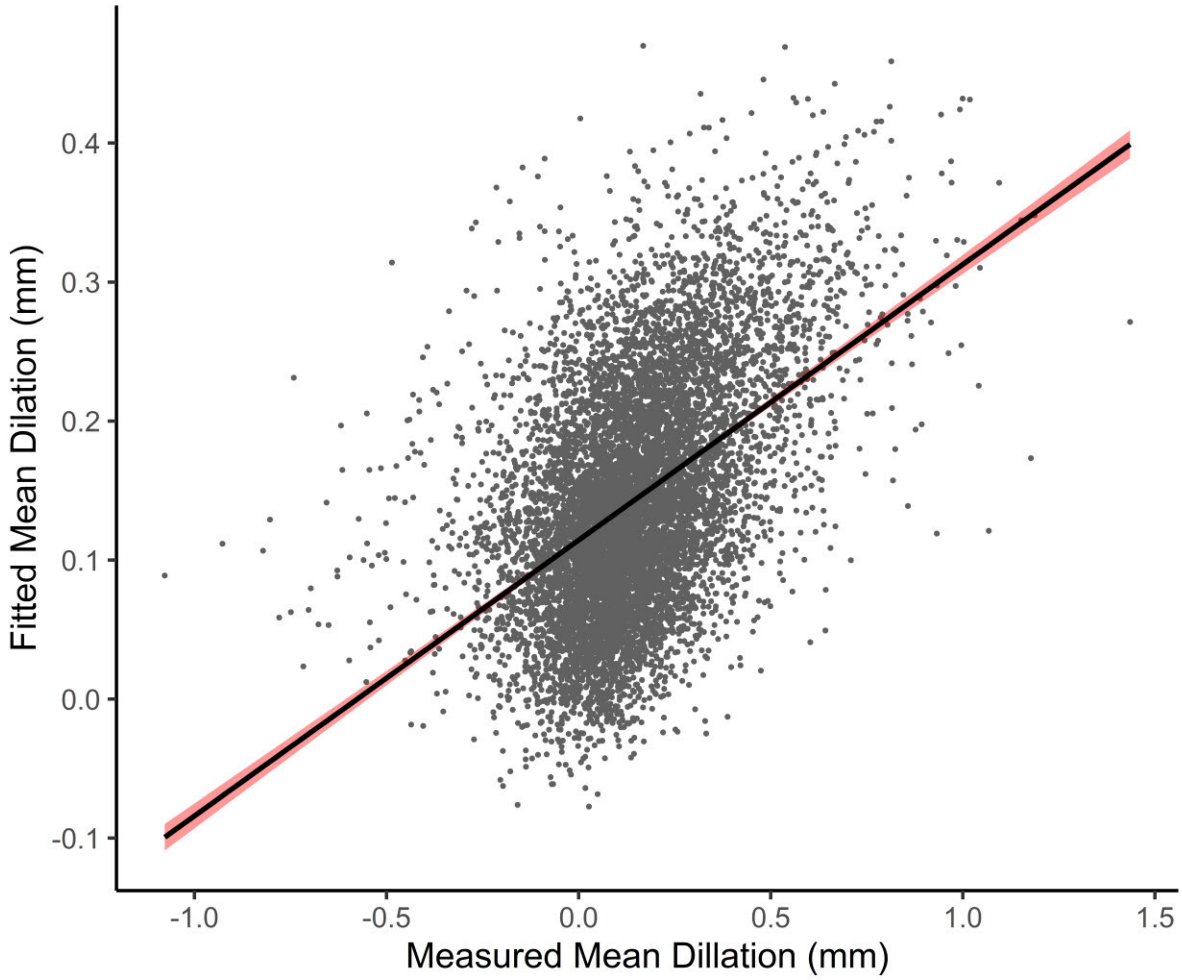
Model fit		
R^2	Marginal (fixed effects)	Conditional (fixed and random effects)
	0.06	0.19

Key: *p values* for fixed effects calculated using Satterthwaites approximations, , * <.05, ** <.01, *** <.001.
Confidence Intervals have been calculated using the Wald method via the *confint.merMod* function in *lme4*.
R^2 values were computed via the *r2* function in the *Performance* package (Lüdecke et al., 2021)
Model equation: meandilation ~ snr * lightlevel + (1 | participant) + (1 | trial_exp) + (1 | sentence)
Base levels for treatment contrasts are SNR -3 dB, L1, and SNR -3 dB*L1.
SE = Standard Error, CI = confidence interval, SD = standard deviation

Figure 18 shows the linear trend line and confidence interval (in red) in a scatter plot of the mean dilation values estimated in the MEM and the measured mean dilation values. The scattering around the trendline aligns with the R^2, that is, there is unexplained variance in the measured peak dilation values which could not be used to estimate the fitted values. Figure 19 shows the same relationship but is split by condition. There is more scattering around the trendline in L1 compared to the other light levels. This indicates that there was more variance measured peak dilation values in these conditions.

Figure 18

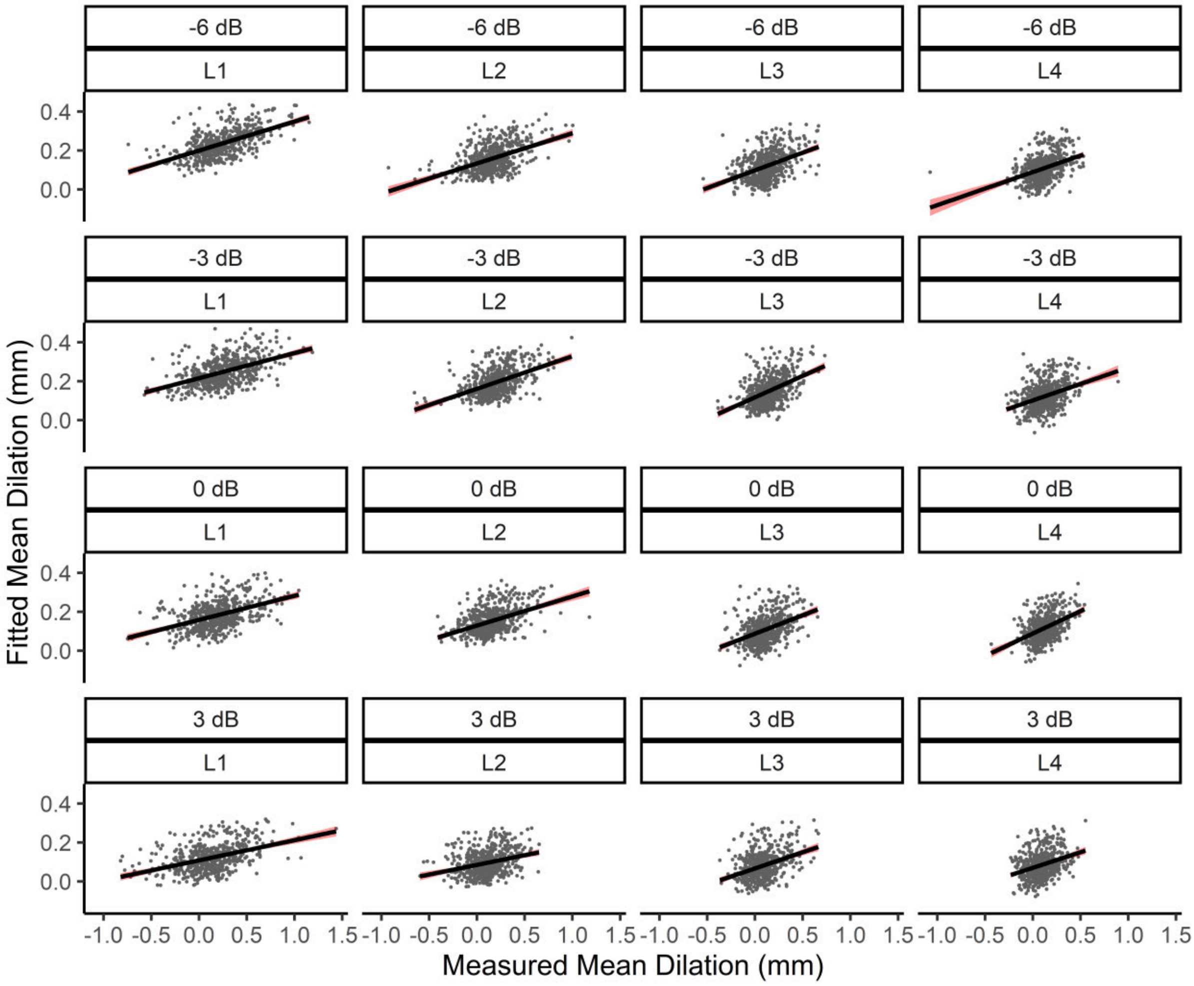
Scatterplot of Fitted Mean Dilation Values by the Measured Mean Dilation Values



Note. This figure shows a scatter plot of mean dilation values fitted by the "Trial-Level Mean Dilation MEM: Final Model" (y axis) and the measured mean dilation values (x axis). The trendline and confidence interval (in red) represents the linear relationship between all fitted and measured mean dilation values.

Figure 19

Scatterplot of Fitted Mean Dilation Values by the Measured Mean Dilation Values Split by Condition



Note. This figure shows a scatter plot of mean dilation values fitted by the "Trial-Level Mean Dilation MEM: Final Model" (y axis) and the measured mean dilation values (x axis) split by condition. The trendline and confidence interval (in red) represents the linear relationship between all fitted and measured mean dilation values.

11.4.2.4 Trial-Level Mean Dilation MEM: ANOVA Results

The *anova* function from the *lmerTest* R package (version 3.1-3) (Kuznetsova et al., 2017) was used to compute scaled F-values and denominator degrees of freedom using Satterthwaites' method. In line with the RANOVA in Chapter 9 and the trial-Level peak dilation MEM above (Section 11.4.1), when using trial-Level mean dilation in an MEM, there was a significant main effect of SNR on mean dilation, $F(3, 8333.4) = 89.44$, $p < .001$. There was a significant main effect of light level on peak dilation, $F(3, 3520.8) = 101.47$, $p < .001$. The interaction between light level and SNR on mean dilation was also significant, $F(9, 8347.8) = 6.75$, $p < .001$.

11.4.2.5 Trial-Level Mean Dilation MEM: Post Hoc Analyses

Post hoc comparisons were completed using the *emmeans* R package (Lenth, 2021). The purpose of these post hoc analyses was to compare significant differences in trial-level mean dilation MEM to trial-level peak dilation MEM. While some of the differences between the trial-level mean dilation MEM and trial-level peak dilation MEM were consistent, there were also some differences (see Table 30 and Table 31). The trial-level peak dilation MEM accounted for a larger proportion of total variance when compared to the trial-level mean dilation MEM. Therefore, the averaging of the dilation response may obscure some of the variance that is associated with the variables in the trial-level mean dilation MEM. However, the task-related signal is still measurable in the trial level mean dilation MEM.

Table 30. Trial-Level Mean Dilation MEM, Post Hoc Simple Effects Analyses for Mean Dilation – Light Level as Grouping Variable

Light Level	Comparison (SNR dB)		Mean difference (mm)	95% CI – low	95% CI – high	df	Pairwise *t* statistic	*p*	*Sig.*	*Sig. in 11.4.1.6 (peak)*
L1	-6	-3	-0.01	-0.03	0.01	8468.53	1.08	.28		
L1	-6	0	0.06	0.04	0.08	8468.37	5.16	<.001	*	*
L1	-6	3	0.12	0.10	0.14	8510.22	10.79	<.001	*	*
L1	-3	0	0.07	0.05	0.09	8498.71	6.33	<.001	*	*
L1	-3	3	0.13	0.11	0.15	8403.80	11.97	<.001	*	*
L1	0	3	0.06	0.04	0.08	8505.57	5.77	<.001	*	*
L2	-6	-3	-0.04	-0.06	0.01	8485.83	3.26	.001	*	*
L2	-6	0	0.006	-0.01	0.03	8460.46	0.59	.56		*
L2	-6	3	0.07	0.04	0.09	8482.30	6.07	<.001	*	*
L2	-3	0	0.04	0.02	0.06	8492.18	3.84	<.001	*	
L2	-3	3	0.10	0.08	0.12	8474.91	9.25	<.001	*	*
L2	0	3	0.06	0.04	0.08	8423.43	5.50	<.001	*	*
L3	-6	-3	-0.03	0.008	0.05	8490.06	2.69	.01		
L3	-6	0	0.01	-0.009	0.03	8432.73	1.17	.24		*

L3	-6	3	0.04	0.02	0.06	8506.31	3.65	<.001	*	*
L3	-3	0	0.04	0.02	0.06	8501.20	3.85	<.001	*	
L3	-3	3	0.07	0.05	0.09	8388.48	6.29	<.001	*	*
L3	0	3	0.03	0.006	0.05	8454.59	2.48	.01		
L4	-6	-3	-0.02	-0.04	0.006	8461.46	1.39	.17		
L4	-6	0	-0.004	-0.03	0.02	8491.62	-0.38	.70		
L4	-6	3	0.03	0.005	0.05	8490.22	2.40	.02		*
L4	-3	0	0.01	-0.01	0.03	8440.54	0.99	.32		
L4	-3	3	0.04	0.02	0.06	8507.67	3.83	<.001	*	
L4	0	3	0.03	0.009	0.05	8488.64	2.78	.01		*

Note. p is not Bonferroni corrected. * in *Sig.* column indicates significant differences after correcting for multiple comparisons (e.g., *p* is significant at 0.001). df = degrees of freedom (method: Kenward-roger). df are fractional because whole-plot and subplot variations are combined when standard errors are estimated.

Table 31. Trial Level Mean Dilation MEM, Post Hoc Simple Effects Analyses for Mean Dilation – SNR as Grouping Variable

SNR (dB)	Comparison (light level)		Mean difference (mm)	95% CI – low	95% CI – high	df	*Pairwise t statistic*	*p*	*Sig.*	*Sig. in 11.4.1.6 (peak)*
-6	L1	L2	0.08	0.06	0.10	7398.12	7.08	<.001	*	*
-6	L1	L3	0.12	0.09	0.14	7162.11	10.40	<.001	*	*
-6	L1	L4	0.13	0.11	0.15	7630.92	11.40	<.001	*	*
-6	L2	L3	0.04	0.02	0.06	7425.09	3.46	<.001	*	*
-6	L2	L4	0.05	0.03	0.07	6953.13	4.48	<.001	*	*
-6	L3	L4	0.01	-0.01	0.03	7760.63	1.07	.28		
-3	L1	L2	0.06	0.03	0.08	7454.97	5.04	<.001	*	*
-3	L1	L3	0.10	0.08	0.12	6838.67	9.006	<.001	*	*
-3	L1	L4	0.13	0.10	0.15	7417.87	11.46	<.001	*	*
-3	L2	L3	0.04	0.02	0.07	7221.60	4.008	<.001	*	*
-3	L2	L4	0.07	0.05	0.09	6933.51	6.39	<.001	*	*
-3	L3	L4	0.03	0.005	0.05	7520.84	2.39	.02		*
0	L1	L2	0.03	0.007	0.05	7771.49	2.57	.01		*
0	L1	L3	0.07	0.05	0.09	8183.51	6.57	<.001	*	*

0	L1	L4	0.07	0.05	0.09	7888.70	6.04	<.001	*	*
0	L2	L3	0.04	0.02	0.07	7938.42	4.04	<.001	*	*
0	L2	L4	0.04	0.02	0.06	8368.61	3.56	<.001	*	*
0	L3	L4	-0.005	-0.03	0.02	8299.50	-0.47	.64		
3	L1	L2	0.02	0.00	0.05	7979.46	2.17	.03		*
3	L1	L3	0.04	0.01	0.06	8166.93	3.26	.001	*	*
3	L1	L4	0.03	0.01	0.06	7923.90	3.11	.002		*
3	L2	L3	0.01	-0.01	0.03	7903.84	1.11	.27		
3	L2	L4	0.01	-0.01	0.03	8352.39	0.95	.34		*
3	L3	L4	-0.002	-0.02	0.02	8251.16	-0.18	.86		

Note. *p* is not Bonferroni corrected. * in *Sig.* column indicates significant differences after correcting for multiple comparisons (e.g., *p* is significant at 0.001). df = degrees of freedom (method: Kenward-roger). df are fractional because whole-plot and subplot variations are combined when standard errors are estimated.

11.4.3 Bayesian Regression Modelling

The results of the Bayesian regression modelling are presented in Table 32. These results align with the results of the trial-level peak dilation MEM in Section 11.4.1.4. This confirmed that the deletion of the model terms that resulted in a singular fit (i.e., slopes) did not lead to distorted beta estimates.

Table 32. Final Trial-Level Peak Dilation Bayesian Regression Model Output Table for Study 1C

Population-level Effects							
	Estimate	SE	95% BCI - low	95% BCI - High	Rhat	Bulk_ESS	Tail_ESS
Intercept	0.56	0.03	0.5	0.61	1.01	447	668
SNR -6 dB	-0.02	0.01	-0.05	0.01	1	2290	3131
SNR 0 dB	-0.07	0.01	-0.1	-0.04	1	1903	3144
SNR 3 dB	-0.15	0.02	-0.18	-0.12	1	1543	2255
L2	-0.12	0.02	-0.16	-0.08	1	1579	2704
L3	-0.21	0.02	-0.25	-0.17	1	1119	2126
L4	-0.26	0.02	-0.3	-0.22	1	874	1860
SNR-6:L2	-0.02	0.02	-0.06	0.01	1	2558	3064
SNR 0 dB:L2	0.03	0.02	-0.01	0.06	1	2359	3132
SNR 3 dB:L2	0.04	0.02	0	0.08	1	2591	3394
SNR -6 dB:L3	0	0.02	-0.04	0.03	1	2540	3264

SNR 0 dB:L3	0.02	0.02	-0.01	0.06	1	2526	3419
SNR 3 dB:L3	0.07	0.02	0.04	0.11	1	2321	2845
SNR -6 dB:L4	0.01	0.02	-0.03	0.05	1	2639	3270
SNR 0 dB:L4	0.07	0.02	0.04	0.11	1	2691	2920
SNR 3 dB:L4	0.11	0.02	0.07	0.14	1	2425	3020

Random Effects		
	SD	95% BCI
light level * SNR | Participant (Intercept and slope)	0.16	0.13 - 0.20
Sentence (Intercept)	0.03	0.02 - 0.03
light level * SNR | Trial Number (Intercept and slope)	0.03	0.01 - 0.04
Residual	0.2	

Notes: Family: gaussian
Links: mu = identity; sigma = identity
Formula: peakdilation ~ 1 + snr * lightlevel + (1 + snr * lightlevel | participant) + (1 | sentence) + (1 + snr * lightlevel | trial_exp)
Data: pup_data (Number of observations: 8749)
Samples: 4 chains, each with iter = 2000; warmup = 1000; thin = 1; total post-warmup samples = 400
BCI = Bayesian Credible Interval
Rhat = R-hat convergence diagnostic, values of 1 or smaller indicate that the MCMC chains mixed well (i.e., the between- and within-chain estimates agree) (Vehtari et al., 2019)
bulk_ESS = Bulk Effective Sample Size and tail_ESS = Tail Effective Sample Size – these values should be at least 100 to indicate that the estimates of the posterior are reliable (Vehtari et al., 2019)

11.5 Discussion

11.5.1 Summary of Results

The aim of Study 1C was to examine the effects of light level and SNR on trial-level peak and mean dilation using mixed-effects modelling. Fundamentally, the use of trial-level pupil data in MEM (where participant, sentence and trial number can be accounted for) produced estimates for the effects of light level and SNR on peak dilation that were consistent with those reported for signal-averaged data in Chapters 9 and 10.

The results were also consistent between trial-level MEMs using peak and mean dilation. Because the trial-level peak dilation MEM was able to account for a larger proportion of total variance, only the results of the trial-level peak dilation MEM are discussed further. The results of the trial-level peak dilation MEM were further verified by the Bayesian regression analysis, computed using the true maximal random effects structure justified by the design (this analysis did not remove effects that resulted in singularity) (Barr et al., 2013; Gelman & Hill, 2007; McElreath, 2020).

The conditional R^2 value in Table 26 showed that the inclusion of light level and SNR as fixed effects (with an interaction term) and participant, trial number and sentence as random effects explained 39% of the total variance in peak dilation in the trial-level peak dilation MEM. Thus, there was still 61% of total variance left unexplained. This is likely due to additional variance that comes with using the trial-level data. However, the additional variance that has been introduced by using trial-level data has not buried the effects of light level, SNR or the interaction term.

Overall, the results are consistent with the earlier findings that effect of SNR on peak dilation is larger in dimmer light levels even when trial-level peak dilation is used. A greater number of significant differences in peak dilation between SNR conditions in the dimmer light levels (1 and 2), and fewer significant differences in peak dilation between SNR conditions in the brighter light levels (3 and 4) were found in the trial-level peak dilation MEM compared to in the signal-averaged peak dilation MEM.

Furthermore, more adverse SNR conditions (SNR -6 dB and SNR -3 dB) led to more significant differences in peak dilation between light levels in the trial-level peak dilation MEM.

These findings provided evidence that signal-averaging in previous chapters did not result in any major distortions in effect estimates. Furthermore, these findings suggest that the task-relevant signal is measurable in trial-level peak and mean dilation data when using appropriate statistical techniques like mixed-effects modelling. Therefore, these techniques may be preferred over signal averaging in applications where trial-level variations are of primary interest.

As in Chapter 10, the trial-level peak dilation MEM reported here, and the signal-averaged peak dilation MEM reported in Section 10.4 were not nested models, therefore, the results cannot be compared statistically. However, they can be compared descriptively and theoretically by examining the patterns of significant results. The next section compares the significant effects computed in the trial-level peak dilation MEM reported here, the signal-averaged peak dilation MEM reported in Section 10.4, and the RANOVA reported in Section 9.4.4.

11.5.2 The Effects of SNR and Light Level on Peak Dilation Using Trial-Level Pupil Data

In general, the results of the trial-level peak dilation MEM were comparable to the signal-averaged peak dilation MEM (and the peak dilation RANOVA). Though there were some differences between the number of significant differences in the trial-level peak dilation MEM. Table 33 shows the number of significant differences between the post hoc comparison for all peak dilation analyses reported in Study 1.

Table 33. Comparison of The Number of Significant Differences in Post Hoc Analyses Between RANOVA (Chapter 9), Signal-Averaged Peak Dilation MEM (Chapter 10), and Trial-Level Peak Dilation MEM.

	Number of significant differences		
SNR as grouping variable – Differences between light levels			
	Peak dilation RANOVA	Signal-averaged peak dilation MEM	Trial-level peak dilation MEM
SNR -6 dB	5	4	5
SNR -3 dB	5	4	6
SNR 0 dB	4	4	5
SNR 3 dB	1	2	4
Total	15	14	20
Light level as grouping variable – Differences between SNR conditions			
	Peak dilation RANOVA	Signal-averaged peak dilation MEM	Trial-level peak dilation MEM
L1	3	4	5
L2	2	3	5
L3	2	1	3
L4	1	0	2
Total	8	8	15

Table 33 highlights that there were a greater number of significant differences in post hoc comparisons when trial-level peak dilation data was analysed using MEM. It is possible that more subtle differences in peak dilation were found to be significant in the trial-level peak dilation MEM. This indicates that the power to detect significant conditions effects may have been greater for the MEM conducted using trial-level peak dilation data.

The trial-level MEM showed that dimmer light levels led to greater sensitivity of peak dilation to distinguish between SNR conditions. All comparisons of peak dilation between SNR levels at L1 were significantly different (except the comparison between SNR -6 dB and SNR -3 dB). Similarly, all comparisons of peak dilation between SNR levels at L2 were significantly different (except the comparison between SNR -3 dB and SNR 0 dB). L3 and L4 showed more significant differences in peak dilation between SNR levels when compared to the signal-averaged peak dilation MEM. However, there were still fewer significant differences between SNR levels in these dimmer light levels (L3: three differences, L4: two differences) compared to L1 and L2 (Table 33).

The result indicating that more difficult SNR conditions resulted in a greater number of differences between light levels was measurable in trial-level data. However, the benefit of using trial-level peak dilation data was less clear when compared to the signal-averaged peak dilation data (see Table 15 and Table 28). In the trial-level peak dilation MEM, at SNR -3 dB, all comparisons between light levels were significant. This condition showed the largest peak dilation across light levels. Therefore, individuals may have expended the most effort in SNR -3 dB, even though it was not the most adverse condition (see Section 9.5 for details on this finding). In the trial-level peak dilation MEM, at SNR 3 dB (the easiest listening condition, with the smallest peak dilation), there were four significant differences between light levels. Therefore, it was still the case that the easiest listening condition resulted in the fewest significant differences between light levels, but it was less clear that peak dilation measured in easier conditions was less affected by light level.

The results of the trial-level peak dilation MEM compared with the signal-averaged peak dilation MEM demonstrated that it was possible to exploit the flexibility of MEMs and construct a model that can analyse trial-level pupil data, successfully. However, in this dataset, the overall conclusions that can be drawn from the trial-level peak dilation MEM do not differ from the RANOVA and the signal-averaged peak dilation MEM.

11.5.3 Accounting for Additional Random Effects

Important information about the variance that exists in large, unaveraged, pupil data may be overlooked during the signal-averaging process (Volpert-Esmond et al., 2018). Specifically, in the current dataset, important (measurable) information about how TEPRs may change over time, and how the unique sentences may have affected TEPRs was neglected when signal-averaging was employed. These random effects could not be included in a MEM when using signal-averaged data.

The flexibility of mixed-effects modelling allowed for the simultaneous inclusion of participant, sentence, and trial number as random effects in the final model (11.4.1.4). Each of these effects significantly improved the model fit. Therefore, they were important effects to include in the trial-level analysis.

Overall, 39% of the total variance in peak dilation in the trial-level peak dilation MEM was accounted for. Of that, 12% of the total variance was accounted for by the fixed effects (light level, SNR, and the interaction). When included as random effects, sentence accounted for 2.6% of the total variance in model and trial number accounted for 3.3% of the total variance in the final model reported in Section 11.4.1.4. Therefore, there was only a small amount of variance being partitioned from the error term due to sentence and trial number.

Nevertheless, it was demonstrated that variance associated with sentence and trial number was quantifiable at the trial-level. Furthermore, the inclusion of these random effects increased the power of the analysis to detect the fixed effects in an analysis using trial-level pupil data. Chapter 14 builds on this information and examines the effects of trial number as a fixed effect for the purposes of making inferences about time-on-task effects at the trial level.

The analyses reported here highlight an opportunity to explore the use of trial-level pupil data using mixed-effects modelling in future research. This may be preferred over signal-averaging in applications where trial-level variations are of primary interest.

11.5.4 The Use of Trial-level Pupil Data

Despite the common use of TEPRs to measure listening effort, the use of trial-level pupil data in inferential analyses has received little attention. The current exploration has demonstrated that trial-level pupil data collected in the current experimental paradigm contains measurable and task relevant information.

Trial-level measurements of peak and mean dilation may have been the result of random noise, or another factor which was not measured. However, if this was the case, and if the task-relevant signal was not measurable in the trial-level pupil data, the results of the trial-level peak dilation MEM would not have been comparable to the more standard methods that were employed in Chapters 9 and 10. By using the trial-level pupil data, the power of the analysis to detect the effects of light level, SNR, and their interaction on trial-level peak dilation values may have increased.

Due to the presence of a measurable, task-relevant signal in the trial-level pupil data, when these data were used in conjunction with the structure of the MEM (and the additional random effects that could be accounted for), the precision in the estimates may have been increased and smaller differences were found to be significant (Detry & Ma, 2016).

The use of trial-level pupil data does come at the cost of increasing the variance in the dataset. The total variance in the dataset increased from 0.15 for the signal-averaged pupil data to 0.27 when trial-level pupil data was used. However, this is not necessarily a weakness. By using the trial-level data, all the variance that exists within the data was acknowledged, rather than averaging across participants/trials and losing this information. From this analysis, it can be concluded that sentence and trial number contributed variance to trial-level peak and mean dilation data and that variance could be partitioned from the error term by using mixed-effects modelling. This enhances understanding of factors that contribute to pupil dynamics when individuals are engaged in an effortful listening task.

As mentioned, there are many factors which can contribute to pupil dynamics, and which are unrelated to the experimental task. Perhaps, if some additional random effects and/or fixed effects (e.g., age, working memory capacity) were measured, the trial-level peak dilation MEM would have explained a larger proportion of the total variance in the trial-level peak dilation, while still benefitting from the additional statistical power that may be achieved through the model structure (Detry & Ma, 2016; Quené & van den Bergh, 2004).

In summary, these results support previous research in which unaveraged, trial-level pupil data was used (Clewett et al., 2020; Clewett et al., 2018; Cohen Hoffing et al., 2020; Leuchs et al., 2017; Wetzel et al., 2020). These findings extend previous findings by demonstrating that trial-level pupil data measured during a speech-in-noise task with varying light levels can be used in analyses and produce results that are comparable to signal-averaged data but may be better able to detect differences in peak dilation between conditions.

11.5.5 Caveats and Future Research

The benefits of analysing trial-level pupil data will depend on the aim of the research. For some studies, information regarding the variance in the trial-level data are inconsequential compared to the improvement in SNR (described in Section 11.1.1) that can be achieved through signal-averaging.

The quantification of the effect size and direction of effects for each participant relative to the mean (using random effects error terms) and the examination of individual differences could not be achieved using the frequentist MEM reported here (Mirman & Klein, 2014). This is because all random slopes resulted in singularity and were removed from the model. Therefore, if the aim is to assess these effects, researchers may choose to use Bayesian regression modelling instead of frequentist MEMs, as Bayesian analyses can provide valid estimates for the maximal random effects structure, regardless of singular fits that may be calculated in lmer models (see section 11.3.5.4).

Trial-level peak dilation latency was not examined in this chapter, but this would be a beneficial exploration in future research due to the impact that variable response latency may have on signal averaging. Examination of trial-level peak dilation latency using similar methods to those used in the current chapter may enhance understanding of the time course of trial-level pupil responses.

This study represents an initial exploration of the use of trial-level peak and mean dilation data in listening effort paradigms. Even though comparable results were achieved between the different analyses, this may not be possible in other experimental paradigms. It is possible that the task-relevant signal was more prominent within the individual trials in the present study because light level and SNR had a relatively large effect on the pupil. The combination of light level and listening effort affecting peak dilation, may have led to more prominent responses in the trial-level pupil data that would not be present in a task that was less challenging and where light level was not varied.

Based on the assumptions of signal-averaging reported in Section 11.1.1, further examination of the presence of systematic variation in pupil responses across trials is warranted. The current analyses have shown that trial number (time-on-task) and sentence accounted for some of the variance in the trial-level peak and mean dilation data, but this did not significantly distort model estimates for signal-averaged data. However, it is possible that more influential systematic variation may be present in other tasks and experimental paradigms and may lead to distorted signal-averaged pupil responses.

While the current analyses yielded interesting results and has provided some of the groundwork necessary for the use of trial-level pupil data in listening effort research, further research should be carried out to examine if noisy pupil data can be analysed with mixed-effects modelling in other auditory paradigms that may not have as robust responses.

11.5.6 Conclusion

In summary, Study 1C demonstrated that despite the greater noise inherent in trial-level pupil data, appropriate statistical analyses (mixed-effects modelling) can produce clear and expected patterns of results, consistent with those from analysis of signal-averaged data. Using trial-level pupil data in MEMs did not substantially alter the inferences that could be made with respect to the impact of light level, SNR and their interaction on peak dilation or mean dilation when compared to results obtained from a RANOVA or MEM using signal-averaged data. However, the mixed-effects modelling performed on trial-level data were able to account for some of the variance in the trial-level pupil data associated with sentence and trial number and were able to detect additional differences between the conditions tested.

Collectively, Study 1 has shown that light level affects TEPRs during a speech-in-noise task and that light level and SNR interact in their effects on peak and mean dilation and these effects are measurable in trial-level peak and mean dilation data. However, there is still uncertainty regarding the mechanisms that led to the relationships between light level and TEPRs. This is addressed in Study 2, Chapter 12.

12 STUDY 2 – THE MEDIATION EFFECT OF BASELINE DIAMETER IN THE RELATIONSHIP BETWEEN LIGHT LEVEL AND PEAK AND MEAN DILATION

12.1 Background and Aims

The analyses reported in Study 1 (A, B, and C) revealed that dimmer light levels led to larger TEPRs (peak and mean dilation) during a speech-in-noise task. This chapter aims to unpack the findings from Study 1. It is possible that baseline diameter may have acted as a mediator in the relationship between light level and TEPRs.

Resting pupil diameter is largely determined by environmental light conditions. Pupil diameter can also be influenced by tonic levels of arousal or an individual’s control state (Gilzenrat et al., 2010) (see Section 5.1.2). Additionally, Tsukahara et al. (2016) and Tsukahara and Engle (2021) reported that baseline diameter was related to cognitive ability (e.g., fluid intelligence and working memory capacity).

In both Tsukahara et al. (2016) and Tsukahara and Engle (2021), fluid intelligence was measured using the Raven's Advanced Progressive Matrices, letter sets, and number series and working memory capacity was measured using the advanced versions of the operation span, symmetry span, and rotation span tasks (see Tsukahara & Engle, 2021 for details). Individuals with higher fluid intelligence (and to a lesser extent, working memory capacity) typically showed significantly larger baseline diameter than individuals with lower fluid intelligence. This was the case even when controlling for age, ethnicity, and drug substances.

Tsukahara and Engle (2021) proposed that the relationship between greater cognitive ability and larger baseline diameter may be reflective of stronger functional connectivity through optimal levels of norepinephrine and tonic LC activation. Individuals with greater cognitive ability may have been in a more “task-ready” state during passive baseline recording and may have greater control of LC functioning modes and therefore, greater control over the use of explorative and exploitative control states. Tsukahara and Engle (2021) did not measure TEPRs but it is possible that cognitive ability may be reflected in TEPRs and this may be related to resting baseline diameter (e.g., Hayes & Petrov, 2016).

Gilzenrat et al. (2010) reported that there was a relationship between baseline diameter and peak dilation during an auditory odd-ball task. Measured in the same light level, they found that participants showing smaller baseline diameters typically also showed better performance and larger peak dilation, whereas participants with large baseline diameters showed poorer performance and smaller peak dilation (Experiment 1A). They concluded that this reflected task-relevant tonic and phasic activity in the LC. The LC plays a central role in the cognitive control of behaviour and projects to the prefrontal cortex – a brain area that is crucial for higher order cognitive abilities (Jung & Haier, 2007; Kane & Engle, 2002). Therefore, Gilzenrat et al. (2010) suggested that the inverse relationship between baseline diameter and peak dilation may reflect LC mediated regulation of internal control states related to task-engagement, such as exploration or exploitation of task (Aston-Jones & Cohen, 2005).

To confirm that their findings were not the result of the LIV (introduced in Section 6.1.2), Gilzenrat et al. (2010) performed an additional study (experiment 1B) in which they manipulated light level to cause either larger or smaller baseline diameters. They did not find a difference in peak dilation between light levels even though baseline diameter differed. Consequently, they concluded that the relationship reported in Experiment 1A was likely due to internal cognitive control states mediated by the LC, rather than the more mechanical LIV interpretation. This finding also suggests that peak dilation may be affected by changes in baseline diameter due to control state but may not be affected by changes in baseline diameter due to light level.

As described in Section 6.4.2, Reilly et al. (2019) manipulated light level to examine whether TEPRs were dependent on baseline diameter in a target detection task and a visual word monitoring task. They found similar peak dilation across light levels (i.e., both baseline diameter conditions) and concluded that peak dilation was unaffected by light-induced changes to baseline diameter in those tasks.

Therefore, the findings by Reilly et al. (2019) may support the findings by Gilzenrat et al. (2010) in which peak dilation was relatively unaffected by light-induced changes in baseline diameter. However, the findings of Study 1 reported here showed that light level affected TEPRs in a speech-in-noise task.

It is not known whether the results reported in Study 1 were due to a direct effect of light level on TEPRs, or if light level led to changes in TEPRs via a relationship between baseline diameter and TEPRs. Mediation analysis could clarify the issue. Specifically, the aim of Study 2 was to use mediation analysis to investigate whether baseline diameter mediated the relationship between light level and TEPRs.

12.1.1 Mediation Analysis

Mediation analysis can illuminate relationships between an initial “cause” variable, and the subsequent “effect” variable via the effect of an intermediary variable (Baron & Kenny, 1986; Vuorre & Bolger, 2018). Experimental questions and hypotheses involving mediation occur frequently across experimental studies. However, there are fewer applications of mediation analysis in cognitive psychology when compared to other branches of psychology (i.e., social psychology). Vuorre and Bolger (2018) postulated that this could be due to cognitive psychology’s common use of repeated-measures data.

Mediation analysis has typically relied on linear regression models (Judd & Kenny, 1981). This method is not suitable for repeated-measures data because it violates the independence assumption underlying linear models. As demonstrated in Chapters 10 and 11, MEMs are becoming increasingly common in the analysis of data with repeated-measures and these models have been successfully applied to assess mediation when data contain repeated-measures across individuals (within-subjects designs).

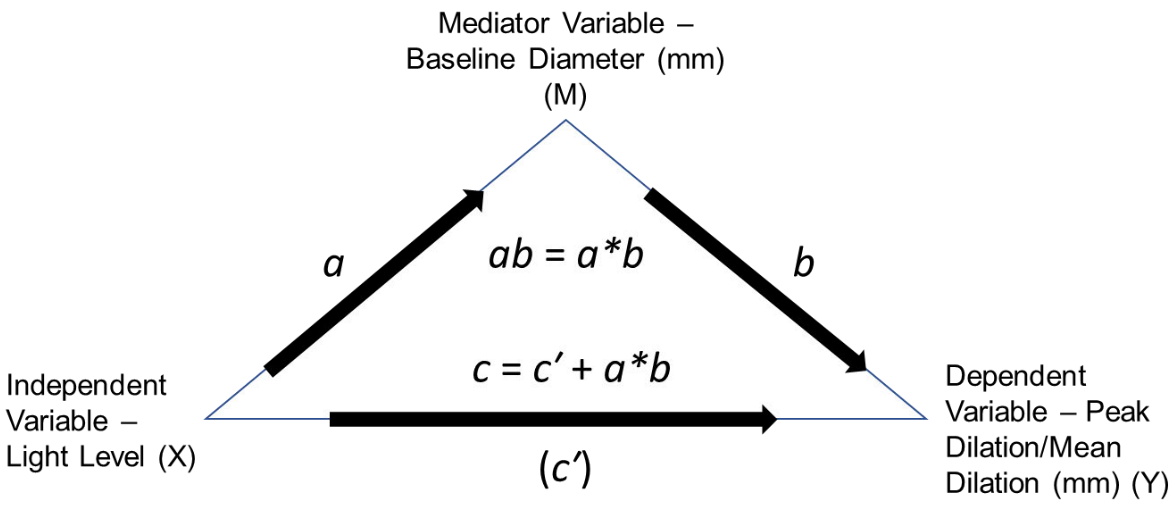
A mediation model is depicted in Figure 20. The initial variable is the independent variable (X) (here, X is light level), the subsequent effect variable is the dependent variables (Y) (here, Y is pupil dilation), and the intermediary variable is the mediation variable (M) (here, M is baseline diameter). Label *a* represents the causal pathway between X and M. Label *b* represents the causal pathway between M and Y. Label *c′* represents the causal pathway between X and Y, having accounted for the effect of M (i.e., the direct effect). These parameters are used to compute the indirect effect (*ab*) and the total effect of X on Y *(c*) (Vuorre & Bolger, 2018, p. 2126).

For a variable to fully mediate a relationship, it must satisfy the following criteria (Baron & Kenny, 1986, p. 1176):

1. Variation in M can be accounted for by different levels of X (path *a*).
2. Variation in Y can be accounted for by variation in M (path *b*).
3. The previously significant relationship between X and Y is no longer significant when paths *a* and *b* are controlled. Full statistical mediation occurs when pathway *c′* = 0.

Figure 20

Diagram of Traditional Mediation Model for the Current Paradigm



Note. This figure shows the paths and estimates computed by a traditional mediation model. Adapted from “Within-subject mediation analysis for experimental data in cognitive psychology and neuroscience”, M. Vuorre, N. Bolger, 2018, *Behavior Research Methods*, 50, p. 2126.

It is clear from the baseline diameter analysis in Section 9.4.3 that light level (X) had a significant effect on baseline diameter (M) (path a). It is possible that the variation in baseline diameter (M) rather than the change in light level (X), may have led to variation in pupil dilation (Y, path b) (Figure 20). If changes in baseline diameter completely explained the relationship between light level and pupil dilation, then the indirect effect (*ab*) would be close in size to *c* because they would be explaining the same effect, whereas the direct effect (*c′*) would be zero (or close to zero). If *c* was larger than *ab*, *c′* would also be larger than zero and would indicate that light level led to variance in peak dilation that was not due to baseline diameter (Vuorre & Bolger, 2018).

12.1.2 Bayesian Estimation

There are multiple ways to analyse mediation in within-subjects designs. However, MEMs are perhaps the most parsimonious and applicable method, as argued by Vuorre and Bolger (2018, p. 2129). In the current study, Bayesian estimation was used instead of the more traditional maximum likelihood estimation (such as those employed in Chapters 10 and 11) as it offers several benefits, particularly for multilevel mediation analyses as outlined by Vuorre and Bolger (2018). For example, maximum likelihood estimation methods rely on assumptions about the sampling distributions of the estimated parameters. In contrast, Bayesian analyses offer interpretable representations of uncertainty by providing full posterior probability distributions of plausible values for each parameter (Kruschke, 2014; Vuorre & Bolger, 2018). This is important when the analysis involves transformations of estimated parameters at multiple levels, like within-subject mediation analyses, mainly because the uncertainty that is associated with the estimated parameters is carried forward and applied to those transformations (Vuorre & Bolger, 2018).

A major difference between Bayesian and the more traditional, frequentist analyses is in the meaning of “probability”. As stated by Heino et al. (2018), “frequentists consider probability as long-run frequency from a very long (or infinite) sequence of repetitions. In Bayesian statistics, the probability is a measure of uncertainty associated with unknown quantities, such as the parameters in a model" (p. 52). Thus, parameter estimates are interpreted differently. There are three main components to a Bayesian analysis: prior distribution, likelihood function and posterior distribution. Information provided by observed data (the likelihood), is used to update prior beliefs (prior distribution) and forms the posterior beliefs about the parameters (posterior distribution) (Heino et al., 2018). Prior beliefs are updated by examining the likelihood of a dataset given the values of the parameters of interest (van Ravenzwaaij et al., 2018).

The prior information that can be incorporated into the models may come from personal expertise in the field of study or from knowledge about the constraints that exist within data (Heino et al., 2018). For example, when a researcher knows that parameter values are unlikely to exceed certain limits, that information can be incorporated into the prior information. This can result in greater sampling efficiency and less variance in the model estimates.

In Bayesian inference, examination of the likelihood of the data, given parameter values, is an important step in updating prior beliefs and making inferences. Often, this likelihood cannot be assessed for each combination of parameter values because posterior distributions are often too difficult to obtain via analytic calculations (Vuorre & Bolger, 2018). Bayesian inference makes use of computer-driven sampling methods such as Markov Chain Monte-Carlo (MCMC) sampling to overcome this issue (van Ravenzwaaij et al., 2018). MCMC sampling draws random samples from the prior distribution to develop a posterior distribution which is then used to estimate effects.

In summary, Bayesian estimation provides a fitting framework and is particularly suited to address within-subjects mediation effects.

12.2 Significance

The mediation effect of baseline diameter in the relationship between light level and pupil dilation has not been previously assessed. Therefore, this chapter provides a contribution to knowledge by using Bayesian mediation analysis techniques to investigate the underlying mechanisms that led to the effects of light level on pupil dilation reported in Study 1. Knowledge gains in this area may lead to enhanced understanding of the mechanisms and influences involved in pupil dilation during a speech-in-noise task.

12.3 Methods

12.3.1 Participants

Information related to the participants can be found in Chapter 8 - General Methods, Section 8.1.

12.3.2 Materials

Information related to the materials can be found in Chapter 8 - General Methods, Section 8.2.

12.3.3 Procedure

Information related to the procedure can be found in Chapter 8 - General Methods, Section 8.3.

12.3.4 Pupil Data Pre-Processing

The TEPRs analysed in this chapter are trial-level peak dilation and trial-level mean dilation data. Information related to the pupil data pre-processing can be found in Chapter 8 - General Methods, Section 8.4. See Chapter 10 for pupil data pre-processing details related to trial-level pupil data. Code specific to

In this study, baseline diameter varies at two levels: (a) both between conditions and trials, within-subjects, and (b) on average, between participants. In the current investigation, the within-subject process was of primary interest. Therefore, it was appropriate to within-subject mean centre the baseline diameter data to isolate the within-subject component and capture changes relative to each person's own average baseline diameter (Vuorre & Bolger, 2018). This was achieved using the *isolate* function in the *bmlm* R package (Vuorre & Bolger, 2018). This was the only difference in the data used in this chapter and Chapter 11.

12.3.5 Data Analysis

Data analyses were completed in R open source software (version 4.0.1) (R Core Team, 2020) and R Studio (version 1.3.1073) (RStudio Team, 2020). Analysis codes are available at https://osf.io/am6uv/.

The *brms* R package (version 2.7.0) (Bürkner, 2017, 2018) was used to conduct a series of Bayesian mediation analyses.

Samples were derived by Hamiltonian Monte Carlo sampling (Stan Development Team, 2021). Four chains with 5000 iterations (2500 warm-up samples for each chain) and 10000 post-warmup samples were used. Unstandardized effect estimates (posterior means, *b*) and 95% Bayesian Credible Intervals (BCI) are reported. Noninformative default priors were used to improve sampling efficiency.

12.3.5.1 Random Effects

As in Chapter 11, participant, sentence, and trial number were included as random effects in each mediation model.

12.3.5.2 Mediation Models

There are a series of steps involved in traditional mediation analyses (Baron & Kenny, 1986; James & Brett, 1984; Judd & Kenny, 1981) and these apply to Bayesian mediation analyses as well. The steps are as follows:

1. Demonstrate that light level is related to pupil dilation. For this step, peak and mean dilation are used as criterion variables in Bayesian regression equations and light level is used as the predictor. The result of these analyses represents path *c* (i.e., the total effect) and establishes that there is an effect that may be mediated. This effect was already established in Chapter 11.
2. Demonstrate that light level is related to baseline diameter (mediator). For this step, baseline diameter is used as the criterion variable in the Bayesian regression equation and light level as a predictor. The result of this is the path *a* estimate (Figure 20). This step essentially involves treating the mediator as if it were an outcome variable.
3. Demonstrate that the baseline diameter affects the pupil dilation. Peak and mean dilation are used as criterion variables in Bayesian regression equations and light level and baseline diameter are predictors. The result of this is the path *b* estimate (Figure 20). It is not sufficient just to correlate baseline diameter with the peak and mean dilation, as baseline diameter and peak and mean dilation may be correlated because they are both caused by the causal variable: light level. Therefore, the effect of light level on pupil dilation must be controlled in establishing the effect of the baseline diameter on pupil dilation.

To examine whether baseline diameter mediated the relationship between light level and peak and mean dilation, the above steps were followed in six separate mediation analyses for data at SNR -3 dB for peak and mean dilation. Examination of mediation at single SNR levels ensures that the interaction effect reported in Study 1 was not present in the analyses. SNR -3 dB was selected because this condition showed the largest dilation values as well as the largest effect of light level on pupil dilation in Study 1. Another six mediation analyses were computed for data collected in SNR 3 dB to investigate whether the results were consistent in a less adverse listening condition.

To evaluate the mediating role of baseline diameter on the relationships between light level and peak and mean dilation (Figure 20) the total (*c*), direct (*c'*), and indirect (*ab*) effects and the associated Bayesian credible intervals (BCIs) are reported. The results of path *a* and path *b* are also reported.

In the mediation models, the direct effect measures the extent to which peak and mean dilation change when the light level increases by one unit (e.g., increases from L1 to L2) and baseline diameter remains unchanged. In contrast, the indirect effect measures the extent to which peak and mean dilation change when the light level is held fixed and baseline diameter changes by the amount it would have changed had the light level increased by one unit. Mediation exists when the uncertainty interval (95% BCI) for the indirect effect is sufficiently small to rule out 0 as a likely population value (Vuorre & Bolger, 2018).'

Mediation models are outlined in the below sections. Mediation models with A in the name examine the mediating effect of baseline diameter in peak dilation and mediation models with B in the name examine the mediating effect of baseline diameter in mean dilation.

12.3.5.2.1 Mediation model 1A (MM1A) and 1B (MM1B):

MM1A examined the mediating effect of baseline diameter for the relationship between L1 versus L2 and peak dilation at SNR -3 dB. MM1B examined the mediating effect of baseline diameter for the relationship between L1 versus L2 and mean dilation at SNR -3 dB.

12.3.5.2.2 Mediation model 2A (MM2A) and 2B (MM2B)

MM2A examined the mediating effect of baseline diameter for the relationship between L2 versus L3 and peak dilation at SNR -3 dB. MM2B examined the mediating effect of baseline diameter for the relationship between L2 versus L3 and mean dilation at SNR -3 dB.

12.3.5.2.3 Mediation model 3A (MM3A) and 3B (MM3B)

MM3A examined the mediating effect of baseline diameter for the relationship between L3 versus L4 and peak dilation at SNR -3 dB. MM1B examined the mediating effect of baseline diameter for the relationship between L3 versus L4 and mean dilation at SNR -3 dB.

12.3.5.2.4 Mediation model 4A (MM4A) and 4B (MM4B):

MM4A examined the mediating effect of baseline diameter for the relationship between L1 versus L2 and peak dilation at SNR 3 dB. MM4B examined the mediating effect of baseline diameter for the relationship between L1 versus L2 and mean dilation at SNR 3 dB.

12.3.5.2.5 Mediation model 5A (MM5A) and 5B (MM5B)

MM5A examined the mediating effect of baseline diameter for the relationship between L2 versus L3 and peak dilation at SNR 3 dB. MM5B examined the mediating effect of baseline diameter for the relationship between L2 versus L3 and mean dilation at SNR 3 dB.

12.3.5.2.6 Mediation model 6A (MM6A) and 6B (MM6B)

MM6A examined the mediating effect of baseline diameter for the relationship between L3 versus L4 and peak dilation at SNR 3 dB. MM6B examined the mediating effect of baseline diameter for the relationship between L3 versus L4 and mean dilation at SNR 3 dB.

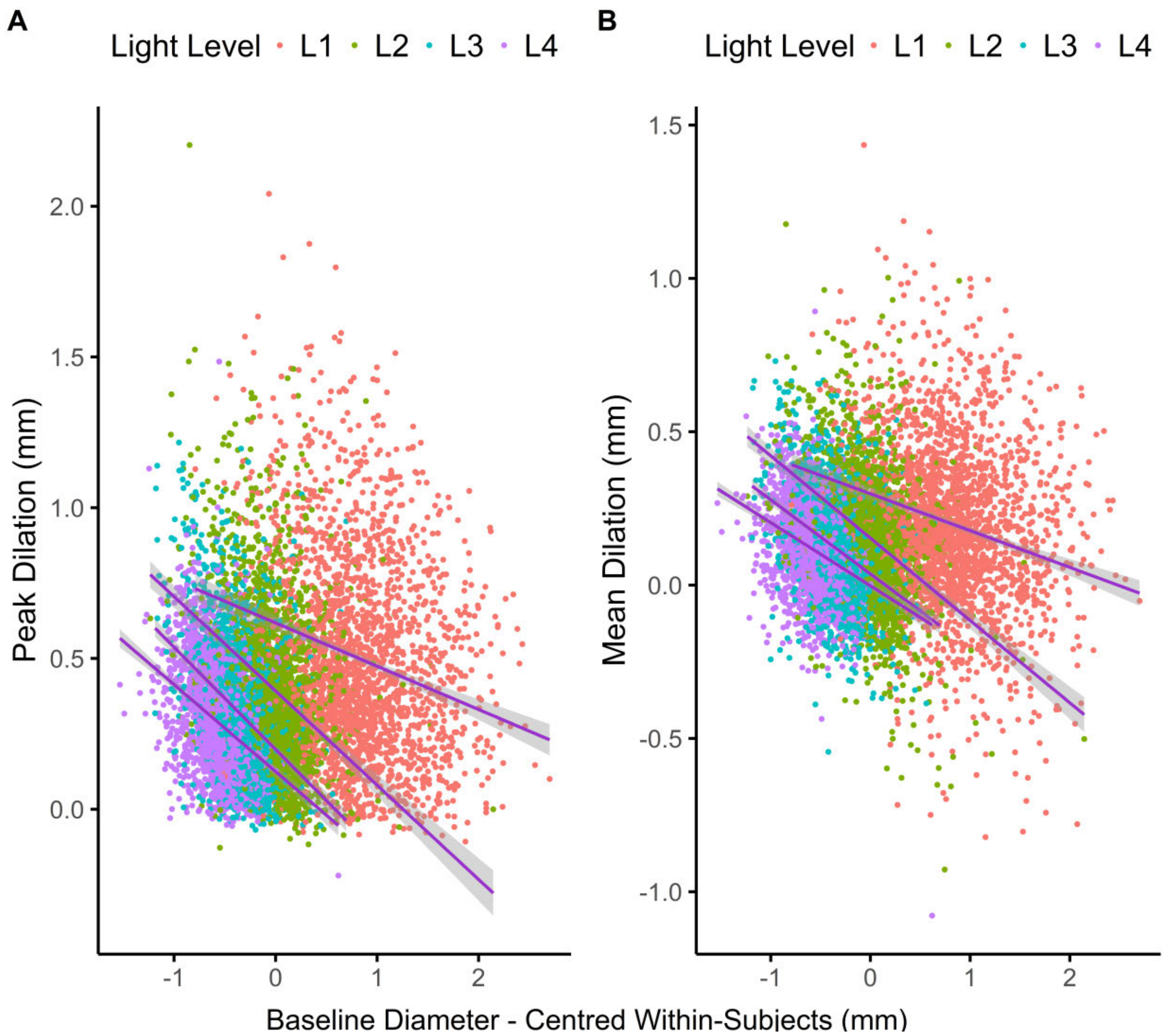
12.4 Results

Rhat convergence diagnostic values of 1 or smaller indicate that the MCMC chains mixed well (i.e., the between and within chain estimates agree) and that the number of iterations was sufficient to estimate the effects (Vehtari et al., 2019). Rhat values of 1 were achieved in all mediation models reported below.

Figure 21 shows the relationship between (A) baseline diameter (centred within-subjects) and peak dilation and (B) baseline diameter (centred within-subjects) and mean dilation. There was a consistent, negative relationship between baseline diameter and peak dilation, and baseline diameter and mean dilation at each light level.

Figure 21

Scatter Plots of the Relationship Between Baseline Diameter, and Peak and Mean dilation



Note. This figure displays scatter plots showing the negative relationships between (A) baseline diameter centred within-subjects (x axis) and peak dilation (y axis) and (B) baseline diameter centred within-subjects (x axis) and mean dilation (y axis) across light levels (legend).

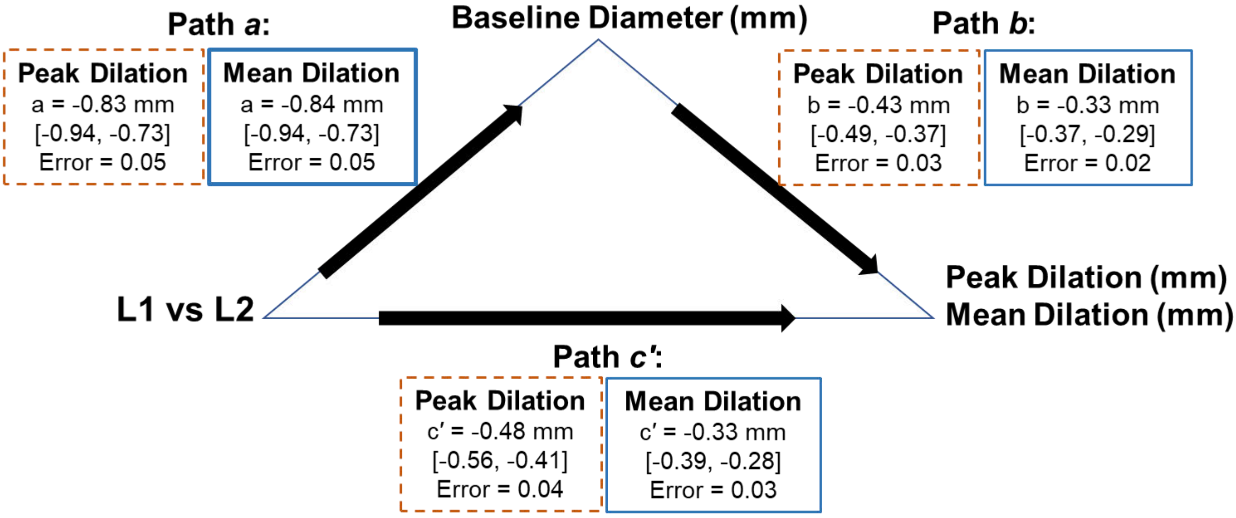
In Figure 22, Figure 23, and Figure 24, path *a* represents the causal pathway between light level and baseline diameter, path *b* represents the causal pathway between baseline diameter and peak dilation/mean dilation, and path *c′* represents the causal pathway between light level and peak dilation, having accounted for baseline diameter.

12.4.1 MM1A and MM1B

Path results for MM1A (peak dilation: orange, dashed line boxes) and MM1B (mean dilation: blue, solid line boxes) are presented in Figure 22.

Figure 22

Mediation Model Pathway Results for MM1A and MM1B



Note. This figure shows the mediation model pathway results for MM1A (peak dilation) in the orange, dashed line boxes and MM1B (mean dilation) in the blue, solid line boxes. The pattern of results obtained was similar for peak and mean dilation

For MM1A, the mean value of the posterior distribution for the total effect of light level on peak dilation (*c*) was -0.12 mm [BCI = -0.18 mm, -0.07 mm]. Therefore, MM1A estimated that peak dilation was, on average, 0.12 mm smaller in L2, compared to L1 (this was the same estimate computed in the trial-level peak dilation

MEM in Chapter 11). MM1B estimated that mean dilation was, on average, 0.06 mm [BCI = -0.1, -0.02] smaller in L2, compared to L1. The total effect parameters for peak and mean dilation were partitioned into two components: path *c′* (the direct effect) and *ab* (the indirect effect) (see Section 12.1.1 for details).

MM1A estimated that one unit change in light level (from dimmer L1 to brighter L2) made baseline diameter 0.83 mm smaller (path *a*), and one unit increase in baseline diameter made peak dilation 0.43 mm smaller (path *b*) when controlling for the effect of light level on peak dilation. The indirect effect (*ab*) was 0.36 mm [BCI = 0.3 mm and 0.43 mm]. Zero was unlikely to be a population value for the indirect effect. Therefore, it was estimated that peak dilation increased by 0.36 mm in L2, compared to L1, through the changes in baseline diameter. These effects are consistent with the effects estimated in MM1B for mean dilation (Figure 22).

The mean value of the posterior distribution for the direct effect (path *c′*) of light level on peak dilation was -0.48 mm [BCI = -0.56 mm and -0.41 mm]. A direct effect of light level on peak dilation remained after taking baseline diameter into account. Therefore, the relationship between light level and peak dilation was partially mediated by baseline diameter.

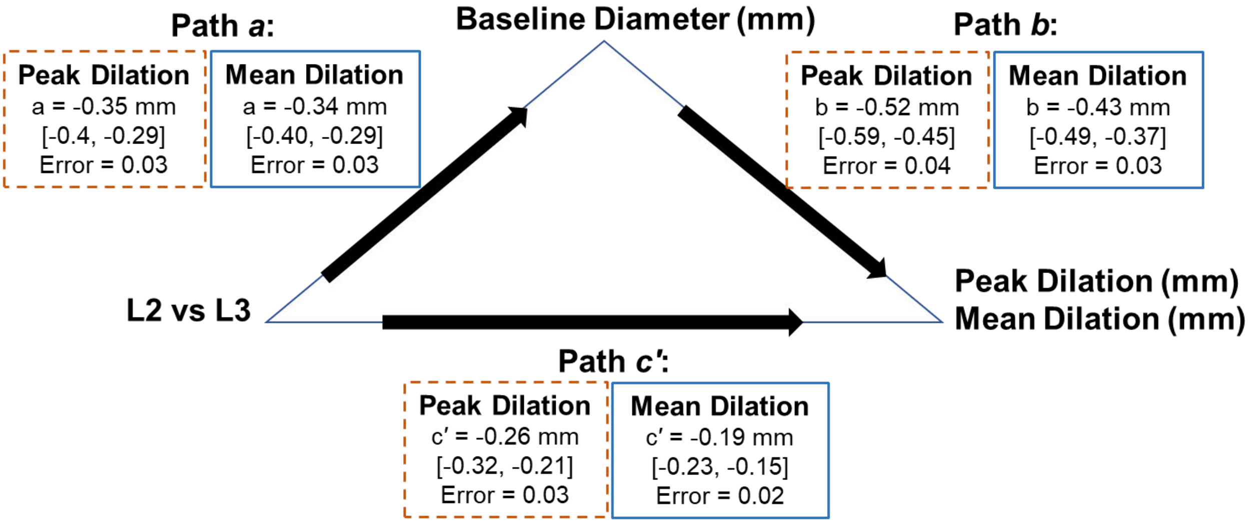
Because path *c′* (the direct effect) was larger than path *c* (the total effect), baseline diameter mediated the relationship between L1 and L2 and this led to a reduction in peak dilation. When the effect of baseline diameter was accounted for in path *c′*, the effect of light level on peak dilation was larger. These effects were consistent with the effects estimated by MM1B for mean dilation (Figure 22).

12.4.2 MM2A and MM2B

Pathway results for MM2A (peak dilation: orange, dashed line boxes) and MM2B (mean dilation: blue, solid line boxes) are presented in Figure 23.

Figure 23

Mediation Model Pathway Results for MM2A and MM2B



Note. This figure shows the mediation model results for MM2A (peak dilation) in the orange, dashed line boxes and MM2B (mean dilation) in the blue, solid line boxes. The pattern of results obtained was similar for peak and mean dilation.

A similar pattern of results was found for MM2A and MM2B when using L2 and L3 as the levels of the independent variable. For MM2A, the mean value of the posterior distribution for the total effect of light level on peak dilation (*c*) was -0.09 mm [BCI = -0.13 mm, -0.04 mm]. Therefore, it was estimated that peak dilation responses were, on average, 0.09 mm smaller in L3, compared to L2 (this was similar to the estimate computed in the trial-level peak dilation MEM in Chapter 11). MM2B estimated that mean dilation was, on average, 0.04 mm [BCI = -0.07 mm, -0.01 mm] smaller in L3, compared to L2. The total effect parameters for peak and mean dilation were partitioned into path *c'* and *ab* components.

MM2A estimated that one unit change in light level (from dimmer L2 to brighter L3) made baseline diameter 0.35 mm smaller (path *a*), and one unit increase in baseline diameter made peak dilation 0.52 mm smaller (path *b*) when controlling for the effect of light level on peak dilation. The indirect effect (*ab*) was 0.18 mm [BCI = 0.14 mm and 0.22 mm]. Zero was unlikely to be a population value for the indirect effect. Therefore, it was estimated that peak dilation increased by 0.18 mm in light level 3, compared to light level 2, through the changes in baseline diameter. These effects are consistent with the effects estimated in MM2B for mean dilation (Figure 23).

The mean value of the posterior distribution for the direct effect (path *c′*) of light level on peak dilation was -0.26 mm [BCI = -0.32 mm and -0.21 mm]. A direct effect of light level on peak dilation remained after taking baseline diameter into account. Therefore, the relationship between light level and peak dilation was partially mediated by baseline diameter.

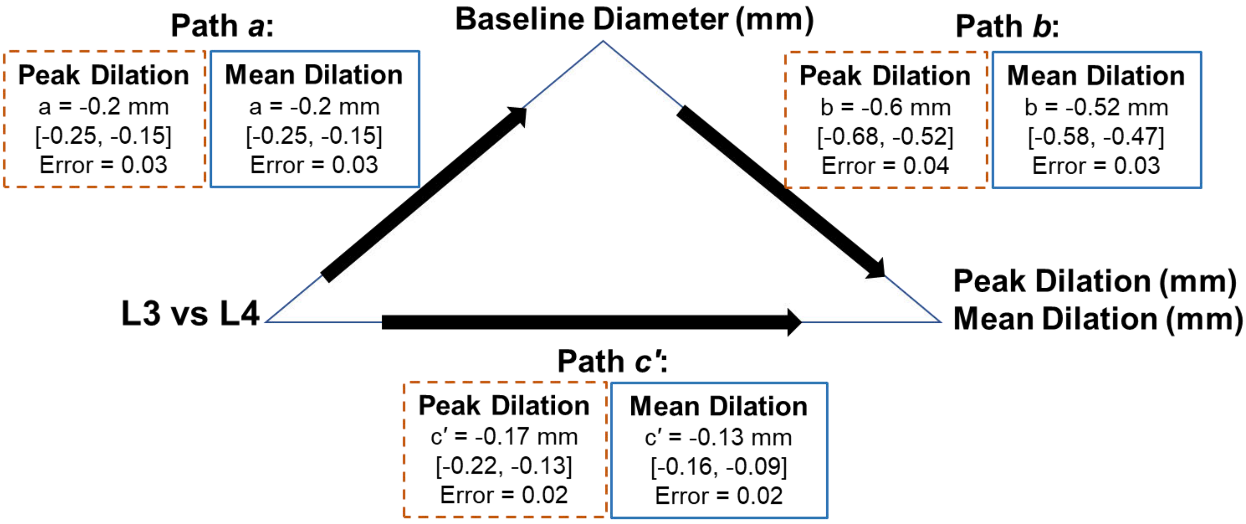
In line with MM1A and MM1B, path *c′* (the direct effect) was larger than *c* (the total effect) in MM2A and MM2B. Therefore, the mediation effect of baseline diameter led to a reduction in peak dilation. When baseline diameter was accounted for in path *c′*, the effect of light level on peak dilation was larger (Figure 23).

12.4.3 MM3A and MM3B

Pathway results for MM3A (peak dilation: orange, dashed line boxes) and MM2B (mean dilation: blue, solid line boxes) are presented in Figure 24.

Figure 24

Mediation Model Pathway Results for MM3A and MM3B



Note. This figure shows the mediation model results MM3A (peak dilation) in the orange, dashed line boxes and MM3B (mean dilation) in the blue, solid line boxes. The pattern of results obtained was similar for peak and mean dilation.

A similar pattern of results was found for MM3A and MM3B when using L3 and L4 as the levels of the independent variable. For MM3A, the mean value of the posterior distribution for the total effect of light level on peak dilation (*c*) -0.05 mm [BCI = -0.09 mm, -0.02 mm]. Therefore, it was estimated that peak dilation responses were, on average, 0.05 mm smaller in L3, compared to L4 (this was similar to the estimate computed in the trial-level peak dilation MEM in Chapter 11). MM3B estimated that mean dilation was, on average, 0.02 mm [BCI = -0.05 mm, 0.01 mm] smaller in L4 versus L3. However, because zero was likely to be a population value, MM3B estimated that the total effect of light level on mean dilation in MM3B was negligible. The total effect parameters for peak and mean dilation were partitioned into path *c′* and *ab* components.

MM3A estimated that one unit change in light level (from dimmer L3 to brighter L4) made baseline diameter 0.2 mm smaller (path *a*), and one unit increase in baseline diameter made peak dilation 0.6 mm smaller (path *b*) when controlling for the effect of light level on peak dilation. The indirect effect (*ab*) was 0.12 mm [BCI = 0.09 mm and 0.16 mm]. Zero was unlikely to be a population value for the indirect effect. Therefore, peak dilation increased by 0.12 mm in light level 4, compared to light level 3, through the changes in baseline diameter (Figure 24).

The mean value of the posterior distribution for the direct effect (path *c′*) of light level on peak dilation was -0.17 mm [BCI = -0.22 mm, -0.13 mm]. A direct effect of light level on peak dilation remained after taking baseline diameter into account. Therefore, the relationship between light level and peak dilation was partially mediated by baseline diameter.

Despite the lack of the total effect estimate for MM3B, a direct effect of light level on mean dilation was estimated after accounting for baseline diameter (-0.13 mm [BCI = -0.16 mm, -0.09 mm]).

In line with MM1A and MM1B and MM2A and MM2B, path *c′* (the direct effect) was larger than *c* (the total effect) in MM3A and MM3B. Therefore, the mediation effect of baseline diameter led to a reduction in peak dilation. When baseline diameter was accounted for in path *c′*, the effect of light level on peak dilation was larger. In MM3B, there was a direct effect of light level on mean dilation, only when baseline diameter was accounted for.

12.4.4 Mediation Models for SNR 3 dB

Similar patterns of partial mediation were found for MM4A and MM4B (Table 34), MM5A and MM5B (Table 35), and MM6A and 6B (Table 36). This confirmed that the mediating effect of baseline diameter on peak and mean dilation reported above was also present in an easier listening condition. The mediating effect was also present in the comparison between L3 and L4 for peak dilation in MM6A, where the total effect indicated that zero was a likely population value. This estimate is consistent with the nonsignificant difference between L3 and L4 in previous analyses (Chapters 9, 10, and 11). MM5B and MM6B also had total effect estimates of zero, where the mediation effect was present.

Table 34. Population-Level Effects for MM4A and 4B: The Role of Baseline Diameter as a Mediator in the Relationship Between L1 and 2 for Peak and Mean Dilation at SNR 3 dB

Population-Level Effects		Estimate (mm)	SE	Lower-95% BCI	Upper-95% BCI
MM4A (peak dilation)					
	Path *a*	-0.71	0.07	-0.84	-0.58
	Path *b*	-0.44	0.04	-0.51	-0.37
	Direct effect (*c'*)	-0.4	0.04	-0.48	-0.31
	Indirect effect (*ab*)	0.31	0.04	0.24	0.39
	Total effect (*c*)	-0.08	0.03	-0.14	-0.03
MM4B (mean dilation)					
	Path *a*	-0.72	0.07	-0.85	-0.58
	Path *b*	-0.39	0.03	-0.46	-0.33
	Direct effect (*c'*)	-0.3	0.04	-0.39	-0.23
	Indirect effect (*ab*)	0.28	0.03	0.22	0.35
	Total effect (*c*)	-0.02	0.03	-0.08	0.03

Table 35. Population-Level Effects for MM5A and 5B: The Role of Baseline Diameter as a Mediator in the Relationship Between L2 and 3 for Peak and Mean Dilation at SNR 3 dB

Population-Level Effects		Estimate (mm)	SE	Lower-95% BCI	Upper-95% BCI
MM5A (peak dilation)					
	Path *a*	-0.3	0.03	-0.36	-0.24
	Path *b*	-0.51	0.04	-0.59	-0.44
	Direct effect (*c'*)	-0.21	0.03	-0.26	-0.16
	Indirect effect (*ab*)	0.15	0.02	0.12	0.19
	Total effect (*c*)	-0.06	0.02	-0.1	-0.02
MM5B (mean dilation)					
	Path *a*	-0.3	0.03	-0.36	-0.24
	Path *b*	-0.46	0.03	-0.51	-0.41
	Direct effect (*c'*)	-0.16	0.02	-0.2	-0.12
	Indirect effect (*ab*)	0.14	0.02	0.11	0.17
	Total effect (*c*)	-0.02	0.02	-0.05	0.01

Table 36. Population-Level Effects for MM6A and 6B: The Role of Baseline Diameter as a Mediator in the Relationship Between L3 and 4 for Peak and Mean Dilation at SNR 3 dB

Population-Level Effects		Estimate (mm)	SE	Lower-95% BCI	Upper-95% BCI
MM6A (peak dilation)					
	Path *a*	-0.18	0.02	-0.22	-0.14
	Path *b*	-0.57	0.04	-0.65	-0.49
	Direct effect (*c'*)	-0.13	0.03	-0.18	-0.07
	Indirect effect (*ab*)	0.1	0.01	0.08	0.13
	Total effect (*c*)	-0.02	0.02	-0.06	0.02
MM6B (mean dilation)					
	Path *a*	-0.18	0.02	-0.22	-0.15
	Path *b*	-0.51	0.03	-0.57	-0.45
	Direct effect (*c'*)	-0.09	0.02	-0.13	-0.06
	Indirect effect (*ab*)	0.09	0.01	0.07	0.12
	Total effect (*c*)	0	0.01	-0.03	0.03

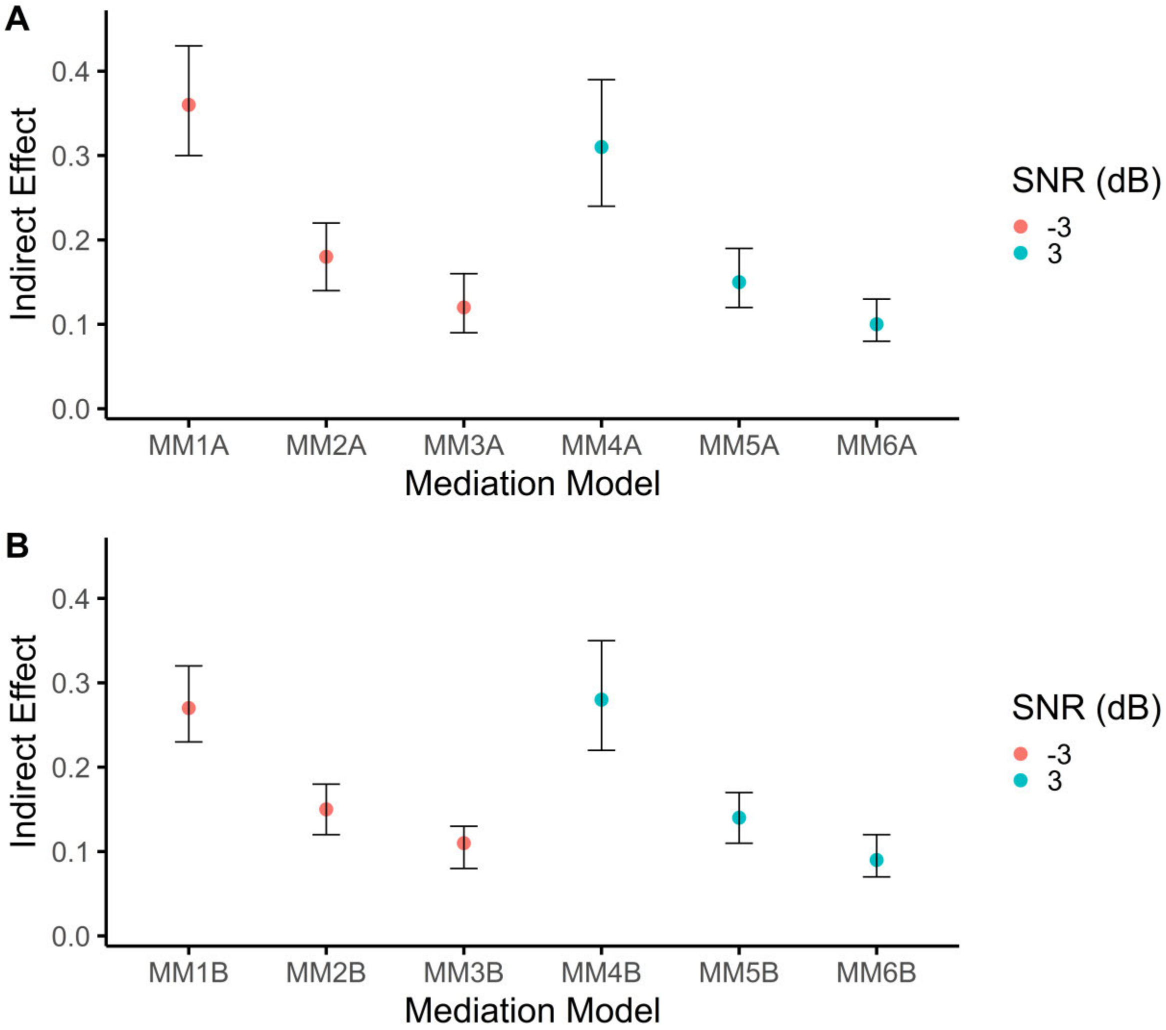
12.4.5 Mediation Results Summary

In all models, baseline diameter partially mediated the relationships between light level and peak and mean dilation, and the direct effects of light level on peak and mean dilation were larger after controlling for baseline diameter. In relationships where the estimated total effect was credibly zero, there were direct effects of light level on peak and mean dilation after accounting for the effects of baseline diameter. Overall, changes in light level and baseline diameter were associated with changes in peak and mean dilation, but these associations were in opposite directions. This was also the case in a less effortful listening condition (SNR 3 dB), where there was a smaller effect of light level on peak dilation (see Chapters 9, 10, and 11).

In models that tested the same changes in light levels (i.e., MM1A and MM1B vs. MM4A and MM4B, MM2A and MM2B vs. MM5A and MM5B, MM3A and MM3B vs. MM6A and MM6B, respectively), indirect effects were similar between SNR 3 dB and SNR -3 dB (Figure 25). Indirect effects were also consistent between peak dilation (Figure 25, A) and mean dilation (Figure 25, B). MM1A and MM1B and MM4A and MM4B showed the largest indirect effect of baseline diameter across SNRs for peak and mean dilation. This showed that there was a larger mediation effect of baseline diameter in models examining the two dimmest light level conditions (L1 and L2), regardless of SNR condition. The degree of mediation was affected by light level, but not by SNR in both peak and mean dilation.

Figure 25

Comparison of Indirect Effects Across Mediation Models



Note. This figure shows a comparison of indirect effects (i.e., mediated effects, *ab*) across models for SNR -3 dB (red) and SNR 3 dB (blue) for peak dilation (A) and mean dilation (B). In other words, the points show the estimated change in peak and mean dilation between light levels that was due to baseline diameter. Error bars represent the BCI for the parameter.

12.5 Discussion

12.5.1 Summary of Results

The analyses reported in Study 1 revealed that dimmer light levels led to larger TEPRs (peak and mean dilation). The aim of Study 2 was to use mediation analysis to ascertain whether baseline diameter mediated the effect of light level reported in Study 1 (Parts A, B and C).

Partial mediation was found for all mediation models using trial-level peak and mean dilation data. This was the case for both SNR 3 dB and SNR -3 dB. Light level and baseline diameter were both related to peak and mean dilation. However, the directions in which changes in light level and changes in baseline diameter were associated with peak and mean dilation were opposing.

These results are summarised schematically in the figures below.

Figure 26 depicts the well-established effect of light level on baseline diameter which was replicated in the current study. Baseline diameter was smaller in brighter light levels (A) and larger in dimmer light levels (B).

Figure 26

The Relationships Between Brighter (A) and Dimmer (B) Light Levels and Baseline Diameter

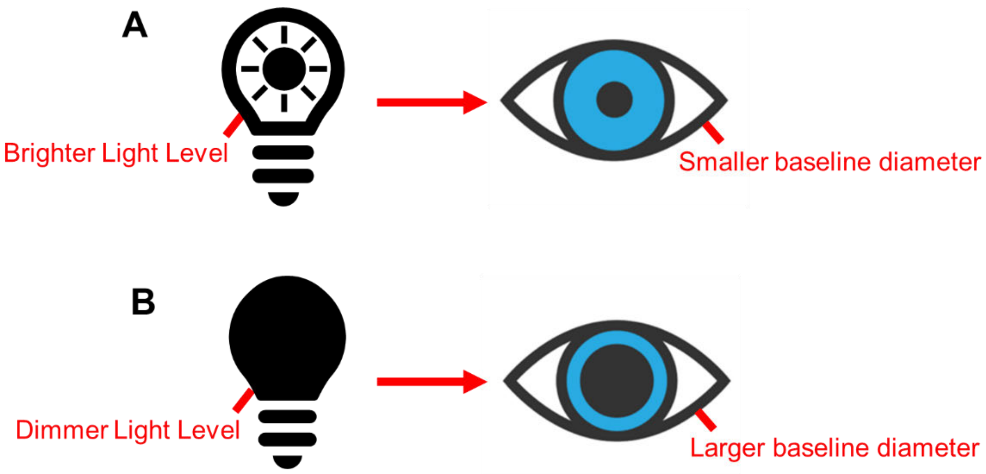


Figure 27 depicts the relationship between light level and peak and mean dilation (as found in Chapters 9, 10, and 11). Peak and mean dilation were smaller in brighter light levels (A) and larger in dimmer light levels (B).

Figure 27

The Relationships Between Brighter (A) and Dimmer (B) Light Levels and Peak and Mean Dilation

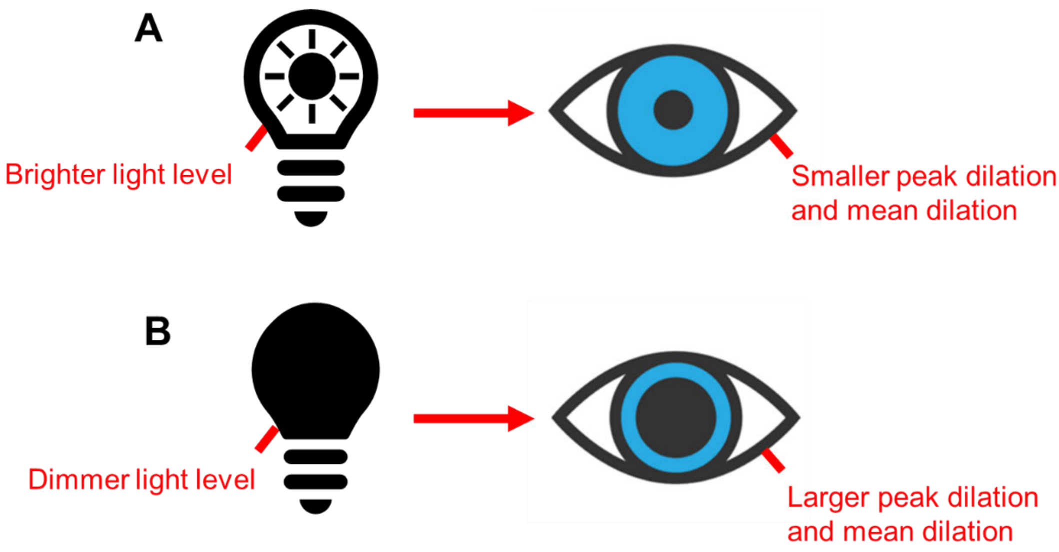
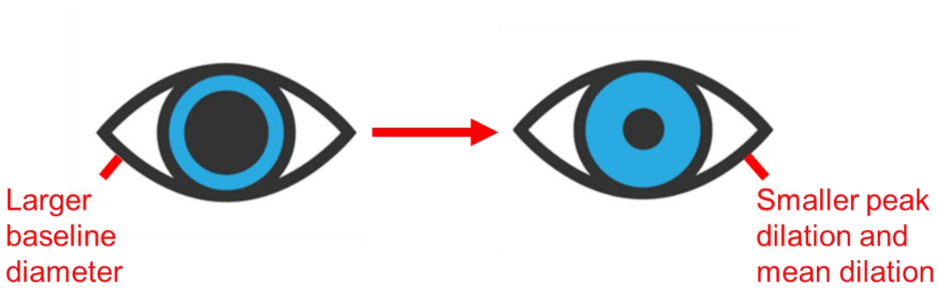


Figure 28 depicts the relationship between baseline diameter and peak and mean dilation, when the effects of light level on peak and mean dilation were accounted for (e.g., path b in the mediation model). A unit increase in baseline diameter led to a decrease in peak and mean dilation.

Figure 28

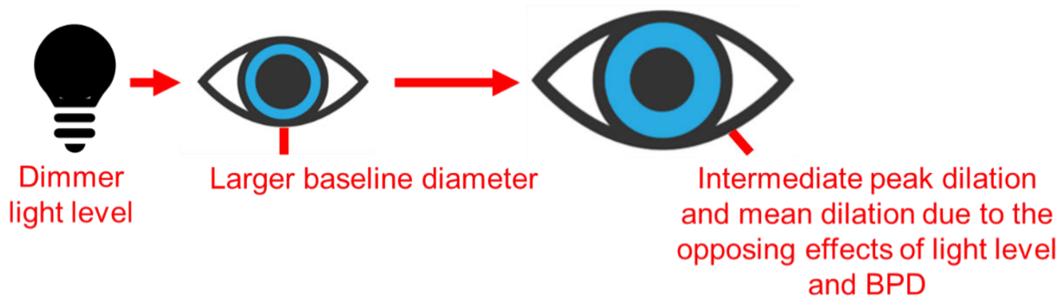
The Relationship Between Baseline Diameter and Peak and Mean Dilation



As illustrated in Figure 27 and Figure 28, the mediation analyses showed that light level and baseline diameter were oppositely related to peak and mean dilation (Figure 29). Specifically, the relationship between baseline diameter and peak and mean dilation suppressed some of the effect of light level. When the relationships between baseline diameter and peak and mean dilation were accounted for, the effect of light level on peak and mean dilation was larger.

Figure 29

The Simultaneous Effects of Light Level and Baseline Diameter on Peak and Mean Dilation



The results were consistent for peak and mean dilation and thus, will be discussed collectively and referred to as "pupil dilation" in this discussion. The novel findings will be discussed in four parts:

- First, the relationship between baseline diameter and peak dilation is discussed.
- Second, the mediatory effect of baseline diameter and potential explanations for the findings are discussed
- Third, the limitations of the current analyses are discussed.
- Fourth, the implications for research and clinical use are discussed.

12.5.2 The Relationship Between Baseline Diameter and Pupil Dilation

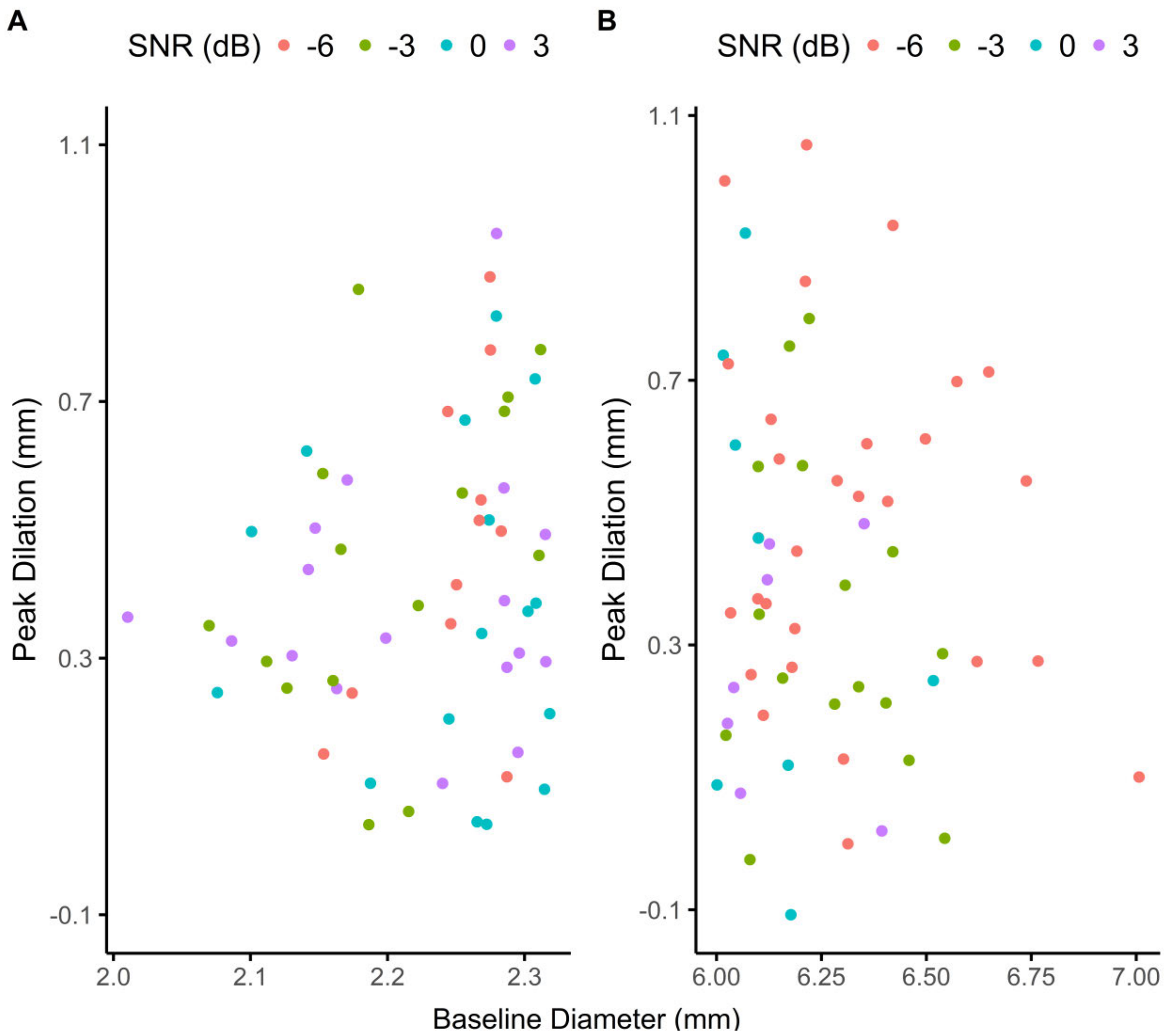
An increase in baseline diameter was associated with a decrease in pupil dilation when the effects of light level on pupil dilation were accounted for. This relationship was found in all mediation models tested (shown in path *b*, in Figure 22, Figure 23, and Figure 24). Two potential explanations for these findings are: (1) restriction of task-evoked dilation due to the dynamic range of the pupil, or (2) that the pupil size may be reflective of the regulation of control state (arousal modulation, as predicted by Adaptive Gain Theory, see Section 5.1.2) during listening.

The pattern of effects in which larger baseline diameter was associated with smaller pupil dilation is congruent with physiological reactivity predictions set out in the LIV (Lacey, 1956). The dynamic range of the pupil has been suggested as the primary factor in the LIV. The LIV posits that higher initial values will lead to smaller subsequent responses due to a ceiling effect (Wilder, 1957). Conversely, smaller initial values will lead to larger subsequent responses because there is more room for an increased response (Wilder, 1957). A more recent reconceptualization of the LIV suggests that the inverse relationship between baseline and response in physiological systems will likely *only* occur when the initial value is at the upper limits of the system's range (Jin, 1992).

Baseline diameter never reached the upper limit of the pupil's range in the current study. Standard pupil diameter range is typically 2 mm to 8 mm (Spector, 1990). The largest trial-level baseline diameter recorded was 7.01 mm (M = 3.55, SD = 0.74). Figure 30 shows a scatter plot of the 60 smallest and largest trial-level baseline diameter values and the associated trial-level peak dilation values. The range of values for peak dilation is comparable between the smallest and largest baseline diameters. Therefore, it is unlikely that the pupil's dynamic range was a limiting factor that led to the relationship reported here. If the dynamic range of pupil diameter were a factor, Figure 30 should have shown differences in the range of peak dilation values for the smallest and largest baseline diameters.

Figure 30

Smallest (A) and Largest (B) Baseline Diameter Values and Associated Peak Dilation Values



Note. This figure is a subset of the trial-level pupil data showing that the 60 smallest (A) and largest (B) recorded baseline diameter values had a comparable spread of peak dilation values. This provides evidence that LIV was not the likely reason for the negative relationship between baseline diameter and peak dilation. If the LIV was responsible for this relationship.

Gilzenrat et al. (2010) (experiment 1A) observed that larger baseline diameter values were associated with smaller peak dilation values and vice versa. They suggested that the observed pupil dynamics reflected changes in control state via LC activity. In

Adaptive Gain Theory, an individual's control state determines the extent to which task-engagement or disengagement behaviours are favoured (Gilzenrat et al., 2010). For example, in an exploitative control state, neural responsivity to task-relevant stimuli is increased which may lead to better task performance (Jepma & Nieuwenhuis, 2011). On the other hand, in an explorative control state, there is a less discriminative increase in neural responsivity which may lead to degraded performance, disengagement, and may allow for processing of non-task-relevant stimuli (Jepma & Nieuwenhuis, 2011). Therefore, in the current study, variation in LC activity and therefore, control state across samples in the current study may explain the relationship between baseline diameter and pupil dilation. This suggestion is speculative but may be testable and could be verified in further studies.

To verify that their findings were not due to the dynamic range of the pupil, Gilzenrat et al. (2010) also systematically adjusted light level to manipulate baseline diameter in a follow up experiment[19]. Light and dark conditions were adjusted for each participant, such that the difference in baseline diameter between the conditions was always 1.4 mm. Therefore, different light levels may have been used between participants. They found no evidence that the pupil’s dynamic range was a factor in the results found in experiment 1A, nor did they find that different light levels affected peak dilation. This is congruent with findings of previous research (Bradshaw, 1969). Therefore, Gilzenrat et al. (2010) concluded that the systematic relationship between baseline diameter and peak dilation reflected variation in control state and LC activity.

More recently, when the effects of light level on tonic pupil diameter and peak dilation were examined, Peysakhovich et al. (2017) found evidence that supported the conclusion of Gilzenrat et al. (2010). However, they did not examine a relationship concerning baseline diameter and pupil dilation. Rather, tonic pupil diameter was defined as the mean diameter between 1 – 3 s, post-stimulus (Peysakhovich et al., 2017, p. 42). This measure does not represent a pre-stimulus baseline diameter. They did not justify their operationalisation of mean diameter post-stimulus for use as tonic pupil diameter. Therefore, It is difficult to ascertain consistency in the findings by Gilzenrat et al. (2010) and by Peysakhovich et al.

[19] More details of this study can be found in Chapter 6.

(2017). Nevertheless, in the absence of luminance effects on peak dilation, Peysakhovich et al. (2017) found larger post-stimulus mean diameter was associated with smaller peak dilation.

Changes in baseline diameter and light level had opposing associations with pupil dilation in the current study. This may suggest that there were separate mechanisms which affected the way in which baseline diameter and light level contributed to pupil dilation. It is possible that the inverse relationship between baseline diameter and pupil dilation (when the effects of light level on pupil dilation were accounted for) are consistent with the relationship function proposed by Gilzenrat et al. (2010).

The relationship between baseline diameter and pupil dilation could represent a “top-down” relationship with LC activity related to control state and performance regulation during the speech-in-noise task as predicted by Adaptive Gain Theory. However, LC activity was not directly studied here and a direct, causal link between LC activity and the autonomic nuclei that control pupil size has not been reliably established (Joshi & Gold, 2020; Nieuwenhuis et al., 2011). Neuroimaging studies targeting the LC may be able to shed light on this suggestion.

This section has focused on path *b* in the mediation models. Explanations for the relationship between baseline diameter and pupil dilation, when the effects of light level on pupil dilation were accounted for, were provided. When *a* and *b* paths were combined into the *ab* parameter, it suggested that changes in baseline diameter were associated with pupil dilation in the opposite direction to the direct effect of changes in light level. For example, it was estimated that peak dilation increased 0.36 mm in light level 2 when compared to light level 1 due to changes in baseline diameter in MM1A. These mediation effects are discussed in more detail below.

12.5.3 The Mediatory Effect of Baseline Diameter on Pupil Dilation

The mediation analyses indicated that baseline diameter partially mediated the relationships between light level and pupil dilation. Based on the results reported here, the findings reported in Study 1 may have been influenced by negative associations between baseline diameter and pupil dilation. The largest mediation effect was found in MM1A and MM1B when L1 and L2 were used. Of the light levels compared, this comparison had the smallest absolute difference in lux measurement but the largest difference in pupil dilation between the light levels tested. The mediation effect of baseline diameter existed to a similar extent for both SNR -3 dB and SNR 3 dB. Therefore, the size of the mediation effect was not influenced by SNR condition, but it was influenced by the light level tested (Figure 25). The difference between the two dimmest light levels appeared to show the greatest mediated effect.

The results of the mediation analyses suggest that the within-subjects effect of light level on pupil dilation may have been partially *suppressed* by the relationship between baseline diameter and TEPRs at fixed light levels. MacKinnon et al. (2007) termed this “inconsistent mediation”. This happens when the direct effect and the indirect effect are opposite in sign (e.g., + or -), that is, changes in the independent variable and resultant changes in the mediator variable affect the dependent variable in opposing directions. This resulted in a diminished total effect and explains why the direct effect of light level on pupil dilation was larger than the total effect.

The effect of light level on pupil dilation reported in Study 1 is consistent with the total effect values estimated in the mediation models here. This is because the models in Study 1 did not account for the association between baseline diameter and pupil dilation. The mediation analyses revealed that light level may have a larger effect on pupil dilation than originally estimated in Study 1 when the relationship between baseline diameter and pupil dilation is accounted for.

Bradshaw (1969) and Peysakhovich et al. (2017) attempted to directly examine the effect of light level on peak dilation. They both concluded that light level did not affect peak dilation. The findings of Bradshaw (1969) have been used as evidence that peak dilation is independent of baseline diameter and light level. This finding is frequently cited in pupillometry resource guides and review papers (e.g., Beatty, 1982b; Beatty & Lucero-Wagoner, 2000). Since then, many researchers have based their experimental design and comparisons between studies on this finding. The results of the current study suggest that this finding may have been oversimplified.

Gilzenrat et al. (2010) and Reilly et al. (2019) examined the relationship between baseline diameter and peak dilation by manipulating light level. They found no differences in peak dilation between the large baseline condition and small baseline condition (as induced by light level). In Chapter 9, it was suggested that these results could be due to the lesser cognitive load that was imposed by the tasks used in these studies. The mediation analyses presented here may provide further evidence for this explanation.

In previous studies, the suppression effect of baseline diameter may have obscured the effect of light level on pupil dilation, because TEPRs may not have been as robust as they were in the current study. This may be due to the smaller amount of cognitive load imposed by the tasks in previous studies (Gilzenrat et al., 2010; Reilly et al., 2019), when compared to the speech-in-noise task used in the current study. For example, the tasks used in Gilzenrat et al. (2010) and Reilly et al. (2019) were target detection tasks in which participants performed at ceiling. The speech-in-noise task used here resulted in performance range from 8.62% correct to 93.94% correct and would have required a broader range of cognitive processes such as working memory, selective attention, and the use of linguistic and contextual cues. The largest effect of light level on pupil dilation was found in the most adverse SNR conditions in Study 1 (SNR -3 dB and SNR -6 dB).

Support for this interpretation can be found in MM6A which evaluated the mediation effect of baseline diameter in peak dilation between L3 and L4, measured at SNR 3 dB. In this mediation model, 0 was a likely population value for the total effect, that is, there were no differences in peak dilation between these conditions. However, once the relationship between baseline diameter and peak dilation was accounted for, there was a direct effect of light level on peak dilation.

Similar results were found regarding mean dilation in MM3B and MM6B. For example, there were no total effects, but there were direct effects of light level when baseline diameter relationships were accounted for. Additionally, total effects of light level on pupil dilation were still observed in more adverse SNR conditions even when the suppressing effect of baseline diameter was not accounted for.

Reilly et al. (2019) concluded that peak dilation was independent of baseline diameter in target detection tasks. However, the current findings suggest that not only could there be a relationship between baseline diameter and peak dilation, but there may also be a direct effect of light level on the peak dilation that is not mediated by baseline diameter in a speech-in-noise task.

The suggestion that a suppressive effect of baseline diameter on pupil dilation may have masked the effect of light level in previous research is speculative and requires verification (this is discussed in more detail in Section 12.5.4). Given the preliminary and exploratory nature of the current findings, the novel analysis methods that were applied, and the differences in methodologies between this study and previous research, further research regarding the relationships between light level, baseline diameter, and pupil dilation is required. This research provided clear evidence that pupil dilation may not be independent of light level or baseline diameter. Potential limitations of the methods that were used in this chapter and alternative explanations for the current results are discussed below.

12.5.4 Limitations

Mediation analyses allow researchers to investigate potential causal pathways of a relationship through intervening variables (Vuorre & Bolger, 2018). The purpose of the mediation analyses in the current study was to assess the ability of baseline diameter to statistically account for the relationship between light level and pupil dilation. A potential limitation of the research reported here is that a mediation model is a theoretical model that implies *causal* relationships (Pirlott & MacKinnon, 2016, p. 3). For example, in a simple mediation model, it is implied that the independent variable causes a change in the mediator and that effect then causes a change in the dependent variable.

However, Pirlott and MacKinnon (2016) have pointed out that providing statistical evidence of mediation does not provide causal evidence for a mediation relationship (p. 4). In order to infer causality of both the independent and mediator variables, Shadish et al. (2002) detailed three experimental design requirements: (1) the causal variable precedes the dependent variable in time (i.e., temporal precedence), (2) the causal variable and the dependent variable vary together (i.e., covariation), and (3) there are no other plausible explanations to account for the relationship between the causal variable and the dependent variable.

In the current experimental paradigm, all participants experienced all four light levels. Light level stimuli were directly manipulated, subsequently baseline diameter was measured and following that, peak dilation value was measured. The results reported here enable causal interpretation of the light level to pupil dilation (i.e., X to Y) and light level to baseline diameter (i.e., X to M) relationships. These relationships satisfy the criteria for implying causality of the independent variable as outlined by Shadish et al. (2002): temporal precedence, covariation, and elimination of alternative explanations.

However, it may not be correct to assume that the changes in baseline diameter measured here had a causal effect on pupil dilation (i.e., M to Y). Mediation analysis also assumes that the mediator variable is independent of unmeasured factors that might affect the dependent variable (Green et al., 2010). Unless the researcher measures and controls the other factors that could affect the relationship between the mediator variable and the dependent variable, there is a risk of misattribution of the mediated effect.

In the current study, light-induced changes in baseline diameter were not the only changes in baseline diameter that may have influenced the relationship between baseline diameter and pupil dilation. Trial number may have affected the relationship between baseline diameter and peak dilation (e.g., due to fatigue) in the current investigation. Therefore, trial number was included in the model as a random effect to account for this potential effect. However, based on the literature reviewed in Section 12.1 and 12.5.2, factors such as arousal or control state may also affect baseline diameter and the relationship between baseline diameter and pupil dilation. These factors were not measured or controlled in the current study, and may in fact be difficult to measure accurately for the purposes of establishing causal relationships (MacKinnon & Pirlott, 2015).

The indirect effect is the product of the *a* and *b* path coefficients in the model and quantifies the mediated effect. For example, in MM1A, the indirect effect represents the estimate of the combined effects of the change in light level (from L1 to L2) on baseline diameter and all other non-light-induced changes to baseline that led to changes in peak dilation. The indirect effect is intended to represent the degree to which light level affected pupil dilation via changes in baseline diameter. However, disentangling the effects of light-induced changes to baseline diameter from the effects of other phenomena that may have led to changes in pupil dilation cannot be achieved using the current methods. Uncontrolled factors such as arousal or control state may well have contributed to the indirect effect reported here. Therefore, any causal conclusions about the M to Y path (i.e., the baseline diameter to peak dilation path) and the mediation analyses in general may not be valid (Green et al., 2010; Judd et al., 2001).

As such, the earlier suggestion that competing relationships between light level and baseline diameter may have led to null effects of light level on peak dilation in previous research may also not be valid. Due to differences in methodologies used in the present study compared to other studies (e.g., Bradshaw, 1969; Gilzenrat et al., 2010; Peysakhovich et al., 2017; Reilly et al., 2019), conclusions about the relationship between baseline diameter and pupil dilation cannot be confidently stated. The uncertainty regarding the factors involved in the relationship between baseline diameter and pupil dilation suggests that more research is needed to understand the mechanisms involved before conclusive effect patterns can be delineated.

Gilzenrat et al. (2010) found a relationship between baseline diameter and peak dilation as predicted by Adaptive Gain Theory (Aston-Jones & Cohen, 2005), but peak dilation was unaffected when light level was manipulated to induce specific baseline diameters in their participants. Despite the clear effects of light level on pupil dilation in the current study, the findings by Gilzenrat et al. (2010) may suggest that peak dilation is relatively insensitive to light-induced changes in baseline diameter and the relationship between baseline diameter and pupil dilation found here may be reflective of another phenomenon (e.g., control state). The findings by Bradshaw (1969), Peysakhovich et al. (2017), and Reilly et al. (2019) may also suggest that pupil responses are relatively insensitive to light-induced changes to baseline diameter but may be more affected by a different underlying relationship between baseline diameter and pupil response.

The current mediation analysis cannot distinguish between the possibility that: (1) light-induced changes to baseline diameter affected pupil dilation, or (2) pupil dilation is relatively insensitive to light-induced changes in baseline diameter but is affected by different relationship between baseline diameter and pupil dilation, for instance, control state.

In summary, any conclusion based on the current results which suggests light-induced changes in baseline diameter mediated the relationship between light level and pupil dilations via a suppression effect should be regarded as tentative. Further research using the mediation analysis approach, but possibly employing alternative methods to characterise the relationship between baseline diameter and pupil dilation, may help to clarify the relationships between light level, baseline diameter, and task-evoked pupil dilation.

12.5.5 Implications

The relationships between light level, baseline diameter, and/or task-evoked pupil dilation have been examined in a number of past studies (e.g., Bradshaw, 1969; Gilzenrat et al., 2010; Jepma & Nieuwenhuis, 2011; Książek et al., 2021; Peysakhovich et al., 2015; Peysakhovich et al., 2017; Reilly et al., 2019; Steinhauer et al., 2004; Wang, Kramer, et al., 2018). These studies have focused on a combination of two out of the three variables. However, a statistical mediation analysis approach involving all three variables, as described here, has not previously been reported.

Considering the current findings, it may not be appropriate to use light level to manipulate baseline diameter, especially when the goal is to examine the relationship between baseline diameter and pupil dilation. Brighter light levels appear to attenuate task-related pupil dilation, particularly in challenging listening conditions.

The relationship between light level, baseline diameter, and pupil dilation may be complex.

Evidence from the current investigation suggests that there is an inverse relationship between baseline diameter and pupil dilation during a speech-in-noise task. This conflicts with the results of Reilly et al. (2019) who concluded that peak dilation was independent of light-induced baseline diameter, that is, the peak dilation response function scaled linearly to baseline diameter. The results of the current study reinforce that the issue of baseline dependence in the TEPR needs further examination. The current study suggested that TEPRs may be affected by baseline diameter and that light level may independently affect TEPRs.

Mathôt et al. (2018) acknowledged that it is unlikely that TEPRs are independent of baseline diameter. However, regardless of pupil scaling, they recommended that subtractive baseline correction should still be performed over alternatives (i.e., divisive) due to the increased statistical power that subtractive correction provides. Therefore, when examining baseline-corrected TEPRs, it is recommended that researchers include measures of baseline diameter in their analyses. Further investigation of the baseline diameter to TEPR relationship in various tasks should take place before any conclusions are made and standard practice recommendations are set.

Previous research has sometimes included a method of baseline standardisation whereby the light levels are individually set so that the baseline diameter is in the middle of the participant's dynamic range for the test session. This leads to different lighting conditions between participants. While this approach may "standardise" baseline diameter, it may not make the subsequent pupil response independent from its initial baseline size as previously claimed (e.g., Zekveld et al., 2011). Furthermore, the research reported here suggests that varying light levels could result in deviations in TEPRs between participants and experimental laboratories that are not due to the task under examination, and that the effect of light level may also vary based on the amount of effort an individual expends in the task. Therefore, these results may suggest that researchers should control for baseline diameter statistically, rather than adjusting light levels individually for each participant.

Additionally, light level should be relatively dim at eye level, to ensure that TEPRs are most sensitive to variation in listening effort.

12.5.6 Conclusion

Study 2 indicated that baseline diameter partially mediated the relationship between light level and pupil dilation. These results also indicated that changes in light level influenced pupil dilation in the opposite direction to corresponding changes in baseline diameter. In the present study, changes in baseline diameter acted as an inconsistent mediator in the relationship between light level and pupil dilation.

The direct effects of light level were larger when within-subject changes in baseline diameter were accounted for in the current study. Building on the techniques described here, additional research should be conducted to ascertain if the results reported here replicate in different tasks, and in different laboratories. Further research may elucidate the mechanisms at play in the relationships between light level, baseline diameter, and task-evoked pupil dilation so that standard recommendations can be established for appropriate research design, analysis, and clinical practice.

13 STUDY 3A – TASK-RELATED FATIGUE AND ENGAGEMENT DURING THE SPEECH-IN-NOISE TASK

13.1 Background and Aims

Like listening effort, listening-related fatigue is a multidimensional phenomenon which can extend across physical, mental, emotional, and social domains (Davis et al., 2021, p. 458). The multidimensional nature of listening-related fatigue can make it difficult to measure precisely. Listening-related fatigue is typically assessed using methods similar to those used to assess listening effort, that is, subjective, behavioural/performance, and physiological methods (McGarrigle et al., 2014) (see Chapter 4).

The first aim of Study 3A was to examine subjective judgements of fatigue and task-engagement during performance of the speech-in-noise task used for the research presented in this thesis. The second aim was to assess whether there was an average performance decrement as a function of time block, and whether performance was linked to subjective feelings of fatigue and/or task-disengagement.

Due to the frequent reports of excessive listening effort and fatigue that clients make to audiologists, consideration of subjective fatigue and task-engagement during task performance is important. It was predicted that subjective fatigue would increase, and subjective task-engagement would decrease over the test session. Furthermore, it was predicted that average performance would decrease over the test session. It was also predicted that the subjective ratings and performance would be related. Therefore, fatigue was assessed subjectively and behaviourally in the current chapter. The findings of this chapter informed the content of Chapter 14, in which the effects of time-on-task on peak dilation, mean dilation, and baseline diameter are reported.

13.1.1 Subjective Measurement of Fatigue

Subjective measurement is one method by which researchers and clinicians can assess listening-related fatigue. As discussed in Section 4.1, subjective measures rely on an individual's provision of accurate self-reports regarding their mental state and experiential feelings. Subjective measures have the benefit of giving a certain level of control to individuals and acknowledging that they are the "experts" when it comes to their own experiences. Subjective, anecdotal reports of excessive listening effort and fatigue to audiologists led to extensive research on these phenomena (Pichora-Fuller et al., 2016).

However, subjective measures have some significant drawbacks related to response bias, interpersonal differences in effort and fatigue thresholds, and/or the use of heuristic response strategies (McGarrigle et al., 2014; Moore & Picou, 2018) (see Section 4.1 for more detail on these biases). Furthermore, there are often discrepancies between subjective and objective measures of listening effort and fatigue, which may suggest that these measures reflect different underlying phenomena (McGarrigle et al., 2014).

Moore and Picou (2018) recently distinguished two types of instruments (e.g., questionnaires, rating scales) used for assessing subjective fatigue: validated and unvalidated. Validated instruments are those that have been subject to formal testing, have known psychometric properties, and have accessible normative data. They are also rigid and cannot be easily modified for specific research purposes. Unvalidated instruments have not been subjected to the same amount of testing but can be more readily customised to suit the specific aims of the study and/or clinical presentation.

For example, Moore et al. (2017) used the validated "Profile of Mood States" (McNair, 1971) and several unvalidated but task-specific rating scales developed based on the Motivational Control Theory of Fatigue. Only responses to the unvalidated, task-specific rating scales were correlated with the neural activity related to decreased arousal (e.g., reduced N1 amplitude in the event-related potential). This might mean that, although the Profile of Mood States is a validated tool for measuring fatigue, the unvalidated rating scales were more sensitive to task-related fatigue effects in individuals without hearing loss and this may be attributable to the scales' task-specific nature. Wording of items/questions, response mode and specificity of the instruments can all affect how a participant interprets and responds on the measures.

Despite the drawbacks of subjective measurement, understanding how individuals feel during and after specific listening situations is valuable for research and clinical audiology. The overarching goal as researchers of hearing science and audiological clinicians is to improve quality of life for individuals. To do this, it is important to understand how individuals feel during specific listening situations and if those feelings have any relationship to task performance which may indicate fatigue. Examination of these factors is important for ongoing research in the area. Furthermore, understanding how individuals feel during task performance may help interpretation of any clinical measure of listening effort.

13.1.2 Subjective Fatigue During Listening

Acute mental fatigue can refer to transient reactions to periods of intense cognitive effort (e.g., a challenging listening situation) (van der Linden, 2011). The literature associated with time-on-task and fatigue effects in TEPRs was reviewed in Section 6.5. As described, fatigue reactions may also involve the subjective experience of tiredness, exhaustion, and/or lack of energy, and they may also involve cognitive performance decrements (Bess & Hornsby, 2014) (e.g., fewer words identified in a speech-in-noise task).

Compared to individuals without hearing loss, both children and adults with hearing loss or hearing difficulties report more severe feelings of fatigue (Alhanbali et al., 2017; Bess & Hornsby, 2014; Hornsby & Kipp, 2016; Hornsby et al., 2014). As such, much of the research that examines listening-related fatigue focuses on these clinically relevant populations (e.g., Alhanbali et al., 2017; Hicks & Tharpe, 2002; Hornsby, 2013; Hornsby & Kipp, 2016; Hornsby et al., 2014; Picou et al., 2019).

However, transient, task-related fatigue likely affects all individuals to some extent during and after sustained effortful listening and this may affect an individual's energy levels, performance, and TEPRs during the measurement of listening effort.

There is considerably less attention given to the experience and performance costs of task and listening-related fatigue in samples of individuals without hearing loss or hearing difficulties. However, in a sample of individuals without hearing loss, Moore et al. (2017) found a relationship between subjective, task-relevant fatigue measures and physiological responses recorded during an auditory task. More recently, McGarrigle, Rakusen, et al. (2021) examined subjective tiredness from listening during a speech-in-noise task in a sample of individuals without hearing loss. They found that tiredness increased as a function of time block during the task. This finding was replicated in both young and old listeners in McGarrigle, Knight, et al. (2021). This may indicate that their participants experienced listening-related fatigue during the testing. Interestingly, despite subjective reports of tiredness from listening, performance improved over the course of the test session in both McGarrigle, Rakusen, et al. (2021) and McGarrigle, Knight, et al. (2021).

Performance decrement during and/or after performance on a cognitive task can also be used as a measure of fatigue (DeLuca, 2005). van der Linden et al. (2003) examined the effects of fatigue on executive functioning and found that after 2 hr of working on cognitively demanding tasks, participants showed compromised executive control which may lead to errors and sub-optimal performance. However, subjective fatigue rarely correlates with cognitive performance (DeLuca, 2005). Despite this, it is possible that time spent engaged in an effortful listening task could lead to subjective fatigue and that this may be related to performance decrement. There is currently a dearth in research examining the effect of time-on-task on speech recognition performance during a speech-in-noise task.

The task used in Moore et al. (2017) was 50 min in duration and the tasks that were used in McGarrigle, Rakusen, et al. (2021) and in McGarrigle, Knight, et al. (2021) were approximately 40 minutes in duration. The speech-in-noise task used in this thesis lasted approximately 1 hr. It is conceivable that subjective fatigue was experienced during this task and that this may be reflected in subjective, performance, and physiological measures. If pupillometry is to be used in clinical audiology, it is important to understand how fatigue manifests in individuals and if this is reflected in subjective experience and performance during the task and in the related TEPR measures (addressed in Study 3B, Chapter 14).

13.2 Significance

Subjective fatigue, subjective task-engagement and performance during a speech-in-noise task was examined in this chapter. If pupillometry is to be used in clinical audiology, it is important to understand how fatigue and task-engagement during listening tasks manifest in individuals and whether this is reflected in both performance during the task and in the related TEPR measures.

13.3 Hypotheses

- H1: Subjective fatigue will increase as a function of time block
- H2: Subjective task-engagement will decrease as a function of time block
- H3: The total number of words correctly identified in a time block will decrease as a function of time block
- H4: The total number of words correct in a time block will be negatively correlated with subjective task-engagement
- H5: The total number of words correct in a time block will be negatively correlated with subjective fatigue scores

13.4 Methods

Information related to the methods can be found in Chapter 8 - General Methods.

13.4.1 Participants

Information related to the participants can be found in Chapter 8 - General Methods, Section 8.1.

13.4.2 Materials

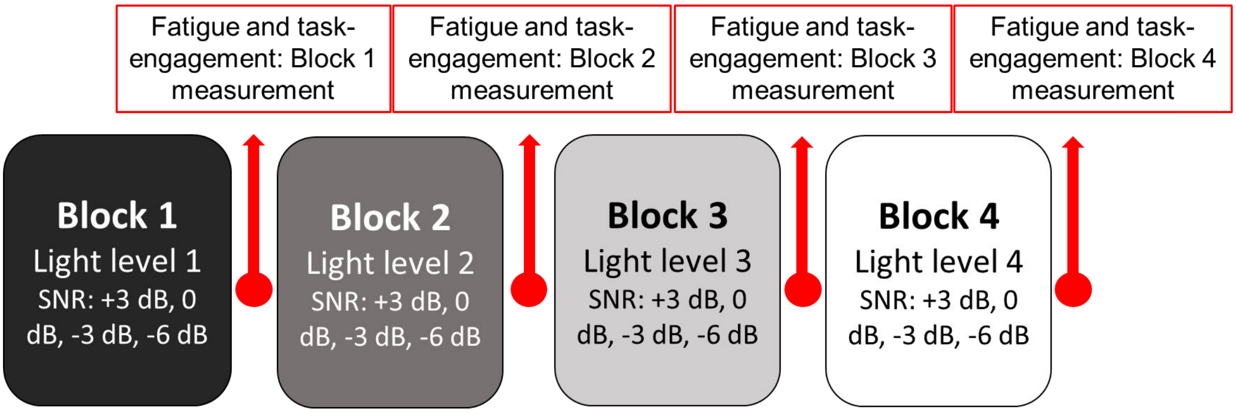
Information related to the materials can be found in Chapter 8 - General Methods, Section 8.2.

13.4.2.1 Fatigue and Task-Engagement Measures

The additional materials used in this chapter relate to the subjective fatigue and task-engagement measures. Subjective fatigue and task-engagement were measured after each block (four times throughout the speech-in-noise task, Figure 31) by asking “How fatigued do you feel right now?” and “How engaged in the task are you right now?”. These questions were similar to the unvalidated, task-relevant scales used in Hopstaken et al. (2015a) and Hopstaken et al. (2015b).

Figure 31

Depiction of Fatigue and Task-Engagement Measurements



Note. This figure shows the procedure for measuring fatigue and task-engagement over the four blocks of the speech-in-noise task. Each block involved 16 sentences at each SNR at one light level. Fatigue and task-engagement was measured after each block. See Appendix 6 for the specific condition order for each participant.

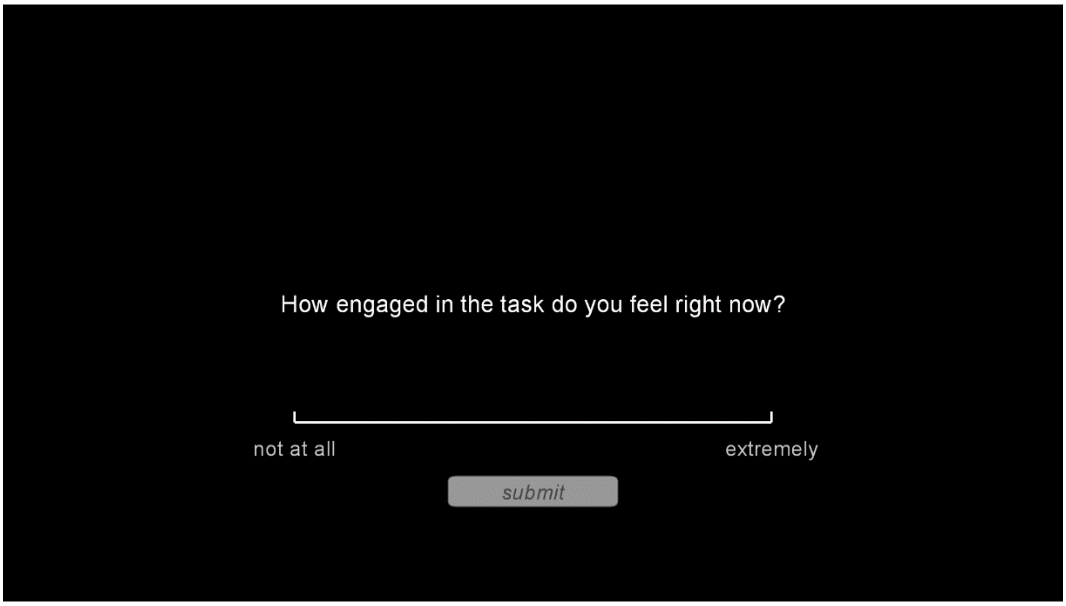
Questions were presented on the computer screen via the PsychoPy program using visual analogue scales ranging from 0 – 100 (Figure 32 and Figure 33). The ends of the scales were labelled with “not at all” and “extremely”. Participants were unable to see the precise number on the scale that they had selected. If a participant can remember the number that they rated themselves on a previous block, they may be more likely to rate themselves differently on a subsequent block based on what is being asked of them. Therefore, this may limit susceptibility to demand characteristics in responses.

Figure 32

Example of the Fatigue Scale Used

Figure 33

Example of the Task-Engagement Scale Used



To analyse the relationships between time block, performance and subjective ratings, the performance scores for each condition were aggregated by block. The total number of key words correctly identified in a time block was used as the performance score. Therefore, each participant had four performance scores, each of which included performance on sentence lists presented at SNR 3 dB, 0 dB, -3 dB, and -6 dB. There were 50 key words per list. The total performance scores were out of 200.

13.4.3 Procedure

Information related to the procedure can be found in Chapter 8 - General Methods, Section 8.3.

13.4.4 Data Analyses

Data analyses and visualisations were completed in R open source software (version 4.0.1) (R Core Team, 2020) and R Studio (version 1.3.1073) (RStudio Team, 2020). Analysis and visualisation codes are available at https://osf.io/am6uv/.

Data visualisations were completed using ggplot2 (version 3.3.3) (Wickham, 2016). Assumption checks were completed using the *rstatix* package (version 0.7.0) (Kassambara, 2021).The *ez* R package (version 4.4-0) (Lawrence, 2016) was used to compute three one-way RANOVAs to examine:

- H1 – The effects of time block on subjective task-engagement,
- H2 – The effects of time block on subjective fatigue,
- H3 – The effects of time-block on performance during the speech in noise task.

The Greenhouse-Geisser correction was used to correct for violations of sphericity. For significant effects, post hoc main effects analyses were computed using the *rstatix* package (version 0.7.0) (Kassambara, 2021).

RANOVAs were used over MEMs for these analyses because, due to the nature of the data, MEMs are unnecessarily more computationally complex and would compute the same results. Additionally, the dataset did not contain any instances of missing data which would justify the use of MEMs.

The R package *rmcorr* (version 0.4.3) (Bakdash & Marusich, 2017) was used to compute two repeated-measures correlation analyses to examine the relationships between subjective fatigue and performance (H4) and subjective task-engagement and performance (H5) over the course of the speech-in-noise task.

Traditional Pearson correlation coefficients assume independence of observations. Data with repeated measurements (i.e., observations measured within-subjects at two or more time-points) violate this assumption. Therefore, when using this method, observations for the same individuals must be averaged together to avoid this violation. This may mask important information about the within-subjects relationships between observations. The *rmcorr* package (Bakdash & Marusich, 2017) can assess the extent to which measures provide related information while accounting for within-subject dependence of observations. This analysis has recently been used to assess similar relationships (McGarrigle, Rakusen, et al., 2021).

13.5 Results

13.5.1 Assumption Checks

Inspection of QQ plots and the Shapiro-Wilks test indicated that subjective fatigue data was not normally distributed (p = .002). No outliers were identified.

Inspection of QQ plots and the Shapiro-Wilks test indicated that subjective task-engagement data was not normally distributed (p = <.001). There were nine outliers identified but none were deemed “extreme”.

Inspection of QQ plots and the Shapiro-Wilks test indicated that performance data was normally distributed (p = >.05). One outlier was identified but was not deemed “extreme”.

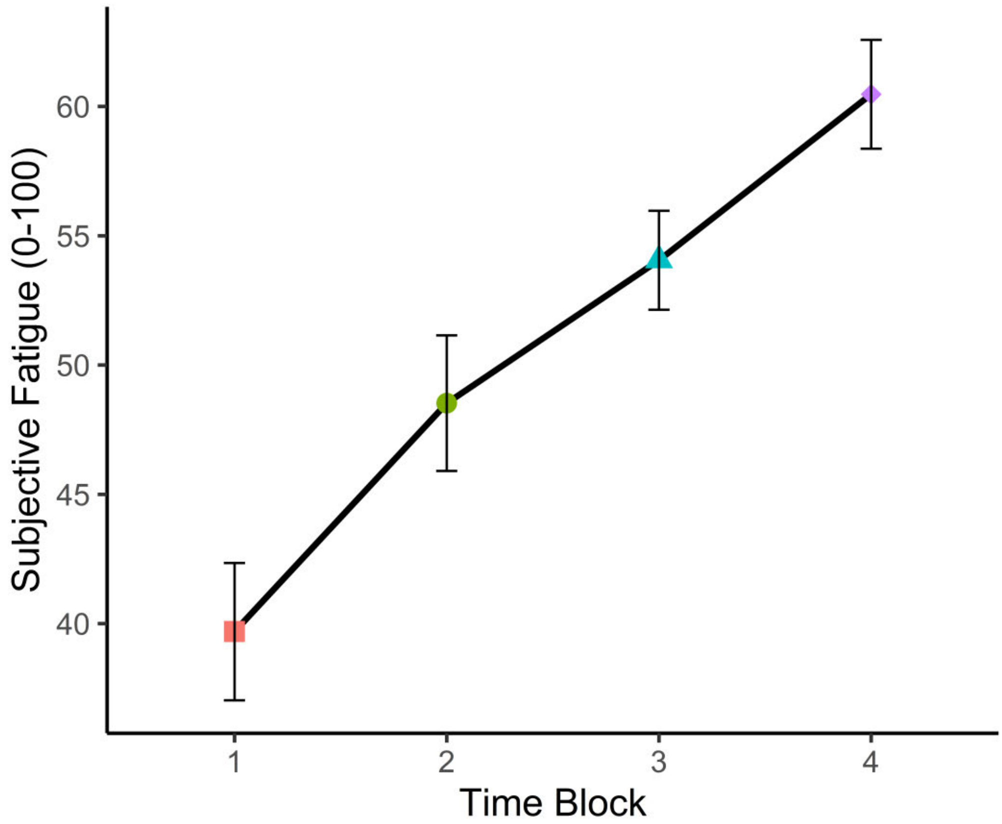
To account for assumption violations in the RANOVAs, a significance level of < .025 was considered significant (Keppel, 1991). Repeated-measures correlations used the bootstrapping method in *rmcorr* to more robustly determine effect sizes (Bakdash & Marusich, 2017, p. 9).

13.5.2 Time Block and Subjective Fatigue

Figure 34 displays the mean fatigue scores as a function of time block and within-subject standard errors. There was a significant main effect of time block on fatigue, $F(2.63, 91.96) = 14.1$, p <.001, $\eta^2_g = .09$). Degrees of freedom and p values incorporate the Greenhouse-Geisser correction.

Figure 34

Mean Subjective Fatigue as a Function of Time Block



Note. Error bars represent within-subject standard error.

Post hoc main effects analysis was conducted to examine significant differences in fatigue as a function of time block (Table 37). Bonferroni correction for six comparisons (alpha is significant at p = .008) was applied. There were significant differences in fatigue between time block 1 and time block 3, time block 1 and time block 4, and time block 2 and time block 4.

Table 37. Post Hoc Main Effects Analysis for Subjective Fatigue as a Function of Time Block

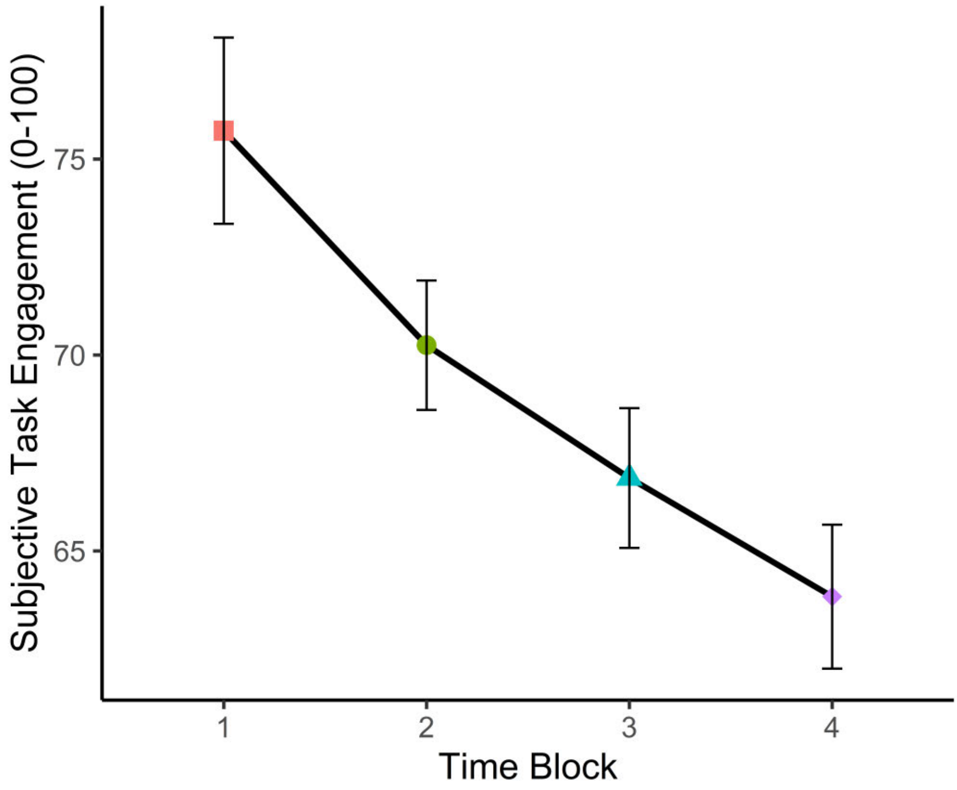
Time block comparison		Mean difference	95% CI – low	95% CI – high	Pairwise *t* statistic	*p*
1	2	-8.83	-16.64	-1.02	-2.30	0.028
1	3	-14.36	-21.38	-7.34	-4.15	<.001
1	4	-20.78	-27.48	-14.08	-6.30	<.001
2	3	-5.53	-11.78	0.72	-1.80	0.081
2	4	-11.94	-19.21	-4.68	-3.34	0.002
3	4	-6.42	-11.40	-1.43	-2.61	0.013

13.5.3 Time block and Subjective Task-Engagement

Figure 35 displays the mean task-engagement scores as a function of time block and within-subject standard errors. There was a significant main effect of time block on task-engagement, $F(2.21, 77.19) = 6.96$, $p = .001$, $\eta^2_g = .04$). Degrees of freedom and *p* values incorporate the Greenhouse-Geisser correction.

Figure 35

Mean Subjective Task-Engagement as a Function of Time Block



Note. Error bars represent within-subject standard error.

Post hoc main effects analysis was conducted to examine significant differences in task-engagement as a function of time block (Table 38). Bonferroni correction for six comparisons (alpha is significant at p = .008) was applied. The only significant difference in task-engagement was between the first time block (1) and the last time block (4).

Table 38. Post Hoc Main Effects Analysis for Subjective Task-Engagement as a Function of Time Block

Time block comparison	Mean difference	95% CI – low	95% CI – high	Pairwise t statistic	*p*

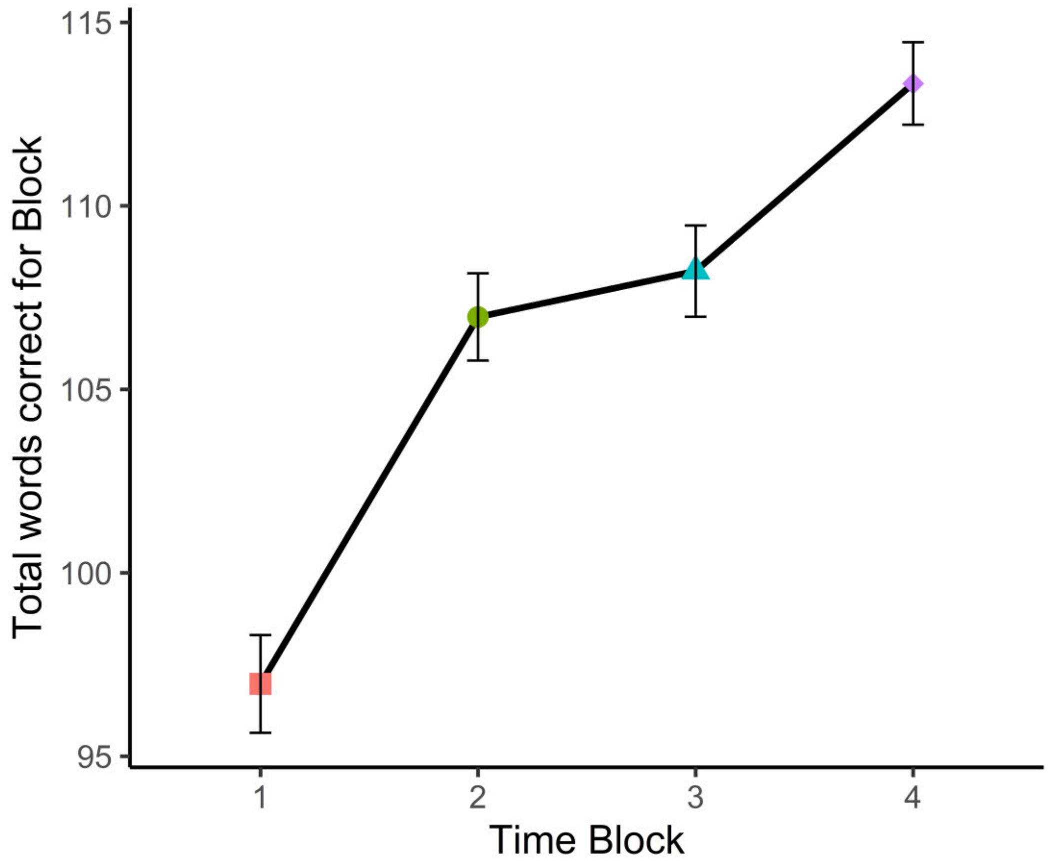
1	2	5.47	0.27	10.68	2.13	0.04
1	3	8.86	1.96	15.76	2.61	0.013
1	4	11.89	5.47	18.31	3.76	<.001
2	3	3.39	-1.31	8.08	1.47	0.152
2	4	6.42	0.88	11.96	2.35	0.024
3	4	3.03	-0.95	7.01	1.54	0.131

13.5.4 Performance and Time Block

Figure 36 displays the mean performance scores as a function of time block and within-subject standard errors. There was a significant main effect of time block on performance, $F(2.73, 95.59) = 31.32$, $p < .001$, $\eta^2_g = .26$. Degrees of freedom and p values incorporate the Greenhouse-Geisser correction.

Figure 36

Mean Performance Scores as a Function of Time Block



Note. Error bars represent within-subject SE.

Post hoc main effects analyses were conducted to examine significant differences in Performance as a function of time block (Table 39). Bonferroni correction for six comparisons (alpha is significant at p = .008) was applied. All comparisons except the comparison between time block 2 and 3 were significant.

Table 39. Post Hoc Main Effects Analysis for Performance as a Function of Time Block

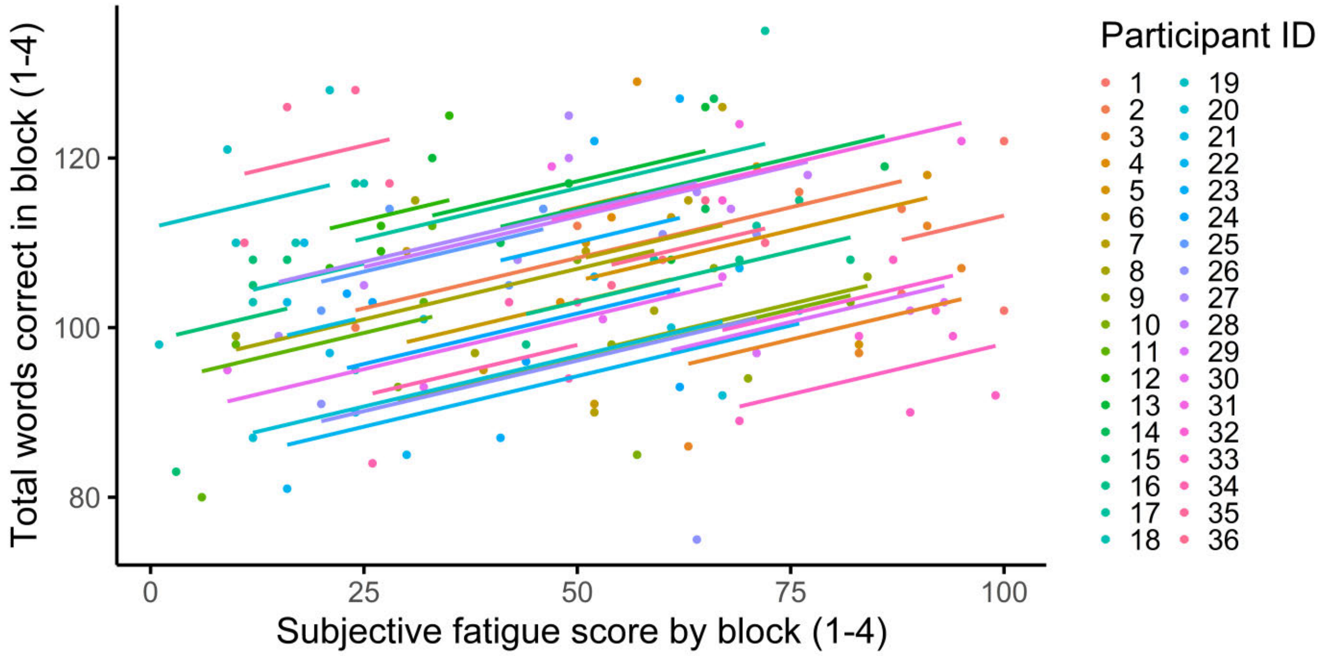
Time block comparison		Mean difference	95% CI – low	95% CI – high	Pairwise t statistic	p
1	2	-10	-13.59	-6.41	-5.66	<.001
1	3	-11.25	-14.80	-7.7	-6.43	<.001
1	4	-16.36	-20.22	-12.5	-8.61	<.001
2	3	-1.25	-5.05	2.55	-0.67	0.51
2	4	-6.36	-9.31	-3.41	-4.38	<.001
3	4	-5.11	-8.36	-1.86	-3.2	0.003

13.5.5 The Relationship Between Performance and Subjective Fatigue

Figure 37 displays the results of the repeated-measures correlation analysis. There was a significant, positive relationship between total number of words correct for a block and subjective fatigue ratings in the same block, $r(107) = .39$ [lower 95% CI = 0.29, upper 95% CI = 0.52], $p = <.001$. As total words correct increased, subjective fatigue also increased.

Figure 37

The Relationship Between Subjective Fatigue Score by Block and Total Number of Words Correct by Block



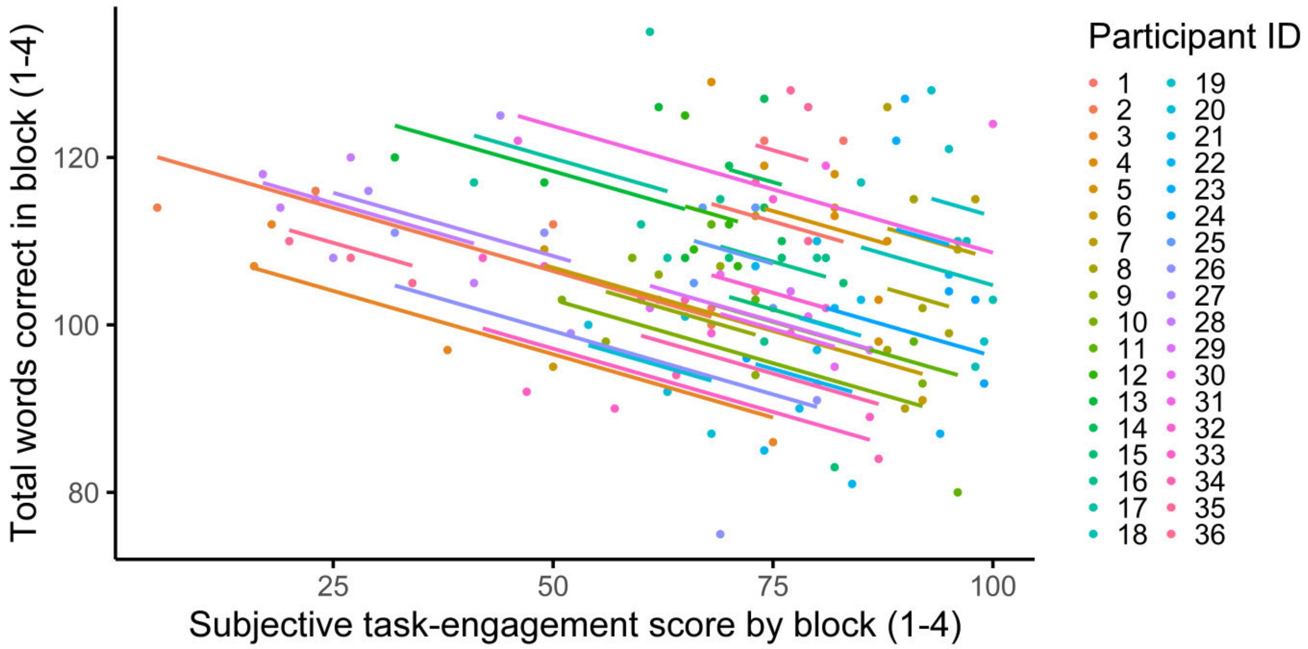
Note. This figure shows the relationship between subjective fatigue score by block (1-4) (x axis) and total number of words correct in a block (1-4) (y axis) by participant (legend). Each participant responded to fatigue question four times throughout the test session, corresponding to the four time blocks. These ratings were then correlated with their aggregated performance scores for each time block.

13.5.6 The Relationship Between Performance and Subjective Task-Engagement

Figure 38 displays the results of the repeated-measures correlation analysis. Results of the repeated-measures correlation analysis indicated that there was a significant negative relationship between total number of words correct for a block and subjective task-engagement ratings in the same block, $r(107) = -.38$ [lower 95% CI = -0.53, upper 95% CI = -0.21], p = <.001. As total words correct increased, subjective task-engagement decreased.

Figure 38

The Relationship Between Subjective Task-Engagement Score by Block and Total Number of Words Correct by Block



Note. This figure shows the between subjective task-engagement score by block (1-4) (x axis) and total number of words correct in a block (1-4) (y axis) by participant (legend). Each participant responded to the task-engagement question four times throughout the test session, corresponding to the four time blocks. These ratings were then correlated with their aggregated performance scores for each time block.

13.6 Discussion

13.6.1 Results Summary

The first aim of Study 3A was to examine subjective fatigue and task-engagement during the speech-in-noise task used in the current thesis. The second aim was to assess whether there was a performance decrement across participants as a function of time block and if this was related to subjective fatigue and/or task-engagement.

As expected, subjective fatigue increased, and subjective task-engagement decreased as a function of time block. Therefore, H1 and H2 were supported. On the other hand, H3 was not supported as performance increased as a function of time block. Furthermore, as subjective fatigue increased, performance also increased. Additionally, as subjective task-engagement decreased, performance increased. Thus, H4 and H5 were not supported. These results are discussed in more detail below.

13.6.2 The Relationship Between Subjective Fatigue and Task-Engagement, and Time block

As expected, subjective fatigue increased as a function of time block, in the current study. Hopstaken et al. (2015a) and Hopstaken et al. (2015b) also found that subjective fatigue increased as a function of time block in a visual n-back task. McGarrigle, Rakusen, et al. (2021) and McGarrigle, Knight, et al. (2021) found subjective tiredness from listening increased as a function of time block in a speech-in-noise task. Post analyses revealed that only the comparisons between time block 1 and time block 3 and time block 1 and time block 4 were significant. The effect of time-on-task on fatigue was significant after participants completed two blocks of the task (i.e., after approximately 30 minutes of the speech-in-noise task).

Subjective task-engagement decreased as a function of time block. This finding aligns with Moore et al. (2017) who found that subjective task-engagement decreased from pre-to-post task in an auditory choice paradigm. Hopstaken et al. (2015a) and Hopstaken et al. (2015b) also found that subjective engagement decreased as a function of time block in a visual n-back task. However, in the current study, post hoc analyses revealed that only the comparison between time block 1 and time block 4 was significant. Therefore, while there was an effect of time-on-task on subjective task-engagement, the effect was not significant until participants had completed three blocks of the task (i.e., after approximately 45 minutes of performance in the speech-in-noise task). Therefore, subjective fatigue developed more quickly than subjective task-disengagement in the current study.

While there were significant effects of time block on subjective fatigue and task-engagement in the current paradigm as outlined above, the amount of variance explained by time block in these measures was small. Effect sizes for the effects of time block on subjective fatigue (η^2_g = .09) and task-engagement (η^2_g = .04) were comparatively smaller than in both Hopstaken et al. (2015a) (fatigue: η_p^2 = .67, task-engagement: η_p^2 = .62, respectively) and Hopstaken et al. (2015b) (fatigue: η_p^2 = .46, task-engagement: η_p^2 = .54, respectively). Therefore, time block explained more variance in subjective fatigue and task-engagement ratings in a visual n-back task, lasting 2 hr in duration, than it did in the current study. The visual n-back task that was used in Hopstaken et al. (2015a) and Hopstaken et al. (2015b) was specifically selected to induce fatigue and involved sustained effort. The speech-in-noise task used in the present study was made up of multiple transient trials and participants had brief breaks after each trial and condition. This could be why the effect sizes in Hopstaken et al. (2015a) and Hopstaken et al. (2015b) were larger than in the current study.

Effect sizes reported here were also smaller than those reported in McGarrigle, Rakusen, et al. (2021) (η_p^2 = .75) and McGarrigle, Knight, et al. (2021) (η_p^2 = .19) where subjective tiredness from listening was examined as a function of time block in a speech-in-noise task that was similar to the task used here. The speech-in-noise tasks used in the current study and in McGarrigle, Rakusen, et al. (2021) and

McGarrigle, Knight, et al. (2021) were also similar in duration. Therefore, time block may affect judgements about tiredness from listening during a speech-in-noise task more than time block affects judgements of fatigue and task-engagement.

In line with the FUEL (Pichora-Fuller et al., 2016), the decrease of task-engagement may be attributable to a reduction in motivation as a protective mechanism to avoid further fatigue (Hockey, 2011, 2013). This is supported by the subjective fatigue results. Subjective fatigue significantly increased after time block 3 from time block 1, and again after time block 4. The development of significantly greater fatigue after time block 3, may have influenced the significant reduction in task-engagement after time block 4. Therefore, these results indicate that during speech-in-noise tasks individuals may subjectively experience some task-disengagement after a period of effortful listening and this could be due to the development of subjective fatigue. However, due to the susceptibility of subjective measurements to various biases (e.g., response bias) and the effect sizes that were reported, these results should be interpreted with caution.

It is possible that the subjective fatigue and task-engagement results reported above were largely due to response bias, specifically – demand characteristics (Gawron, 2016). Because participants were asked “How fatigued do you feel right now?” and “How engaged in the task are you right now?” four times throughout the test session, they may have altered their responses to conform to what they believed the purpose of the experiment was. This issue was circumvented as much as possible by using a slider bar where the participants were unable to see the number score, they had given. Due to the effortful nature of the task, they may have been distracted enough to forget where they had placed the slider bar, in the previous block.

Due to the drawbacks associated with subjective measurements (e.g., response biases), they are typically used in conjunction with performance or behavioural measures. Performance decline over the course of a task may be used as an indication of fatigue and may support subjective reports (DeLuca, 2005). However, this was not the case in the current dataset – performance increased with increasing time-on-task.

13.6.3 The Relationship Between Performance and Time Block

Performance measured over the course of a task is a common metric used to assess fatigue and/or listening effort (Hornsby et al., 2016; McGarrigle et al., 2014). Performance increased with increasing time-on-task in the current study. These findings likely indicate a practice effect. It might also indicate that participants did not have enough practice on the task before beginning. However, if this was the case, as there was a consistent increase in performance by time block (over the 1 hr task), it would likely be impractical to perform a practice task until performance stabilises.

The performance results may also support that the subjective results reported above were due to response bias. Participants became better at identifying the words of sentences presented in background noise as the task progressed, by time block (except from time block 2 to time block 3). Typically, when an individual is becoming fatigued during a task (and potentially disengaging from the task as a result), performance will decrease (DeLuca, 2005). This was not the case in the current dataset.

The performance findings reported here conflicted with two studies where it was reported that while subjective fatigue increased and subjective task-engagement decreased, performance also decreased as a function of time-on-task (Hopstaken et al., 2015a, 2015b; Hopstaken et al., 2016). However, these studies used a visual n-back task with 2 hr duration. The difference in findings reported here and those using the n-back task may be due to the different nature of the tasks (e.g., auditory vs. visual) and the different task durations that were used (e.g., 1 hr vs. 2 hr)

When performance as a function of time block was examined in an auditory paradigm similar to the one used in the current study, the performance results seem more consistent with the current findings. McGarrigle, Rakusen, et al. (2021) and McGarrigle, Knight, et al. (2021) also found that speech recognition performance (in a speech-in-noise task) increased as a function of time block. The effect size for this finding ($\eta_p^2 = .52$) was larger than in the current study ($\eta^2_g = .26$). Time block did not affect performance as much in the current study as it did in McGarrigle, Rakusen, et al. (2021) but it was still a relatively large effect.

In line with predictions set out in the Motivational Control Theory of Fatigue (Hockey, 2011, 2013), Boksem and Tops (2008) argued that fatigue leads to a reassessment of the costs and rewards of expending effort and engaging in a task. From this view, performance is compromised when the costs of expending effort in the task outweigh the benefits of sustaining effort. It is possible that despite subjective reports of fatigue, participants persisted and continued to expend effort to maintain (and improve) their performance. If this was the case, the pupil parameters may show an increase over the course of the test session, indicating greater effort expenditure (addressed in Chapter 14).

Motivation plays an important role in effort expenditure, task-engagement, and fatigue. The performance findings reported here may indicate that individuals were sufficiently motivated to do well in the speech-in-noise task, despite the costs of sustained effortful listening. The motivational potential of a task has an important influence on whether individuals stay engaged in a task when experiencing fatigue (Boksem et al., 2006). If an individual is not motivated to perform well, task-engagement would likely decrease when fatigue is experienced. Because performance increased with increasing time-on-task, it may be assumed that participants stayed relatively engaged in the task, despite their subjective reports of task-disengagement and this may have been due to their level of motivation. However, because data on motivation during the task were not collected during this task, this cannot be directly examined or concluded.

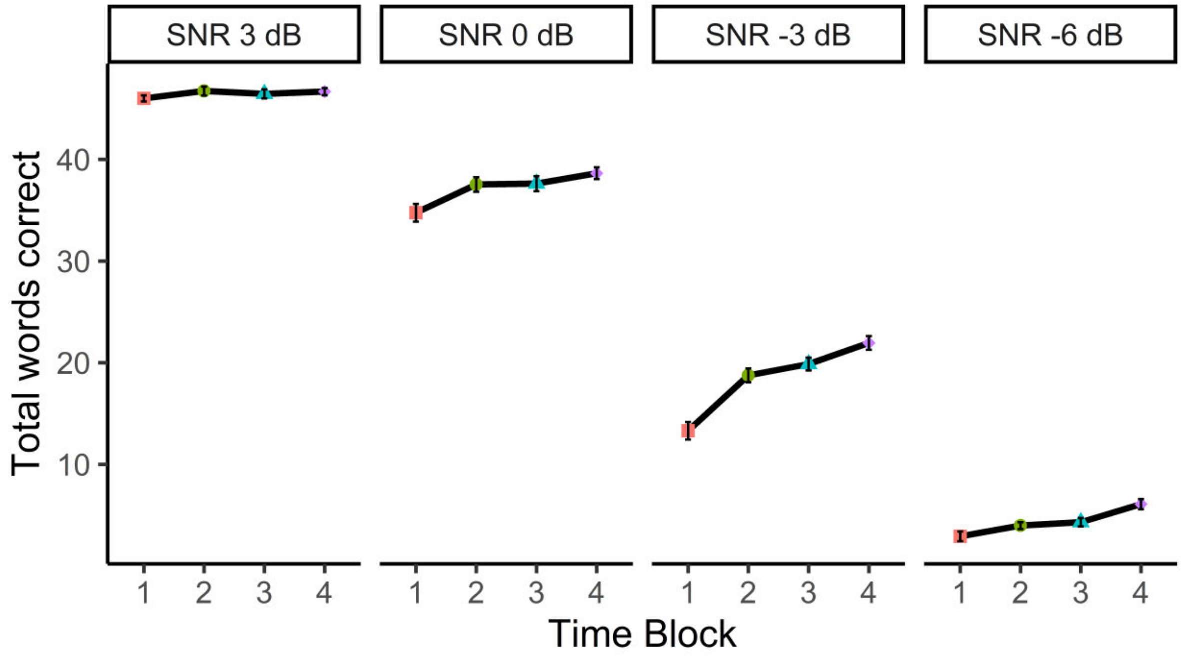
Participants experienced every SNR condition in each time block (Appendix 6). It is possible that participants became better at the task in the easier SNR conditions over the course of the session but became worse in the more difficult conditions (perhaps due to fatigue, task-disengagement, and/or loss of motivation). Differences in responses between the SNR conditions within time blocks may explain why participant’s aggregated performance score increased as a function of time block. When faced with fatigue, participants may have allocated more effort to conditions in which they were more successful and withdrawn effort from the conditions in which they were less successful.

This explanation may be supported by the pupillometry studies in which the inverted-U relationship between pupil dilation and speech intelligibility was reported (Ohlenforst, Zekveld, Lunner, et al., 2017). These findings indicate that the relationship between listening demand and effort expenditure has an upper limit. For example, effort may be withdrawn in tasks that are too demanding, where success is low/impossible. However, direct evidence of the effect of task-success on performance during a speech-in-noise task as a function of time-on-task is lacking. Due to the aggregation of performance scores by block, variation in performance as function of SNR and time block was not directly examinable in the current study, but if this was the case, the subjective and performance measures may have aligned.

Figure 39 depicts the relationship between SNR and performance across time blocks. The placement of the condition within the time block is variable. Despite this, it appears that performance still increased over the test session, even in the more challenging SNR conditions. It would be beneficial to explicitly examine this effect in where SNR presentation is standardised across participants and the position of conditions within time blocks can be specified with greater certainty.

Figure 39

The Relationship Between Performance and Time Block by SNR Condition



Note. This figure shows the average words correct by time block for a given an SNR condition within that time block. Error bars represent within subject standard error.

There has been little examination of the effects of time-on-task on performance in speech-in-noise tasks. This is an area that deserves more attention, particularly due to the findings that performance improved as a function of time-on-task reported here and in McGarrigle, Rakusen, et al. (2021) and McGarrigle, Knight, et al. (2021). This may not be the case for individuals with hearing loss. Due to the increased effort that may be required of individuals with hearing loss to improve performance over time in a speech-in-noise task, a reassessment of the costs associated with task performance may lead to withdrawal, rather than persistence in this group. Further research should verify this claim.

13.6.4 The Relationship Between Performance and Subjective Task-Engagement and Fatigue

Performance was related to subjective fatigue and task-engagement in the current study, but in the opposite direction to that which was expected. Performance increased with increasing subjective fatigue and performance increased with decreasing subjective task-engagement. This was the first study to examine within-subject correlations between subjective fatigue and performance, and task-engagement and performance over multiple time blocks during a speech-in-noise task.

These findings may provide further support for the presence of practice effects in the performance data and response bias in the subjective data of the current dataset. However, they may also highlight the importance of motivational factors in sustained task performance, when subjectively fatigued and disengaged.

It was evident in the relationship between performance and subjective task-engagement that participants did not disengage from the task as much as they thought they did, and their feelings of fatigue did not lead to a decrease in performance. Alhanbali et al. (2019) reported that subjective measures of effort often do not align with behavioural/performance measures as they may tap into different dimensions of the complex construct. The multi-dimensional nature of fatigue may mean that subjective and behavioural measures of fatigue are susceptible to similar discrepancies, and this may be why the current predictions were not supported.

According to DeLuca (2005), it is common for subjective measures of fatigue and performance to be unrelated and that this may reflect that cognitive fatigue and cognitive performance have different underlying mechanisms. The physiological costs of fatigue (e.g., higher systolic blood pressure, more widespread cerebral activation) are said to be more related to the perception of fatigue, rather than performance (DeLuca, 2005). However, the findings reported here show that subjective measures of fatigue and task engagement were positively and negatively (respectively) correlated to averaged by-block performance score.

In the current study, it may be that participants continued to expend effort in the task, to an extent that enabled them to improve their average by-block performance, even though they were feeling fatigued and/or disengaged. Therefore, the participants showed a willingness to perform well in the task, at the expense of feeling fatigued. Relationships depicted in Figure 39 may also support this interpretation.

13.6.5 Conclusion

The aims of Study 3A were: (1) to assess the effect of time block on subjective fatigue, task-engagement, and performance during the speech-in-noise task, and (2) to examine the relationships between these measures. As predicted, subjective fatigue increased, and subjective task-engagement decreased as a function of time-on-task. However, contrary to predictions, performance increased as a function of time-on-task, consistent with a practice effect.

While the patterns of subjective task-engagement and subjective fatigue were significantly related to performance, the direction of the relationship was the reverse of what was predicted. The relationship between subjective measures and performance indicated that the subjective results may have been affected by demand characteristics. In the current dataset, subjective and performance measures of fatigue did not align. More research is needed to elucidate the relationship between listening-related fatigue and performance during speech-in-noise tasks.

Because participants became better at the task over time, examination of time-on-task effects on TEPRs may reveal if the participants experienced a decrease in physiological arousal, consistent with their subjective reports of fatigue and task-disengagement, regardless of their increasing performance. The increase in performance may be attributable to motivation to do well, and/or a practice effect. The effects of time-on-task on TEPRs are examined in the next chapter.

14 STUDY 3B – THE RELATIONSHIP BETWEEN TIME-ON-TASK AND TRIAL-LEVEL PUPIL PARAMETERS

14.1 Background and Aims

In addition to subjective judgements of fatigue and potential performance decrements after a period of sustained effortful listening, fatigue may also manifest physiologically. For instance, the evidence reviewed in Section 6.5 suggests that time-on-task can affect TEPRs in listening tasks and that this may be associated with the subjective experience of fatigue and/or task-disengagement.

Alhanbali et al. (2020) examined whether performance and subjective, task-related fatigue and effort predicted baseline diameter during an auditory digits-in-noise task. They found a relationship between smaller baseline diameter and higher scores on a distinct dimension of the “Visual Analogue Scale to Evaluate Fatigue Severity” scale (Shahid et al., 2012) related to tiredness and drowsiness. Peak dilation measures were not related to fatigue. However, it has been reported that both baseline diameter (Hopstaken et al., 2015a) and peak dilation (Hopstaken et al., 2015b) decreased in size as a function of time-on-task and were also related to measures of subjective fatigue. Furthermore, previous findings indicate that mean TEPRs decreased as a function of time block (McGarrigle, Rakusen, et al., 2021) and trial number (McGarrigle, Knight, et al., 2021) in a speech-in-noise task. Specifically, based on the review in Section 6.5, it may be expected that pupil parameters (peak dilation, mean dilation, and baseline diameter) will decrease as a function of time-on-task and this may be related to fatigue in the current study.

Due to the finding reported in Chapter 13 that performance increased on average by time block, participants may have experienced a learning effect when performing the speech-in-noise task (Foroughi et al., 2017; Sibley et al., 2019). Participants may have improved their performance while maintaining similar effort levels. This would be reflected in consistent measures of task-related effort (peak and mean dilation) across the test session.

On the other hand, it is possible that the task became easier for the participants over the course of the test session by way of a learning effect and they expended less effort while also improving their performance. (Foroughi et al., 2017; Sibley et al., 2019). Peak and mean dilation are measured during stimulus processing and should reveal if participants expended less effort to process the stimuli over time. If this is the case, there should only be a decrease in measures of task-related effort (i.e., peak and mean dilation) and not in baseline diameter. However, it is possible that baseline diameter may decrease over the test session due to a decrease in arousal that is consistent with the task getting easier over time, but this effect has not been established.

Another possibility is that the improvement in performance reflects that participants expended more effort to improve their performance across the test session. This would be supported by an increase in measures of task-related effort (i.e., peak, and mean dilation) and not in baseline diameter, rather than the expected decrease that may be associated with fatigue.

The primary aim of this chapter was to assess if time-on-task effects were present in pupil parameters (peak dilation, mean dilation, and baseline diameter) at the trial-level using mixed-effects modelling. The results of this may elucidate whether the physiological manifestation of fatigue in TEPRs at the trial-level coincides with subjective feelings of fatigue and task-engagement (reported in Chapter 13) for individuals without hearing loss over the course of a speech-in-noise task (1 hr).

The trial-level MEMs reported in Chapter 11 suggested that some variance in trial-level peak and mean dilation was attributable to time-on-task via inclusion of trial number as a random effect. To draw inferences about how trial number affected trial-level pupil parameters, trial number was examined as a fixed effect in this chapter.

The exploratory analyses reported here may confirm the existence of time-on-task effects in trial-level pupil data and may clarify the time-on-task effect patterns. Time-on-task effects have not yet been examined in trial-level pupil parameters. For pupillometry to be used in clinical audiology, it is necessary to examine how the common TEPRs are affected by time-on-task in a clinically relevant speech-in-noise task. Assessment of these effects at the trial-level may lead to enhanced understanding of the factors that affect TEPRs.

14.2 Significance

The findings reported here are significant for research and potential clinical application of pupillometry. Specifically, these findings may: (1) clarify whether individuals performing a speech-in-noise task show decreased arousal via pupil parameters, consistent with fatigue and/or task-disengagement at the trial-level, and (2) add to the growing body of literature regarding consistency between subjective and physiological measurement of fatigue.

14.3 Methods

14.3.1 Participants

Information related to the participants can be found in Chapter 8 - General Methods, Section 8.1.

14.3.2 Materials

Information related to the materials can be found in Chapter 8 - General Methods, Section 8.2.

14.3.3 Procedure

Information related to the procedure can be found in Chapter 8 - General Methods, Section 8.3.

14.3.4 Pupil Data Pre-Processing

Information related to the pupil data pre-processing can be found in Chapter 8 - General Methods, Section 8.4 and information related to the trial data pre-processing can be found in Chapter 11, Section 11.3.4.

14.3.5 Data Analyses

Data analyses and visualisations were completed in R open source software (version 4.0.1) (R Core Team, 2020) and R Studio (version 1.3.1073) (RStudio Team, 2020). Analysis and visualisation codes are available at https://osf.io/am6uv/.

As in Chapter 10 and 11, the model building process and results are reported based on recommendations set out in the recent paper: “Best practice guidance for linear mixed-effects models in psychological science” (Meteyard & Davies, 2020).

Data visualisations were completed using *ggplot2* (version 3.3.3) (Wickham, 2016). A series of 3 MEMs were computed using the *lmer* function of the *lme4* R package (version 3.5.2) (Bates et al., 2015) and *p* values from summary outputs for MEMs were obtained via the *lmerTest* package (version 3.1.0) (Kuznetsova et al., 2017). Model comparisons were conducted using likelihood ratio tests. The default variance-covariance structure specified by *lmer* models (unstructured) was used.

14.3.5.1 Fixed Effects

As in Chapter 11, the fixed effects for the MEMs in this chapter were light level, SNR, and the interaction between light level and SNR. Results reported in Chapter 11 indicated that trial number (as a random effect) improved the model fit for both peak and mean dilation. For example, trial number accounted for 3.3% of the total variance in the Trial-Level Peak Dilation MEM (11.4.1).

The trial number reflects the order in which the specific trial occurred across the test session. For example, “1” in the trial number variable marks the first trial, and “256” in the trial number variable marks the last trial across the entire test session.

To examine the effect of time-on-task, trial number was included as a fixed effect in the MEMs reported in the current chapter. Therefore, these models also estimated the main effect (marginal) variance explained by trial number. This is the variance in peak dilation, mean dilation, and baseline diameter explained by trial number averaged across light levels and SNRs.

14.3.5.2 Random Effects

The maximal random effects structure justified by the design that resulted in a non-singular fit was used for each MEM. Details regarding the random effects structure for each *lmer* model tested and the model comparison results can be found in Table 40 (peak dilation), Table 41 (mean dilation), and Table 42 (baseline diameter).

For peak and mean dilation, the model structure was the same used in the final models in Chapter 11, except “trial number” was removed as a random and included as a fixed effect (Table 40 and Table 41, respectively). For baseline diameter, the model building process started with the null model and all models justified by the design were tested (Table 42).

Table 40. Model Comparison and the Model Building/Selection Process for Peak Dilation with Trial Number as a Fixed Effect

	Sampling Units	N total observations = 8749, N Participants = 36, N Sentence = 256											
	Model specification	**Model name**	**Nested / simpler Model**	**Fixed Effects added**		**Random Effects**		**Model fit**				**LRT Test against nested**	
						Participant	**Sentence**	**AIC**	**BIC**	**LL**	**df**	**df**	**X^2**
1	Final Model from previous analysis, Section 11.4.1.4	MEM_fat 1	-	Light level x SNR		Intercept and random slope by light level	intercept	-2120.3	-1922.1	1088.1	28	-	-
	Note: Trial number deleted as random effect in MEM_fat1												
2	Addition of trial number as fixed effect	MEM_fat 2	MEM_fat 1	Trial number		Intercept and random slope by light level	intercept	-2076.3	-73.63	1321.2	283	255	466.08 ***
	Note. AIC = Aikake Information Criterion, BIC = Bayesian Information Criterion, LL = LogLikelihood, LRT – Likeilhood Ratio Test. * = *p* <.05, ** = *p* <.01, *** = *p* <.001. Final model formula: peakdilation ~ snr * lightlevel + trial_exp + (1 + lightlevel|participant) + (1|sentence), pup_data, REML = FALSE												

Table 41. Model Comparison and the Model Building/Selection Process for Mean Dilation with Trial Number as a Fixed Effect

	Sampling Units	N total observations = 8749, N Participants = 36, N Sentence = 256											
	Model specification	**Model name**	**Nested / simpler Model**	**Fixed Effects added**		**Random Effects**		**Model fit**				**LRT Test against nested**	
						Participant	**Sentence**	**AIC**	**BIC**	**LL**	**df**	**df**	**X2**
1	Final Model from previous analysis, Section 11.4.2.3	MEM_fat 1	-	Light level x SNR		Intercept	intercept	-4664.9	-4530.4	2351.4	19	-	-
	Note: Trial number deleted as random effect in MEM_fat1												
2	Addition of trial number as fixed effect	MEM_fat 2	MEM_fat 1	Trial number		Intercept	intercept	-4720.8	-2781.8	2634.4	274	255	565.91 ***
	Note. AIC = Aikake Information Criterion, BIC = Bayesian Information Criterion, LL = LogLikelihood, LRT – Likeilhood Ratio Test. * = *p* <.05, ** = *p* <.01, *** = *p* <.001. Final model formula: meandilation ~ snr * lightlevel + trial_exp + (1|participant) + (1|sentence), pup_data, REML = FALSE												

Table 42. Model Comparison and the Model Building/Selection Process for Baseline Diameter with Trial Number as a Fixed Effect

	Sampling Units		N total observations = 8749, N Participants = 36, N Sentence = Not included because sentences started after baseline diameter was measured										
	Model specification	**Model name**	**Nested / simpler Model**	**Fixed Effects added**		**Random Effects**		**Model fit**				**LRT Test against nested**	
						Participants	**Sentence**	**AIC**	**BIC**	**LL**	**df**	**df**	**X2**
1	Participant random effect	MEM_BLnull	-	-		intercept	NA	16697.6	16718.8	-8345.8	3	-	-
2	Addition of light level as fixed effect	MEM_BL1	MEM_BL null	Light level		intercept	NA	5865	5907.5	-2926.5	6	3	10839 ***
3	Addition of SNR as fixed effect	MEM_BL2	MEM_BL 1	SNR		intercept	NA	5765.1	5828.8	-2873.6	9	3	105.93 ***

4	Addition of Two-way interaction	MEM_ BL3	MEM_BL 2	Light level x SNR		Intercept	NA	5732.9	5860.2	-2848.4	18	9	50.26 ***
5	Addition of trial number as a fixed effect	MEM_ BL4	MEM_BL 3	Trial number		Intercept	NA	4989.5	6921.4	-2221.7	273	255	1253.4 ***
6	Addition of random slope for light level by participant	MEM_ BL5	MEM_BL 4	-		Intercept and random slope by light level	NA	2525.8	4521.4	-980.91	282	9	2481.7 ***
7	Addition of random slope for SNR by participant	MEM_ BL6	MEM_BL 5	-		Intercept and random slope by light level and SNR	NA	2291.2	4414.2	-845.6	300	18	270.61 ***
	Note. Random slope for the interaction between light level x SNR by participant resulted in singularity. Therefore, this term was deleted. AIC = Aikake Information Criterion, BIC = Bayesian Information Criterion, LL = LogLikelihood, LRT – Likeilhood Ratio Test. * = *p* <.05, ** = *p* <.01, *** = *p* <.001. Final model formula: baseline ~ lightlevel * snr + trial_exp + (1 + lightlevel + snr|participant), pup_data, REML = FALSE, control = lmerControl(optimizer = "bobyqa")												

14.4 Results

As the fixed effects of light level, SNR and the light level/SNR interaction were reported in Chapter 11, this results section focuses solely on the main effect of trial number on peak dilation, mean dilation, and baseline diameter.

14.4.1 Trial-level Peak Dilation and Time-On-Task

14.4.1.1 Model Assumptions

Inspection of the residual plots indicated linearity. However, like the model in Chapter 11, they also indicated some heteroskedasticity (fanning pattern). While MEMs assume homoskedasticity, model estimates are usually robust to violations of this assumption (Schielzeth et al., 2020). Therefore, it should be noted that the final model reported in the following section showed some heteroskedasticity in the residuals. This was not assumed to have negatively affected the model estimates. Additionally, QQ plots indicated slight non-normality of residuals. However, MEMs are also robust to this type of violation (Gelman & Hill, 2007; Knief & Forstmeier, 2020; Schielzeth et al., 2020).

14.4.1.2 Model Convergence

All *lmer* models for peak dilation converged.

14.4.1.3 Trial-Level Peak Dilation and Time-On-Task MEM: Final Model Results

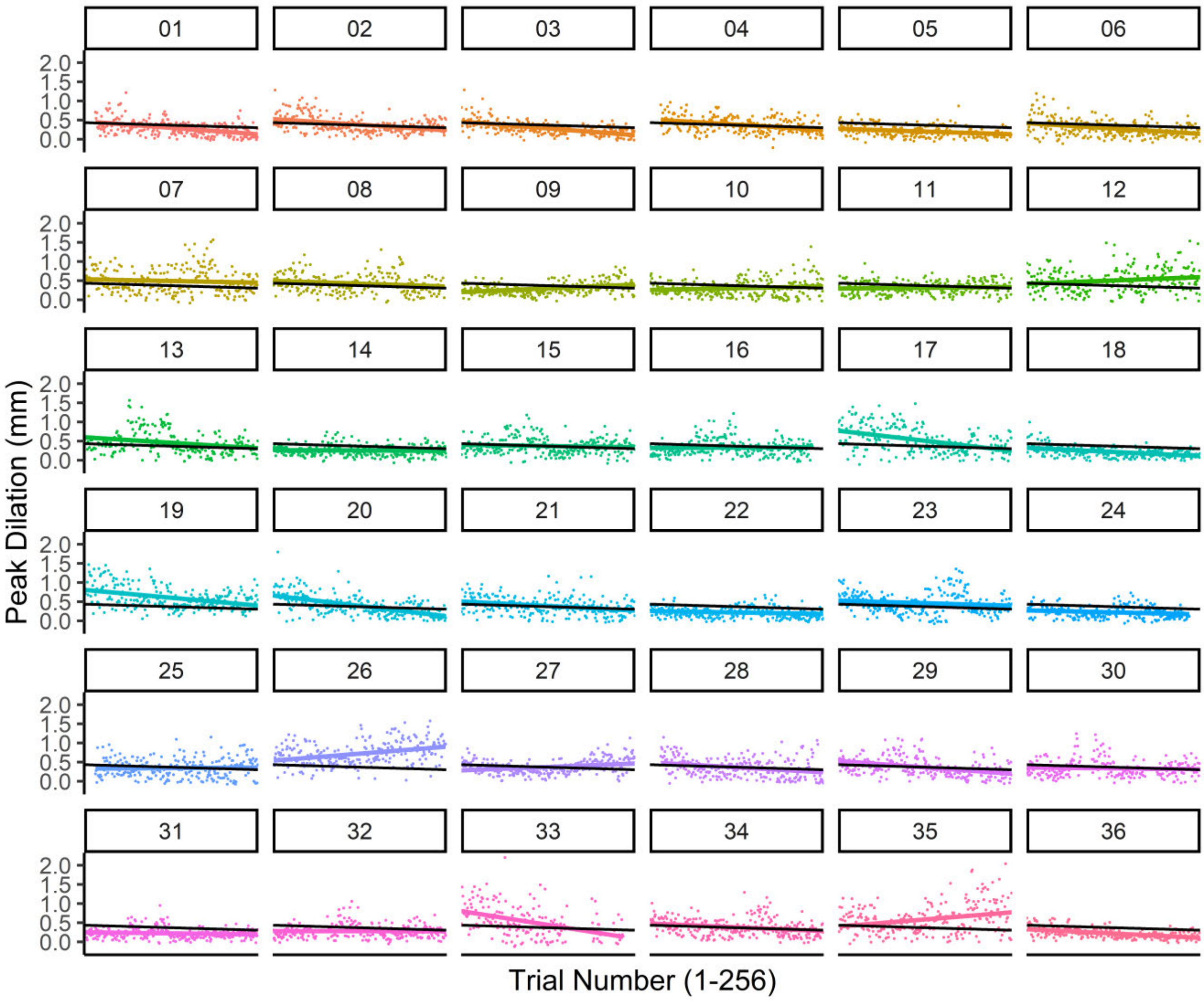
The final model indicated that, when averaged over light level and SNR, trial-level peak dilation was typically smaller when compared to the base level (Trial 1) over the task (i.e., trial 2 – 256) (see Appendix 7 for the model output).

The *anova* function from the *lmerTest* R package (version 3.1-3) (Kuznetsova et al., 2017) was used to compute scaled F-values and denominator degrees of freedom using Satterthwaites' method. Trial number had a significant effect on peak dilation, $F(255, 1870.8) = 2.13$, $p < .001$.

Figure 40 shows the raw peak dilation data for each participant across all trials. Even with the variance associated with light level, SNR and their interaction existing in the data, most participants showed a slight decrease in peak dilation over the course of the experiment. Figure 41 shows the linear trend line and confidence interval (in red) in a scatter plot of the peak dilation values estimated in the MEM and the measured peak dilation values.

Figure 40

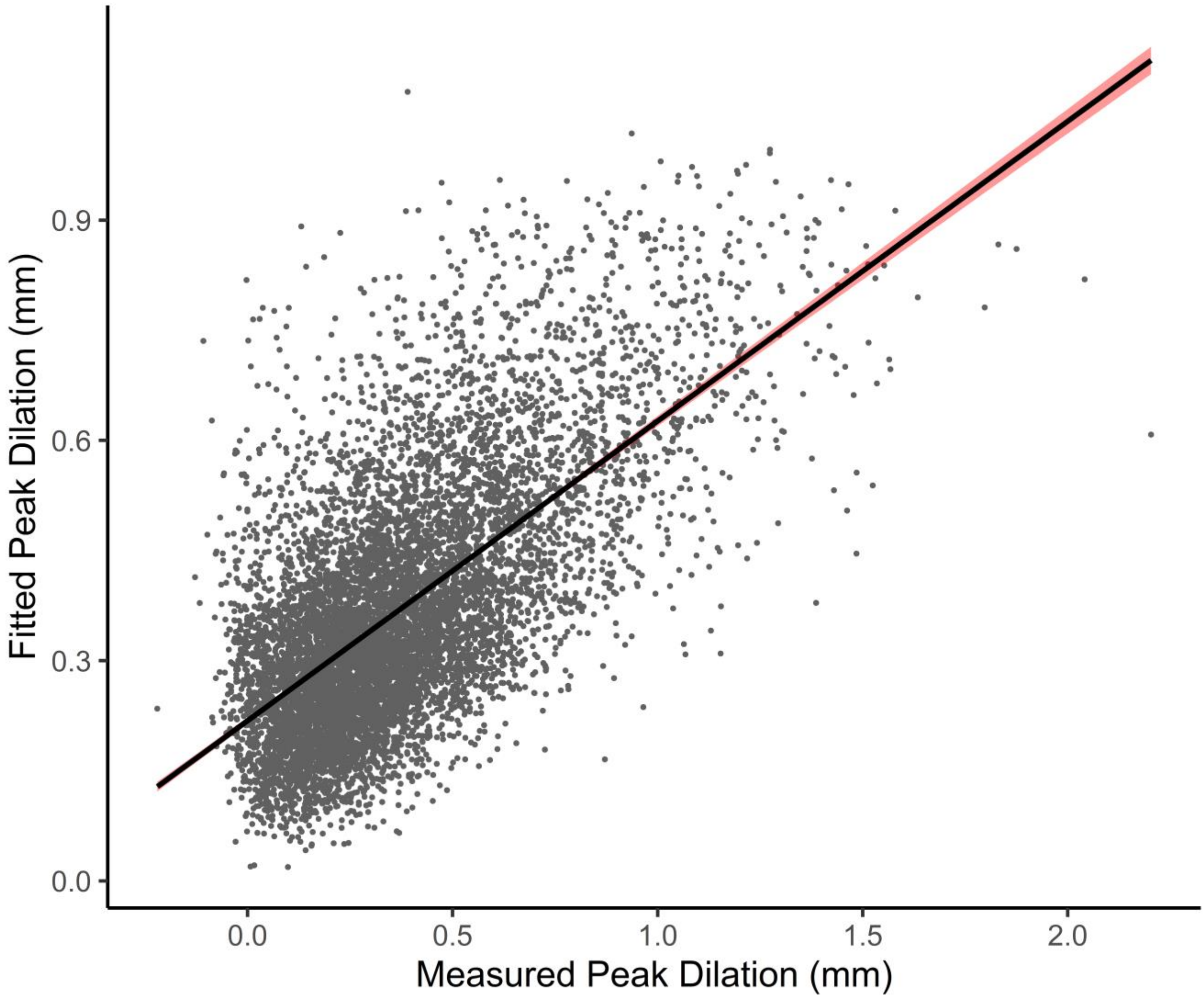
Peak Dilation by Time-On-Task (Trial Number) for Each Participant



Note. This figure shows peak dilation as a function of time-on-task (trial number) for each participant. These data are the raw measurements of peak dilation. Each coloured data point shows peak dilation for a single trial and the coloured line shows the trend for each participant. The black line shows average peak dilation across all participants versus trial number.

Figure 41

Scatter Plot of Fitted Peak Dilation Values Versus Measured Peak Dilation Values



Note. This figure shows a scatter plot of the peak dilation values fitted by the Trial-Level Peak Dilation and Time-On-Task MEM (y axis) versus the measured peak dilation values (x axis). The trendline and confidence interval (in red) represents the linear relationship between all fitted and measured peak dilation values.

R^2 values were computed using the *performance* R package (Lüdecke et al., 2021). The included fixed effects accounted for 17% of the variance in the model. The fixed effects and random effects combined accounted for 41% of the variance in the model.

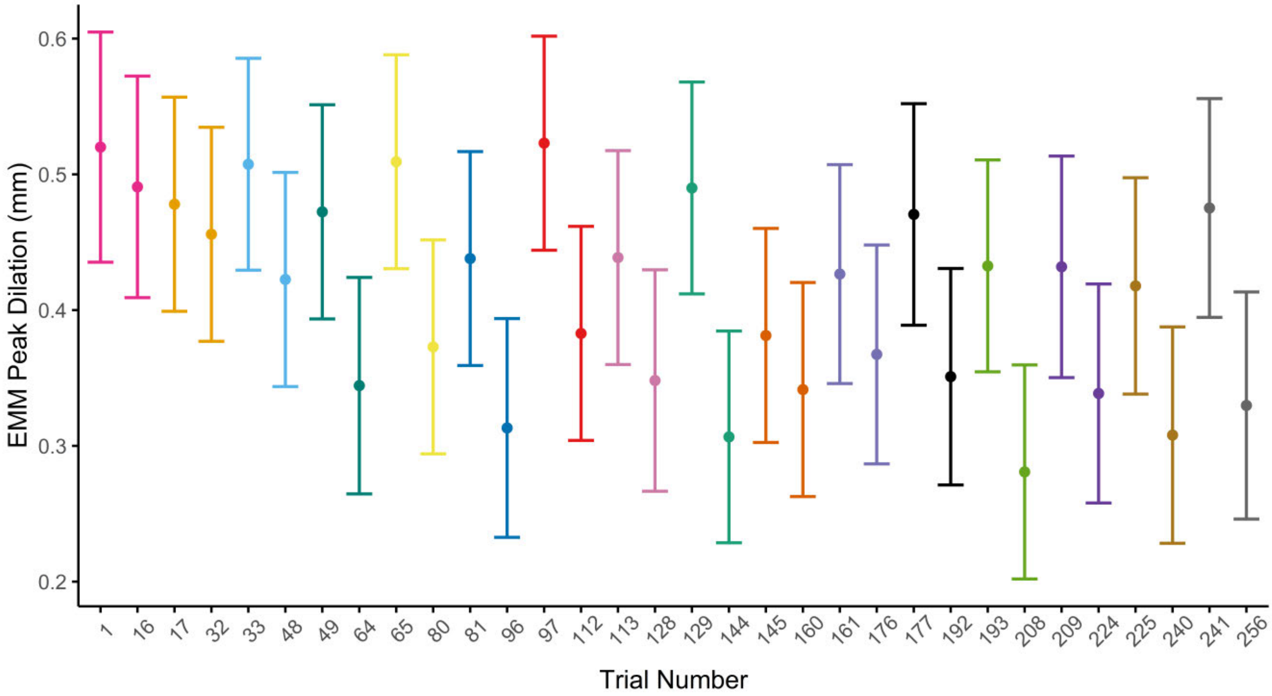
14.4.1.4 Trial-Level Peak Dilation and Time-On-Task MEM: Post Hoc Analyses

Estimated marginal means (EMMs) were derived from the *emmeans* R package (Lenth, 2021) for the first and last trial of each condition (*n* = 16). For example, averaged over light levels and SNR, EMMs were compared for trial 1 and trial 16, trial 17 and trial 32, trial 33 and trial 48, and so forth. This was done to examine if a time-on-task effect on peak dilation occurred at the trial-level within each condition. Within-condition time-on-task effects may be most important for practical purposes. Furthermore, the time-on-task effect may be clearer when examined in this way as light level and SNR do not vary for a given condition. The order of light level and SNR presentation varied between participants (see Appendix 6) which may explain the wide confidence intervals depicted in Figure 42.

Figure 42 shows the EMMs comparisons for peak dilation. The colours indicate a specific comparison of peak dilation between the first and last trial of each condition. In all comparisons, the EMM for the last trial in the condition is smaller than the first trial of the condition, regardless of the light level or SNR condition, or position in the test session (i.e., 1-256). Despite this, the overlap in the confidence intervals for between comparisons indicates that most of these differences may not be significant.

Figure 42

Peak Dilation EMMs for the First and Last Trial of Each Condition



Note. This figure shows EMMs and 95% confidence intervals for peak dilation averaged over light levels and SNR levels extracted from the final model (14.4.1.3). The different colours indicate the comparisons of interest (i.e., the first trial of a condition and the last trial of a condition across the 16 light level x SNR conditions).

Pairwise comparisons testing the difference between means for each trial comparison are presented in Table 43. Peak dilation was significantly smaller in trial 144 compared to 129 and in trial 208 compared to trial 193.

Table 43. Post Hoc Pairwise Tests on the Difference Between Peak Dilation EMMs for Each Trial Comparison

Trial Comparison	Condition number	Mean difference (mm)	SE	df	Pairwise *t* statistic	*p*-value	*Sig.*
1 - 16	1	0.03	0.05	8151.89	0.556	.58	
17 - 32	2	0.02	0.05	7914.17	0.451	.65	
33 - 48	3	0.08	0.05	7938.26	1.740	.08	
49 - 64	4	0.13	0.05	7966.49	2.585	.01	
65 - 80	5	0.14	0.05	7949.99	2.775	.006	
81 - 96	6	0.12	0.05	7998.31	2.502	.01	
97 - 112	7	0.14	0.05	7984.49	2.853	.004	
113 - 128	8	0.09	0.05	8037.81	1.801	.07	
129 - 144	9	0.18	0.05	7922.92	3.785	<.001	*
145 - 160	10	0.04	0.05	7972.09	0.811	.42	
161 - 176	11	0.06	0.05	8059.86	1.170	.24	
177 - 192	12	0.12	0.05	7968.32	2.360	.02	
193 - 208	13	0.15	0.05	7888.63	3.109	.002	*
209 - 224	14	0.09	0.05	8037.34	1.830	.07	
225 - 240	15	0.11	0.05	7966.18	2.204	.03	
241 - 256	16	0.15	0.05	8101.96	2.806	.005	

Note. *p* is not Bonferroni corrected. * in *Sig.* column indicates significant differences after correcting for multiple comparisons (e.g., *p* is significant at 0.003).

14.4.2 Trial-level Mean Dilation and Time-On-Task

14.4.2.1 Model Assumptions

Inspection of residual plots indicated linearity and homoscedasticity. QQ plots indicated slight non-normality of residuals. However, MEMs are typically robust to this type of violation (Gelman & Hill, 2007; Knief & Forstmeier, 2020; Schielzeth et al., 2020).

14.4.2.2 Model Convergence

All *lmer* models for mean dilation converged.

14.4.2.3 Trial-level Mean Dilation and Time-On-Task MEM: Final Model Results

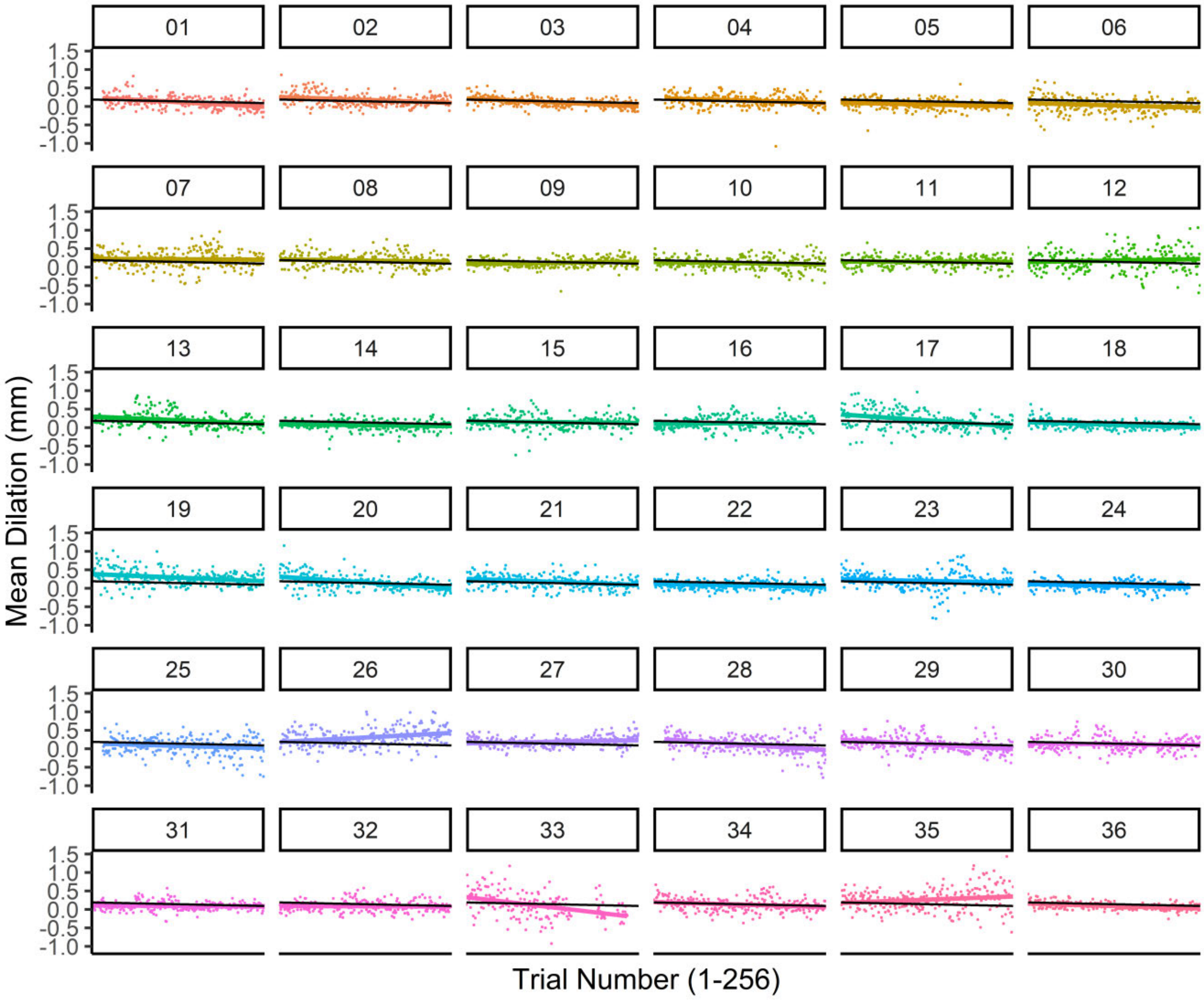
The final model indicated that, when averaged over levels of light level and SNR, trial-level mean dilation was typically smaller when compared to the base level (Trial 1) over the task (i.e., trial 2 – 256) (see Appendix 8 for the model output).

The *anova* function from the *lmerTest* R package (version 3.1-3) (Kuznetsova et al., 2017) was used to compute scaled F-values and denominator degrees of freedom using Satterthwaites' method. Trial number had a significant effect on mean dilation, $F(255, 3195.3) = 2.04$, $p < .001$.

Figure 43 shows the raw mean dilation data for each participant across all trials. Even with the variance associated with light level, SNR and their interaction existing in the data, most participants showed a slight decrease in mean dilation over the course of the experiment. Figure 44 shows the linear trend line and confidence interval (in red) in a scatter plot of the mean dilation values estimated in the MEM and the measured mean dilation values.

Figure 43

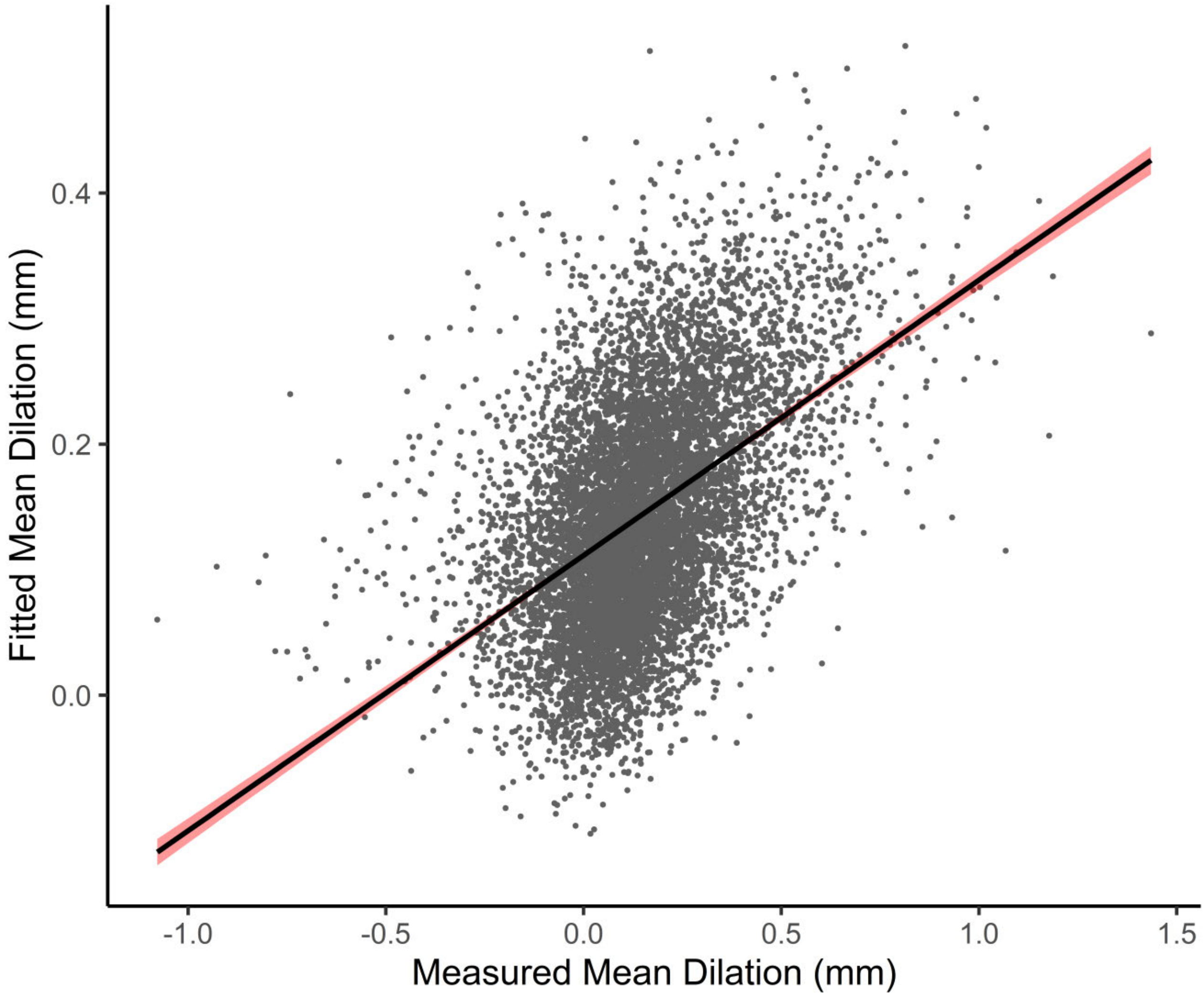
Mean Dilation by Time-On-Task (Trial Number) for Each Participant



Note. This figure shows mean dilation as a function of time-on-task (trial number) for each participant. These data are the raw measurements of mean dilation. Each coloured data point shows mean dilation for a single trial and the coloured line shows the trend for each participant. The black line shows average mean dilation across all participants versus trial number.

Figure 44

Scatter Plot of Fitted Mean Dilation Values Versus Measured Mean Dilation Values



Note. This figure shows a scatter plot of the mean dilation values fitted by the Trial-Level Mean Dilation and Time-On-Task MEM (y axis) versus the measured mean dilation values (x axis). The trendline and confidence interval (in red) represents the linear relationship between all fitted and measured mean dilation values.

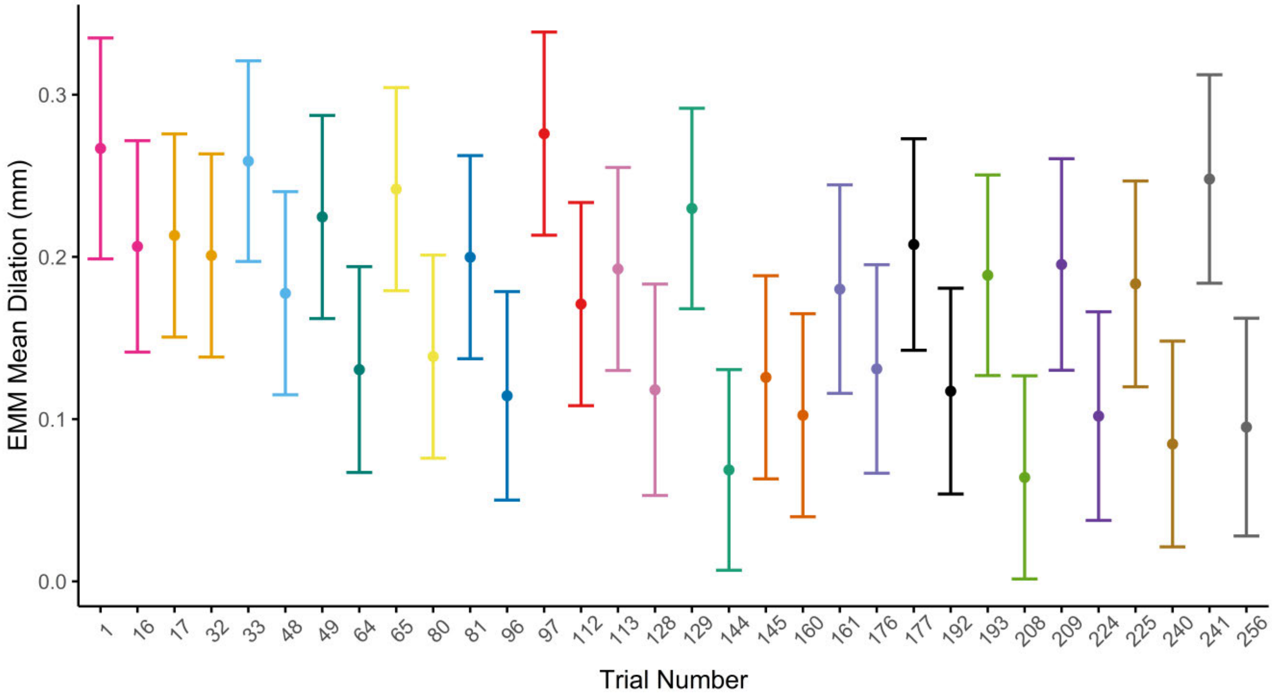
R^2 values were computed using the *performance* R package (Lüdecke et al., 2021). The fixed effects included accounted for 12% of the variance in the model. The fixed effects and random effects combined accounted for 22% of the variance in the model.

14.4.2.4 Trial-Level Mean Dilation and Time-On-Task MEM: Post Hoc Analyses

Figure 45 shows the EMMs comparisons for mean dilation. The colours indicate a specific comparison between the first and last trial of each condition. In all comparisons, the EMM for the last trial in the condition is smaller than the first trial of the condition, regardless of the light level or SNR condition, or the position in the test session (i.e., 1-256). Despite this, the overlap in the confidence intervals between comparisons indicates that most of these differences may not be significant.

Figure 45

Mean Dilation EMMs for the First and Last Trial of Each Condition



Note. This figure shows EMMs and 95% confidence intervals for mean dilation averaged over light levels and SNR levels extracted from the final model (14.4.2.3). The different colours indicate the comparisons of interest (i.e., the first trial of a condition and the last trial of a condition across the 16 light level x SNR conditions).

Pairwise comparisons testing the difference between means for each trial comparison are presented in Table 44. Mean dilation was significantly smaller in trial 144 compared to 129, in trial 208 compared to trial 193, and in trial 256 compared to trial 241.

Table 44. Post Hoc Pairwise Tests on the Difference Between Mean Dilation EMMs for Each Trial Comparison

Trial comparison	Condition number	Mean difference (mm)	SE	df	Pairwise *t* statistic	*p*-value	*Sig.*
1 - 16	1	0.06	0.05	8191.90	1.32	.19	
17 - 32	2	0.01	0.04	7936.52	0.29	.77	
33 - 48	3	0.08	0.04	7955.92	1.92	.05	
49 - 64	4	0.09	0.04	7987.42	2.19	.03	
65 - 80	5	0.10	0.04	7972.25	2.42	.02	
81 - 96	6	0.09	0.04	8022.60	1.97	.05	
97 - 112	7	0.11	0.04	8006.62	2.46	.01	
113 - 128	8	0.07	0.04	8060.02	1.71	.09	
129 - 144	9	0.16	0.04	7940.10	3.83	<.001	*
145 - 160	10	0.02	0.04	7989.99	0.55	.58	
161 - 176	11	0.05	0.04	8088.20	1.12	.26	
177 - 192	12	0.09	0.04	7998.26	2.06	.04	
193 - 208	13	0.12	0.04	7905.97	2.94	.003	*
209 - 224	14	0.09	0.04	8069.47	2.11	.03	
225 - 240	15	0.10	0.04	7986.99	2.28	.02	
241 - 256	16	0.15	0.05	8137.53	3.40	.001	*

Note. p is not Bonferroni corrected. * in *Sig.* column indicates significant differences after correcting for multiple comparisons (e.g., *p* is significant at 0.003).

14.4.3 Trial-Level Baseline Diameter and Time-On-Task

14.4.3.1 Model Assumptions

Inspection of residual plots indicated linearity and homoscedasticity. QQ plots indicated slight non-normality of residuals. However, MEMs are typically robust to this type of violation (Gelman & Hill, 2007; Knief & Forstmeier, 2020).

14.4.3.2 Model Convergence

The final baseline *lmer* model did not converge on the first run. To address this, the control parameters were adjusted with an optimiser as suggested by Brown (2021). The "all_fit" function from the *afex* R package (Singmann et al., 2021) was used to assess which optimiser was suitable. The "bobyqa" optimiser led to model convergence.

14.4.3.3 Trial-Level Baseline Diameter and Time-On-Task: Final Model Results

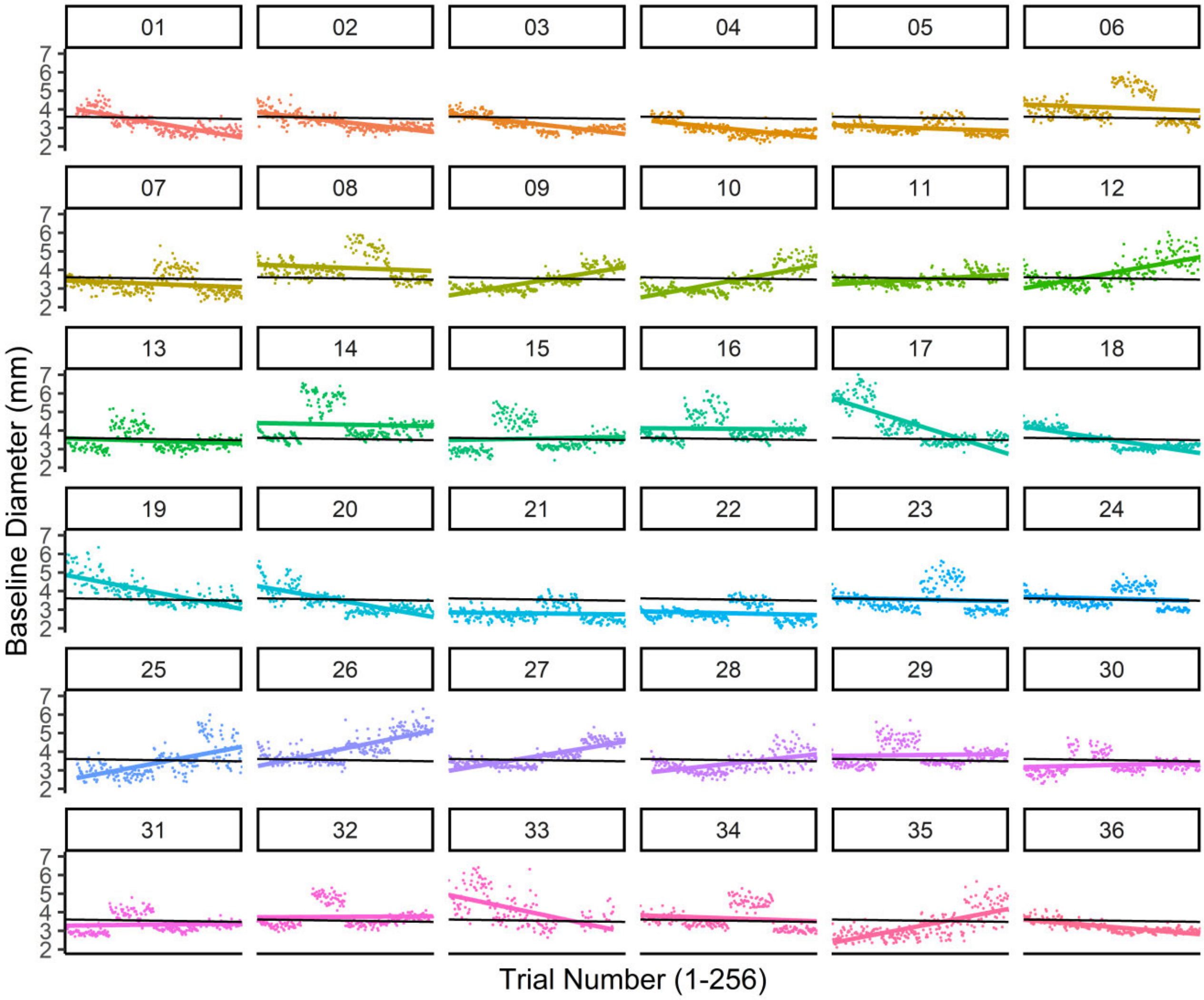
The final model indicated that, when averaged over levels of light level and SNR, trial-level baseline diameter was typically smaller when compared to the base level (Trial 1) over the task (i.e., trial 2 -- 256) (see Appendix 9 for the model output).

The *anova* function from the *lmerTest* R package (version 3.1-3) (Kuznetsova et al., 2017) was used to compute scaled F-values and denominator degrees of freedom using Satterthwaites' method. Trial number had a significant effect on baseline diameter, $F(255, 1947.6) = 7.31$, $p < .001$.

Figure 46 shows the raw baseline diameter data for each participant across all trials. The clustering of data for individual participants can be attributed to the large effect of light level on baseline diameter. Figure 47 shows the linear trend line and confidence interval (in red) in a scatter plot of the mean dilation values estimated in the MEM and the measured mean dilation values.

Figure 46

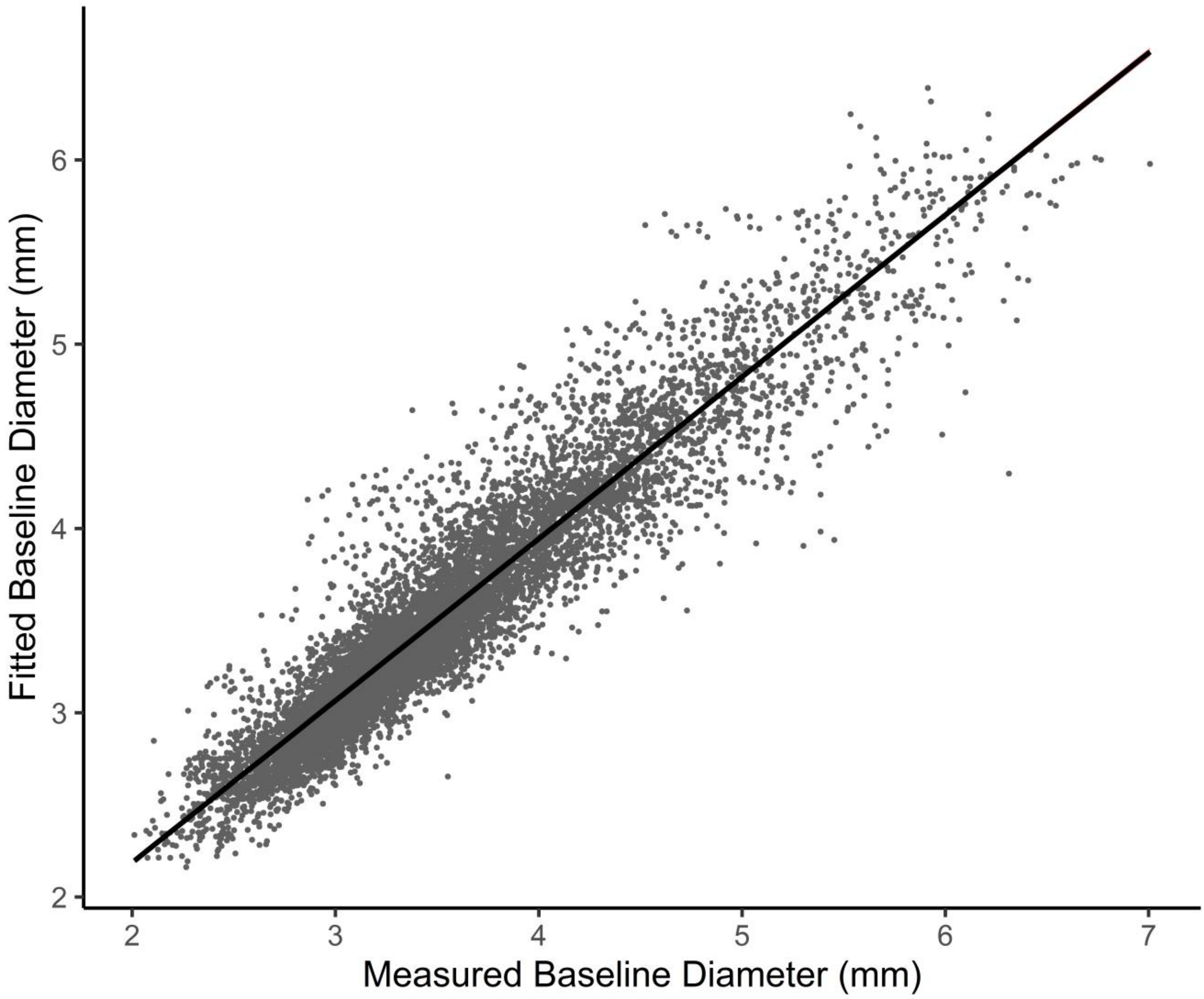
Baseline Diameter by Time-On-Task (Trial Number) for Each Participant



Note. This figure shows baseline diameter as a function of time-on-task (trial number) for each participant. These data are the raw measurements of baseline diameter. Each coloured data point shows baseline diameter for a single trial and the coloured line shows the trend for each participant. The black line shows average baseline diameter across all participants versus trial number.

Figure 47

Scatter Plot of Fitted Baseline Diameter Values Versus Measured Baseline Diameter Values



Note. This figure shows a scatter plot of the baseline diameter values fitted by the Trial-Level Baseline Diameter and Time-On-Task MEM (y axis) versus the measured baseline diameter values (x axis). The trendline and confidence interval (in red) represents the linear relationship between all fitted and measured baseline diameter values.

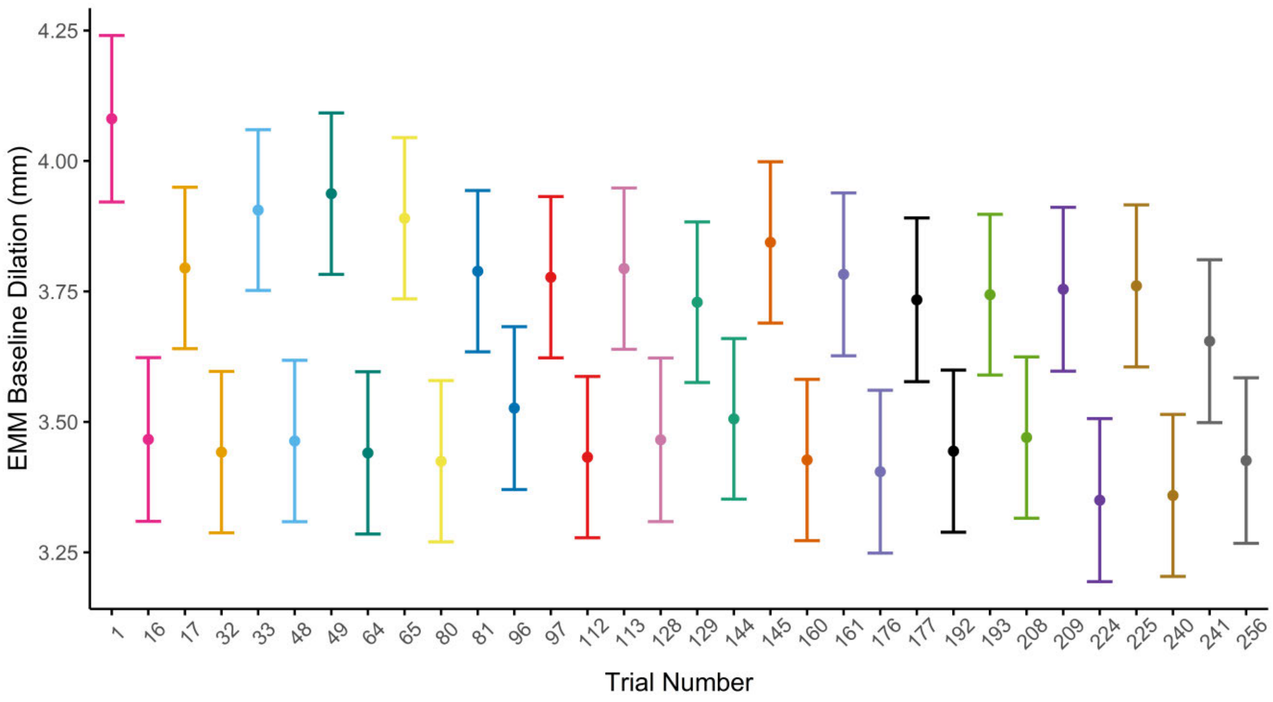
R^2 values were computed using the *performance* R package (Lüdecke et al., 2021). The fixed effects included accounted for 54.3% of the variance in the model. The fixed effects and random effects combined accounted for 87.8% of the variance in the model. Because there was less variance left unexplained in the baseline diameter model, it was able to estimate the measured baseline diameter data with more precision (Figure 47).

14.4.3.4 Trial-Level Baseline Diameter and Time-On-Task MEM: Post Hoc Analyses

Figure 48 shows the EMMs comparisons for baseline diameter. The colours indicate a specific comparison between the first and last trial of each condition. In all comparisons, the baseline diameter EMM for the last trial in the condition was smaller than the first trial of the condition, regardless of the light level or SNR condition, or position in the test session (i.e., 1-256). There is less overlap in the confidence intervals for EMM comparisons than in the peak and mean dilation models. The within-condition time-on-task effect is more pronounced for baseline diameter.

Figure 48

Baseline Diameter EMMs for the First and Last Trial of Each Condition



Note. This figure shows EMMs and 95% confidence intervals for baseline diameter averaged over light levels and SNR levels extracted from the final model (14.4.3.3). The different colours indicate the comparisons of interest (i.e., the first trial of a condition and the last trial of a condition across the 16 light level x SNR conditions).

Pairwise comparisons testing the difference between means for each trial comparison are presented in Table 45. Baseline diameter was significantly smaller in last trial of every condition, when compared to the first trial.

Table 45. Post Hoc Pairwise Tests on the Difference Between Baseline Diameter EMMs for Each Trial Comparison

Trial comparison	Condition number	Mean difference (mm)	SE	df	Pairwise *t* statistic	*p*-value	*Sig.*
1 - 16	1	0.61	0.07	8500.70	9.30	<.001	*
17 - 32	2	0.35	0.06	8495.00	5.73	<.001	*
33 - 48	3	0.44	0.06	8497.11	7.23	<.001	*
49 - 64	4	0.50	0.06	8500.89	8.00	<.001	*
65 - 80	5	0.47	0.06	8495.00	7.56	<.001	*
81 - 96	6	0.26	0.06	8498.49	4.20	<.001	*
97 - 112	7	0.34	0.06	8495.00	5.60	<.001	*
113 - 128	8	0.33	0.06	8504.01	5.20	<.001	*
129 - 144	9	0.22	0.06	8495.00	3.68	<.001	*
145 - 160	10	0.42	0.06	8500.66	6.77	<.001	*
161 - 176	11	0.38	0.06	8499.56	5.96	<.001	*
177 - 192	12	0.29	0.06	8498.86	4.57	<.001	*
193 - 208	13	0.27	0.06	8497.96	4.48	<.001	*
209 - 224	14	0.40	0.06	8496.58	6.32	<.001	*
225 - 240	15	0.40	0.06	8503.12	6.42	<.001	*
241 - 256	16	0.23	0.07	8500.90	3.52	<.001	*

Note. p is not Bonferroni corrected. * in *Sig.* column indicates significant differences after correcting for multiple comparisons (e.g., *p* is significant at 0.003).

14.5 Discussion

14.5.1 Results Summary

The main aim of this chapter was to assess time-on-task effects in trial-level peak dilation, mean dilation, and baseline diameter measures. Averaged over light levels and SNR, a main effect of trial number was found for all pupil parameters tested. Therefore, the current results indicated that trial number did affect trial-level pupil parameters in a speech-in-noise task lasting approximately 1 hr.

Peak dilation, mean dilation, and baseline diameter were typically smaller in other trials compared to trial 1. This may indicate a time-on-task effect across the test session. However, given that the exact position of significant differences in trials across the test session could not be identified, these results should be interpreted with caution. The presence of a time-on-task effect across the whole one-hour test session could not be identified conclusively.

EMMs showed that peak dilation, mean dilation, and baseline diameter were typically larger in the first trial of each condition, than in the last trial of each condition across the test session. However, post hoc analyses revealed that only two of these comparisons were significantly different for peak and mean dilation. Conversely, all comparisons of the first and last trial of each condition were significant for baseline diameter.

These results provide an indication that time-on-task effects may exist within-conditions in trial-level pupil parameters measured in the current paradigm and that baseline diameter may be more sensitive to these effects. These results are discussed in more detail below.

14.5.2 Trial-Level Peak and Mean Dilation and Time-On-Task

The models that included trial number as a fixed effect were able to account for a larger proportion of the total variance than the models reported in Chapter 11. Regarding peak dilation, when trial number was included as a fixed effect, the fixed effects accounted for 17% of the total variance in the model. However, when trial number was included as a random effect (and not as a fixed effect), the fixed effects only accounted for 12% (Section 11.4.1). Overall, for peak dilation, the proportion of total variance explained by the combination of fixed and random effects in the model increased when trial number was included as a fixed effect (41%), rather than as a random effect (39%) (Section 11.4.1).

The findings were similar with respect to mean dilation. The inclusion of trial number as a fixed effect increased the proportion of total variance explained by the fixed effects from 6% in the previous model (Section 11.4.2), to 12%. Overall, the proportion of total variance explained by the combination of fixed and random effects in the mean dilation model increased when trial number was included as a fixed effect (22%), rather than as a random effect (19%) (Section 11.4.2).

The additional proportion of total variance explained by the models when trial number was included as a fixed effect in peak and mean dilation models (2% and 3%, respectively) was small and may not be of practical significance. However, the inclusion of trial number as a fixed effect rather than a random effect meant that the effects of trial number on TEPRs could be examined in more detail.

Averaged over light levels and SNR levels, peak and mean dilation were typically smaller in subsequent trials when compared to trial 1 (Appendix 7 and Appendix 8). These findings align somewhat with previous research where relationships between smaller peak dilation and time-on-task were identified (Hopstaken et al., 2015b; McGarrigle et al., 2017b; McGarrigle, Knight, et al., 2021; McGarrigle, Rakusen, et al., 2021). McGarrigle, Knight, et al. (2021) recently reported similar effects in older and younger listeners but averaged trial-level TEPRs across participant groups. The findings reported here extend this by demonstrating that a relationship between TEPRs and time-on-task might also exist in un-averaged, trial-level pupil data. This has not previously been examined in a speech-in-noise task.

The decrease in trial-level peak and mean dilation may be related to decreased physiological arousal related to time-on-task and potentially, fatigue. However, the overall main effect of trial number means that significant differences between trials may have occurred between any of the trial numbers, hence, at any point throughout the test session. Therefore, it cannot be concluded that these findings represent an effect of time-on-task over the whole 1 hr speech-in-noise test session. The presence of this effect needs to be confirmed in more detailed follow up studies.

EMMs for peak dilation (Figure 42) and mean dilation (Figure 45) measured in the first trial of each condition were consistently larger than the last trial of the condition (*n* = 16) regardless of where that trial occurred across the total experiment. However, there was a substantial overlap in the 95% confidence intervals of the EMM comparisons. Post hoc analyses revealed that the differences between EMMs in trial comparisons for both peak and mean dilation were only significant for two comparisons which occurred in the latter half of the test session (Table 43 and Table 44).

Trial number improved the model fit, in that the proportion of total variance explained in the model increased. Therefore, task duration may be an important consideration in the use of pupillometry in research and for potential clinical use. Because pupil responses may decrease in size, with increasing time-on-task, care must be taken when attributing smaller pupil responses entirely to effort reduction. Due to the potential impact of fatigue effects on peak and mean dilation, this relationship should be examined in more detail.

The large confidence intervals with respect to the peak and mean dilation EMMs (Figure 42, Figure 45, respectively) likely reflect the presence of light level and SNR effects in the data. Because levels of these variables were counterbalanced between trials, conditions, and participants, each EMM was averaged across all participants and includes all levels of light level and SNR. Therefore, it would also be beneficial to examine time-on-task effects at trial-level without light level and SNR effects in the data as these may obscure an underlying effect of time-on-task within-conditions and across the whole test session.

14.5.3 Trial-Level Baseline Diameter and Time-On-Task

There was also a significant effect of trial number on baseline diameter. Averaged across light levels and SNR levels, baseline diameter was typically smaller in other trials when compared to trial 1 (Appendix 9). This effect was more consistent in baseline diameter than peak and mean dilation. This finding supports earlier findings by Hopstaken et al. (2015a) who found that baseline diameter decreased as a function of time block in a visual paradigm, and was also associated with increased subjective fatigue and decreased task-engagement. Additionally, these results support earlier findings by Zekveld et al. (2010) who demonstrated that baseline diameter decreased over the course of the first and last speech-in-noise task in their experimental paradigm. The current findings extend these previous findings and provide evidence for a time-on-task effect in baseline diameter at the trial-level in a speech-in-noise task.

The final model explained a large proportion of the total variance in baseline diameter (87.8%). The proportion of total variance explained by the fixed effects increased by 3% when trial number was included in the model, compared to a model without trial number. Therefore, the proportion of total variance in baseline diameter explained by the inclusion of trial number was small and may not be of practical significance. Despite this, it was still an important factor to include in the model due to the improvement in the model fit.

The effect of trial number on baseline diameter may have been masked by the larger effects of light level. This is evidenced by the R^2 values which indicated that the majority of variance in baseline diameter can be attributed to light level and individual differences. R^2 values only increased marginally with the inclusion of trial number as a fixed effect. In Figure 46, the participants who displayed increasing baseline diameter (in the raw data) over the test session typically experienced the dimmer light levels at the end of the experiment. Like peak and mean dilation reported above, the large confidence intervals for the EMMs (Figure 48) likely reflect the presence of light level and SNR effects in the data. Therefore, baseline diameter should be examined in a paradigm where light level is not manipulated. Such a study could use the same experimental stimuli as the current study, but without varying the light level and SNR. This would show how TEPRs are affected by time-on-task during a speech-in-noise task while light level and SNR are held constant.

As with the peak and mean dilation results reported above, the overall main effect of trial number on baseline diameter means that significant differences between trials may have occurred between any of the trial numbers, hence, at any point throughout the test session. Therefore, it cannot be categorically concluded that these findings represent an effect of time-on-task over the whole test session.

EMMs regarding baseline diameter (Figure 48) measured in the first trial of a condition were consistently larger compared to the last trial of a condition. This effect was consistent regardless of where the condition occurred across the total experiment. Therefore, time-on-task may affect baseline diameter after 16 trials of a speech-in-noise task.

Compared to peak and mean dilation, there was substantially less overlap in the 95% confidence intervals for baseline diameter (Figure 48). Furthermore, the post hoc analyses revealed that differences between EMMs in trial comparisons for baseline diameter were all significant. This may indicate that baseline diameter is more sensitive to within-condition time-on-task effects than peak and mean dilation.

This suggestion is supported by Alhanbali et al. (2020), if the time-on-task effects reported here were due to fatigue. Alhanbali et al. (2020) reported that larger baseline diameters were associated with less tiredness and better performance, but peak dilation was not. This also suggests that baseline diameter may be a more sensitive measure of time-on-task effects associated with fatigue.

The EMMs depicted in Figure 48 provide evidence that the delay between conditions was sufficient to allow baseline diameter (and maybe the participant) to largely recover from time-on-task effects and may also suggest that the delay between trials may have been too short, hence, the time-on-task effect. The delay between trials and conditions were not tightly controlled. However, after completion of a trial (i.e., after the response), subsequent trials were delayed by approximately 3 s. The delay between conditions was approximately 12 – 14 seconds. This was due to the time needed to set up and initiate the next condition in the counterbalanced sequence (Appendix 6) and the additional time allocated for accommodation to a new light level. Therefore, baseline diameter was able to recover to a greater degree between-conditions than within-conditions. Although not statistically tested, Figure 48 also shows that, on average, baseline diameter may also have decreased over the test session. The significance of this effect should be examined in future studies.

Time-on-task effects, both within-condition and across multiple conditions, may be larger for individuals with hearing loss as they may be more susceptible to fatigue (Alhanbali et al., 2017), especially when adequate rest time between trials is not provided. Winn et al. (2018) noted the importance of allowing enough time between trials for the pupil to return to baseline. The results reported here may indicate that time intervals between trials were sufficient to ensure that peak dilation responses did not affect baseline diameter measurements on subsequent trials but may have been insufficient to allow adequate recovery between trials. If peak dilation responses were affecting baseline diameter values in subsequent trials, baseline diameter values would likely have been larger than the previous trial. The decline in baseline diameter after 16 trials of the speech-in-noise task may indicate fatigue. It would be beneficial to examine the effects of the duration of between trial delay, as longer and more controlled delays between trials may alleviate some of the time-on-task effects reported here.

The difference between the first and last trial of each condition in baseline diameter may be consistent with reduced physiological arousal which participants were not able to adequately recover from, within-conditions. This may be consistent with a fatigue effect. Alhanbali et al. (2020) suggested that baseline diameter may have potential as an objective, physiological measure of fatigue during listening tasks. This suggestion is addressed in section (14.5.6) below.

It is possible that the decline in baseline diameter reflects a progressive reduction in an anticipatory response rather than a fatigue response. In line with findings reported here (decrease in baseline diameter) and in Chapter 13 (performance improvement), Ayasse and Wingfield (2020) found that baseline diameter declined over trials in a sentence comprehension task, while performance increased. Ayasse and Wingfield (2020) also reported that there was a steeper rate of decline in baseline diameter for participants with greater hearing difficulty compared to those with less hearing difficulty. They reasoned that this was because participants with greater hearing difficulty had larger baseline diameters at the beginning of the experimental session. This could be due to anticipatory arousal that accompanies performing speech perception tasks for this population. The decline in baseline diameter as a function of trial in the present study could have been due to a progressive decrease in this arousal as the test session unfolded and as performance improved. It is possible that this effect could partly explain the current results. Because we examined the within-condition decline in baseline diameter (Trial 1 compared to Trial 16), it may be of interest to examine the differences between the first trial of each condition to further assess differences in baseline diameter across the whole test session.

More research is needed to clarify the existence and the specific patterns of the time-on-task and/or fatigue effect in baseline diameter and this should be done without varying light level. Due to the possible relationship between baseline diameter and peak and mean dilation reported in Chapter 12, it would also be of interest to examine if/how fatigue effects this relationship.

14.5.4 Comparison with Subjective and Performance data

The above findings indicate that individuals might have experienced a small reduction in physiological arousal within each condition of the speech-in-noise task. There was also marginal evidence that this may also occur over a speech-in-noise task of 1 hr duration. These results may be consistent with a time-on-task effect due to fatigue (Hopstaken et al., 2015a, 2015b; McGarrigle et al., 2017b). However, as noted earlier, this cannot be confirmed conclusively based on the current results.

Nevertheless, the suggestion that the time-on-task effects in pupil parameters were due to fatigue was supported by the subjective reports of increased fatigue and decreased task-engagement reported by the participants (see Chapter 13). However, even though the participants may have been experiencing subjective and physiological fatigue, this was not apparent in their average performance which improved across time blocks. This may be due to a practice effect.

Another aspect of the practice effect might be that, even in the face of some fatigue, participants were sufficiently motivated to do well in the task. Therefore, while participants might have experienced a degree of subjective and physiological fatigue, their performance was not compromised.

Based on the FUEL, Alhanbali et al. (2020) noted three reasons that baseline diameter and pupil responses may decrease in an effortful listening task: (1) decreased listening demands, (2) a lack of task engagement due to poor motivation, or (3) the development of fatigue (p. 2). The listening demands of the speech-in-noise task used in the current study were consistent between participants and condition order was counterbalanced to avoid order effects. Thus, it is unlikely that decreased listening demands led to the results reported here. Due to the average improvement in performance across the test session, it was also unlikely that participants disengaged from the task due to poor motivation. Therefore, it was possible that the results demonstrated slightly reduced physiological arousal within each condition and (potentially) across the test session. This might be consistent with fatigue, even though it did not compromise performance. Participants may have been sufficiently motivated to improve their performance, even though they may have been experiencing fatigue.

Because subjective fatigue and task-engagement, as well as performance were measured across the whole test session, it is difficult to ascertain consistency between the findings of Chapter 13 and the decrease in TEPRs within each condition. Future research could examine the within-condition effects in more detail and directly compare this to the performance on individual trials to produce a more detailed description of these effects.

14.5.5 An Alternative Explanation

It is possible that the decrease in TEPRs within each condition and from trial 1 over the course of the test session reflects a systematic decrease in effort, that is, as participants became more familiar with the task, they had to expend less effort to succeed. This possibility is supported by the finding that, on average, performance in the speech-in-noise task improved as a function of time block (Chapter 13). Foroughi et al. (2017) and Sibley et al. (2019) reported evidence that maximum pupil size and peak dilation (respectively) decreased as a function of trial number in spatial orientation tasks. In both studies, overall performance also improved. Foroughi et al. (2017) and Sibley et al. (2019) exclusively interpreted this finding as a reflection of reduced effort due to task learning, rather than fatigue.

However, if the results reported here could be explained by a practice/learning effect that led to less effort expenditure within each condition and/or over the course of the test session, there should have been a more pronounced effect of time-on-task on peak and mean dilation when compared to baseline diameter. Peak and mean dilation are task-relevant measures of effort. Baseline diameter values are more reflective of an individual’s tonic arousal levels (Unsworth & Robison, 2018) and were measured before stimulus presentation in the current study. Therefore, if a learning effect that resulted in less effort required to maintain/improve performance was present in the current data, the decrease in TEPRs should have been more prominent in peak and mean dilation measurements.

In the current study, baseline diameter was typically smaller in other trials compared to trial 1. The time-on-task effect was more pronounced within each condition for baseline diameter than for peak and mean dilation. This may indicate that tonic arousal levels decreased within each condition and over the course of the test session. Alhanbali et al. (2020) also found an association between baseline diameter and subjective fatigue, but no association between peak dilation and subjective fatigue. Therefore, despite by-block performance improvements over the test session, the results of the current study reflect a time-on-task effect within each condition and over the test session that is more consistent with decreased physiological arousal and fatigue, than less effort expenditure related to a learning effect.

14.5.6 Implications

The above findings have implications for the measurement of listening effort in listening tasks. Firstly, the LC norepinephrine system may be involved in fatigue during task performance. Secondly, evidence that there is systematic variation in trial-level pupil data within-conditions was provided.

Adaptive Gain Theory is perhaps the most accepted theory of the LC norepinephrine system (Aston-Jones & Cohen, 2005). Based on this theory, it has been proposed that the LC norepinephrine system regulates task-engagement by increasing or decreasing norepinephrine levels and this may be reflected in pupil dynamics (Gilzenrat et al., 2010; Hopstaken et al., 2015a, 2015b; Jepma & Nieuwenhuis, 2011).

Examination of LC norepinephrine system functioning modes typically involve comparing the tonic (larger baseline and smaller peak dilation) and phasic modes (smaller baseline and larger peak dilation). Hopstaken (2016) reported the combined findings from their previous work (Hopstaken et al., 2015a, 2015b) and suggested that there was another mode of LC norepinephrine system functioning that has not typically been addressed in studies using Adaptive Gain Theory, namely a “disengaged mode” (p. 121). In this mode, with increasing fatigue, they posit that there is a decrease in baseline and stimulus-evoked levels of norepinephrine, and that this is reflected in pupil dynamics (smaller baseline diameter and smaller peak dilation).

This mode of functioning may be reflected in the current data and may be related to time-on-task. For example, peak dilation, mean dilation, and baseline diameter all typically decreased from trial 1 across the test session and within-condition for baseline diameter. However, because average performance increased as a function of time block, it may not be accurate to assume that participants became “disengaged”, despite their subjective reports. The subjective reports indicating that participants became less engaged in the task over time may reflect a response bias where they assumed that they should become less engaged in the task because they were being explicitly asked about it.

As mentioned in Section 13.6.4, the aggregation of performance over time block may have concealed more nuanced information about how fatigue manifested in performance over the course of the test session between SNR conditions. Future research could include studies aimed at examining whether individuals allocate more effort to the task in easier listening conditions and withdraw effort in harder listening conditions when fatigued, as shown in TEPRs, performance measures, and subjective measures.

The second implication relates to possible systematic variation in trial-level pupil data. The results of the trial-level modelling reported in Chapter 11 indicated that the variance associated with trial number could be accounted for and partitioned from the error term when included as a random effect. The results reported in this chapter build on that finding, by modelling trial number as a fixed effect. This enabled a more detailed examination of how trial number affected trial-level pupil parameters.

Evidence of systematic variation in trial-level pupil data that is related to time-on-task, and potentially, fatigue over the test session and within each condition was provided here. As such, there are implications for the practice of signal-averaging pupil traces. However, due to the congruency of results reported in Chapter 9 (signal-averaged data) and Chapter 11 (trial-level data), it doesn’t appear that signal-averaging pupil traces led to distorted effect estimates. However, further research examining systematic variation in trial-level pupil data should be carried out to ensure that signal-averaging does not distort effect estimates in different types of tasks and under different conditions. Additionally, trial-level peak dilation latency was not examined in this study. This should be examined in future studies.

The findings reported here highlight that characterisation and consideration of how fatigue manifests in pupil responses are necessary for future research and clinical applications of pupillometry. Examination of how fatigue develops based on task type and length is also needed, if pupillometry is to be used in clinical audiology as a measure of listening effort. Researchers and clinicians would need to be aware of and familiar with the typical time-on-task/fatigue response in pupil trajectories over the course of the test session and within-conditions to avoid misattributing smaller aggregated responses to effort reduction when in fact, it may be a fatigue response.

Based on the findings reported here, as well as those reported in Zekveld et al. (2010) and Alhanbali et al. (2020), further investigation into baseline diameter within-conditions and over a test session is warranted. Baseline diameter measurement may provide a more sensitive measurement of time-on-task effects during listening effort assessments, when compared to peak dilation or mean dilation responses. Therefore, the inclusion of baseline diameter measures during listening effort assessments may provide a measurement of fatigue and at the same time, could be used to signal when a participant/client needs a rest and/or whether the delay between trials allows for adequate recovery.

Any examinations of fatigue in trial-level pupil parameters will require careful planning. This is because, for example: (1) the large number of trials typically needed when measuring listening effort using pupillometry and (2) issues with multiple testing, it may not be feasible to test every trial against the other for significant differences to examine fatigue response functions. Therefore, careful consideration of planned comparisons between trials to examine possible patterns and the specific time course of time-on-task/fatigue effects is also recommended.

14.5.7 Limitations

The aim of this chapter was to explore the effects of time-on-task on peak dilation, mean dilation, and baseline diameter during a speech-in-noise task. Several limitations of the current investigation were identified. The data were collected using an experimental design that simultaneously examined effects of light level, SNR, and fatigue. The effects of light level and SNR and the counterbalanced presentation of conditions may have masked fatigue effects across the whole test session. Thus, there was not sufficient evidence to claim that there was a decrease in physiological arousal across the test session that was consistent with fatigue.

The post hoc analyses were computed to examine the within-condition effect of time-on-task as this may be practically relevant for a clinical measure of listening effort. Furthermore, by performing post hoc analyses in this manner, the pairwise comparisons would not be affected by changing light level and SNR within each participant (although light levels and SNRs varied between participants). This limited the comparisons that can be drawn between the time-on-task findings and the subjective fatigue, subjective task-engagement, and performance data which were measured by-block. Future examination of pupil parameters, as well as subjective fatigue, subjective task-engagement, and performance within-conditions may enhance understanding of the effects reported here.

Overall, the findings have provided preliminary evidence that there was an effect of time-on-task on pupil parameters within-conditions (comprising 16 trials) during a speech-in-noise task. This effect was most prominent in baseline diameter comparisons. There might also be an effect of time-on-task in trial-level pupil parameters over the course of the whole test session. It is possible that the findings were associated with subjective fatigue and task-engagement measured by-block over the course of the speech-in-noise task, despite the improvement in performance. However, more research is needed to clarify the existence of these effects. The limitations identified above can be addressed in follow up studies and may provide important information regarding time-on-task effects in trial-level TEPRs. Any future research carried out to examine time-on-task effects should be done without variations to light level and SNR, to ensure the time-on-task effect is not confounded by these variables.

14.5.8 Conclusion

The above results extend the findings of Zekveld et al. (2010), Hopstaken et al. (2015a), Hopstaken et al. (2015b), Alhanbali et al. (2020), McGarrigle, Rakusen, et al. (2021), and McGarrigle, Knight, et al. (2021) by providing preliminary evidence that the effect of time-on-task and fatigue may also be present in trial-level pupil data during a speech-in-noise task when peak dilation, mean dilation, and baseline diameter are used as measures. Furthermore, the results provide evidence that significant time-on-task effects may develop in baseline diameter, within-conditions of a speech-in-noise task as short as 16 trials. The within-condition effects may be due to limited recovery time between trials.

The results reported in this chapter provide further support for the involvement of the LC norepinephrine system in the development of fatigue across and within-conditions but also show that these effects may not coincide with performance decrements when performance is measured by block (reported in Chapter 13). Future research is necessary to rectify the limitations of the current study and to draw informed conclusions about the time-on-task and fatigue response in trial-level pupil data.

15 GENERAL DISCUSSION

The studies reported in this thesis aimed to address multiple gaps in the growing body of research concerning the use of TEPRs to measure listening effort. An original contribution to knowledge is provided through the finding that light level affected TEPRs in a speech-in-noise task, that there is an interaction between the effects of light level and SNR on TEPRs. Evidence for these effects was established using both traditional signal-averaged pupil data (Study 1A and Study 1B) and trial-level pupil data (Study 1C).

Furthermore, Study 2 provided an additional original contribution to knowledge by using novel analysis techniques to unpack the mechanisms involved in pupil dilation under different light levels. This revealed that baseline diameter was an inconsistent mediator in the relationship between light level and pupil dilation during a speech-in-noise task.

Subjective and behavioural fatigue during the speech-in-noise task was assessed in Study 3A. Subjective task-engagement decreased with increasing time-on-task, whereas subjective fatigue and performance increased. Study 3B provided evidence for an effect of time-on-task on TEPRs. This supports previous findings but shows that this effect may also apply to pupil parameters measured at the trial-level and within-conditions of a speech-in-noise task.

The general discussion below is divided according to the general effects and themes that arose in this thesis. Findings related to light level and TEPRs are addressed first (15.1). Subsequently, the time-on-task and fatigue findings are discussed (15.2). Additionally, there are separate sections which provide discussions about the use of trial-level pupil data (15.3), effect sizes (15.4), and general limitations of the current work and future directions (15.5).

15.1 Light Level and TEPRs

In Study 1, peak and mean dilation were found to be larger in dimmer light levels. Additionally, light level and SNR interacted in their effects on peak and mean dilation such that light level had a greater effect on pupil dilation in more adverse SNR conditions. Furthermore, brighter light levels obscured differences in pupil dilation between SNR conditions (Figure 11, Figure 12). These findings align with those reported by Peysakhovich et al. (2015) who used a task that was different to the speech-in-noise task used here (i.e., digit recall task). In contrast, these findings are not congruent with the findings of Steinhauer et al. (2004), Wang, Kramer, et al. (2018) and Książek et al. (2021) who also reported effects of light level but all reported larger pupil dilation in their bright conditions, compared to dark conditions. Wang, Kramer, et al. (2018) and Książek et al. (2021) used a speech-in-noise task which was similar to the task used in the current thesis. Therefore, methodological differences in how light levels were manipulated (e.g., screen luminance here vs. room illumination previously) and the degree of light levels used (four mid-range levels here vs. bright and dark previously) may explain these disparate results. Future research into this relationship should be done to verify this claim.

In Study 1, the sensitivity of peak dilation to detect differences in listening effort between the SNR conditions was reduced under brighter light levels (as manipulated by a computer screen). Most eye trackers require computers to record measurements, and to synchronise stimulus presentation and TEPR measurement via appropriate platforms (e.g., PsychoPy). Therefore, it is recommended that researchers and clinicians use relatively dim ambient lighting and visual fields to maximise sensitivity of the measure.

Based on the current findings, the most sensitive measure of listening effort was achieved in L1. When measured at eye level (from a standard point on a chin rest), 60 cm away from the computer screen, the mean lux measurement for L1 was 21.98 (SD = 1.29). It is possible that this light level allowed the dilator muscles to dilate in response to cognitive effort to a greater extent than was possible in brighter light levels, where the constrictor muscles were more engaged. This was recently suggested by Winn et al. (2018) but had not previously been empirically tested in a listening effort paradigm.

The results of Study 1 also indicated that the effect of light level on peak dilation was larger when more effort was expended. In the least effortful listening condition, peak and mean dilation showed fewer significant differences between the light levels, and there were more differences between light levels in more effortful listening condition. These findings align with Peysakhovich et al. (2015) who found that peak dilation was greater in dimmer screen luminance conditions and that peak dilation was more affected by screen luminance, only in the most challenging memory task.

Section 9.5 provides potential explanations for the results of Study 1 based on the relative contributions of the SNS and PNS. For example, in the most difficult and dimmest condition, pupil dilation may have been caused by sympathetic stimulation of the dilator muscles and parasympathetic inhibition leading to relaxation of the constrictor muscles. On the other hand, in the brightest condition of the current study, the parasympathetic activity reaching the constrictor muscles (induced by the bright light) may have blocked the sympathetically driven pupil dilation and may have led to relaxation of the dilator muscles. Moreover, effort-related central activity may not have been powerful enough to inhibit the parasympathetic activity at the Edinger-Westphal nucleus. This may have impeded parasympathetic contributions to TEPRs (Joshi & Gold, 2020). The results of this study, therefore, speak to the potential interacting mechanisms underlying TEPRs.

The main purpose of Study 1B was to investigate the use of mixed-effects modelling using signal-averaged pupil data. The results of Study 1B (Chapter 10) replicated the results of the RANOVA in Study 1A. They also demonstrated the benefits of using mixed-effects modelling for the analysis of repeated-measures data. Additionally, Study 1B showed that time-on-task (condition order) was an important factor that could be beneficial to include in statistical modelling. The inclusion of this factor led to the discovery of more subtle effects of light level and SNR in the current study because it partitioned more variance from the error term.

Baseline diameter was explored as a potential source for the effect of light level on peak and mean dilation in Study 2 (Chapter 12). The findings suggest that changes in light level and the resultant changes in baseline diameter may have affected peak and mean dilation in opposing directions. The novel finding that baseline diameter acted as an inconsistent mediator in the relationships between light level and pupil dilation was reported. This has important implications for research; however, it also suggests further studies are warranted to elucidate the relationships between pupil dilation and changes in baseline diameter that both are and are not due to light level in effortful listening tasks.

The reasons for the relationships between baseline diameter and peak and mean dilation were not directly assessed. However, the findings of Gilzenrat et al. (2010) and Jepma and Nieuwenhuis (2011) suggest that the relationship between baseline diameter and pupil dilation may be reflective of LC activity modulating control state during task performance. Specifically, it may be due to trade-offs between exploration (task-disengagement) and exploitation (task-engagement) control states via LC activity and norepinephrine secretion (see Section 5.1.2). Gilzenrat et al. (2010) also reported that this relationship was not present when baseline diameter was manipulated via light level. This may suggest that pupil dilation is not related to light-induced changes in baseline diameter but is more related to LC activity-induced changes in baseline diameter.

Crucially, the mediation analyses were not able to differentiate the effects of light-induced changes in baseline diameter and changes in baseline due to other phenomena in the relationship between light level and pupil dilation. More research is needed to verify the physiological mechanisms responsible for the current results. Such research may benefit from the use of neuroimaging techniques, as well as pupillometry, to measure brain activity directly and indirectly, respectively. This would enhance understanding of how brain activity is reflected in pupil dilation during effortful listening and what factors contribute to this pupil response.

The lack of an effect of light level that has been previously reported (Bradshaw, 1969; Gilzenrat et al., 2010; Reilly et al., 2019) may have arisen due to the inconsistent mediator effect of baseline diameter found here, or pupil dilation may be relatively insensitive to light-induced changes in baseline diameter. The results reported here suggest that the relationship between baseline diameter and pupil dilation may be reflective of control state via LC activity, as predicted by Adaptive Gain Theory. The underlying physiological and neural mechanisms that led to this relationship have not been directly examined and therefore, further research is required to clarify these aspects.

Overall, based on the results reported in the current thesis, it is recommended that researchers and clinicians use dim light levels when assessing listening effort using pupillometry. Furthermore, and in line with recommendations set out in Tsukahara and Engle (2021), it can also be recommended that researchers report light conditions in enough detail that the lighting conditions can be replicated in different laboratories and/or clinics, or at the very least, the differences between the laboratories and/or clinics are known. This would involve reporting:

- Measurements of the light that reaches the participant/clients eyes (e.g., measurements of lux from approximate eye level).

- The precise location of measurement devices (when measuring lux).
- Measurements of multiple sources of light in the environment (e.g., computer screen, light bulbs, other equipment).
- Precise descriptions of the instruments and units used for measurements (including any unit conversions).

Detailed reporting of these factors will enable researchers and clinicians to accurately understand and/or replicate conditions which may lead to better replication studies and more comparable results between laboratories and clinics.

15.2 Time-On-Task and Fatigue

The results of Study 3A and 3B (Chapters 13 and 14, respectively) address the development of fatigue and time-on-task effects subjectively, behaviourally, and physiologically during the speech-in-noise task. Study 3A provided evidence that subjective fatigue and task-disengagement may develop during a speech-in-noise task lasting 1 hr. However, there was no behavioural effect consistent with fatigue or task-disengagement. Performance improved over the duration of the speech-in-noise task. This may have indicated a practice effect. It is possible that the subjective fatigue and task-engagement ratings were affected by response bias. This could be the reason why the subjective and behavioural measures of fatigue did not align. Conversely, by aggregating the performance scores by block, more detailed information about performance by condition and by trial was lost.

Study 3B provided evidence that trial number (time-on-task) significantly affected peak dilation, mean dilation, and baseline diameter. Based on visual inspection of the raw pupil data there was a small decrease across all TEPRs over the course of the test session (Figure 40, Figure 43, Figure 46). Averaged across levels of light level and SNR, Peak dilation, mean dilation, and baseline diameter were typically smaller in other trials compared to trial 1 (Appendix 7, Appendix 8, and Appendix 9). This provided marginal support for the subjective findings reported in Study 3A (Chapter 13) The findings reported in Study 3B (Chapter 14) may represent a reduction in physiological arousal that is consistent with fatigue. However, these results should be interpreted with caution. Specific conclusions regarding time-on-task effects over the whole test session cannot be drawn based on the analyses in Chapter 14. The exact position of significant differences in trials across the test session could not be identified.

Post hoc analyses were conducted to examine if significant time-on-task patterns existed within-conditions. EMMs were consistently smaller at the end of each condition, than at the beginning of the condition, regardless of the effect of light level or SNR. However, the differences in EMMs were only consistently significant for baseline diameter. This may represent a time-on-task effect, related to fatigue, that manifests after 16 trials of a speech-in-noise task and is most prominent in baseline diameter when compared to peak and mean dilation. Monitoring of baseline diameter during effortful listening may provide researchers and/or clinicians with an indication of time-on-task effects related to fatigue. Furthermore, because baseline diameter was able to recover between conditions but not between trials, this finding suggests that trial spacing is an important consideration for research and clinical applications of pupillometry.

Study 3B provided preliminary evidence that a physiological fatigue effect may be present at the trial-level in peak dilation, mean dilation, and baseline diameter and this may align with subjective feelings of fatigue and task-engagement measured by time block, even if the effects do not align with performance measures by time block. These findings also may indicate that there are systematic variations in trial-level pupil data which may have implications for signal-averaging. Future research should examine the effects of systematic variation in trial-level pupil data in more detail, across more tasks and include measures of peak dilation latency as this measure was not examined here.

15.3 The Use of Trial-Level Pupil Data

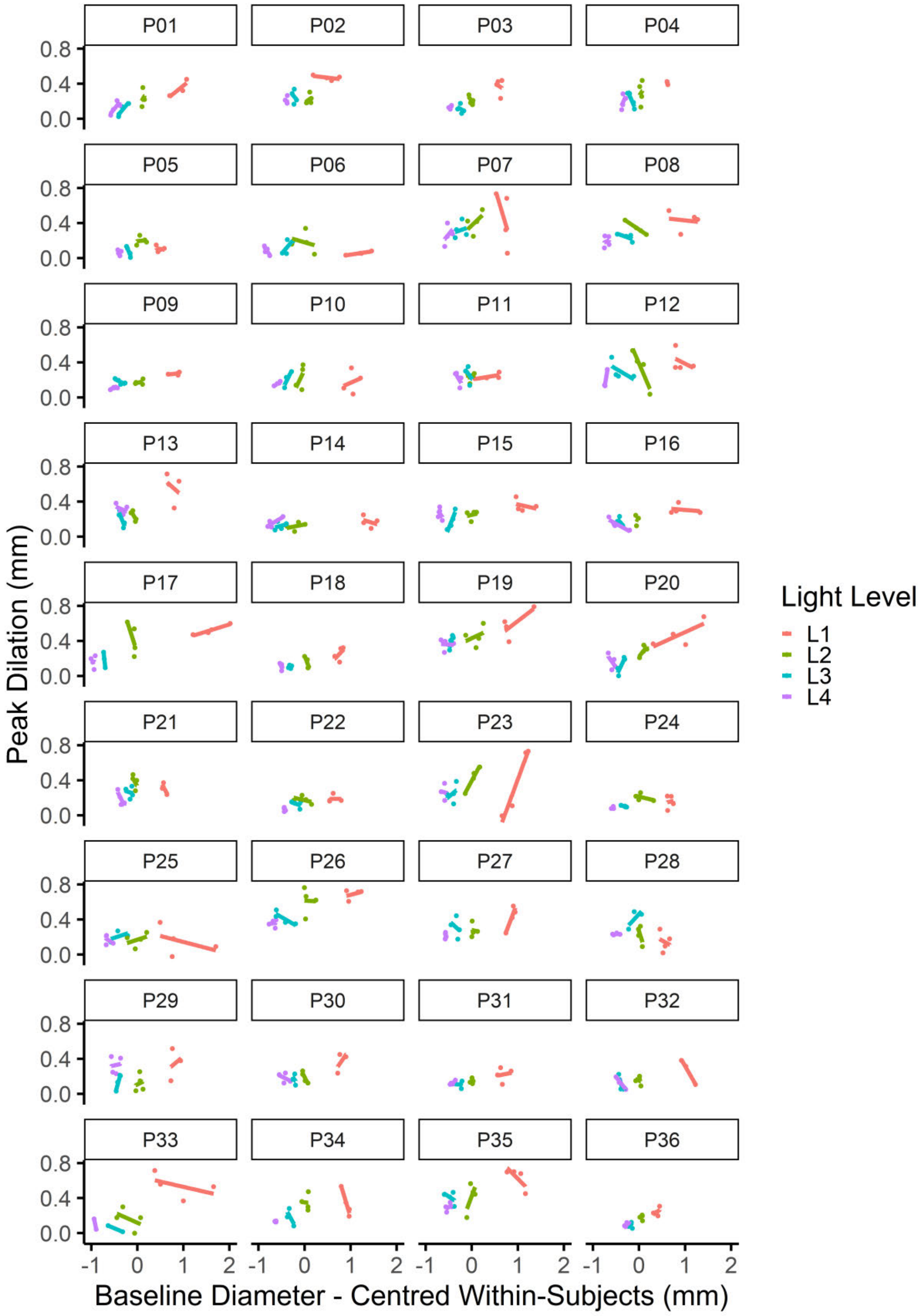
Study 1C examined the use of trial-level pupil data and mixed-effects modelling where additional random effects could be accounted for andverified that the effects of light level and SNR were measurable in trial-level pupil data (peak and mean dilation). This contributed to the growing body of findings which demonstrate that TEPRs can be measured from trial-level pupil waveforms without having to signal-average data prior to analysis (Clewett et al., 2020; Clewett et al., 2018; Cohen Hoffing et al., 2020; Leuchs et al., 2017; Wetzel et al., 2020) and extends those findings to a listening effort paradigm. This is discussed in more detail in Section 15.3.

Traditional approaches to the use of pupil data suggest that only when signal-averaging is implemented will a researcher get a reliable estimate of listening-related TEPRs, which is separate from other task-irrelevant pupil fluctuations (Winn et al., 2018). However, Section 11.1.1 describes some of the potential disadvantages associated with signal-averaging pupil data. When data was averaged over light levels and SNRs, a main effect of trial number was reported in Chapter 14. This could indicate the systematic variation in trial-level TEPRs. However, it does not appear that systematic variation in trial-level pupil data distorted signal-averaged TEPRs in Chapters 9 and 10. It would be beneficial to further examine the presence of systematic variation in trial-level TEPRs to see if (and in what tasks) systematic variation leads to distorted TEPRs. This would be particularly important if trial-level variation was not linear. Furthermore, examination of the effects of variable response latency on the signal-averaging of pupil data is still required.

Visualisation of signal-averaged pupil data (Figure 49) compared to trial-level pupil data (Figure 50) shows why consideration of the trial-level pupil data might be important. It can be seen in Figure 50 that the trial-level pupil data shows a more consistent negative relationship between baseline diameter and peak dilation for each light level and each participant. This might be due to control state as suggested in Chapter 12. On the other hand, the signal-averaged pupil data shows a more random relationship between baseline diameter and peak dilation for each light level and each participant (Figure 49). This may be due to systematic variability in the raw pupil traces distorting the signal-averaged pupil values. The effects that led to the relationship between baseline diameter and peak dilation in Figure 50 were lost when the trials within each condition were averaged together in Figure 49.

Figure 49

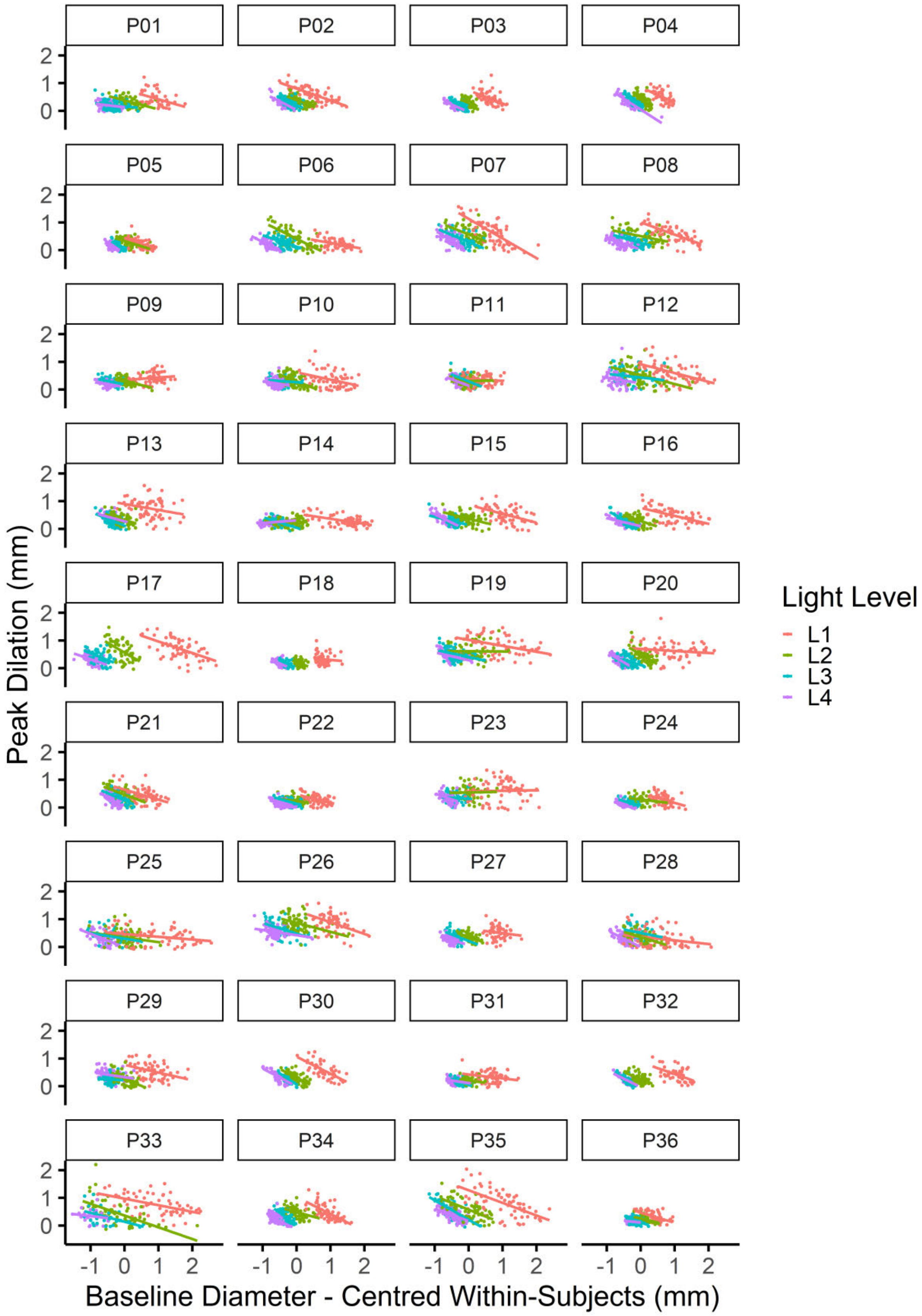
The Relationship Between Baseline Diameter and Peak Dilation Using Signal-Averaged Pupil Data



Note. This figure shows the relationship between baseline diameter (centred within-subjects) (mm) and peak dilation (mm) during a speech-in-noise task for each participant using signal-averaged pupil data. The different colours represent the different light levels in the task. The four points of the same colour are the four SNR conditions.

Figure 50

The Relationship Between Baseline Diameter and Peak Dilation Using Trial-Level Pupil Data



Note. This figure shows the relationship between baseline diameter (centred within-subjects) (mm) and peak dilation (mm) during a speech-in-noise task for each participant using trial-level pupil data. The different colours represent the different light levels in the task. The four points of the same colour are the four SNR conditions.

By using trial-level pupil data with mixed-effects modelling, multiple sources of variance in trial-level pupil data can be accounted for simultaneously. The use of trial-level pupil data enabled the novel examination of the effect of trial number of TEPRs which had not previously been examined in a speech-in-noise paradigm.

If using trial-level pupil data, MEMs should be used in place of traditional RANOVAs to make use of the partial pooling methods that are inherent to MEMs and to account for as many sources of variance in the data as is justified by the experimental design. When a larger proportion of total variance is partitioned from the error term, the power of the analysis to detect the fixed effects may increase. This may enable detection of more subtle effects.

However, trial-level pupil data may not be viable in other experimental paradigms, for instance, when pupil responses are not as robust as in Study 1 and 2 reported in the current thesis. The conclusions that can be drawn in Study 3 may support this. While there was an effect of trial number on trial-level TEPRs, the proportion of total variance explained by trial number was small. Despite the significant main effect of trial number, it is possible that the effect of fatigue was too small to show up in trial-level pupil data and/or the effect might have been dominated by the effects of light level and SNR (or other task-irrelevant fluctuations). Therefore, despite the presence of a measurable task-relevant pupil signal in the trial-level pupil data analysed in Study 1 and 2, it is possible that other experimental paradigms may not be able to detect TEPRs in trial-level pupil data.

The studies reported in this thesis provide a novel approach for using trial-level pupil data in the quantification of listening effort and in the examination of time-on-task effects. However, if using trial-level pupil data, signal-averaging should still be performed for comparative purposes and data visualisation should be used to assess if the trial-level pupil data are viable or not. Fundamentally, the feasibility of using trial-level pupil data will depend on the objectives of the research.

15.4 Effect Sizes

Measures of statistical significance (the *p* value) can be used to confirm how likely the pattern of results between the independent and dependent variable is, assuming no relationship exists in the population (Lakens, 2013). However, it cannot provide information of the size of the effect. Standardised statistical effect sizes are important for understanding the degree to which an intervention/manipulation affects a dependent variable, and therefore, the extent to which findings have practical consequences (Lakens, 2013).

Researchers can take two approaches when reporting effect sizes; (1) they can report effect sizes that are generalisable and independent of research design, or (2) they can report effect sizes relevant to the statistical test that was used (Lakens, 2013). In the current thesis, generalised eta squared was the effect size computed for the analyses which used RANOVA (Chapter 9 and Chapter 13). This effect size estimate excludes variation from other additional factors in analyses which enables comparison with other studies where those factors were not manipulated, including between subjects analyses (Olejnik & Algina, 2003). This makes generalised eta squared the most appropriate tool for performing meta-analyses (Lakens, 2013). On the other hand, partial eta squared can only be used to compare effects between studies with similar experimental designs and eta squared is not suitable to compare effect sizes between studies, generally (Lakens, 2013).

The use of different effect size estimates between Study 1 and previous research affected the degree to which effects could be compared. For example Peysakhovich et al. (2015) reported eta squared values for their effect size estimates. This measure is appropriate for comparing effects *within* a study but is inefficient when comparing between studies (Lakens, 2013). This is because calculation of eta squared values relies on the total variability in a study (all eta squared values sum to 100%), which depends on study design and increases when other variables are included in an analysis (Lakens, 2013).

For repeated-measures designs with a single factor, generalised eta squared, and partial eta squared are the same (Bakeman, 2005). Therefore, it was appropriate to compare generalised eta squared values in Study 3A to partial eta squared values reported in the literature (Section 13.6.2 and 13.6.3). For all other repeated-measures or mixed designs, generalised eta squared will be smaller than partial eta squared (Bakeman, 2005). Despite this, generalised eta squared values seem to be a better measure for comparing effect sizes between studies due to the ability to account for different research designs (Olejnik & Algina, 2003).

Considering this, the effect sizes reported in the current thesis can reveal information about the variance that may be explained by the factors under investigation, but how these estimates compare to the different estimates reported literature is more complex. As suggested by Lakens (2013), choices about which effect size to report should depend on the research questions under investigation, but these choices may affect how comparable the estimates are between studies.

The use of generalised eta squared in the current thesis may make comparisons to future research where the effects of light level on TEPRs are examined between individuals with and without hearing loss (in a between-subjects design) more straight forward. This would be dependent on those studies also reporting generalised eta squared estimates of effect size, rather than partial eta squared, or eta squared.

15.5 Limitations and Future Directions

There were several limitations in the studies and the findings reported in the current thesis. Firstly, explanations of effects that involved inferences about brain activity (in the LC or otherwise) based on TEPRs should be treated with caution because brain activity was not directly measured. TEPRs may not reflect LC activity in as simple a way as is often postulated in the literature. For example, McGinley et al. (2015) did not link pupil size to a particular brain structure, but instead linked pupil size to a more general concept; brain state. Winn et al. (2018) suggested that this may have been deliberately done to avoid erroneously implicating any specific brain structure without evidenced justification. Joshi et al. (2016) reported relationships between TEPRs and the LC, but also reported relationships between TEPRs and the inferior and superior colliculus and anterior and posterior cingulate cortex in monkeys. More recently, dorsal raphe serotonergic neurons (involved in learning, memory and affect) have been associated with pupil size changes in mice (Cazettes et al., 2021). Therefore, there are likely many brain factors that contribute to TEPRs, and it is important to reiterate that the explanations in the current thesis which relate TEPRs to brain activity and specific brain structures and processes are speculative. More research into human brain structures, mechanisms, chemistry, and physiology are needed to validate the explanations that relate TEPRs to specific neural processes.

Another limitation is that there were several extraneous variables that may have affected autonomic arousal of the participants which were not controlled for. Participants were asked to avoid caffeinated beverages before their scheduled test session. However, this is not easily controlled. If participants did have a caffeinated beverage before the test session, this may have affected pupil responses (Abokyi et al., 2017).

Additionally, some individuals participated in this study after a full day of occupational performance (with varying degrees of cognitive and physical labour). This may have influenced participant's subjective, behavioural and physiological responses (Sluiter et al., 2003). Furthermore, the test sessions were scheduled between the hours of 7:00 am to 8:00 pm, and therefore, varied between participants. Time of day may have affected pupil responses (Eggert et al., 2012) and subjective fatigue scores (Ferguson et al., 2012).

Baseline diameter is also affected by multiple factors. When examining how cognitive ability was reflected in baseline diameter, Tsukahara et al. (2016) also examined the effects of ethnicity, age (in years), college student status, nicotine (in the last 10 h), medications (that might affect attention and memory in the last 24 h), gender (male/female), handedness (right/left), caffeine (in the last 8 h), alcohol (more than 2 drinks in the last 24 h), and sleep (hours of sleep in the preceding night) on baseline diameter. They found that ethnicity, age, college student status, nicotine, and medication were significantly related to baseline diameter. Taken together, these demographic variables explained 15 % of the total variance in baseline diameter. Age showed the strongest association. Therefore, there are several factors that may contribute to baseline diameter. More research is needed to ascertain if these variables are also associated with variance in signal-averaged and trial-level TEPRs. If they are, it would also be important to understand the mechanisms responsible for that association.

Pupil size and pupil dilation range is affected by age (Bitsios et al., 1996; Piquado et al., 2010). Generally, older adults have smaller pupil sizes and their range of possible pupil sizes is restricted when compared with younger people (Piquado et al., 2010). This is called senile miosis. Therefore, measurements of the older adult pupil may lead to an underestimation of the amount of cognitive effort that is expended in a task, due to the smaller pupil size and restricted reactivity in this age group, when compared with measurements from younger pupils. Participants aged between 18-40 were recruited for the studies reported in the current thesis to avoid large differences in dynamic range of the pupil between participants. However, these differences may still exist.

Due to the potential for pupillometry to be used for clinical measure of listening effort, future research should seek to replicate the light level by SNR interaction in older individuals with and without hearing loss. These populations may experience increased listening effort due to their hearing loss, or natural cognitive decline (Pichora-Fuller et al., 2016). Recent findings also suggest that older adults show a more sustained pattern of effortful listening than younger adults (McGarrigle, Knight, et al., 2021). Therefore, the effects of light level and SNR may be more pronounced, or otherwise different for older individuals and/or those with hearing loss and/or cognitive decline.

The examination of time-on-task effects reported in Study 3B (Chapter 14) would have benefitted from a separate investigation in which other variables were not studied simultaneously (i.e., in the same design). This would have provided more conclusive evidence regarding time-on-task effects in trial-level pupil parameters, where the effect of light level was not influencing the pupil signal. However, the investigations reported in Chapter 14 do provide a base for further research. Evidence of an effect of trial number on trial-level TEPRs was provided. Therefore, this may inform future investigations into time-on-task effects by prompting questions about the number of trials (or the amount of time) it takes before an effect of time-on-task becomes statistically significant in TEPRs.

Additionally, baseline diameter was shown to be most sensitive to time-on-task within conditions when compared with peak and mean dilation. This finding could be used to prompt investigations into adequate delay times between trials and conditions during assessments. Moreover, when examining pupil dilation as a measure of listening effort, concurrent measurement of time sensitive baseline diameter during listening tasks could be examined. These measurements may serve as a marker for time-on-task/fatigue effects and improve listening effort assessments in which pupillometry is used.

Growth curve analysis provides an opportunity to examine the time course of fatigue during effortful listening. The application of growth curve analysis to pupillometric data is gaining in popularity (e.g., Geller et al., 2019; Kuchinsky et al., 2014; Kuchinsky et al., 2013; McGarrigle et al., 2017b; McLaughlin & Engen, 2020; McLaughlin et al., 2021; Winn et al., 2015). Growth curve analysis uses orthogonal parameters to describe the shape of curves in data (Kuchinsky et al., 2013). Therefore, linear, and nonlinear changes in pupil responses can be analysed over time. For this reason, growth curve analysis may be particularly suited to the examination of time-on-task effects in TEPRs, as they can model changes in the shape and timing of TEPRs across a test session and at the individual trial level (McGarrigle, Knight, et al., 2021). Recent findings by McGarrigle, Knight, et al. (2021) highlighted the additional information that can be gained from studying pupil responses using growth curve analysis. They found differences in the steepness of pupil responses between older and younger listeners, suggesting that there are differences in how these age groups sustain effort expenditure in speech-in-noise over time.

In summary, while there were many other factors that may have contributed to pupil responses reported in the current thesis, these findings and the associated limitations provide avenues for further investigation. Additionally, the results reported in the current thesis should be subject to replication studies in different laboratories, and verification with other sample groups (older individuals with and without hearing loss). This may lead to a better understanding of TEPRs and aid clinical applications.

16 CONCLUSION

The overarching aim of this thesis was to contribute to the growing body of work examining TEPRs as a measure of listening effort.

The findings show that light level does affect TEPRs in a speech-in-noise task. Evidence of an interaction effect between light level and SNR was presented in Study 1. This has implications for research and any future clinical implementation. When using pupillometry to measure listening effort, environmental light levels should be dim (including dim computer screens) and light levels should be reported in detail. This thesis also provided evidence that the use of trial-level pupil data may provide an opportunity for researchers to acknowledge the full variance structure within TEPRs which may lead to better interpretation of these responses.

The first analyses treating baseline diameter as a mediator in the relationship between light level and TEPRs were reported in Study 2. These revealed that baseline diameter acted as a partial and inconsistent mediator in the relationship between light level and pupil dilation by suppressing some of the effect. Changes in light level and baseline diameter affected pupil dilation in opposing ways. This suggests a need for further research into the relationships between baseline diameter and pupil responses.

Study 3 provided evidence that individuals may have experienced subjective fatigue and task-disengagement over the course of the speech-in-noise task, but this did not align with performance. Study 3 also provided the first evidence that trial-level baseline diameter was sensitive to within-condition time-on-task effects. Furthermore, there was evidence that trial number affected trial-level TEPRs (when averaged over the effects of light level and SNR), but more research is needed to assess fatigue effects across test sessions.

A clinical measure of listening effort might complement current audiological assessment tools (see Section 6.3). Task-related pupil dilation may be used as a clinical measure of listening effort in the future. While the use of pupillometry to measure listening effort in clinical audiology shows promise, standard clinical techniques and protocols do not exist yet. Enhanced understanding of TEPRs and the factors that affect them in listening tasks will assist in the development of such techniques and protocols. This thesis provides an important contribution towards the goal of establishing a clinical measure of listening effort.

17 REFERENCES

Abokyi, S., Owusu-Mensah, J., & Osei, K. A. (2017). Caffeine intake is associated with pupil dilation and enhanced accommodation. *Eye*, *31*(4), 615-619. https://doi.org/10.1038/eye.2016.288

Afanador, N. L., Tran, T., Blanchet, L., & Baumgartner, R. (2021). *mvdalab: Multivariate data analysis laboratory*. In (Version 1.5) R Package. https://cran.r-project.org/web/packages/mvdalab/index.html

Alain, C., Arnott, S. R., & Picton, T. W. (2001). Bottom-up and top-down influences on auditory scene analysis: evidence from event-related brain potentials. *Journal of Experimental Psychology: Human Perception and Performance*, *27*(5), 1072-1089. https://doi.org/10.1037//0096-1523.27.5.1072

Alhanbali, S., Dawes, P., Lloyd, S., & Munro, K. J. (2017). Self-reported listening-related effort and fatigue in hearing-impaired adults. *Ear and Hearing*, *38*(1), e39-e48. https://doi.org/10.1097/aud.0000000000000361

Alhanbali, S., Dawes, P., Millman, R. E., & Munro, K. J. (2019). Measures of listening effort are multidimensional. *Ear and Hearing*, *40*(5), 1084-1097. https://doi.org/10.1097/AUD.0000000000000697

Alhanbali, S., Munro, K. J., Dawes, P., Carolan, P. J., & Millman, R. E. (2020). Dimensions of self-reported listening effort and fatigue on a digits-in-noise task, and association with baseline pupil size and performance accuracy. *International Journal of Audiology*, *Advance online publication*, 1-11. https://doi.org/10.1080/14992027.2020.1853262

Alnæs, D., Sneve, M. H., Espeseth, T., Endestad, T., van de Pavert, S. H. P., & Laeng, B. (2014). Pupil size signals mental effort deployed during multiple object tracking and predicts brain activity in the dorsal attention network and the locus coeruleus. *Journal of Vision*, *14(4)*(1), 1-20. https://doi.org/10.1167/14.4.1

American Speech-Language-Hearing Association. (2015). *Audiology Information Series: Type, Degree, and Configuration of Hearing Loss*. https://www.asha.org/uploadedFiles/AIS-Hearing-Loss-Types-Degree-Configuration.pdf

Aminihajibashi, S., Hagen, T., Foldal, M. D., Laeng, B., & Espeseth, T. (2019). Individual differences in resting-state pupil size: Evidence for association between working memory capacity and pupil size variability. *International Journal of Psychophysiology*, *140*, 1-7. https://doi.org/10.1016/j.ijpsycho.2019.03.007

Aston-Jones, G., & Cohen, J. D. (2005). An integrative theory of locus coeruleus-norepinephrine function: Adaptive gain and optimal performance. *Annual Review of Neuroscience*, *28*(1), 403-450. https://doi.org/10.1146/annurev.neuro.28.061604.135709

Audacity Team. (2019). *Audacity(R): Free Audio Editor and Recorder (version 2.2.2)*. In https://audacityteam.org

Ayasse, N. D., & Wingfield, A. (2020). Anticipatory baseline pupil diameter is sensitive to differences in hearing thresholds [Brief Research Report]. *Frontiers in Psychology*, *10*(2947), 1-7. https://doi.org/10.3389/fpsyg.2019.02947

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. https://doi.org/10.1016/j.jml.2007.12.005

Baddeley, A. (2010). Working memory. *Current Biology*, *20*(4), R136-R140. https://doi.org/10.1016/j.cub.2009.12.014

Bafna, T., & Hansen, J. P. (2021). Mental fatigue measurement using eye metrics: A systematic literature review. *Psychophysiology*, *58*(6), 1-23. https://doi.org/10.1111/psyp.13828

Bagiella, E., Sloan, R. P., & Heitjan, D. F. (2000). Mixed-effects models in psychophysiology. *Psychophysiology*, *37*(1), 13-20. https://doi.org/10.1111/1469-8986.3710013

Bakdash, J. Z., & Marusich, L. R. (2017). Repeated measures correlation. *Frontiers in Psychology*, *8*(456), 1-13. https://doi.org/10.3389/fpsyg.2017.00456

Bakeman, R. (2005). Recommended effect size statistics for repeated measures designs. *Behavior Research Methods*, *37*(3), 379-384. https://doi.org/10.3758/BF03192707

Baldock, J., Kapadia, S., & van Steenbrugge, W. (2019). The task-evoked pupil response in divided auditory attention tasks. *Journal of the American Academy of Audiology*, *30*(4), 264-272. https://doi.org/10.3766/jaaa.17060

Barnett, A. G., Koper, N., Dobson, A. J., Schmiegelow, F., & Manseau, M. (2010). Using information criteria to select the correct variance–covariance structure for longitudinal data in ecology. *Methods in Ecology and Evolution*, *1*(1), 15-24. https://doi.org/10.1111/j.2041-210X.2009.00009.x

Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*(6), 1173–1182. https://doi.org/10.1037/0022-3514.51.6.1173

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. https://doi.org/10.1016/j.jml.2012.11.001

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1-48. https://doi.org/doi:10.18637/jss.v067.i01

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., Dai, B., Scheipl, F., Grothendieck, G., Green, P., Fox, J., Bauer, A., & Krivitsky, P. N.

(2021). *Package 'lme4'* [Reference Manual]. https://cran.r-project.org/web/packages/lme4/lme4.pdf

Bear, M. F., Connors, B. W., & Paradiso, M. A. (2016). *Neuroscience: Exploring the Brain* (4th ed.). Wolters Kluwer.

Beatty, J. (1982a). Phasic not tonic pupillary responses vary with auditory vigilance performance. *Psychophysiology*, *19*(2), 167-172. https://doi.org/10.1111/j.1469-8986.1982.tb02540.x

Beatty, J. (1982b). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological Bulletin*, *91*(2), 276-292. https://doi.org/10.1037/0033-2909.91.2.276

Beatty, J., & Lucero-Wagoner, B. (2000). The pupillary system. In John T. Cacioppo, Louis G. Tassinary, & Gary G. Berntson (Eds.), *Handbook of Psychophysiology* (pp. 142-162). Cambridge University Press.

Benarroch, E. E. (1993). The central autonomic network: Functional organization, dysfunction, and perspective. *Mayo Clinic Proceedings*, *68*(10), 988-1001. https://doi.org/10.1016/S0025-6196(12)62272-1

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. https://doi.org/10.3109/03005367909078884

Berridge, C. W., & Waterhouse, B. D. (2003). The locus coeruleus–noradrenergic system: Modulation of behavioral state and state-dependent cognitive processes. *Brain Research Reviews*, *42*(1), 33-84. https://doi.org/10.1016/S0165-0173(03)00143-7

Bess, F. H., & Hornsby, B. W. Y. (2014). Commentary: listening can be exhausting--fatigue in children and adults with hearing loss. *Ear and Hearing*, *35*(6), 592-599. https://doi.org/10.1097/AUD.0000000000000099

Bitsios, P., Prettyman, R., & Szabadi, E. (1996). Changes in autonomic function with age: A study of pupillary kinetics in healthy young and old people. *Age and Ageing*, *25*(6), 432-438. https://doi.org/10.1093/ageing/25.6.432

Boisgontier, M. P., & Cheval, B. (2016). The anova to mixed model transition. *Neuroscience & Biobehavioral Reviews*, *68*, 1004-1005. https://doi.org/10.1016/j.neubiorev.2016.05.034

Boksem, M. A. S., Meijman, T. F., & Lorist, M. M. (2006). Mental fatigue, motivation and action monitoring. *Biological Psychology*, *72*(2), 123-132. https://doi.org/10.1016/j.biopsycho.2005.08.007

Boksem, M. A. S., & Tops, M. (2008). Mental fatigue: Costs and benefits. *Brain Research Reviews*, *59*(1), 125-139. https://doi.org/10.1016/j.brainresrev.2008.07.001

Bönitz, H., Lunner, T., Finke, M., Fiedler, L., Lyxell, B., Riis, S. K., Ng, E., Valdes, A. L., Büchner, A., & Wendt, D. (2021). How do we allocate our resources when listening and memorizing speech in noise? A pupillometry study. *Ear and Hearing*, *42*(4), 846-859. https://doi.org/10.1097/AUD.0000000000001002

Boucsein, W. (2012). *Electrodermal activity*. Springer Science & Business Media.

Bouret, S., & Sara, S. J. (2005). Network reset: A simplified overarching theory of locus coeruleus noradrenaline function. *Trends in Neuroscience*, *28*(11), 574-582. https://doi.org/10.1016/j.tins.2005.09.002

Bradshaw, J. L. (1969). Background light intensity and the pupillary response in a reaction time task. *Psychonomic Science*, *14*(6), 271-272. https://doi.org/10.3758/BF03329118

Brännström, K. J., Rudner, M., Carlie, J., Sahlén, B., Gulz, A., Andersson, K., & Johansson, R. (2021). Listening effort and fatigue in native and non-native primary school children. *Journal of Experimental Child Psychology*, *210*, 1-18. https://doi.org/10.1016/j.jecp.2021.105203

Bregman, A. S. (1994). *Auditory scene analysis: The perceptual organization of sound*. MIT press.

Brehm, J. W., & Self, E. A. (1989). The intensity of motivation. *Annual Review of Psychology*, *40*(1), 109-131. https://doi.org/10.1146/annurev.ps.40.020189.000545

Brewster, K. K., Ciarleglio, A., Brown, P. J., Chen, C., Kim, H.-O., Roose, S. P., Golub, J. S., & Rutherford, B. R. (2018). Age-related hearing loss and its association with depression in later life. *The American Journal of Geriatric Psychiatry*, *26*(7), 788-796. https://doi.org/10.1016/j.jagp.2018.04.003

Brons, I., Houben, R., & Dreschler, W. A. (2014). Effects of noise reduction on speech intelligibility, perceived listening effort, and personal preference in hearing-impaired listeners. *Trends in Hearing*, *18*, 1-10. https://doi.org/10.1177/2331216514553924

Brown, V. A. (2021). An introduction to linear mixed-effects modeling in R. *Advances in Methods and Practices in Psychological Science*, *4*(1), 1-19. https://doi.org/10.1177/2515245920960351

Burgess, R. C. (2019). Magnetoencephalography for localizing and characterizing the epileptic focus. In K. H. Levin & P. Chauvel (Eds.), *Handbook of Clinical Neurology* (Vol. 160, pp. 203-214). Elsevier. https://doi.org/10.1016/B978-0-444-64032-1.00013-8

Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using stan [Bayesian inference; multilevel model; ordinal data; MCMC; Stan; R]. *Journal of Statistical Software*, *80*(1), 1-28. https://doi.org/10.18637/jss.v080.i01

Bürkner, P.-C. (2018). Advanced Bayesian multilevel modeling with the R package brms. *The R Journal*, *10*(1), 395-411. https://doi.org/doi:10.32614/RJ-2018-017

Burle, B., Spieser, L., Roger, C., Casini, L., Hasbroucq, T., & Vidal, F. (2015). Spatial and temporal resolutions of EEG: Is it really black and white? A scalp current density view. *International Journal of Psychophysiology*, *97*(3), 210-220. https://doi.org/10.1016/j.ijpsycho.2015.05.004

Cairns, R. B. (1986). Phenomena lost: Issues in the study of development. In J. Valsiner (Ed.), *The Individual Subject and Scientific Psychology* (pp. 97-111). Springer US. https://doi.org/10.1007/978-1-4899-2239-7_5

Carlyon, R. P., Cusack, R., Foxton, J. M., & Robertson, I. H. (2001). Effects of attention and unilateral neglect on auditory stream segregation. *Journal of Experimental Psychology: Human Perception and Performance*, *27*(1), 115-127. https://doi.org/10.1037/0096-1523.27.1.115

Cazettes, F., Reato, D., Morais, J. P., Renart, A., & Mainen, Z. F. (2021). Phasic activation of dorsal raphe serotonergic neurons increases pupil size. *Current Biology*, *31*(1), 192-197. https://doi.org/10.1016/j.cub.2020.09.090

Chapman, C. R., Oka, S., Bradshaw, D. H., Jacobson, R. C., & Donaldson, G. W. (1999). Phasic pupil dilation response to noxious stimulation in normal volunteers: Relationship to brain evoked potentials and pain report. *Psychophysiology*, *36*(1), 44-52. https://doi.org/10.1017/s0048577299970373

Clewett, D., Gasser, C., & Davachi, L. (2020). Pupil-linked arousal signals track the temporal organization of events in memory. *Nature Communications*, *11*(4007), 1-14. https://doi.org/10.1038/s41467-020-17851-9

Clewett, D. V., Huang, R., Velasco, R., Lee, T. H., & Mather, M. (2018). Locus coeruleus activity strengthens prioritized memories under arousal. *The Journal of Neuroscience*, *38*(6), 1558-1574. https://doi.org/10.1523/jneurosci.2097-17.2017

Cohen, D., & Cuffin, B. N. (1983). Demonstration of useful differences between magnetoencephalogram and electroencephalogram. *Electroencephalography and Clinical Neurophysiology*, *56*(1), 38-51. https://doi.org/10.1016/0013-4694(83)90005-6

Cohen Hoffing, R. A., Lauharatanahirun, N., Forster, D. E., Garcia, J. O., Vettel, J. M., & Thurman, S. M. (2020). Dissociable mappings of tonic and phasic pupillary features onto cognitive processes involved in mental arithmetic. *PLoS ONE*, *15*(3), e0230517. https://doi.org/10.1371/journal.pone.0230517

Costa, Vincent D., & Rudebeck, Peter H. (2016). More than meets the eye: The relationship between pupil size and locus coeruleus activity. *Neuron*, *89*(1), 8-10. https://doi.org/10.1016/j.neuron.2015.12.031

Danermark, B., & Gellerstedt, L. C. (2004). Psychosocial work environment, hearing impairment and health. *International Journal of Audiology*, *43*(7), 383-389. https://doi.org/10.1080/14992020400050049

Davis, H., Schlundt, D., Bonnet, K., Camarata, S., Bess, F. H., & Hornsby, B. (2021). Understanding listening-related fatigue: Perspectives of adults with hearing loss. *International Journal of Audiology*, *60*(6), 458-468. https://doi.org/10.1080/14992027.2020.1834631

de Morree, H. M., Szabó, B. M., Rutten, G. J., & Kop, W. J. (2013). Central nervous system involvement in the autonomic responses to psychological distress. *Netherlands Heart Journal*, *21*(2), 64-69. https://doi.org/10.1007/s12471-012-0351-1

DeLuca, J. (2005). Fatigue, cognition, and mental effort. In J. DeLuca (Ed.), *Fatigue as a window to the brain* (pp. 37-59). MIT Press. https://doi.org/10.7551/mitpress/2967.001.0001

Detry, M. A., & Ma, Y. (2016). Analyzing repeated measurements using mixed models. *Journal of the American Medical Association*, *315*(4), 407-408. https://doi.org/10.1001/jama.2015.19394

Dimitrijevic, A., Smith, M. L., Kadis, D. S., & Moore, D. R. (2019). Neural indices of listening effort in noisy environments. *Scientific Reports*, *9*(11278), 1–10. https://doi.org/10.1038/s41598-019-47643-1

Downs, D. W. (1982). Effects of hearing aid use on speech discrimination and listening effort. *Journal of Speech and Hearing Disorders*, *47*(2), 189-193. https://doi.org/doi:10.1044/jshd.4702.189

Eckstein, M. K., Guerra-Carrillo, B., Miller Singley, A. T., & Bunge, S. A. (2017). Beyond eye gaze: What else can eyetracking reveal about cognition and cognitive development? *Developmental Cognitive Neuroscience*, *25*, 69-91. https://doi.org/10.1016/j.dcn.2016.11.001

Edwards, B. (2007). The future of hearing aid technology. *Trends in Amplification*, *11*(1), 31-45. https://doi.org/10.1177/1084713806298004

Edwards, B. (2016). A model of auditory-cognitive processing and relevance to clinical applicability. *Ear and Hearing*, *37*, 85S-91S. https://doi.org/10.1097/aud.0000000000000308

Eggert, T., Sauter, C., Popp, R., Zeitlhofer, J., Danker-Hopfe, H., Research, o. b. o. t. w. g. V. o. t. G. S. f. S., & Medicine, S. (2012). The pupillographic sleepiness test in adults: Effect of age, gender, and time of day on pupillometric variables. *American Journal of Human Biology*, *24*(6), 820-828. https://doi.org/10.1002/ajhb.22326

Einhäuser, W. (2017). The pupil as marker of cognitive processes. In Q. Zhao (Ed.), *Computational and Cognitive Neuroscience of Vision* (pp. 141-169). Springer. https://doi.org/10.1007/978-981-10-0213-7_7

Feng, T., Chen, Q., & Xiao, Z. (2018). Age-related differences in the effects of masker cuing on releasing chinese speech from informational masking [Original Research]. *Frontiers in Psychology*, *9*(1922). https://doi.org/10.3389/fpsyg.2018.01922

Ferguson, S. A., Paech, G. M., Sargent, C., Darwent, D., Kennaway, D. J., & Roach, G. D. (2012). The influence of circadian time and sleep dose on subjective fatigue ratings. *Accident Analysis & Prevention*, *45S*, 50-54. https://doi.org/10.1016/j.aap.2011.09.026

Fiedler, L., Seifi Ala, T., Graversen, C., Alickovic, E., Lunner, T., & Wendt, D. (2021). Hearing aid noise reduction lowers the sustained listening effort during continuous speech in noise — A combined pupillometry and EEG study. *Ear and Hearing*, *Advance online publication*, 1-12. https://doi.org/10.1097/aud.0000000000001050

Foroughi, C. K., Sibley, C., & Coyne, J. T. (2017). Pupil size as a measure of within-task learning. *Psychophysiology*, *54*(10), 1436-1443. https://doi.org/10.1111/psyp.12896

Francis, A. L., Bent, T., Schumaker, J., Love, J., & Silbert, N. (2021). Listener characteristics differentially affect self-reported and physiological measures of effort associated with two challenging listening conditions. *Attention, Perception & Psychophysics*, *83*(4), 1818-1841. https://doi.org/10.3758/s13414-020-02195-9

Francis, A. L., & Love, J. (2020). Listening effort: Are we measuring cognition or affect, or both? *WIREs Cognitive Science*, *11*(e1514), 1-27. https://doi.org/10.1002/wcs.1514

Francis, A. L., & Oliver, J. (2018). Psychophysiological measurement of affective responses during speech perception. *Hearing Research*, *369*, 103-119. https://doi.org/10.1016/j.heares.2018.07.007

Gage, N. M., & Baars, B. J. (2019). Observing the brain. In N. M. Gage & B. J. Baars (Eds.), *Fundamentals of Cognitive Neuroscience* (2nd ed., pp. 53-97). Academic Press. https://doi.org/10.1016/B978-0-12-803813-0.00003-9

Gagné, J.-P., Besser, J., & Lemke, U. (2017). Behavioral assessment of listening effort using a dual-task paradigm: A review. *Trends in Hearing*, *21*, 1-27. https://doi.org/10.1177/2331216516687287

Gallagher, N. E., & Woodside, J. V. (2018). Factors affecting hearing aid adoption and use: A qualitative study. *Journal of the American Academy of Audiology*, *29*(4), 300-312. https://doi.org/10.3766/jaaa.16148

Gatehouse, S., & Noble, W. (2004). The speech, spatial and qualities of hearing scale (ssq). *International Journal of Audiology*, *43*(2), 85-99. https://doi.org/10.1080/14992020400050014

Gawron, V. J. (2016). Overview of self-reported measures of fatigue. *The International Journal of Aviation Psychology*, *26*(3-4), 120-131. https://doi.org/10.1080/10508414.2017.1329627

Gazzaley, A., & Nobre, A. C. (2012). Top-down modulation: Bridging selective attention and working memory. *Trends in Cognitive Sciences*, *16*(2), 129-135. https://doi.org/10.1016/j.tics.2011.11.014

Geller, J., Landrigan, J.-F., & Mirman, D. (2019). A pupillometric examination of cognitive control in taxonomic and thematic semantic memory. *Journal of Cognition*, *2*(1), 1-10. https://doi.org/10.5334/joc.56

Geller, J., Winn, M. B., Mahr, T., & Mirman, D. (2020). GazeR: A package for processing gaze position and pupil size data. *Behavior Research Methods*, *52*(5), 2232-2255. https://doi.org/10.3758/s13428-020-01374-8

Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press.

Gelman, A., Hill, J., & Yajima, M. (2012). Why we (usually) don't have to worry about multiple comparisons. *Journal of Research on Educational Effectiveness*, *5*(2), 189-211. https://doi.org/10.1080/19345747.2011.618213

Gilzenrat, M. S., Nieuwenhuis, S., Jepma, M., & Cohen, J. D. (2010). Pupil diameter tracks changes in control state predicted by the adaptive gain theory of locus coeruleus function. *Cognitive, Affective, & Behavioral Neuroscience*, *10*(2), 252-269. https://doi.org/10.3758/CABN.10.2.252

Giuliani, N. P., Brown, C. J., & Wu, Y.-H. (2020). Comparisons of the sensitivity and reliability of multiple measures of listening effort. *Ear and Hearing*, *42*(2), 465-474. https://doi.org/10.1097/AUD.0000000000000950

Glover, G. H. (2011). Overview of functional magnetic resonance imaging. *Neurosurgery Clinics of North America*, *22*(2), 133–139. https://doi.org/10.1016/j.nec.2010.11.001

Gosselin, P. A., & Gagné, J.-P. (2011). Older adults expend more listening effort than young adults recognizing speech in noise. *Journal of Speech, Language, and Hearing Research*, *54*(3), 944-958. https://doi.org/10.1044/1092-4388(2010/10-0069

Green, D. P., Ha, S. E., & Bullock, J. G. (2010). Enough already about “Black Box” experiments: Studying mediation is more difficult than most scholars suppose. *The ANNALS of the American Academy of Political and Social Science*, *628*(1), 200-208. https://doi.org/10.1177/0002716209351526

Gustafson, S., McCreery, R., Hoover, B., Kopun, J. G., & Stelmachowicz, P. (2014). Listening effort and perceived clarity for normal-hearing children with the use of digital noise reduction. *Ear and Hearing*, *35*(2), 183-194. https://doi.org/10.1097/01.aud.0000440715.85844.b8

Hasson, D., Theorell, T., Liljeholm-Johansson, Y., & Canlon, B. (2009). Psychosocial and physiological correlates of self-reported hearing problems in male and female musicians in symphony orchestras. *International Journal of Psychophysiology*, *74*(2), 93-100. https://doi.org/10.1016/j.ijpsycho.2009.07.009

Hayes, T. R., & Petrov, A. A. (2016). Pupil diameter tracks the exploration–exploitation trade-off during analogical reasoning and explains individual differences in fluid intelligence. *Journal of Cognitive Neuroscience*, *28*(2), 308-318. https://doi.org/10.1162/jocn_a_00895

Hecker, M. H. L., Stevens, K. N., & Williams, C. E. (1966). Measurements of reaction time in intelligibility tests. *The Journal of the Acoustical Society of America*, *39*(6), 1188-1189. https://doi.org/10.1121/1.1910013

Heino, M. T. J., Vuorre, M., & Hankonen, N. (2018). Bayesian evaluation of behavior change interventions: A brief introduction and a practical example. *Health Psychology and Behavioral Medicine*, *6*(1), 49-78. https://doi.org/10.1080/21642850.2018.1428102

Hess, E. H., & Polt, J. M. (1964). Pupil size in relation to mental activity during simple problem-solving. *Science*, *143*(3611), 1190-1192. https://doi.org/10.1126/science.143.3611.1190

Hetu, R., Jones, L., & Getty, L. (1993). The impact of acquired hearing impairment on intimate relationships: Implications for rehabilitation. *Audiology*, *32*(6), 363-381. https://doi.org/10.3109/00206099309071867

Hetu, R., Riverin, L., Lalande, N., Getty, L., & St-Cyr, C. (1988). Qualitative analysis of the handicap associated with occupational hearing loss. *British Journal of Audiology*, *22*(4), 251-264. https://doi.org/10.3109/03005368809076462

Heyl, V., & Wahl, H.-W. (2012). Managing daily life with age-related sensory loss: Cognitive resources gain in importance. *Psychology and Aging*, *27*(2), 510-521. https://doi.org/10.1037/a0025471

Hicks, C. B., & Tharpe, A. M. (2002). Listening effort and fatigue in school-age children with and without hearing loss. *Journal of Speech, Language, and Hearing Research*, *45*(3), 573-584. https://doi.org/10.1044/1092-4388(2002/046)

Hintze, J. (2015). *PASS 14 Power Analysis and Sample Size Software* In (Version 14.0.2) NCSS, LLC.

Hockey, R. (2011). A motivational control theory of cognitive fatigue. In P. L. Ackerman (Ed.), *Cognitive fatigue: Multidisciplinary perspectives on current research and* *future applications* (pp. 167-187). American Psychological Association. https://doi.org/10.1037/12343-008

Hockey, R. (2013). *The psychology of fatigue: Work, effort and control*. Cambridge University Press.

Holman, J. A., Drummond, A., Hughes, S. E., & Naylor, G. (2019). Hearing impairment and daily-life fatigue: A qualitative study. *International Journal of Audiology*, *58*(7), 408-416. https://doi.org/10.1080/14992027.2019.1597284

Holube, I., Haeder, K., Imbery, C., & Weber, R. (2016). Subjective listening effort and electrodermal activity in listening situations with reverberation and noise. *Trends in Hearing*, *20*, 1-15. https://doi.org/10.1177/2331216516667734

Hopstaken, J. (2016). *Conquering fatigue: The battle for engagement* [Ph.D. thesis, Erasmus University Rotterdam]. http://hdl.handle.net/1765/93180

Hopstaken, J. F., van der Linden, D., Bakker, A. B., & Kompier, M. A. J. (2015a). A multifaceted investigation of the link between mental fatigue and task disengagement. *Psychophysiology*, *52*(3), 305-315. https://doi.org/10.1111/psyp.12339

Hopstaken, J. F., van der Linden, D., Bakker, A. B., & Kompier, M. A. J. (2015b). The window of my eyes: Task disengagement and mental fatigue covary with pupil dynamics. *Biological Psychology*, *110*, 100-106. https://doi.org/10.1016/j.biopsycho.2015.06.013

Hopstaken, J. F., van der Linden, D., Bakker, A. B., Kompier, M. A. J., & Leung, Y. K. (2016). Shifts in attention during mental fatigue: Evidence from subjective, behavioral, physiological, and eye-tracking data. *Journal of Experimental Psychology: Human Perception and Performance*, *42*(6), 878-889. https://doi.org/10.1037/xhp0000189

Hornsby, B. W. Y. (2013). The effects of hearing aid use on listening effort and mental fatigue associated with sustained speech processing demands. *Ear and Hearing*, *34*(5), 523-534. https://doi.org/10.1097/AUD.0b013e31828003d8

Hornsby, B. W. Y., Camarata, S., Cho, S.-J., Davis, H., McGarrigle, R., & Bess, F. H. (2021). Development and validation of the Vanderbilt Fatigue Scale for Adults (VFS-A). *Psychological assessment*, *33*(8), 777-788. https://doi.org/10.1037/pas0001021

Hornsby, B. W. Y., & Kipp, A. M. (2016). Subjective ratings of fatigue and vigor in adults with hearing loss are driven by perceived hearing difficulties not degree of hearing loss. *Ear and Hearing*, *37*(1), e1-e10. https://doi.org/10.1097/AUD.0000000000000203

Hornsby, B. W. Y., Naylor, G., & Bess, F. H. (2016). A taxonomy of fatigue concepts and their relation to hearing loss. *Ear and Hearing*, *37* (Suppl 1), 136S-144S. https://doi.org/10.1097/AUD.0000000000000289

Hornsby, B. W. Y., Werfel, K., Camarata, S., & Bess, F. H. (2014). Subjective fatigue in children with hearing loss: Some preliminary findings. *American Journal of Audiology*, *23*(1), 129-134. https://doi.org/10.1044/1059-0889(2013/13-0017)

Houben, R., van Doorn-Bierman, M., & Dreschler, W. A. (2013). Using response time to speech as a measure for listening effort. *International Journal of Audiology*, *52*(11), 753-761. https://doi.org/10.3109/14992027.2013.832415

Hudspeth, A. (1997). How hearing happens. *Neuron*, *19*(5), 947-950. https://doi.org/10.1016/s0896-6273(00)80385-2

Hyönä, J., Tommola, J., & Alaja, A.-M. (1995). Pupil dilation as a measure of processing load in simultaneous interpretation and other language tasks. *The Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, *48*(3), 598-612. https://doi.org/10.1080/14640749508401407

James, L. R., & Brett, J. M. (1984). Mediators, moderators, and tests for mediation. *Journal of Applied Psychology*, *69*(2), 307-321. https://doi.org/10.1037/0021-9010.69.2.307

Jänig, W. (2013). Autonomic nervous system. In R. F. Schmidt & G. Thews (Eds.), *Human Physiology* (2nd ed., pp. 333 - 369). Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-642-73831-9

Jayakody, D. M. P., Almeida, O. P., Speelman, C. P., Bennett, R. J., Moyle, T. C., Yiannos, J. M., & Friedland, P. L. (2018). Association between speech and high-frequency hearing loss and depression, anxiety and stress in older adults. *Maturitas*, *110*, 86-91. https://doi.org/10.1016/j.maturitas.2018.02.002

Jepma, M., & Nieuwenhuis, S. (2011). Pupil diameter predicts changes in the exploration–exploitation trade-off: Evidence for the adaptive gain theory. *Journal of Cognitive Neuroscience*, *23*(7), 1587-1596. https://doi.org/10.1162/jocn.2010.21548

Jin, P. (1992). Toward a reconceptualization of the law of initial value. *Psychological Bulletin*, *111*(1), 176–184. https://doi.org/10.1037/0033-2909.111.1.176

Johnson, P. C. D. (2014). Extension of Nakagawa & Schielzeth's R2GLMM to random slopes models [10.1111/2041-210X.12225]. *Methods in Ecology and Evolution*, *5*(9), 944-946. https://doi.org/https://doi.org/10.1111/2041-210X.12225

Joos, K. M., & Melson, M. R. (2012). Control of the Pupil. In D. Robertson, I. Biaggioni, G. Burnstock, P. A. Low, & J. F. R. Paton (Eds.), *Primer on the Autonomic Nervous System* (pp. 239-242). Academic Press. https://doi.org/10.1016/B978-0-12-386525-0.00049-4

Joshi, S., & Gold, J. I. (2020). Pupil size as a window on neural substrates of cognition. *Trends in Cognitive Sciences*, *24*(6), 466-480. https://doi.org/10.1016/j.tics.2020.03.005

Joshi, S., Li, Y., Kalwani, Rishi M., & Gold, Joshua I. (2016). Relationships between pupil diameter and neuronal activity in the locus coeruleus, colliculi, and cingulate cortex. *Neuron*, *89*(1), 221-234. https://doi.org/10.1016/j.neuron.2015.11.028

Ju, W. (2018). The Autonomic Nervous System. In G. Bains, I. Barany, M. Shcherbina, & S. Lee (Eds.), *Neuroscience: Canadian 1st Edition Open Textbook*. Pressbooks. http://neuroscience.openetext.utoronto.ca/about/

Judd, C. M., & Kenny, D. A. (1981). Process analysis: Estimating mediation in evaluating research. *Evaluation Review*, *5*(5), 602–619. https://doi.org/10.1177/0193841X8100500502

Judd, C. M., Kenny, D. A., & McClelland, G. H. (2001). Estimating and testing mediation and moderation in within-subject designs. *Psychological Methods*, *6*(2), 115-134. https://doi.org/10.1037/1082-989X.6.2.115

Judd, C. M., Westfall, J., & Kenny, D. A. (2012). Treating stimuli as a random factor in social psychology: A new and comprehensive solution to a pervasive but largely ignored problem. *Journal of Personality and Social Psychology*, *103*(1), 54-69. https://doi.org/10.1037/a0028347

Jung, R. E., & Haier, R. J. (2007). The Parieto-Frontal Integration Theory (P-FIT) of intelligence: Converging neuroimaging evidence. *The Behavioral and Brain Sciences*, *30*(2), 135-154. https://doi.org/10.1017/s0140525x07001185

Kahneman, D. (1973). *Attention and effort*. Prentice-Hall.

Kahneman, D., & Beatty, J. (1966). Pupil diameter and load on memory. *Science*, *154*(3756), 1583-1585. https://doi.org/10.1126/science.154.3756.1583

Kahneman, D., & Beatty, J. (1967). Pupillary responses in a pitch-discrimination task. *Perception & Psychophysics*, *2*(3), 101-105. https://doi.org/10.3758/BF03210302

Kane, M. J., & Engle, R. W. (2002). The role of prefrontal cortex in working-memory capacity, executive attention, and general fluid intelligence: An individual-differences perspective. *Psychonomic Bulletin & Review*, *9*(4), 637-671. https://doi.org/10.3758/BF03196323

Kardon, R. (2011). Regulation of light through the pupil. In L. A. Levin, S. F. E. Nilsson, J. ver Hoeve, S. M. Wu, P. L. Kaufman, & A. Alm (Eds.), *Adler's Physiology of the Eye* (11th ed., pp. 502-526). Elsevier Health Sciences.

Kassambara, A. (2021). *rstatix: Pipe-friendly framework for basic statistical tests*. In (Version 0.7.0) R package. https://CRAN.R-project.org/package=rstatix

Keidser, G., Ching, T., Dillon, H., Agung, K., Brew, C., Brewer, S., Fisher, M., Foster, L., Grant, F., & Storey, L. (2002). The National Acoustic Laboratories (NAL) CDs of speech and noise for hearing aid evaluation: Normative data and potential applications. *The Australian and New Zealand Journal of Audiology*, *24*(1), 16-35. https://doi.org/10.1375/audi.24.1.16.31112

Keppel, G. (1991). *Design and analysis: A researcher's handbook* (3rd ed.). Prentice-Hall, Inc.

Kidd, G., Mason, C. R., Richards, V. M., Gallun, F. J., & Durlach, N. I. (2008). Informational Masking. In W. A. Yost, A. N. Popper, & R. R. Fay (Eds.), *Auditory Perception of Sound Sources* (pp. 143-189). Springer US. https://doi.org/10.1007/978-0-387-71305-2_6

Knief, U., & Forstmeier, W. (2020). Violating the normality assumption may be the lesser of two evils [Preprint]. *bioRxiv*, 1-33. https://doi.org/10.1101/498931

Koelewijn, T., de Kluiver, H., Shinn-Cunningham, B. G., Zekveld, A. A., & Kramer, S. E. (2015). The pupil response reveals increased listening effort when it is difficult to focus attention. *Hearing Research*, *323*, 81-90. https://doi.org/10.1016/j.heares.2015.02.004

Koelewijn, T., Shinn-Cunningham, B. G., Zekveld, A. A., & Kramer, S. E. (2014). The pupil response is sensitive to divided attention during speech processing. *Hearing Research*, *312*, 114-120. https://doi.org/10.1016/j.heares.2014.03.010

Koelewijn, T., Zekveld, A. A., Festen, J. M., & Kramer, S. E. (2012). Pupil dilation uncovers extra listening effort in the presence of a single-talker masker. *Ear and Hearing*, *33*(2), 291-300. https://doi.org/10.1097/AUD.0b013e3182310019

Koelewijn, T., Zekveld, A. A., Festen, J. M., & Kramer, S. E. (2014). The influence of informational masking on speech perception and pupil response in adults with hearing impairment. *The Journal of the Acoustical Society of America*, *135*(3), 1596-1606. https://doi.org/10.1121/1.4863198

Koelewijn, T., Zekveld, A. A., Festen, J. M., Rönnberg, J., & Kramer, S. E. (2012). Processing load induced by informational masking is related to linguistic abilities. *International Journal of Otolaryngology*, *2012*(865731), 1-11. https://doi.org/10.1155/2012/865731

Koelewijn, T., Zekveld, A. A., Lunner, T., & Kramer, S. E. (2018). The effect of reward on listening effort as reflected by the pupil dilation response. *Hearing Research*, *367*, 106-112. https://doi.org/10.1016/j.heares.2018.07.011

Koelewijn, T., Zekveld, A. A., Lunner, T., & Kramer, S. E. (2021). The effect of monetary reward on listening effort and sentence recognition. *Hearing Research*, *406*(108255), 1-9. https://doi.org/10.1016/j.heares.2021.108255

Koss, M. C. (1986). Pupillary dilation as an index of central nervous system alpha 2-adrenoceptor activation. *Journal of Pharmacological Methods*, *15*(1), 1-19. https://doi.org/10.1016/0160-5402(86)90002-1

Kramer, S. E., Kapteyn, T. S., Festen, J. M., & Kuik, D. J. (1997). Assessing aspects of auditory handicap by means of pupil dilatation. *Audiology*, *36*(3), 155-164. https://doi.org/10.3109/00206099709071969

Kramer, S. E., Kapteyn, T. S., & Houtgast, T. (2006). Occupational performance: Comparing normally-hearing and hearing-impaired employees using the Amsterdam Checklist for Hearing and Work. *International Journal of Audiology*, *45*(9), 503-512. https://doi.org/10.1080/14992020600754583

Kramer, S. E., Teunissen, C. E., & Zekveld, A. A. (2016). Cortisol, chromogranin a, and pupillary responses evoked by speech recognition tasks in normally hearing and hard-of-hearing listeners: A pilot study. *Ear and Hearing*, *37*, 126S-135S. https://doi.org/10.1097/aud.0000000000000311

Kret, M. E., & Sjak-Shie, E. E. (2019). Preprocessing pupil size data: Guidelines and code. *Behavior Research Methods*, *51*(3), 1336-1342. https://doi.org/10.3758/s13428-018-1075-y

Kruschke, J. (2014). *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan* (2nd ed.). Academic Press.

Książek, P., Zekveld, A. A., Wendt, D., Fiedler, L., Lunner, T., & Kramer, S. E. (2021). Effect of Speech-to-Noise Ratio and Luminance on a Range of Current and Potential Pupil Response Measures to Assess Listening Effort. *Trends in Hearing*, *25*, 1-18. https://doi.org/10.1177/23312165211009351

Kuchinsky, S. E., Ahlstrom, J. B., Cute, S. L., Humes, L. E., Dubno, J. R., & Eckert, M. A. (2014). Speech-perception training for older adults with hearing loss impacts word recognition and effort. *Psychophysiology*, *51*(10), 1046-1057. https://doi.org/10.1111/psyp.12242

Kuchinsky, S. E., Ahlstrom, J. B., Vaden JR., K. I., Cute, S. L., Humes, L. E., Dubno, J. R., & Eckert, M. A. (2013). Pupil size varies with word listening and response selection difficulty in older adults with hearing loss. *Psychophysiology*, *50*(1), 23-34. https://doi.org/10.1111/j.1469-8986.2012.01477.x

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(1), 1-26. https://doi.org/10.18637/jss.v082.i13

Lacey, J. I. (1956). The evaluation of autonomic responses: Toward a general solution [10.1111/j.1749-6632.1956.tb46040.x]. *Annals of the New York Academy of Sciences*, *67*(5), 125-163. https://doi.org/10.1111/j.1749-6632.1956.tb46040.x

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs [Review]. *Frontiers in Psychology*, *4*(863), 1-12. https://doi.org/10.3389/fpsyg.2013.00863

Larsby, B., Hällgren, M., Lyxell, B., & Arlinger, S. (2005). Cognitive performance and perceived effort in speech processing tasks: Effects of different noise backgrounds in normal-hearing and hearing-impaired subjects. *International Journal of Audiology*, *44*(3), 131-143. https://doi.org/10.1080/14992020500057244

Larsen, R. S., & Waters, J. (2018). Neuromodulatory correlates of pupil dilation. *Frontiers in Neural Circuits*, *12*(21), 1-9. https://doi.org/10.3389/fncir.2018.00021

Lawrence, M. A. (2016). *ez: Easy Analysis and Visualization of Factorial Experiments*. In (Version 4.4-0) [R Package]. https://CRAN.R-project.org/package=ez

Lawrence, R. J., Wiggins, I. M., Anderson, C. A., Davies-Thompson, J., & Hartley, D. E. H. (2018). Cortical correlates of speech intelligibility measured using functional near-infrared spectroscopy (fNIRS). *Hearing Research*, *370*, 53-64. https://doi.org/10.1016/j.heares.2018.09.005

Lenth, R. V. (2021). *emmeans: Estimated Marginal Means, aka Least-Squares Means*. In (Version 1.6.1) R package. https://CRAN.R-project.org/package=emmean

Lesica, N. A. (2018). Why do hearing aids fail to restore normal auditory perception? *Trends in Neurosciences*, *41*(4), 174-185. https://doi.org/10.1016/j.tins.2018.01.008

Leuchs, L., Schneider, M., Czisch, M., & Spoormaker, V. I. (2017). Neural correlates of pupil dilation during human fear learning. *Neuroimage*, *147*, 186-197. https://doi.org/10.1016/j.neuroimage.2016.11.072

Little, R. J. A. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association*, *83*(404), 1198-1202. https://doi.org/10.1080/01621459.1988.10478722

Loewenfeld, I. E. (1958). Mechanisms of reflex dilatation of the pupil. *Documenta Ophthalmologica*, *12*(1), 185-448. https://doi.org/10.1007/BF00913471

Loewenfeld, I. E., & Lowenstein, O. (1993). Reflex Dilation. In *The Pupil: Anatomy, Physiology, and Clinical Applications* (pp. 318-395). Butterworth Heinemann.

Lopes da Silva, F. (2013). EEG and MEG: Relevance to neuroscience. *Neuron*, *80*(5), 1112-1128. https://doi.org/10.1016/j.neuron.2013.10.017

Lowenstein, O., & Loewenfeld, I. E. (1950). Role of sympathetic and parasympathetic systems in reflex dilatation of the pupil: Pupillographic studies. *Archives of Neurology & Psychiatry*, *64*(3), 313-340. https://doi.org/10.1001/archneurpsyc.1950.02310270002001

Lowenstein, O., & Loewenfeld, I. E. (1962). The Pupil. In H. Davson (Ed.), *The Eye* (3 ed., pp. 231-265). Academic Press Inc.

Luck, S. J. (2014). *An introduction to the event-related potential technique* (2nd ed.). MIT press.

Lüdecke, D., Ben-Shachar, M. S., Patil, I., Waggoner, P., & Makowski, D. (2021). performance: An R package for assessment, comparison and testing of statistical models. *Journal of Open Source Software*, *6*(60), 1-8. https://doi.org/10.21105/joss.03139

Lunner, T., Rudner, M., & Rönnberg, J. (2009). Cognition and hearing aids. *Scandinavian Journal of Psychology*, *50*(5), 395-403. https://doi.org/10.1111/j.1467-9450.2009.00742.x

Mackersie, C. L., & Calderon-Moultrie, N. (2016). Autonomic nervous system reactivity during speech repetition tasks: Heart rate variability and skin conductance. *Ear and Hearing*, *37*, 118S-125S. https://doi.org/10.1097/aud.0000000000000305

Mackersie, C. L., & Cones, H. (2011). Subjective and psychophysiological indexes of listening effort in a competing-talker task. *Journal of the American Academy of Audiology*, *22*(2), 113-122. https://doi.org/10.3766/jaaa.22.2.6

Mackersie, C. L., MacPhee, I. X., & Heldt, E. W. (2015). Effects of hearing loss on heart rate variability and skin conductance measured during sentence recognition in noise. *Ear and Hearing*, *36*(1), 145-154. https://doi.org/10.1097/AUD.0000000000000091

MacKinnon, D. P., Fairchild, A. J., & Fritz, M. S. (2007). Mediation analysis. *Annual Review of Psychology*, *58*(1), 593-614. https://doi.org/10.1146/annurev.psych.58.110405.085542

MacKinnon, D. P., & Pirlott, A. G. (2015). Statistical approaches for enhancing causal interpretation of the M to Y relation in mediation analysis. *Personality and Social Psychology Review*, *19*(1), 30-43. https://doi.org/10.1177/1088868314542878

Mahr, T. J. (2017, June 22). Plotting partial pooling in mixed-effects models. https://www.tjmahr.com/plotting-partial-pooling-in-mixed-effects-models/

Mandrick, K., Peysakhovich, V., Rémy, F., Lepron, E., & Causse, M. (2016). Neural and psychophysiological correlates of human performance under stress and high mental workload. *Biological Psychology*, *121*, 62-73. https://doi.org/10.1016/j.biopsycho.2016.10.002

Mathôt, S. (2018). Pupillometry: Psychology, physiology, and function. *Journal of Cognition*, *1*(1), 1-16. https://doi.org/10.5334/joc.18

Mathôt, S., Fabius, J., Van Heusden, E., & Van der Stigchel, S. (2018). Safe and sensible preprocessing and baseline correction of pupil-size data. *Behavior Research Methods*, *50*(1), 94-106. https://doi.org/10.3758/s13428-017-1007-2

Matthews, P. M., Honey, G. D., & Bullmore, E. T. (2006). Applications of fMRI in translational medicine and clinical practice. *Nature Reviews Neuroscience*, *7*(9), 732-744. https://doi.org/10.1038/nrn1929

Mattys, S. L., Brooks, J., & Cooke, M. (2009). Recognizing speech under a processing load: Dissociating energetic from informational factors. *Cognitive Psychology*, *59*(3), 203-243. https://doi.org/10.1016/j.cogpsych.2009.04.001

McAuliffe, M. J., Wilding, P. J., Rickard, N. A., & O'Beirne, G. A. (2012). Effect of speaker age on speech recognition and perceived listening effort in older adults with hearing loss. *Journal of Speech, Language, and Hearing Research*, *55*(3), 838-847. https://doi.org/10.1044/1092-4388(2011/11-0101)

McCormack, A., & Fortnum, H. (2013). Why do people fitted with hearing aids not wear them? *International Journal of Audiology*, *52*(5), 360-368. https://doi.org/10.3109/14992027.2013.769066

McCorry, L. K. (2007). Physiology of the autonomic nervous system. *American journal of pharmaceutical education*, *71*(4), 1-11. https://doi.org/10.5688/aj710478

McCulloch, C. E. (2005). Repeated Measures ANOVA, R.I.P.? *CHANCE*, *18*(3), 29-33. https://doi.org/10.1080/09332480.2005.10722732

McElreath, R. (2020). *Statistical rethinking: A Bayesian course with examples in R and Stan*. CRC press.

McGarrigle, R., Dawes, P., Stewart, A. J., Kuchinsky, S. E., & Munro, K. J. (2017a). Measuring listening-related effort and fatigue in school-aged children using pupillometry. *Journal of Experimental Child Psychology*, *161*, 95-112. https://doi.org/10.1016/j.jecp.2017.04.006

McGarrigle, R., Dawes, P., Stewart, A. J., Kuchinsky, S. E., & Munro, K. J. (2017b). Pupillometry reveals changes in physiological arousal during a sustained listening task. *Psychophysiology*, *54*(2), 193-203. https://doi.org/10.1111/psyp.12772

McGarrigle, R., Gustafson, S. J., Hornsby, B. W. Y., & Bess, F. H. (2019). Behavioral measures of listening effort in school-age children: Examining the effects of signal-to-noise ratio, hearing loss, and amplification. *Ear and Hearing*, *40*(2), 381-392. https://doi.org/10.1097/aud.0000000000000623

McGarrigle, R., Knight, S., Rakusen, L., Geller, J., & Mattys, S. (2021). Older adults show a more sustained pattern of effortful listening than young adults. *Psychology and Aging*, *36*(4), 504-519. https://doi.org/10.1037/pag0000587

McGarrigle, R., Munro, K., Dawes, P., Stewart, A., Moore, D., Barry, J., & Amitay, S. (2014). Listening effort and fatigue: what exactly are we measuring? A British Society of Audiology Cognition in Hearing Special Interest Group 'white paper'. *International Journal of Audiology*, *53*(7), 433-440. https://doi.org/10.3109/14992027.2014.890296

McGarrigle, R., Rakusen, L., & Mattys, S. (2021). Effortful listening under the microscope: Examining relations between pupillometric and subjective markers of effort and tiredness from listening. *Psychophysiology*, *58*(e13703), 1-22. https://doi.org/10.1111/psyp.13703

McGinley, Matthew J., David, Stephen V., & McCormick, David A. (2015). Cortical Membrane Potential Signature of Optimal States for Sensory Signal Detection. *Neuron*, *87*(1), 179-192. https://doi.org/10.1016/j.neuron.2015.05.038

McLaughlin, D. J., & Engen, K. J. V. (2020). Task-evoked pupil response for accurately recognized accented speech. *The Journal of the Acoustical Society of America*, *147*(2), 151-156. https://doi.org/10.1121/10.0000718

McLaughlin, D. J., Zink, M. E., Gaunt, L., Brent, S., Van Engen, K. J., Sommers, M. S., & Peelle, J. E. (2021). Pupillometry reveals cognitive demands of lexical competition during spoken word recognition in young and older adults. *Psychonomic Bulletin & Review*, *Advance online publication*. https://doi.org/10.3758/s13423-021-01991-0

McNair, D., Lorr, M., DroppLemn, L. (1971). *Manual Profile of Mood States*. Educational and Industrial Testing Service.

Mele, G., Cavaliere, C., Alfano, V., Orsini, M., Salvatore, M., & Aiello, M. (2019). Simultaneous EEG-fMRI for functional neurological assessment [Review]. *Frontiers in Neurology*, *10*(848), 1-11. https://doi.org/10.3389/fneur.2019.00848

Meteyard, L., & Davies, R. A. I. (2020). Best practice guidance for linear mixed-effects models in psychological science. *Journal of Memory and Language*, *112*(104092), 1-22. https://doi.org/10.1016/j.jml.2020.104092

Middlebrooks, J. C., & Simon, J. Z. (2017). Ear and brain mechanisms for parsing the auditory scene. In J. C. Middlebrooks, J. Z. Simon, A. N. Popper, & R. R. Fay (Eds.), *The Auditory System at the Cocktail Party* (pp. 1-6). Springer International Publishing. https://doi.org/10.1007/978-3-319-51662-2_1

Miles, K., McMahon, C., Boisvert, I., Ibrahim, R., de Lissa, P., Graham, P., & Lyxell, B. (2017). Objective assessment of listening effort: Coregistration of pupillometry and EEG. *Trends in Hearing*, *21*, 1-13. https://doi.org/10.1177/2331216517706396

Milne, A. E., Zhao, S., Tampakaki, C., Bury, G., & Chait, M. (2021). Sustained pupil responses are modulated by predictability of auditory sequences. *The Journal of Neuroscience*, *41*(28), 6116-6127. https://doi.org/10.1523/jneurosci.2879-20.2021

Mirman, D., & Klein, M. (2014). Individual Differences. In *Growth Curve Analysis and Visualization Using R*. CRC Press LLC. http://ebookcentral.proquest.com/lib/flinders/detail.action?docID=1408034

Moore, D. R. (2015). Sources of pathology underlying listening disorders in children. *International Journal of Psychophysiology*, *95*(2), 125-134. https://doi.org/10.1016/j.ijpsycho.2014.07.006

Moore, T. M., Key, A. P., Thelen, A., & Hornsby, B. W. Y. (2017). Neural mechanisms of mental fatigue elicited by sustained auditory processing. *Neuropsychologia*, *106*, 371-382. https://doi.org/10.1016/j.neuropsychologia.2017.10.025

Moore, T. M., & Picou, E. M. (2018). A potential bias in subjective ratings of mental effort. *Journal of Speech, Language, and Hearing Research*, *61*(9), 2405-2421. https://doi.org/10.1044/2018_JSLHR-H-17-0451

Moritz, S., & Bartz-Beielstein, T. (2017). imputeTS: Time series missing value imputation in r. *The R Journal*, *9*(1), 207-218. https://doi.org/10.32614/RJ-2017-009

Murphy, P. R., O'Connell, R. G., O'Sullivan, M., Robertson, I. H., & Balsters, J. H. (2014). Pupil diameter covaries with BOLD activity in human locus coeruleus. *Human Brain Mapping*, *35*(8), 4140-4154. https://doi.org/10.1002/hbm.22466

Murphy, P. R., Robertson, I. H., Balsters, J. H., & O'connell, R. G. (2011). Pupillometry and P3 index the locus coeruleus–noradrenergic arousal function in humans. *Psychophysiology*, *48*(11), 1532-1543. https://doi.org/10.1111/j.1469-8986.2011.01226.x

Nachtegaal, J., Kuik, D. J., Anema, J. R., Goverts, S. T., Festen, J. M., & Kramer, S. E. (2009). Hearing status, need for recovery after work, and psychosocial work characteristics: Results from an internet-based national survey on hearing. *International Journal of Audiology*, *48*(10), 684-691. https://doi.org/10.1080/14992020902962421

Nachtegaal, J., Smit, J. H., Smits, C., Bezemer, P. D., van Beek, J. H. M., Festen, J. M., & Kramer, S. E. (2009). The association between hearing status and psychosocial health before the age of 70 years: Results from an internet-based national survey on hearing. *Ear and Hearing*, *30*(3), 302-312. https://doi.org/10.1097/AUD.0b013e31819c6e01

Nagle, K. F., & Eadie, T. L. (2012). Listener effort for highly intelligible tracheoesophageal speech. *Journal of Communication Disorders*, *45*(3), 235-245. https://doi.org/10.1016/j.jcomdis.2012.01.001

Nieuwenhuis, S., Aston-Jones, G., & Cohen, J. D. (2005). Decision making, the P3, and the locus coeruleus--norepinephrine system. *Psychological Bulletin*, *131*(4), 510-532. https://doi.org/10.1037/0033-2909.131.4.510

Nieuwenhuis, S., De Geus, E. J., & Aston-Jones, G. (2011). The anatomical and functional relationship between the P3 and autonomic components of the orienting response. *Psychophysiology*, *48*(2), 162-175. https://doi.org/10.1111/j.1469-8986.2010.01057.x

Obleser, J., Wöstmann, M., Hellbernd, N., Wilsch, A., & Maess, B. (2012). Adverse listening conditions and memory load drive a common alpha oscillatory network. *The Journal of Neuroscience*, *32*(36), 12376-12383. https://doi.org/10.1523/JNEUROSCI.4908-11.2012

Ohlenforst, B., Wendt, D., Kramer, S. E., Naylor, G., Zekveld, A. A., & Lunner, T. (2018). Impact of SNR, masker type and noise reduction processing on sentence recognition performance and listening effort as indicated by the pupil dilation response. *Hearing Research*, *365*, 90-99. https://doi.org/10.1016/j.heares.2018.05.003

Ohlenforst, B., Zekveld, A. A., Jansma, E. P., Wang, Y., Naylor, G., Lorens, A., Lunner, T., & Kramer, S. E. (2017). Effects of hearing impairment and hearing aid amplification on listening effort: A systematic review. *Ear and Hearing*, *38*(3), 267-281. https://doi.org/10.1097/AUD.0000000000000396

Ohlenforst, B., Zekveld, A. A., Lunner, T., Wendt, D., Naylor, G., Wang, Y., Versfeld, N. J., & Kramer, S. E. (2017). Impact of stimulus-related factors and hearing impairment on listening effort as indicated by pupil dilation. *Hearing Research*, *351*, 68-79. https://doi.org/10.1016/j.heares.2017.05.012

Olejnik, S., & Algina, J. (2003). Generalized eta and omega squared statistics: Measures of effect size for some common research designs. *Psychological Methods*, *8*(4), 434-447. https://doi.org/10.1037/1082-989X.8.4.434

Pals, C., Sarampalis, A., Rijn, H. v., & Başkent, D. (2015). Validation of a simple response-time measure of listening effort. *The Journal of the Acoustical Society of America*, *138*(3), EL187-EL192. https://doi.org/10.1121/1.4929614

Panico, J., & Healey, E. C. (2009). Influence of text type, topic familiarity, and stuttering frequency on listener recall, comprehension, and mental effort. *Journal of Speech, Language, and Hearing Research*, *52*(2), 534-546. https://doi.org/10.1044/1092-4388(2008/07-0238)

Paredes-Gallardo, A., Innes-Brown, H., Madsen, S. M. K., Dau, T., & Marozeau, J. (2018). Auditory stream segregation and selective attention for cochlear implant listeners: Evidence from behavioral measures and event-related potentials [Original Research]. *Frontiers in Neuroscience*, *12*(581), 1-15. https://doi.org/10.3389/fnins.2018.00581

Peelle, J. E. (2014). Methodological challenges and solutions in auditory functional magnetic resonance imaging [Review]. *Frontiers in Neuroscience*, *8*(253), 1-13. https://doi.org/10.3389/fnins.2014.00253

Peelle, J. E. (2018). Listening effort: How the cognitive consequences of acoustic challenge are reflected in brain and behavior. *Ear and Hearing*, *39*(2), 204-214. https://doi.org/10.1097/AUD.0000000000000494

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy.

Behavior Research Methods, *51*(1), 195-203. https://doi.org/10.3758/s13428-018-01193-y

Peysakhovich, V., Causse, M., Scannella, S., & Dehais, F. (2015). Frequency analysis of a task-evoked pupillary response: Luminance-independent measure of mental effort. *International Journal of Psychophysiology*, *97*(1), 30-37. https://doi.org/10.1016/j.ijpsycho.2015.04.019

Peysakhovich, V., Vachon, F., & Dehais, F. (2017). The impact of luminance on tonic and phasic pupillary responses to sustained cognitive load. *International Journal of Psychophysiology*, *112*, 40-45. https://doi.org/10.1016/j.ijpsycho.2016.12.003

Pfleging, B., Fekety, D. K., Schmidt, A., & Kun, A. L. (2016). A model relating pupil diameter to mental workload and lighting conditions. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (pp. 5776–5788). Association for Computing Machinery. https://doi.org/10.1145/2858036.2858117

Pichora-Fuller, K. M., Kramer, S. E., Eckert, M. A., Edwards, B., Hornsby, B. W., Humes, L. E., Lemke, U., Lunner, T., Matthen, M., Mackersie, C. L., Naylor, G., Phillips, N. A., Richter, M., Rudner, M., Sommers, M. S., Tremblay, K. L., & Wingfield, A. (2016). Hearing impairment and cognitive energy: The framework for understanding effortful listening (FUEL). *Ear and Hearing*, *37* (Suppl 1), 5S-27S. https://doi.org/10.1097/aud.0000000000000312

Pichora-Fuller, K. M., & Singh, G. (2006). Effects of age on auditory and cognitive processing: Implications for hearing aid fitting and audiologic rehabilitation. *Trends in Amplification*, *10*(1), 29-59. https://doi.org/10.1177/108471380601000103

Pichora-Fuller, K. M., & Souza, P. E. (2003). Effects of aging on auditory processing of speech. *International Journal of Audiology*, *42*, 1-16. https://doi.org/10.3109/14992020309074638

Picou, E. M., Bean, B., Marcrum, S. C., Ricketts, T. A., & Hornsby, B. W. Y. (2019). Moderate reverberation does not increase subjective fatigue, subjective listening effort, or behavioral listening effort in school-aged children [Original Research]. *Frontiers in Psychology*, *10*(1749), 1-16. https://doi.org/10.3389/fpsyg.2019.01749

Picou, E. M., Moore, T. M., & Ricketts, T. A. (2017). The effects of directional processing on objective and subjective listening effort. *Journal of Speech, Language, and Hearing Research*, *60*(1), 199-211. https://doi.org/10.1044/2016_JSLHR-H-15-0416

Picou, E. M., & Ricketts, T. A. (2018). The relationship between speech recognition, behavioural listening effort, and subjective ratings. *International Journal of Audiology*, *57*(6), 457-467. https://doi.org/10.1080/14992027.2018.1431696

Picou, E. M., Ricketts, T. A., & Hornsby, B. W. Y. (2011). Visual cues and listening effort: Individual variability. *Journal of Speech, Language, and Hearing Research*, *54*(5), 1416-1430. https://doi.org/10.1044/1092-4388(2011/10-0154)

Piquado, T., Isaacowitz, D., & Wingfield, A. (2010). Pupillometry as a measure of cognitive effort in younger and older adults. *Psychophysiology*, *47*(3), 560-569. https://doi.org/10.1111/j.1469-8986.2009.00947.x

Pirlott, A. G., & MacKinnon, D. P. (2016). Design approaches to experimental mediation. *Journal of Experimental Social Psychology*, *66*, 29-38. https://doi.org/10.1016/j.jesp.2015.09.012

Pomplun, M., & Sunkara, S. (2003). Pupil dilation as an indicator of cognitive workload in human-computer interaction. Proceedings of the International Conference on Human-Computer Interactions, Crete, Greece.

Pronk, M., Deeg, D. J. H., Smits, C., van Tilburg, T. G., Kuik, D. J., Festen, J. M., & Kramer, S. E. (2011). Prospective effects of hearing status on loneliness and depression in older persons: Identification of subgroups. *International Journal of Audiology*, *50*(12), 887-896. https://doi.org/10.3109/14992027.2011.599871

Quené, H., & van den Bergh, H. (2004). On multi-level modeling of data from repeated measures designs: A tutorial. *Speech Communication*, *43*(1), 103-121. https://doi.org/10.1016/j.specom.2004.02.004

R Core Team. (2020). *R: A language and environment for statistical computing*. In https://www.R-project.org/

Rabinbach, A. (1992). *The human motor: Energy, fatigue, and the origins of modernity*. University of California Press.

Radiological Society of North America. (2020, June, 15). *Magnetic Resonance, Functional (fMRI) - Brain*. https://www.radiologyinfo.org/en/info.cfm?pg=fmribrain#part_two

Reilly, J., Kelly, A., Kim, S. H., Jett, S., & Zuckerman, B. (2019). The human task-evoked pupillary response function is linear: Implications for baseline response scaling in pupillometry. *Behavior Research Methods*, *51*(2), 865-878. https://doi.org/10.3758/s13428-018-1134-4

Rennies, J., Best, V., Roverud, E., & Kidd, G. (2019). Energetic and informational components of speech-on-speech masking in binaural speech intelligibility and perceived listening effort. *Trends in Hearing*, *23*, 1-21. https://doi.org/10.1177/2331216519854597

Rennies, J., Schepker, H., Holube, I., & Kollmeier, B. (2014). Listening effort and speech intelligibility in listening situations affected by noise and reverberation. *The Journal of the Acoustical Society of America*, *136*(5), 2642-2653. https://doi.org/10.1121/1.4897398

Richter, M. (2016). The moderating effect of success importance on the relationship between listening demand and listening effort. *Ear and Hearing*, *37*, 111S-117S. https://doi.org/10.1097/aud.0000000000000295

Richter, M., Gendolla, G. H., & Wright, R. A. (2016). Three decades of research on motivational intensity theory: What we have learned about effort and what we still don't know. In A. J. Elliot (Ed.), *Advances in Motivation Science* (Vol. 3, pp. 149-186). Elsevier. https://doi.org/10.1016/BS.ADMS.2016.02.001

Rönnberg, J., Holmer, E., & Rudner, M. (2019). Cognitive hearing science and ease of language understanding. *International Journal of Audiology*, *58*(5), 247-261. https://doi.org/10.1080/14992027.2018.1551631

Rönnberg, J., Lunner, T., Zekveld, A., Sörqvist, P., Danielsson, H., Lyxell, B., Dahlström, Ö., Signoret, C., Stenfelt, S., Pichora-Fuller, M. K., & Rudner, M. (2013). The Ease of Language Understanding (ELU) model: Theoretical, empirical, and clinical advances [Review]. *Frontiers in Systems Neuroscience*, *7*(31). https://doi.org/10.3389/fnsys.2013.00031

Rönnberg, J., Rudner, M., Foo, C., & Lunner, T. (2008). Cognition counts: A working memory system for ease of language understanding (ELU). *International Journal of Audiology*, *47*(Suppl 2), S99-S105. https://doi.org/10.1080/14992020802301167

Rönnberg, N., Rudner, M., Lunner, T., & Stenfelt, S. (2014). Assessing listening effort by measuring short-term memory storage and processing of speech in noise. *Speech, Language and Hearing*, *17*(3), 123-132. https://doi.org/10.1179/2050572813Y.0000000033

Rönnberg, N., Stenfelt, S., & Rudner, M. (2011). Testing listening effort for speech comprehension using the individuals' cognitive spare capacity. *Audiology research*, *1*(e22), 82-85. https://doi.org/10.4081/audiores.2011.e22

Rosemann, S., & Thiel, C. M. (2019). The effect of age-related hearing loss and listening effort on resting state connectivity. *Scientific Reports*, *9*(2337), 1-9. https://doi.org/10.1038/s41598-019-38816-z

Rovetti, J., Goy, H., Pichora-Fuller, M. K., & Russo, F. A. (2019). Functional near-infrared spectroscopy as a measure of listening effort in older adults who use hearing aids. *Trends in Hearing*, *23*, 1-22. https://doi.org/10.1177/2331216519886722

RStudio Team. (2020). *RStudio: Integrated Development Environment for R*. In http://www.rstudio.com/

Rudner, M., Lunner, T., Behrens, T., Thorèn, E. S., & Rönnberg, J. (2012). Working memory capacity may influence perceived effort during aided speech recognition in noise. *Journal of the American Academy of Audiology*, *23*(8), 577-589. https://doi.org/10.3766/jaaa.23.7.7

Saito, H., Nishiwaki, Y., Michikawa, T., Kikuchi, Y., Mizutari, K., Takebayashi, T., & Ogawa, K. (2010). Hearing handicap predicts the development of depressive symptoms after 3 years in older community-dwelling japanese. *Journal of the American Geriatrics Society*, *58*(1), 93-97. https://doi.org/10.1111/j.1532-5415.2009.02615.x

Samuels, E. R., & Szabadi, E. (2008). Functional neuroanatomy of the noradrenergic locus coeruleus: Its roles in the regulation of arousal and autonomic function Part I: principles of functional organisation. *Current Neuropharmacology*, *6*(3), 235-253. https://doi.org/10.2174/157015908785777229

Sara, S. J. (2009). The locus coeruleus and noradrenergic modulation of cognition. *Nature Reviews Neuroscience*, *10*(3), 211-223. https://doi.org/10.1038/nrn2573

Sarampalis, A., Kalluri, S., Edwards, B., & Hafter, E. (2009). Objective measures of listening effort: Effects of background noise and noise reduction. *Journal of Speech, Language, and Hearing Research*, *52*(5), 1230-1240. https://doi.org/10.1044/1092-4388(2009/08-0111)

Scarapicchia, V., Brown, C., Mayo, C., & Gawryluk, J. R. (2017). Functional magnetic resonance imaging and functional near-infrared spectroscopy: Insights from combined recording studies. *Frontiers in Human Neuroscience*, *11*(419), 1-12. https://doi.org/10.3389/fnhum.2017.00419

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, *7*(2), 147-177. https://doi.org/10.1037/1082-989X.7.2.147

Schielzeth, H., Dingemanse, N. J., Nakagawa, S., Westneat, D. F., Allegue, H., Teplitsky, C., Réale, D., Dochtermann, N. A., Garamszegi, L. Z., & Araya-Ajoy, Y. G. (2020). Robustness of linear mixed-effects models to violations of distributional assumptions. *Methods in Ecology and Evolution*, *11*(9), 1141-1152. https://doi.org/10.1111/2041-210X.13434

Schielzeth, H., & Nakagawa, S. (2013). Nested by design: Model fitting and interpretation in a mixed model era [10.1111/j.2041-210x.2012.00251.x]. *Methods in Ecology and Evolution*, *4*(1), 14-24. https://doi.org/10.1111/j.2041-210x.2012.00251.x

Seeman, S., & Sims, R. (2015). Comparison of psychophysiological and dual-task measures of listening effort. *Journal of Speech, Language, and Hearing Research*, *58*(6), 1781-1792. https://doi.org/doi:10.1044/2015_JSLHR-H-14-0180

Shadish, W., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin.

Shahid, A., Wilkinson, K., Marcu, S., & Shapiro, C. M. (2012). Visual analogue scale to evaluate fatigue severity (VAS-F). In A. Shahid, K. Wilkinson, S. Marcu, & C. M. Shapiro (Eds.), *STOP, THAT and One Hundred Other Sleep Scales* (pp. 399-402). Springer New York. https://doi.org/10.1007/978-1-4419-9893-4_100

Shukla, A., Harper, M., Pedersen, E., Goman, A., Suen, J. J., Price, C., Applebaum, J., Hoyer, M., Lin, F. R., & Reed, N. S. (2020). Hearing loss, loneliness, and social isolation: A systematic review. *Otolaryngology–Head and Neck Surgery*, *162*(5), 622-633. https://doi.org/10.1177/0194599820910377

Sibley, C., Foroughi, C. K., Moclaire, C. M., King, K. M., Brown, N. L., & Coyne, J. T. (2019). *Additional evidence for pupil size as a measure of within-task learning* Proceedings of the Human Factors and Ergonomics Society Annual Meeting, Seattle, Washington, USA. https://journals.sagepub.com/doi/abs/10.1177/1071181319631173

Silbert, N. H., Jong, K. d., Regier, K., Albin, A., & Hao, Y.-C. (2014). Acoustic properties of multi-talker babble. *The Journal of the Acoustical Society of America*, *135*(4), 2227. https://doi.org/10.1121/1.4877284

Singmann, H., Bolker, B., Westfall, J., Aust, F., & Ben-Shachar, M. S. (2021). *afex: Analysis of Factorial Experiments*. In (Version 0.28-1) [R package]. https://CRAN.R-project.org/package=afex

Slade, K., Kramer, S. E., Fairclough, S., & Richter, M. (2021). Effortful listening: Sympathetic activity varies as a function of listening demand but parasympathetic activity does not. *Hearing Research*, *410*(108348), 2-11. https://doi.org/10.1016/j.heares.2021.108348

Sluiter, J. K., de Croon, E. M., Meijman, T. F., & Frings-Dresen, M. H. W. (2003). Need for recovery from work related fatigue and its role in the development and prediction of subjective health complaints. *Occupational and Environmental Medicine*, *60*(suppl 1), i62-i70. https://doi.org/10.1136/oem.60.suppl_1.i62

Smith, P. G. (2009). Neural regulation of the pupil. In M. D. Binder, N. Hirokawa, & U. Windhorst (Eds.), *Encyclopedia of Neuroscience* (pp. 2597-2601). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-29678-2_3799

Smits, C., & Houtgast, T. (2006). Measurements and calculations on the simple up-down adaptive procedure for speech-in-noise tests. *The Journal of the Acoustical Society of America*, *120*(3), 1608-1621. https://doi.org/10.1121/1.2221405

Solheim, J., Kværner, K. J., & Falkenberg, E.-S. (2011). Daily life consequences of hearing loss in the elderly. *Disability and Rehabilitation*, *33*(22-23), 2179-2185. https://doi.org/10.3109/09638288.2011.563815

Spector, R. H. (1990). The pupils. In W. H. K., H. W. D., & H. J. W. (Eds.), *Clinical Methods: The History, Physical, and Laboratory Examinations* (3rd ed.). Butterworths.

Sreenivasan, K. K., & Jha, A. P. (2007). Selective attention supports working memory maintenance by modulating perceptual processing of distractors. *Journal of Cognitive Neuroscience*, *19*(1), 32-41. https://doi.org/10.1162/jocn.2007.19.1.32

Srinivasan, R., & Nunez, P. L. (2012). Electroencephalography. In V. S. Ramachandran (Ed.), *Encyclopedia of Human Behavior* (2nd ed., pp. 15-23). Academic Press. https://doi.org/https://doi.org/10.1016/B978-0-12-375000-6.00395-5

Stan Development Team. (2021). *Stan modeling language users guide and reference manual*. In (Version 2.27) https://mc-stan.org

Steinhauer, S. R., Siegle, G. J., Condray, R., & Pless, M. (2004). Sympathetic and parasympathetic innervation of pupillary dilation during sustained processing. *International Journal of Psychophysiology*, *52*(1), 77-86. https://doi.org/10.1016/j.ijpsycho.2003.12.005

Stemmler, E. (2020). *R function for Little's test for data missing completely at random*. https://stats-bayes.com/post/2020/08/14/r-function-for-little-s-test-for-data-missing-completely-at-random/

Stenfelt, S., & Rönnberg, J. (2009). The signal-cognition interface: Interactions between degraded auditory signals and cognitive processes. *Scandinavian Journal of Psychology*, *50*(5), 385-393. https://doi.org/10.1111/j.1467-9450.2009.00748.x

Sussman, E. S. (2017). Auditory scene analysis: An attention perspective. *Journal of Speech, Language, and Hearing Research*, *60*(10), 2989-3000. https://doi.org/10.1044/2017_JSLHR-H-17-0041

Svinndal, E. V., Solheim, J., Rise, M. B., & Jensen, C. (2018). Hearing loss and work participation: A cross-sectional study in Norway. *International Journal of Audiology*, *57*(9), 646-656. https://doi.org/10.1080/14992027.2018.1464216

Tomasi, D., Caparelli, E. C., Chang, L., & Ernst, T. (2005). fMRI-acoustic noise alters brain activation during working memory tasks. *Neuroimage*, *27*(2), 377-386. https://doi.org/10.1016/j.neuroimage.2005.04.010

Tryon, W. W. (1975). Pupillometry: A survey of sources of variation. *Psychophysiology*, *12*(1), 90-93. https://doi.org/10.1111/j.1469-8986.1975.tb03068.x

Tsukahara, J. S., & Engle, R. W. (2021). Is baseline pupil size related to cognitive ability? Yes (under proper lighting conditions). *Cognition*, *211*(104643), 1-19. https://doi.org/10.1016/j.cognition.2021.104643

Tsukahara, J. S., Harrison, T. L., & Engle, R. W. (2016). The relationship between baseline pupil size and intelligence. *Cognitive Psychology*, *91*, 109-123. https://doi.org/10.1016/j.cogpsych.2016.10.001

Unsworth, N., Miller, A. L., & Robison, M. K. (2020). Is working memory capacity related to baseline pupil diameter? *Psychonomic Bulletin & Review*. https://doi.org/10.3758/s13423-020-01817-5

Unsworth, N., & Robison, M. K. (2018). Tracking arousal state and mind wandering with pupillometry. *Cognitive, Affective, & Behavioral Neuroscience*, *18*(4), 638-664. https://doi.org/10.3758/s13415-018-0594-4

Unsworth, N., Robison, M. K., & Miller, A. L. (2019). Individual differences in baseline oculometrics: Examining variation in baseline pupil diameter, spontaneous eye blink rate, and fixation stability. *Cognitive, Affective, & Behavioral Neuroscience*, *19*(4), 1074-1093. https://doi.org/10.3758/s13415-019-00709-z

van den Brink, R. L., Murphy, P. R., & Nieuwenhuis, S. (2016). Pupil diameter tracks lapses of attention. *PLoS ONE*, *11*(10), e0165274. https://doi.org/10.1371/journal.pone.0165274

van der Linden, D. (2011). The urge to stop: The cognitive and biological nature of acute mental fatigue. In *Cognitive fatigue: Multidisciplinary perspectives on current* *research and future applications.* (pp. 149-164). American Psychological Association. https://doi.org/10.1037/12343-007

van der Linden, D., Frese, M., & Meijman, T. F. (2003). Mental fatigue and the control of cognitive processes: effects on perseveration and planning. *Acta Psychologica*, *113*(1), 45-65. https://doi.org/10.1016/S0001-6918(02)00150-6

Van Der Meer, E., Beyer, R., Horn, J., Foth, M., Bornemann, B., Ries, J., Kramer, J., Warmuth, E., Heekeren, H. R., & Wartenburger, I. (2010). Resource allocation and fluid intelligence: Insights from pupillometry [10.1111/j.1469-8986.2009.00884.x]. *Psychophysiology*, *47*(1), 158-169. https://doi.org/https://doi.org/10.1111/j.1469-8986.2009.00884.x

Van Engen, K. J., & Peelle, J. E. (2014). Listening effort and accented speech. *Frontiers in Human Neuroscience*, *8*(577), 1-4. https://doi.org/10.3389/fnhum.2014.00577

van Ravenzwaaij, D., Cassey, P., & Brown, S. D. (2018). A simple introduction to Markov Chain Monte–Carlo sampling. *Psychonomic Bulletin & Review*, *25*(1), 143-154. https://doi.org/10.3758/s13423-016-1015-8

Varazzani, C., San-Galli, A., Gilardeau, S., & Bouret, S. (2015). Noradrenaline and dopamine neurons in the reward/effort trade-off: A direct electrophysiological comparison in behaving monkeys. *The Journal of Neuroscience*, *35*(20), 7866-7877. https://doi.org/10.1523/JNEUROSCI.0454-15.2015

Vehtari, A., Gelman, A., Simpson, D., Carpenter, B., & Bürkner, P.-C. (2019). Rank-normalization, folding, and localization: An improved R-hat for assessing convergence of MCMC [Preprint]. *arXiv*. https://doi.org/arXiv:1903.08008v5

Volpert-Esmond, H. I., Merkle, E. C., Levsen, M. P., Ito, T. A., & Bartholow, B. D. (2018). Using trial-level data and multilevel modeling to investigate within-task change in event-related potentials. *Psychophysiology*, *55*(5), e13044. https://doi.org/10.1111/psyp.13044

Volpert-Esmond, H. I., Page-Gould, E., & Bartholow, B. D. (2021). Using multilevel models for the analysis of event-related potentials. *International Journal of Psychophysiology*, *162*, 145-156. https://doi.org/10.1016/j.ijpsycho.2021.02.006

Vossen, H., Van Breukelen, G., Hermens, H., Van Os, J., & Lousberg, R. (2011). More potential in statistical analyses of event-related potentials: a mixed regression approach. *International Journal of Methods in Psychiatric Research*, *20*(3), e56-e68. https://doi.org/10.1002/mpr.348

Vuorre, M., & Bolger, N. (2018). Within-subject mediation analysis for experimental data in cognitive psychology and neuroscience. *Behavior Research Methods*, *50*(5), 2125-2143. https://doi.org/10.3758/s13428-017-0980-9

Wagner, A. E., Nagels, L., Toffanin, P., Opie, J. M., & Başkent, D. (2019). Individual variations in effort: Assessing pupillometry for the hearing impaired. *Trends in Hearing*, *23*, 1-18. https://doi.org/10.1177/2331216519845596

Wagner, A. E., Toffanin, P., & Başkent, D. (2016). The timing and effort of lexical access in natural and degraded speech. *Frontiers in Psychology*, *7*(398), 1-14. https://doi.org/10.3389/fpsyg.2016.00398

Wang, Y., Kramer, S. E., Wendt, D., Naylor, G., Lunner, T., & Zekveld, A. A. (2018). The pupil dilation response during speech perception in dark and light: The involvement of the parasympathetic nervous system in listening effort. *Trends in Hearing*, *22*, 1-11. https://doi.org/10.1177/2331216518816603

Wang, Y., Naylor, G., Kramer, S. E., Zekveld, A. A., Wendt, D., Ohlenforst, B., & Lunner, T. (2018). Relations between self-reported daily-life fatigue, hearing status, and pupil dilation during a speech perception in noise task. *Ear and Hearing*, *39*(3), 573-582. https://doi.org/10.1097/aud.0000000000000512

Wang, Y., Zekveld, A. A., Wendt, D., Lunner, T., Naylor, G., & Kramer, S. E. (2018). Pupil light reflex evoked by light-emitting diode and computer screen: Methodology and association with need for recovery in daily life. *PLoS ONE*, *13*(6), e0197739. https://doi.org/10.1371/journal.pone.0197739

Weisz, N., Hartmann, T., Müller, N., & Obleser, J. (2011). Alpha rhythms in audition: Cognitive and clinical perspectives [Review]. *Frontiers in Psychology*, *2*(73). https://doi.org/10.3389/fpsyg.2011.00073

Wendt, D., Dau, T., & Hjortkjær, J. (2016). Impact of background noise and sentence complexity on processing demands during sentence comprehension [10.3389/fpsyg.2016.00345]. *Frontiers in Psychology*, *7*(345), 1-12. https://www.frontiersin.org/article/10.3389/fpsyg.2016.00345

Wendt, D., Hietkamp, R. K., & Lunner, T. (2017). Impact of noise and noise reduction on processing effort: A pupillometry study. *Ear and Hearing*, *38*(6), 690-700. https://doi.org/10.1097/AUD.0000000000000454

Wendt, D., Koelewijn, T., Książek, P., Kramer, S. E., & Lunner, T. (2018). Toward a more comprehensive understanding of the impact of masker type and signal-to-noise ratio on the pupillary response while performing a speech-in-noise test. *Hearing Research*, *369*, 67-78. https://doi.org/10.1016/j.heares.2018.05.006

Wetzel, N., Einhäuser, W., & Widmann, A. (2020). Picture-evoked changes in pupil size predict learning success in children. *Journal of Experimental Child Psychology*, *192*, 104787. https://doi.org/10.1016/j.jecp.2019.104787

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. In (Version 3.3.3) Springer-Verlag New York. https://ggplot2.tidyverse.org

Wijayasiri, P., Hartley, D. E. H., & Wiggins, I. M. (2017). Brain activity underlying the recovery of meaning from degraded speech: A functional near-infrared spectroscopy (fNIRS) study. *Hearing Research*, *351*, 55-67. https://doi.org/10.1016/j.heares.2017.05.010

Wild, C. J., Yusuf, A., Wilson, D. E., Peelle, J. E., Davis, M. H., & Johnsrude, I. S. (2012). Effortful listening: The processing of degraded speech depends critically on attention. *The Journal of Neuroscience*, *32*(40), 14010-14021. https://doi.org/10.1523/JNEUROSCI.1528-12.2012

Wilder, J. (1957). The law of initial value in neurology and psychiatry. *Journal of Nervous and Mental Disease*, *125*, 73-86. https://doi.org/10.1097/00005053-195701000-00009

Wilder, J. (1958). Modern psychophysiology and the law of initial value. *American Journal of Psychotherapy*, *12*(2), 199-221. https://doi.org/10.1176/appi.psychotherapy.1958.12.2.199

Wilhelm, H. (2011). Disorders of the pupil. In C. Kennard & R. J. Leigh (Eds.), *Handbook of Clinical Neurology* (Vol. 102, pp. 427-466). Elsevier. https://doi.org/https://doi.org/10.1016/B978-0-444-52903-9.00022-4

Winn, M. B., Edwards, J. R., & Litovsky, R. Y. (2015). The impact of auditory spectral resolution on listening effort revealed by pupil dilation. *Ear and Hearing*, *36*(4), e153-e165. https://doi.org/10.1097/aud.0000000000000145

Winn, M. B., & Teece, K. H. (2021). Listening effort is not the same as speech intelligibility score. *Trends in Hearing*, *25*, 1-26. https://doi.org/10.1177/23312165211027688

Winn, M. B., Wendt, D., Koelewijn, T., & Kuchinsky, S. E. (2018). Best practices and advice for using pupillometry to measure listening effort: An introduction for those who want to get started. *Trends in Hearing*, *22*, 1-32. https://doi.org/10.1177/2331216518800869

World Health Organization. (2021). Deafness and hearing loss. https://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss

Xu, J., Wang, Y., Chen, F., & Choi, E. (2011). Pupillary response based cognitive workload measurement under luminance changes. In P. Campos, N. Graham, J. Jorge, N.

Nunes, P. Palanque, & M. Winckler, *Human-Computer Interaction – INTERACT 2011* Human-Computer Interaction – INTERACT, Berlin, Heidelberg.

Yerkes, R. M., & Dodson, J. D. (1908). The Relation of Strength of Stimulus to Rapidity of Habit-Formation. *Journal of Comparative Neurology and Psychology*, *18*, 459-482. https://doi.org/10.1002/cne.920180503

Zekveld, A. A., Festen, J. M., & Kramer, S. E. (2013). Task difficulty differentially affects two measures of processing load: The pupil response during sentence processing and delayed cued recall of the sentences. *Journal of Speech, Language, and Hearing Research*, *56*(4), 1156-1165. https://doi.org/10.1044/1092-4388(2012/12-0058)

Zekveld, A. A., Koelewijn, T., & Kramer, S. E. (2018). The pupil dilation response to auditory stimuli: Current state of knowledge. *Trends in Hearing*, *22*, 1-22. https://doi.org/10.1177/2331216518777174

Zekveld, A. A., & Kramer, S. E. (2014). Cognitive processing load across a wide range of listening conditions: Insights from pupillometry. *Psychophysiology*, *51*(3), 277-284. https://doi.org/10.1111/psyp.12151

Zekveld, A. A., Kramer, S. E., & Festen, J. M. (2010). Pupil response as an indication of effortful listening: The influence of sentence intelligibility. *Ear and Hearing*, *31*(4), 480-490. https://doi.org/10.1097/AUD.0b013e3181d4f251

Zekveld, A. A., Kramer, S. E., & Festen, J. M. (2011). Cognitive load during speech perception in noise: The influence of age, hearing loss, and cognition on the pupil response. *Ear and Hearing*, *32*(4), 498-510. https://doi.org/10.1097/AUD.0b013e31820512bb

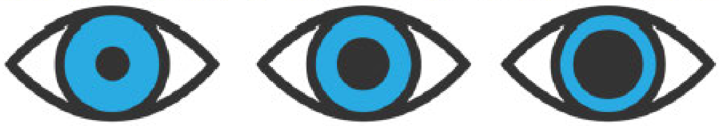
Zekveld, A. A., Rudner, M., Kramer, S. E., Lyzenga, J., & Rönnberg, J. (2014). Cognitive processing load during listening is reduced more by decreasing voice similarity than by increasing spatial separation between target and masker speech. *Frontiers in Neuroscience*, *8*, 88-88. https://doi.org/10.3389/fnins.2014.00088

18 APPENDICES

18.1 Appendix 1: Study Advertisement

Flinders UNIVERSITY

PARTICIPANTS NEEDED



What can your pupils tell us about your listening brain?!

The department of Speech Pathology and Audiology at Flinders University is seeking volunteers for a new research project using state of the art technology to examine pupil dilation during a speech-in-noise task.

What is required?

You will be required to attend a single 2 hour appointment at Flinders University Audiology Clinic. Participation involves having a standard hearing screening test, completion of questionnaires and a standard speech-in-noise task. You will be reimbursed $20 for your time.

Are you eligible to take part?

To participate in this study, you must be aged between 18 – 40, have English as your first language and have no known hearing or cognitive difficulties.

Interested?

Please contact the research group to arrange an appointment via
jennifer.baldock@flinders.edu.au

This study has been reviewed by the Southern Adelaide Clinical Human Research Ethics Committee *v3.1 05/09/2018*

Pupil Dilation & Listening
jennifer.baldock@flinders.edu.au

Pupil Dilation & Listening
jennifer.baldock@flinders.edu.au
Pupil Dilation & Listening
jennifer.baldock@flinders.edu.au
Pupil Dilation & Listening
jennifer.baldock@flinders.edu.au
Pupil Dilation & Listening
jennifer.baldock@flinders.edu.au
Pupil Dilation & Listening
jennifer.baldock@flinders.edu.au
Pupil Dilation & Listening
jennifer.baldock@flinders.edu.au
Pupil Dilation & Listening
jennifer.baldock@flinders.edu.au
Pupil Dilation & Listening
jennifer.baldock@flinders.edu.au
Pupil Dilation & Listening
jennifer.baldock@flinders.edu.au
Pupil Dilation & Listening
jennifer.baldock@flinders.edu.au

18.2 Appendix 2: Pre-Task Questionnaire

Pre-task questionnaire

Participant number: _________

D.O.B.: _________ Age:________

Date: __________ Time: _________

Gender:________

Have you had a hearing test before? **YES / NO**

If yes, what were the results? ____

____

Are you familiar with the BKB sentence test? **YES / NO**

Details: ____

Do you have any hearing difficulties? **YES / NO**

Details:____

____

Do you have a history of ear infections? **YES / NO**

If yes, when was your last ear infection? ____________________________________

Have you had a cold or been sick in the last two weeks? **YES / NO**

If yes, can you describe it? ____

____

Do you experience tinnitus (ringing/buzzing in one or both ears)? **YES / NO**

If yes, can you describe it? ____

____

**Let me know if it affecting concentration during the task

Have you had or do you have any vision difficulties? **YES / NO**

If yes, can you describe it? ____

____

Are they corrected? (e.g. glasses, surgery): _____________________________________

Do you have any medical conditions related to your pupils? **YES / NO**

If yes, what condition? ____

The influence of light level and task-induced fatigue on pupil dilation during listening effort

Has it been treated? **YES / NO**

Treatment? ____

Is English your first language? **YES / NO**

How many years have you lived in Australia? ____________

Have you ever been diagnosed with a speech/language/learning impairment? **YES / NO**

If yes, what impairment? ____

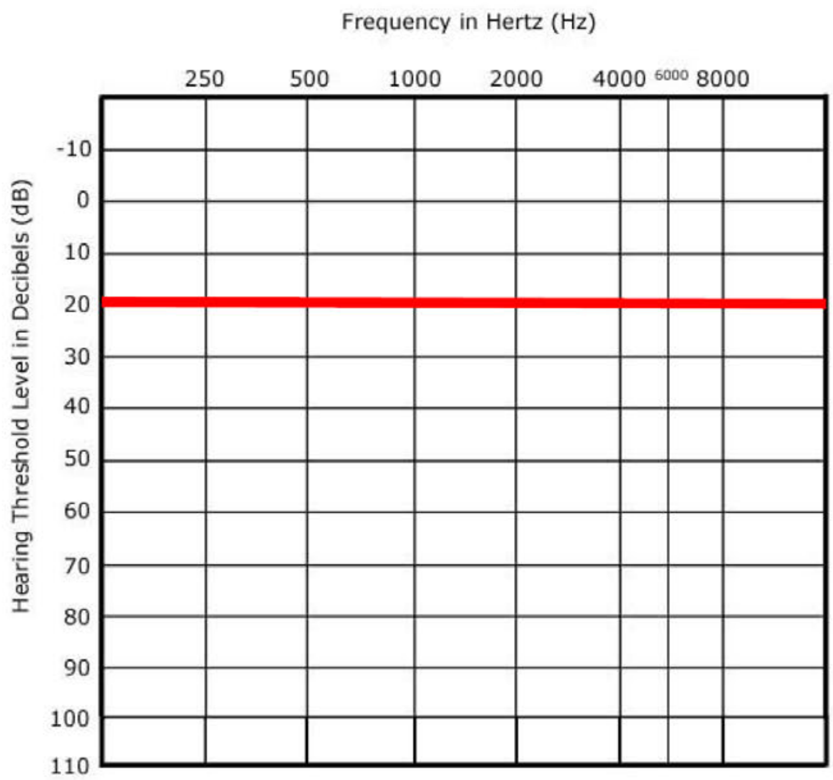
At what age? ________

Is it an ongoing problem? **YES / NO**

Approximately, how many hours of sleep did you get last night? ________

When was your last caffeinated beverage? ________

Audiogram:



Red/Right/Round

The influence of light level and task-induced fatigue on pupil dilation during listening effort

18.3 Appendix 3: Ethics Approval Letter

Office for Research
Flinders Medical Centre
Ward 6C, Room 6A219
Flinders Drive, Bedford Park SA 5042
Tel: (08) 8204 6453
E: Health.SALHNOfficeforResearch@sa.gov.au

Government of South Australia
SA Health
Southern Adelaide Local Health Network

Final Approval for Ethics Application

23 October 2018

Dr Sarosh Kapadia
Speech Pathology and Audiology
Flinders University
BEDFORD PARK SA 5042

Jennifer.baldock@flinders.edu.au

OFR Number: 235.18
HREC reference: HREC/18/SAC/277
Study title: The influence of light level and task-induced fatigue on pupil dilation during listening effort
Chief Investigator: Dr Sarosh Kapadia

Ethics Approval Period: 10 October 2018 - 10 October 2021

The Southern Adelaide Clinical Human Research Ethics Committee (SAC HREC EC00188) have reviewed and provided approval for this application which meets the requirements of the *National Statement on Ethical Conduct in Human Research (2007).*

You are reminded that this letter constitutes **Ethics** approval only. **Ethics approval is one aspect of the research governance process**.

You must not commence this research project at any SA Health sites listed in the application until a Site Specific Assessment (SSA), or Access Request for data or tissue form, has been approved by the Chief Executive or delegate of each site.

The below documents have been reviewed and approved:

- Low and negligible risk application form dated 05 September 2018
- Participant Information Sheet/Consent Form v1 dated 09 August 2018
- Recruitment advert v3.1 dated 05 September 2018
- Pre-task questionnaire v1 dated 21 August 2018
- RSME7 Scales v1 dated 09 August 2018

Page **1** of **2**

Terms and Conditions Of Ethics Approval:

It is essential that researchers adhere to the conditions below and with the *National Statement chapter 5.5.*

Final ethics approval is granted subject to the researcher agreeing to meet the following terms and conditions:

1. The approval only covers the science and ethics component of the application. A SSA will need to be submitted and authorised before this research project can commence at any of the approved sites identified in the application.
2. If University personnel are involved in this project, the Principal Investigator should notify the University before commencing their research to ensure compliance with University requirements including any insurance and indemnification requirements.
3. Compliance with the *National Statement on Ethical Conduct in Human Research* (2007) & the *Australian Code for the Responsible Conduct of Research* (2007).
4. To immediately report to SAC HREC anything that may change the ethics or scientific integrity of the project.
5. Report Significant Adverse events (SAE's) as per SAE requirements available at our website.
6. Submit an annual report on each anniversary of the date of final approval and in the correct template from the SAC HREC website.
7. Confidentiality of research participants MUST be maintained at all times.
8. A copy of the signed consent form must be given to the participant unless the project is an audit.
9. Any reports or publications derived from the research should be submitted to the Committee at the completion of the project.
10. All requests for access to medical records at any SALHN site must be accompanied by this approval email.
11. To regularly review the SAC HREC website and comply with all submission requirements, as they change from time to time.
12. Once your research project has concluded, any new product/procedure/intervention cannot be conducted in the SALHN as standard practice without the approval of the SALHN New Medical Products and Standardisation Committee or the SALHN New Health Technology and Clinical Practice Innovation Committee (as applicable). Please refer to the relevant committee link on the SALHN intranet for further information.

For any queries about this matter, please contact The Office for Research on (08) 8204 6453 or via email to Health.SALHNOfficeforResearch@sa.gov.au

Yours sincerely

A/Professor Bernadette Richards
Chair, SAC HREC

18.4 Appendix 4: The BKB Sentence Lists/Score Sheets

Sentence list 1 track 4

Lighting level:	SNR
Data file name:	Condition order (1st-16th):

Sentence stimuli	**SCORE**
1. The **CLOWN** had a **FUNNY FACE** (Max 3)	
2. The **CAR ENGINE'S RUNNING** (Max 3)	
3. **SHE CUT** with her **KNIFE** (Max 3)	
4. **CHILDREN LIKE STRAWBERRIES** (Max 3)	
5. The **HOUSE** has **NINE ROOMS** (Max 3)	
6. **THEY'RE BUYING** some **BREAD** (Max 3)	
7. The **GREEN TOMATOES** are **SMALL** (Max 3)	
8. **HE PLAYED** with his **TRAIN** (Max 3)	
9. The **POSTMAN SHUT** the **GATE** (Max 3)	
10. **THEY'RE LOOKING AT** the **CLOCK** (Max 4)	
11. The **BAG BUMPS** on the **GROUND** (Max 3)	
12. The **BOY DID a HANDSTAND** (Max 3)	
13. a **CAT SITS ON** the **BED** (Max 4)	
14. The **TRUCK CARRIED FRUIT** (Max 3)	
15. The **RAIN CAME DOWN** (Max 3)	
16. The **ICE CREAM** was **PINK** (Max 3)	
TOTAL SCORE	/50

Additional notes:

Sentence list 2 track 5

Lighting level:	SNR
Data file name:	Condition order (1st-16th):

Sentence stimuli	SCORE
1. The **LADDER'S NEAR** the **DOOR** (Max 3)	
2. **THEY** had a **LOVELY DAY** (Max 3)	
3. The **BALL WENT** into the **NET** (Max 3)	
4. The **OLD GLOVES** are **DIRTY** (Max 3)	
5. **HE CUT** his **FINGER** (Max 3)	
6. The **THIN DOG** was **HUNGRY** (Max 3)	
7. The **BOY KNEW** the **GAME** (Max 3)	
8. The **GRASS GROWS** in **SUMMER** (Max 3)	
9. **SHE'S TAKING** her **COAT** (Max 3)	
10. The **POLICE CHASED** the **CAR** (Max 3)	
11. a **MOUSE RAN DOWN** the **HOLE** (Max 4)	
12. The **LADY'S MAKING** a **TOY** (Max 3)	
13. Some **STICKS WERE UNDER** the **TREE** (Max 4)	
14. The **LITTLE BABY SLEEPS** (Max 3)	
15. **THEY'RE WATCHING** the **TRAIN** (Max 3)	
16. **SCHOOL FINISHED EARLY** (Max 3)	
TOTAL SCORE	/50

Additional notes:

Sentence list 3 track 6

Lighting level:	SNR
Data file name:	Condition order (1st-16th):

Sentence stimuli	**SCORE**
1. The **GLASS BOWL BROKE** (Max 3)	
2. The **DOG PLAYED** with a **STICK** (Max 3)	
3. The **KETTLE'S QUITE HOT** (Max 3)	
4. The **FARMER FEEDS** a **LAMB** (Max 3)	
5. **THEY SAY** some **SILLY THINGS** (Max 4)	
6. The **LADY WORE** a **COAT** (Max 3)	
7. The **CHILDREN** are **WALKING HOME** (Max 3)	
8. **HE NEEDED** his **HOLIDAY** (Max 3)	
9. The **MILK CAME** in a **BOTTLE** (Max 3)	
10. The **MAN CLEANED** his **SHOES** (Max 3)	
11. **THEY ATE** the **LEMON JELLY** (Max 4)	
12. The **BOY'S RUNNING AWAY** (Max 3)	
13. **FATHER LOOKED** at the **BOOK** (Max 3)	
14. **SHE DRINKS** from her **CUP** (Max 3)	
15. The **ROOMS GETTING COLD** (Max 3)	
16. a **GIRL KICKED** the **TABLE** (Max 3)	
TOTAL SCORE	/50

Additional notes:

Sentence list 4 track 7

Lighting level:	SNR
Data file name:	Condition order (1st-16th):

Sentence stimuli	SCORE
1. The **WIFE HELPED** her **HUSBAND** (Max 3)	
2. The **MUSIC** was **VERY LOUD** (Max 3)	
3. The **OLD MAN WORRIES** (Max 3)	
4. a **BOY RAN DOWN** the **PATH** (Max 4)	
5. The **HOUSE** had a **NICE GARDEN** (Max 3)	
6. **SHE SPOKE TO** her **SON** (Max 4)	
7. **THEY'RE CROSSING** the **STREET** (Max 3)	
8. **LEMONS GROW** on **TREES** (Max 3)	
9. **HE FOUND** his **BROTHER** (Max 3)	
10. Some **ANIMALS SLEEP** on **STRAW** (Max 3)	
11. The **JAM JAR** was **FULL** (Max 3)	
12. **THEY'RE KNEELING DOWN** (Max 3)	
13. The **GIRL LOST** her **DOLL** (Max 3)	
14. The **COOK'S MAKING** a **CAKE** (Max 3)	
15. The **CHILD DROPS** the **TOY** (Max 3)	
16. The **MUD STUCK** on his **SHOE** (Max 3)	
TOTAL SCORE	/50

Additional notes:

Sentence list 5 track 8

Lighting level:	SNR
Data file name:	Condition order (1st-16th):

Sentence stimuli	SCORE
1. The **BATH TOWEL** was **WET** (Max 3)	
2. The **MATCHES** were **NEAR** the **SHELF** (Max 3)	
3. **THEY'RE RUNNING PAST** the **HOUSE** (Max 4)	
4. The **TRAIN** had a **BAD CRASH** (Max 3)	
5. The **KITCHEN SINK'S EMPTY** (Max 3)	
6. A **BOY FELL** from the **WINDOW** (Max 4)	
7. **SHE USED** her **SPOON** (Max 3)	
8. The **PARK'S NEAR** the **ROAD** (Max 3)	
9. The **COOK CHOPPED** some **CARROTS** (Max 3)	
10. The **DOG MADE** an **ANGRY NOISE** (Max 4)	
11. **HE'S WASHING** his **FACE** (Max 3)	
12. **SOMEBODY TOOK** the **MONEY** (Max 3)	
13. The **LIGHT WENT OUT** (Max 3)	
14. **THEY WANTED** some **POTATOES** (Max 3)	
15. The **NAUGHTY GIRL'S SHOUTING** (Max 3)	
16. The **COLD MILK'S** in a **JUG** (Max 3)	
TOTAL SCORE	/50

Additional notes:

Sentence list 6 track 9

Lighting level:	SNR
Data file name:	Condition order (1st-16th):

Sentence stimuli	SCORE
1. The **PAINT** **DRIPPED** on the **GROUND** (Max 3)	
2. The **MOTHER** **STIRS** the **SOUP** (Max 3)	
3. **THEY** **LAUGHED** at his **STORY** (Max 3)	
4. **MEN** **WEAR** **LONG** **TROUSERS** (Max 4)	
5. The **SMALL** **BOY** was **ASLEEP** (Max 3)	
6. The **LADY** **GOES** to the **SHOP** (Max 3)	
7. The **SUN** **MELTED** the **SNOW** (Max 3)	
8. The **FATHER'S** **COMING** **HOME** (Max 3)	
9. **SHE** **HAD** her **POCKET** **MONEY** (Max 4)	
10. The **TRUCK** **DROVE** up the **ROAD** (Max 3)	
11. **HE'S** **BRINGING** his **RAINCOAT** (Max 3)	
12. a **SHARP** **KNIFE'S** **DANGEROUS** (Max 3)	
13. **THEY** **TOOK** some **FOOD** (Max 3)	
14. The **CLEVER** **GIRLS** are **READING** (Max 3)	
15. The **BROOM** **STOOD** in the **CORNER** (Max 3)	
16. The **WOMAN** **TIDIED** her **HOUSE** (Max 3)	
TOTAL SCORE	/50

Additional notes:

Sentence list 7 track 10

Lighting level:	SNR
Data file name:	Condition order (1st-16th):

Sentence stimuli	SCORE
1. The **CHILDREN** **DROPPED** the **BAG** (Max 3)	
2. The **DOG** **CAME** **BACK** (Max 3)	
3. The **FLOOR** **LOOKED** **CLEAN** (Max 3)	
4. **SHE** **FOUND** her **PURSE** (Max 3)	
5. The **FRUIT** **LIES** on the **GROUND** (Max 3)	
6. **MOTHER** **BUYS** a **SAUCEPAN** (Max 3)	
7. **THEY** **WASHED** in **COLD** **WATER** (Max 4)	
8. The **YOUNG** **PEOPLE** are **DANCING** (Max 3)	
9. The **BUS** **WENT** **EARLY** (Max 3)	
10. **THEY** had **TWO** **EMPTY** **BOTTLES** (Max 4)	
11. a **BALL'S** **BOUNCING** **ALONG** (Max 3)	
12. The **FATHER** **FORGOT** the **BREAD** (Max 3)	
13. The **GIRL** has a **PICTURE** **BOOK** (Max 3)	
14. The **ORANGE** was **QUITE** **SWEET** (Max 3)	
15. **HE'S** **HOLDING** his **NOSE** (Max 3)	
16. The **NEW** **ROAD'S** on the **MAP** (Max 3)	
TOTAL SCORE	/50

Additional notes:

Sentence list 9 track 12

Lighting level:	SNR
Data file name:	Condition order (1st-16th):

Sentence stimuli	SCORE
1. The **BOOK TELLS** a **STORY** (Max 3)	
2. The **YOUNG BOY LEFT HOME** (Max 4)	
3. **THEY'RE CLIMBING** the **TREE** (Max 3)	
4. **SHE STOOD** near her **WINDOW** (Max 3)	
5. The **TABLE** has **THREE LEGS** (Max 3)	
6. a **LETTER FELL** on the **MAT** (Max 3)	
7. The **FIVE MEN** are **WORKING** (Max 3)	
8. **HE LISTENS TO** his **FATHER** (Max 4)	
9. The **SHOES** were **VERY DIRTY** (Max 3)	
10. **THEY WENT** on **HOLIDAY** (Max 3)	
11. **BABY BROKE** his **MUG** (Max 3)	
12. The **LADY PACKED** her **BAG** (Max 3)	
13. The **DINNER PLATE'S HOT** (Max 3)	
14. The **TRAIN'S MOVING FAST** (Max 3)	
15. The **CHILD DRANK** some **MILK**(Max 3)	
16. The **CAR HIT** a **WALL** (Max 3)	
TOTAL SCORE	/50

Additional notes:

Sentence list 11 track 14

Lighting level:	SNR
Data file name:	Condition order (1st-16th):

Sentence stimuli	SCORE
1. The **MILK BAR** was **EMPTY** (Max 3)	
2. The **DOGS GO** for a **WALK** (Max 3)	
3. **SHE'S WASHING** her **DRESS** (Max 3)	
4. The **LADY STAYED** for **TEA** (Max 3)	
5. The **BUS WAITS** at the **CORNER** (Max 3)	
6. **THEY FINSIHED** the **MEAL** (Max 3)	
7. The **POLICEMAN KNOWS** the **WAY** (Max 3)	
8. The **LITTLE GIRL** was **HAPPY** (Max 3)	
9. **HE WORE** a **YELLOW SHIRT** (Max 4)	
10. **THEY'RE COMING** for **CHRISTMAS** (Max 3)	
11. The **COW ATE** some **HAY** (Max 3)	
12. The **BOY GOT INTO BED** (Max 4)	
13. The **TWO FARMERS** are **TALKING** (Max 3)	
14. **MOTHER PICKED** some **FLOWERS** (Max 3)	
15. a **PEA FELL** off the **PLATE** (Max 3)	
16. The **FATHER WRITES** a **LETTER** (Max 3)	
TOTAL SCORE	/50

Additional notes:

Sentence list 12 track 15

Lighting level:	SNR
Data file name:	Condition order (1st-16th):

Sentence stimuli	SCORE
The **FOOD COSTS** a **LOT** (Max 3)	
The **GIRL'S WASHING** her **HAIR** (Max 3)	
The **FRONT GARDEN** was **PRETTY** (Max 3)	
HE LOST his **HAT** (Max 3)	
The **TAPS** are **ABOVE** the **SINK** (Max 3)	
FATHER PAID at the **GATE** (Max 3)	
SHE'S WAITING for her **BUS** (Max 3)	
The **BREAD VAN'S COMING** (Max 3)	
THEY had some **COLD MEAT** (Max 3)	
The **FOOTBALL GAME'S OVER** (Max 3)	
THEY CARRY some **SHOPPING BAGS** (Max 4)	
The **CHILDREN HELP** the **MILKMAN** (Max 3)	
The **PICTURE CAME** from a **BOOK** (Max 3)	
The **PLUM PUDDING** was **READY** (Max 3)	
The **BOY** had a **TOY DRAGON** (Max 3)	
A **TREE FELL ON** the **HOUSE** (Max 4)	
TOTAL SCORE	/50

Additional notes:

Sentence list 13 track 16

Lighting level:	SNR
Data file name:	Condition order (1st-16th):

Sentence stimuli	SCORE
1. The **FRUIT** **CAME** in a **BOX** (Max 3)	
2. The **HUSBAND** **BRINGS** some **FLOWERS** (Max 3)	
3. **THEY'RE** **PLAYING** in the **PARK** (Max 3)	
4. **SHE** **ARGUED** with her **SISTER** (Max 3)	
5. A **MAN** **TOLD** the **POLICE** (Max 3)	
6. **POTATOES** **GROW** in the **GROUND** (Max 3)	
7. **HE'S** **CLEANING** his **CAR** (Max 3)	
8. The **MOUSE** **FOUND** the **CHEESE** (Max 3)	
9. **THEY** **WAITED** for **ONE** **HOUR** (Max 4)	
10. The **BIG** **DOG** was **DANGEROUS** (Max 3)	
11. The **STRAWBERRY** **JAM** was **SWEET** (Max 3)	
12. The **PLANT** **HANGS** **ABOVE** the **DOOR** (Max 4)	
13. The **CHILDREN** are **ALL** **EATING** (Max 3)	
14. The **BOY** has **BLACK** **HAIR** (Max 3)	
15. The **MOTHER** **HEARD** her **BABY** (Max 3)	
16. The **TRUCK** **CLIMBED** the **HILL** (Max 3)	
TOTAL SCORE	/50

Additional notes:

Sentence list 14 track 17

Lighting level:	SNR
Data file name:	Condition order (1st-16th):

Sentence stimuli	SCORE
1. The **ANGRY MAN SHOUTED** (Max 3)	
2. The **DOG SLEEPS** in a **BASKET** (Max 3)	
3. **THEY'RE DRINKING TEA** (Max 3)	
4. **MOTHER OPENS** the **DRAWER** (Max 3)	
5. An **OLD WOMAN** was at **HOME** (Max 3)	
6. **HE DROPPED** his **MONEY** (Max 3)	
7. **THEY BROKE ALL** the **EGGS** (Max 4)	
8. The **KITCHEN WINDOW** was **CLEAN** (Max 3)	
9. The **GIRL PLAYS** with the **BABY** (Max 3)	
10. The **BIG FISH GOT AWAY** (Max 4)	
11. **SHE'S HELPING** her **FRIEND** (Max 3)	
12. The **CHILDREN WASHED** the **PLATES** (Max 3)	
13. The **POSTMAN COMES EARLY** (Max 3)	
14. The **SIGN SHOWED** the **WAY** (Max 3)	
15. The **GRASS** is **GETTING LONG** (Max 3)	
16. The **MATCH FELL** on the **FLOOR** (Max 3)	
TOTAL SCORE	/50

Additional notes:

Sentence list 15 track 18

Lighting level:	SNR
Data file name:	Condition order (1st-16th):

Sentence stimuli	SCORE
1. A **MAN'S TURNING** the **HANDLE** (Max 3)	
2. The **FIRE** was **VERY HOT** (Max 3)	
3. HE'S SUCKING his **THUMB** (Max 3)	
4. The **SHOP CLOSED** for **LUNCH** (Max 3)	
5. The **DRIVER STARTS** the **ENGINE** (Max 3)	
6. The **BOY HURRIED** to **SCHOOL** (Max 3)	
7. Some **NICE PEOPLE** are **COMING** (Max 3)	
8. SHE BUMPED her **HEAD** (Max 3)	
9. THEY MET SOME FRIENDS (Max 4)	
10. FLOWERS GROW in the **GARDEN** (Max 3)	
11. The **TINY BABY** was **PRETTY** (Max 3)	
12. The **DAUGHTER SET** the **TABLE** (Max 3)	
13. THEY WALKED ACROSS the **GRASS** (Max 4)	
14. The **MOTHER TIED** the **STRING** (Max 3)	
15. The **TRAIN STOPS** at the **STATION** (Max 3)	
16. The **PUPPY PLAYS** with a **BALL** (Max 3)	
TOTAL SCORE	/50

Additional notes:

Sentence list 18 track 21

Lighting level:	SNR
Data file name:	Condition order (1^{st}-16^{th}):

Sentence stimuli	SCORE
1. The **CAKE SHOP'S OPENING** (Max 3)	
2. **THEY LIKE ORANGE MARMALADE** (Max 4)	
3. The **MOTHER SHUT** the **WINDOW** (Max 3)	
4. **HE'S SKATING WITH** his **FRIEND** (Max 4)	
5. The **MEAT PIE** was **GOOD** (Max 3)	
6. **RAIN FALLS** from **CLOUDS** (Max 3)	
7. **SHE TALKED** to her **DOLL** (Max 3)	
8. **THEY PAINTED** the **WALL** (Max 3)	
9. The **TOWEL DROPPED** on the **FLOOR** (Max 3)	
10. The **DOG'S EATING** some **MEAT** (Max 3)	
11. A **BOY BROKE** the **FENCE** (Max 3)	
12. The **YELLOW PEARS** were **LOVELY** (Max 3)	
13. The **POLICE HELP** the **DRIVER** (Max 3)	
14. The **LEAVES LAY** on the **ROOF** (Max 3)	
15. The **LADY WASHED** the **SHIRT** (Max 3)	
16. The **CUP HANGS** on a **HOOK** (Max 3)	
TOTAL SCORE	/50

Additional notes:

Sentence list 19 track 22

Lighting level:	SNR
Data file name:	Condition order (1st-16th):

Sentence stimuli	SCORE
1. The **FAMILY** **LIKE** **FISH** (Max 3)	
2. **SUGAR'S** **VERY** **SWEET** (Max 3)	
3. The **BABY** **LAY** on a **RUG** (Max 3)	
4. The **WASHING** **MACHINE** **BROKE** (Max 3)	
5. **THEY'RE** **CLEARING** the **TABLE** (Max 3)	
6. The **CLEANER** **SWEPT** the **FLOOR** (Max 3)	
7. A **BUTCHER** **SELLS** **MEAT** (Max 3)	
8. The **BATH** **WATER** was **WARM** (Max 3)	
9. **HE'S** **REACHING** for his **SPOON** (Max 3)	
10. **SHE** **HURT** her **HAND** (Max 3)	
11. The **FIREMAN** **DRIVES** a **RED** **TRUCK** (Max 4)	
12. The **BOY** **SLIPPED** **ON** the **STAIRS** (Max 4)	
13. **THEY'RE** **STAYING** for **SUPPER** (Max 3)	
14. The **GIRL** **HELD** a **MIRROR** (Max 3)	
15. The **CUP** was **BY** the **SAUCER** (Max 3)	
16. The **LADIES** **WENT** to **CHURCH** (Max 3)	
TOTAL SCORE	/50

Additional notes:

Sentence list 21 track 24

Lighting level:	SNR
Data file name:	Condition order (1st-16th):

Sentence stimuli	SCORE
1. The **COFFEE POT** was **EMPTY** (Max 3)	
2. The **DOG DRANK** from a **BOWL** (Max 3)	
3. A **GIRL CAME** into the **ROOM** (Max 3)	
4. **THEY'RE PUSHING** an **OLD CAR** (Max 4)	
5. The **CAT CAUGHT** a **MOUSE** (Max 3)	
6. The **ROAD GOES UP** a **HILL** (Max 4)	
7. **SHE MADE** her **BED** (Max 3)	
8. **BANANAS** are **YELLOW FRUIT** (Max 3)	
9. The **COW LIES** on the **GRASS** (Max 3)	
10. The **EGG CUPS** are on the **TABLE** (Max 3)	
11. **HE FRIGHTENED** his **SISTER** (Max 3)	
12. The **CRICKET TEAM'S PLAYING** (Max 3)	
13. The **FATHER PICKED** some **PEARS** (Max 3)	
14. The **KETTLE BOILED QUICKLY** (Max 3)	
15. The **MAN'S PAINTING** a **SIGN** (Max 3)	
16. **THEY LOST** some **MONEY** (Max 3)	
TOTAL SCORE	/50

Additional notes:

18.5 Appendix 5: Order of BKB List Presentation by Participant

Participant number																	
1	1	2	21	3	19	4	18	5	15	6	14	7	13	9	12	11
2	2	3	1	4	21	5	19	6	18	7	15	9	14	11	13	12
3	3	4	2	5	1	6	21	7	19	9	18	11	15	12	14	13
4	4	5	3	6	2	7	1	9	21	11	19	12	18	13	15	14
5	5	6	4	7	3	9	2	11	1	12	21	13	19	14	18	15
6	6	7	5	9	4	11	3	12	2	13	1	14	21	15	19	18
7	7	9	6	11	5	12	4	13	3	14	2	15	1	18	21	19
8	9	11	7	12	6	13	5	14	4	15	3	18	2	19	1	21
9	11	12	9	13	7	14	6	15	5	18	4	19	3	21	2	1
10	12	13	11	14	9	15	7	18	6	19	5	21	4	1	3	2
11	13	14	12	15	11	18	9	19	7	21	6	1	5	2	4	3
12	14	15	13	18	12	19	11	21	9	1	7	2	6	3	5	4
13	15	18	14	19	13	21	12	1	11	2	9	3	7	4	6	5
14	18	19	15	21	14	1	13	2	12	3	11	4	9	5	7	6
15	19	21	18	1	15	2	14	3	13	4	12	5	11	6	9	7
16	21	1	19	2	18	3	15	4	14	5	13	6	12	7	11	9
17	1	2	21	3	19	4	18	5	15	6	14	7	13	9	12	11
18	2	3	1	4	21	5	19	6	18	7	15	9	14	11	13	12
19	3	4	2	5	1	6	21	7	19	9	18	11	15	12	14	13
20	4	5	3	6	2	7	1	9	21	11	19	12	18	13	15	14
21	5	6	4	7	3	9	2	11	1	12	21	13	19	14	18	15
22	6	7	5	9	4	11	3	12	2	13	1	14	21	15	19	18
23	7	9	6	11	5	12	4	13	3	14	2	15	1	18	21	19
24	9	11	7	12	6	13	5	14	4	15	3	18	2	19	1	21
25	11	12	9	13	7	14	6	15	5	18	4	19	3	21	2	1
26	12	13	11	14	9	15	7	18	6	19	5	21	4	1	3	2
27	13	14	12	15	11	18	9	19	7	21	6	1	5	2	4	3
28	14	15	13	18	12	19	11	21	9	1	7	2	6	3	5	4
29	15	18	14	19	13	21	12	1	11	2	9	3	7	4	6	5
30	18	19	15	21	14	1	13	2	12	3	11	4	9	5	7	6
31	19	21	18	1	15	2	14	3	13	4	12	5	11	6	9	7
32	21	1	19	2	18	3	15	4	14	5	13	6	12	7	11	9
33	1	2	21	3	19	4	18	5	15	6	14	7	13	9	12	11
34	2	3	1	4	21	5	19	6	18	7	15	9	14	11	13	12
35	3	4	2	5	1	6	21	7	19	9	18	11	15	12	14	13
36	4	5	3	6	2	7	1	9	21	11	19	12	18	13	15	14

18.6 Appendix 6: Order of Condition by Participant (Counterbalancing)

Participant		Block 1-L1					Block 2-L2						Block 3-L4					Block 4-L3		
1	3	0	-6	-3	B1 QU	0	-3	3	-6	B2 QU	-3	-6	0	3	B3 QU	-6	3	-3	0	B4 QU
SNR used:																				
2	0	-3	3	-6	B1 QU	-3	-6	0	3	B2 QU	-6	3	-3	0	B3 QU	3	0	-6	-3	B4 QU
SNR used:																				
3	-3	-6	0	3	B1 QU	-6	3	-3	0	B2 QU	3	0	-6	-3	B3 QU	0	-3	3	-6	B4 QU
SNR used:																				
4	-6	3	-3	0	B1 QU	3	0	-6	-3	B2 QU	0	-3	3	-6	B3 QU	-3	-6	0	3	B1 QU
SNR used:																				
		Block 1-L2					Block 2-L3					Block 3-L1					Block 4-L4			
5	3	0	-6	-3	B1 QU	0	-3	3	-6	B2 QU	-3	-6	0	3	B3 QU	-6	3	-3	0	B4 QU
SNR used:																				
6	0	-3	3	-6	B1 QU	-3	-6	0	3	B2 QU	-6	3	-3	0	B3 QU	3	0	-6	-3	B4 QU
SNR used:																				
7	-3	-6	0	3	B1 QU	-6	3	-3	0	B2 QU	3	0	-6	-3	B3 QU	0	-3	3	-6	B4 QU
SNR used:																				
8	-6	3	-3	0	B1 QU	3	0	-6	-3	B2 QU	0	-3	3	-6	B3 QU	-3	-6	0	3	B1 QU
SNR used:		Block 1-L3					Block 2-L4					Block 3-L2					Block 4-L1			
9	3	0	-6	-3	B1 QU	0	-3	3	-6	B2 QU	-3	-6	0	3	B3 QU	-6	3	-3	0	B4 QU
SNR used:																				
10	0	-3	3	-6	B1 QU	-3	-6	0	3	B2 QU	-6	3	-3	0	B3 QU	3	0	-6	-3	B1 QU
SNR used:																				
11	-3	-6	0	3	B1 QU	-6	3	-3	0	B2 QU	3	0	-6	-3	B3 QU	0	-3	3	-6	B4 QU
SNR used:																				
12	-6	3	-3	0	B1 QU	3	0	-6	-3	B2 QU	0	-3	3	-6	B3 QU	-3	-6	0	3	B4 QU
SNR used:																				
		Block 1-L4					Block 2-L1					Block 3-L3					Block 4-L2			
13	3	0	-6	-3	B1 QU	0	-3	3	-6	B2 QU	-3	-6	0	3	B3 QU	-6	3	-3	0	B4 QU
SNR used:																				
14	0	-3	3	-6	B1 QU	-3	-6	0	3	B2 QU	-6	3	-3	0	B3 QU	3	0	-6	-3	B1 QU
SNR used:																				
15	-3	-6	0	3	B1 QU	-6	3	-3	0	B2 QU	3	0	-6	-3	B3 QU	0	-3	3	-6	B4 QU
SNR used:																				
16	-6	3	-3	0	B1 QU	3	0	-6	-3	B2 QU	0	-3	3	-6	B3 QU	-3	-6	0	3	B4 QU
SNR used:																				
		Block 1-L1					Block 2-L2						Block 3-L4					Block 4-L3		
17	3	0	-6	-3	B1 QU	0	-3	3	-6	B2 QU	-3	-6	0	3	B3 QU	-6	3	-3	0	B1 QU
SNR used:																				
18	0	-3	3	-6	B1 QU	-3	-6	0	3	B2 QU	-6	3	-3	0	B3 QU	3	0	-6	-3	B4 QU

SNR used:																				
19	-3	-6	0	3	B1 QU	-6	3	-3	0	B2 QU	3	0	-6	-3	B3 QU	0	-3	3	-6	B4 QU
SNR used:																				
20	-6	3	-3	0	B1 QU	3	0	-6	-3	B2 QU	0	-3	3	-6	B3 QU	-3	-6	0	3	B4 QU
SNR used:																				
		Block 1-L2					Block 2-L3					Block 3-L1					Block 4-L4			
21	3	0	-6	-3	B1 QU	0	-3	3	-6	B2 QU	-3	-6	0	3	B3 QU	-6	3	-3	0	B1 QU
SNR used:																				
22	0	-3	3	-6	B1 QU	-3	-6	0	3	B2 QU	-6	3	-3	0	B3 QU	3	0	-6	-3	B4 QU
SNR used:																				
23	-3	-6	0	3	B1 QU	-6	3	-3	0	B2 QU	3	0	-6	-3	B3 QU	0	-3	3	-6	B4 QU
SNR used:																				
24	-6	3	-3	0	B1 QU	3	0	-6	-3	B2 QU	0	-3	3	-6	B3 QU	-3	-6	0	3	B4 QU
SNR used:																				
		Block 1-L3					Block 2-L4					Block 3-L2					Block 4-L1			
25	3	0	-6	-3	B1 QU	0	-3	3	-6	B2 QU	-3	-6	0	3	B3 QU	-6	3	-3	0	B4 QU
SNR used:																				
26	0	-3	3	-6	B1 QU	-3	-6	0	3	B2 QU	-6	3	-3	0	B3 QU	3	0	-6	-3	B4 QU
SNR used:																				
27	-3	-6	0	3	B1 QU	-6	3	-3	0	B2 QU	3	0	-6	-3	B3 QU	0	-3	3	-6	B1 QU
SNR used:																				
28	-6	3	-3	0	B1 QU	3	0	-6	-3	B2 QU	0	-3	3	-6	B3 QU	-3	-6	0	3	B4 QU
SNR used:																				
		Block 1-L4					Block 2-L1					Block 3-L3					Block 4-L2			
29	3	0	-6	-3	B1 QU	0	-3	3	-6	B2 QU	-3	-6	0	3	B3 QU	-6	3	-3	0	B4 QU
SNR used:																				
30	0	-3	3	-6	B1 QU	-3	-6	0	3	B2 QU	-6	3	-3	0	B3 QU	3	0	-6	-3	B4 QU
SNR used:																				
31	-3	-6	0	3	B1 QU	-6	3	-3	0	B2 QU	3	0	-6	-3	B3 QU	0	-3	3	-6	B1 QU
SNR used:																				
32	-6	3	-3	0	B1 QU	3	0	-6	-3	B2 QU	0	-3	3	-6	B3 QU	-3	-6	0	3	B4 QU
		Block 1-L1					Block 2-L2					Block 3-L4					Block 4-L3			
33	3	0	-6	-3	B1 QU	0	-3	3	-6	B2 QU	-3	-6	0	3	B3 QU	-6	3	-3	0	B4 QU
SNR used:																				
34	0	-3	3	-6	B1 QU	-3	-6	0	3	B2 QU	-6	3	-3	0	B3 QU	3	0	-6	-3	B4 QU
SNR used:																				
35	-3	-6	0	3	B1 QU	-6	3	-3	0	B2 QU	3	0	-6	-3	B3 QU	0	-3	3	-6	B4 QU
SNR used:																				
36	-6	3	-3	0	B1 QU	3	0	-6	-3	B2 QU	0	-3	3	-6	B3 QU	-3	-6	0	3	B1 QU

18.7 Appendix 7: Output from Trial-Level Peak Dilation and Time-On-Task MEM: Final Model Results

```
Output captured by 'eatGet', Version 0.0.9, build 2018-05-15.
User: bald0106, computer: HXD4WF2, R version 4.0.1 (2020-06-06), Time: Tue
Oct 26 09:55:16 2021
Linear mixed model fit by maximum likelihood . t-tests use Satterthwaite's
method ['lmerModLmerTest']
Formula: peakdilation ~ snr * lightlevel + trial_exp + (1 + lightlevel |
participant) + (1 | sentence)
   Data: pup_data

     AIC      BIC   logLik deviance df.resid
 -2076.3    -73.6   1321.2  -2642.3     8466

Scaled residuals:
    Min      1Q  Median      3Q     Max
-4.1259 -0.6089 -0.0575  0.5632  7.8069

Random effects:
 Groups      Name        Variance Std.Dev. Corr
 sentence    (Intercept) 0.000214 0.01463
 participant (Intercept) 0.029530 0.17184
             lightlevel2 0.007615 0.08726  -0.65
             lightlevel3 0.011729 0.10830  -0.83  0.88
             lightlevel4 0.010630 0.10310  -0.92  0.80  0.96
 Residual                0.041750 0.20433
Number of obs: 8749, groups:  sentence, 256; participant, 36

Fixed effects:
                     Estimate  Std. Error          df t value Pr(>|t|)
(Intercept)           0.706609    0.048680  281.734955  14.516  < 2e-16 ***
snr-6                -0.021280    0.012465 8583.843044  -1.707 0.087821 .
snr0                 -0.072126    0.012277 8606.343244  -5.875 4.38e-09 ***
snr3                 -0.146916    0.012519 8402.265696 -11.735  < 2e-16 ***
lightlevel2          -0.117721    0.019069   73.967919  -6.173 3.27e-08 ***
lightlevel3          -0.205811    0.021874   59.997084  -9.409 2.04e-13 ***
lightlevel4          -0.259895    0.021130   62.898130 -12.300  < 2e-16 ***
trial_exp2           -0.067496    0.053509 8198.767597  -1.261 0.207208
trial_exp3           -0.033474    0.053511 8182.850285  -0.626 0.531624
trial_exp4           -0.040879    0.053500 8179.897998  -0.764 0.444833
trial_exp5           -0.023576    0.052678 8152.178171  -0.448 0.654486
trial_exp6           -0.082126    0.052678 8152.207651  -1.559 0.119030
trial_exp7           -0.033247    0.053083 8167.258689  -0.626 0.531124
trial_exp8            0.003440    0.052678 8152.232062   0.065 0.947940
trial_exp9           -0.013727    0.052678 8152.168967  -0.261 0.794414
trial_exp10          -0.108864    0.052678 8152.195187  -2.067 0.038803 *
trial_exp11          -0.101105    0.053085 8166.089401  -1.905 0.056865 .
trial_exp12          -0.113992    0.052678 8152.089187  -2.164 0.030497 *
trial_exp13          -0.108239    0.052678 8152.131953  -2.055 0.039936 *
trial_exp14          -0.133694    0.053515 8181.012775  -2.498 0.012501 *
trial_exp15          -0.053695    0.053085 8166.315621  -1.011 0.311810
trial_exp16          -0.029289    0.052678 8151.892780  -0.556 0.578223
trial_exp17          -0.042048    0.051377 8416.049179  -0.818 0.413144
trial_exp18          -0.114236    0.051989 8093.825068  -2.197 0.028026 *
trial_exp19          -0.098127    0.051990 8094.392805  -1.887 0.059136 .
trial_exp20          -0.042724    0.051990 8094.527026  -0.822 0.411230
trial_exp21          -0.086332    0.051980 8138.808459  -1.661 0.096777 .
trial_exp22          -0.078919    0.051310 8073.696452  -1.538 0.124064
```

trial_exp23	-0.151352	0.051641	8083.211852	-2.931	0.003390	**
trial_exp24	-0.116001	0.051640	8083.943456	-2.246	0.024710	*
trial_exp25	-0.148245	0.051639	8083.971607	-2.871	0.004105	**
trial_exp26	-0.083882	0.051987	8095.206624	-1.614	0.106666	
trial_exp27	-0.043678	0.051983	8122.642045	-0.840	0.400801	
trial_exp28	-0.122949	0.051640	8084.125043	-2.381	0.017294	*
trial_exp29	-0.112014	0.051310	8073.743468	-2.183	0.029057	*
trial_exp30	-0.203169	0.051641	8082.822295	-3.934	8.42e-05	***
trial_exp31	-0.015719	0.051310	8073.616159	-0.306	0.759344	
trial_exp32	-0.064209	0.051642	8083.261446	-1.243	0.213774	
trial_exp33	-0.012643	0.051045	8418.961757	-0.248	0.804379	
trial_exp34	-0.092741	0.051638	8097.697443	-1.796	0.072534	.
trial_exp35	-0.035148	0.051635	8098.706162	-0.681	0.496082	
trial_exp36	-0.041082	0.051979	8124.090417	-0.790	0.429348	
trial_exp37	-0.031730	0.051984	8110.262073	-0.610	0.541626	
trial_exp38	-0.073085	0.051305	8087.970256	-1.425	0.154336	
trial_exp39	-0.073636	0.051973	8121.750928	-1.417	0.156576	
trial_exp40	-0.012110	0.051971	8106.207429	-0.233	0.815759	
trial_exp41	-0.152389	0.051634	8099.690096	-2.951	0.003173	**
trial_exp42	-0.149989	0.051305	8087.898488	-2.923	0.003471	**
trial_exp43	-0.138546	0.051984	8109.911108	-2.665	0.007710	**
trial_exp44	-0.167574	0.051305	8087.903381	-3.266	0.001094	**
trial_exp45	-0.136017	0.051305	8087.983244	-2.651	0.008038	**
trial_exp46	-0.166703	0.051305	8087.926400	-3.249	0.001162	**
trial_exp47	-0.106034	0.051305	8087.828169	-2.067	0.038791	*
trial_exp48	-0.097532	0.051626	8094.794050	-1.889	0.058898	.
trial_exp49	-0.047695	0.051380	8442.169822	-0.928	0.353296	
trial_exp50	-0.106610	0.051305	8086.769111	-2.078	0.037746	*
trial_exp51	-0.107643	0.051305	8086.788007	-2.098	0.035928	*
trial_exp52	-0.100581	0.051305	8086.515923	-1.960	0.049978	*
trial_exp53	-0.085322	0.051305	8086.841061	-1.663	0.096346	.
trial_exp54	-0.128032	0.051638	8096.702532	-2.479	0.013181	*
trial_exp55	-0.126150	0.051305	8086.694301	-2.459	0.013961	*
trial_exp56	-0.138782	0.051983	8107.170713	-2.670	0.007605	**
trial_exp57	-0.169763	0.051635	8097.514265	-3.288	0.001014	**
trial_exp58	-0.162934	0.051635	8097.619693	-3.155	0.001608	**
trial_exp59	-0.161054	0.051986	8108.026533	-3.098	0.001955	**
trial_exp60	-0.190820	0.051305	8086.858845	-3.719	0.000201	***
trial_exp61	-0.169156	0.051305	8086.827271	-3.297	0.000981	***
trial_exp62	-0.148378	0.051305	8086.817359	-2.892	0.003837	**
trial_exp63	-0.176576	0.052352	8120.973506	-3.373	0.000747	***
trial_exp64	-0.175683	0.051975	8105.874317	-3.380	0.000728	***
trial_exp65	-0.010805	0.051863	8045.417924	-0.208	0.834966	
trial_exp66	-0.117668	0.052116	7706.378301	-2.258	0.023985	*
trial_exp67	-0.160452	0.052116	7706.363410	-3.079	0.002086	**
trial_exp68	-0.151246	0.052116	7706.376775	-2.902	0.003717	**
trial_exp69	-0.113487	0.052451	7739.274686	-2.164	0.030519	*
trial_exp70	-0.106647	0.052807	7759.114495	-2.020	0.043465	*
trial_exp71	-0.158171	0.052460	7724.241354	-3.015	0.002577	**
trial_exp72	-0.155451	0.052460	7724.448343	-2.963	0.003053	**
trial_exp73	-0.190176	0.052455	7725.483147	-3.625	0.000290	***
trial_exp74	-0.146708	0.052460	7724.062601	-2.797	0.005177	**
trial_exp75	-0.211119	0.052116	7706.547468	-4.051	5.15e-05	***
trial_exp76	-0.176695	0.052116	7706.494194	-3.390	0.000701	***
trial_exp77	-0.175742	0.052116	7706.591646	-3.372	0.000750	***
trial_exp78	-0.188984	0.052116	7706.486105	-3.626	0.000289	***
trial_exp79	-0.204658	0.052116	7706.571302	-3.927	8.68e-05	***
trial_exp80	-0.147138	0.052116	7706.517757	-2.823	0.004766	**
trial_exp81	-0.082062	0.051864	8045.349525	-1.582	0.113635	
trial_exp82	-0.154263	0.052468	7738.993878	-2.940	0.003290	**

trial_exp83	-0.147557	0.052117	7707.282463	-2.831	0.004648	**
trial_exp84	-0.185121	0.052462	7724.115501	-3.529	0.000420	***
trial_exp85	-0.133775	0.052117	7707.290336	-2.567	0.010282	*
trial_exp86	-0.216347	0.052117	7707.485046	-4.151	3.34e-05	***
trial_exp87	-0.167420	0.052117	7707.236476	-3.212	0.001322	**
trial_exp88	-0.126108	0.052117	7707.267257	-2.420	0.015555	*
trial_exp89	-0.216388	0.052117	7707.194693	-4.152	3.33e-05	***
trial_exp90	-0.147693	0.052117	7707.282077	-2.834	0.004610	**
trial_exp91	-0.220765	0.052117	7707.216260	-4.236	2.30e-05	***
trial_exp92	-0.168133	0.052461	7723.778356	-3.205	0.001356	**
trial_exp93	-0.153189	0.052452	7739.597829	-2.921	0.003504	**
trial_exp94	-0.202920	0.052117	7707.161390	-3.894	9.96e-05	***
trial_exp95	-0.159941	0.052117	7706.930349	-3.069	0.002156	**
trial_exp96	-0.206823	0.052815	7758.067208	-3.916	9.08e-05	***
trial_exp97	0.002919	0.051873	8038.226628	0.056	0.955126	
trial_exp98	-0.128139	0.052129	7733.704859	-2.458	0.013988	*
trial_exp99	-0.169191	0.052128	7733.721653	-3.246	0.001177	**
trial_exp100	-0.155199	0.053196	7820.126061	-2.918	0.003538	**
trial_exp101	-0.089534	0.052128	7733.832249	-1.718	0.085916	.
trial_exp102	-0.090608	0.052128	7733.800420	-1.738	0.082220	.
trial_exp103	-0.177222	0.052833	7772.197411	-3.354	0.000799	***
trial_exp104	-0.177986	0.052838	7801.560258	-3.369	0.000759	***
trial_exp105	-0.122740	0.052472	7751.877004	-2.339	0.019353	*
trial_exp106	-0.240128	0.052829	7784.868658	-4.545	5.57e-06	***
trial_exp107	-0.146284	0.052128	7733.762768	-2.806	0.005025	**
trial_exp108	-0.218439	0.052472	7752.164128	-4.163	3.18e-05	***
trial_exp109	-0.208091	0.052464	7765.011642	-3.966	7.36e-05	***
trial_exp110	-0.202553	0.052128	7733.845022	-3.886	0.000103	***
trial_exp111	-0.189949	0.052826	7774.639156	-3.596	0.000326	***
trial_exp112	-0.137235	0.052128	7733.919547	-2.633	0.008490	**
trial_exp113	-0.081378	0.051868	8035.360403	-1.569	0.116699	
trial_exp114	-0.115894	0.051797	7690.207783	-2.237	0.025283	*
trial_exp115	-0.187358	0.051797	7690.317089	-3.617	0.000300	***
trial_exp116	-0.123619	0.051797	7690.283316	-2.387	0.017028	*
trial_exp117	-0.062136	0.051797	7690.138205	-1.200	0.230323	
trial_exp118	-0.111097	0.052455	7741.067163	-2.118	0.034209	*
trial_exp119	-0.201329	0.052119	7719.308936	-3.863	0.000113	***
trial_exp120	-0.138289	0.051797	7690.120512	-2.670	0.007605	**
trial_exp121	-0.235382	0.051797	7690.262916	-4.544	5.59e-06	***
trial_exp122	-0.157606	0.051797	7690.281064	-3.043	0.002352	**
trial_exp123	-0.109470	0.051797	7690.206373	-2.113	0.034593	*
trial_exp124	-0.042012	0.052125	7706.569779	-0.806	0.420275	
trial_exp125	-0.212843	0.052462	7738.934460	-4.057	5.02e-05	***
trial_exp126	-0.212014	0.052123	7706.622357	-4.068	4.80e-05	***
trial_exp127	-0.158123	0.051797	7690.257797	-3.053	0.002275	**
trial_exp128	-0.171909	0.053202	7798.065504	-3.231	0.001238	**
trial_exp129	-0.030089	0.051540	7975.235350	-0.584	0.559370	
trial_exp130	-0.157418	0.052118	7667.260068	-3.020	0.002532	**
trial_exp131	-0.165599	0.052117	7667.404943	-3.177	0.001492	**
trial_exp132	-0.183445	0.052118	7666.665421	-3.520	0.000434	***
trial_exp133	-0.132128	0.052117	7667.552203	-2.535	0.011258	*
trial_exp134	-0.151150	0.052117	7667.349731	-2.900	0.003740	**
trial_exp135	-0.273809	0.051798	7650.143317	-5.286	1.28e-07	***
trial_exp136	-0.125748	0.051798	7649.793491	-2.428	0.015219	*
trial_exp137	-0.177001	0.052117	7666.759642	-3.396	0.000687	***
trial_exp138	-0.235359	0.052117	7667.344897	-4.516	6.40e-06	***
trial_exp139	-0.159570	0.051798	7650.042830	-3.081	0.002073	**
trial_exp140	-0.205617	0.052117	7666.782932	-3.945	8.04e-05	***
trial_exp141	-0.195842	0.051798	7650.199682	-3.781	0.000157	***
trial_exp142	-0.238356	0.052460	7687.268057	-4.544	5.62e-06	***

trial_exp143	-0.227788	0.052464	7692.379878	-4.342	1.43e-05	***
trial_exp144	-0.213439	0.051798	7649.985304	-4.121	3.82e-05	***
trial_exp145	-0.138739	0.051866	7981.794436	-2.675	0.007489	**
trial_exp146	-0.171374	0.052117	7666.479778	-3.288	0.001013	**
trial_exp147	-0.160142	0.052113	7680.553908	-3.073	0.002127	**
trial_exp148	-0.220545	0.051798	7651.007096	-4.258	2.09e-05	***
trial_exp149	-0.099245	0.052113	7681.563994	-1.904	0.056895	.
trial_exp150	-0.195913	0.051797	7651.126736	-3.782	0.000157	***
trial_exp151	-0.172460	0.052124	7682.402891	-3.309	0.000942	***
trial_exp152	-0.193860	0.052463	7692.634239	-3.695	0.000221	***
trial_exp153	-0.211558	0.052828	7735.302488	-4.005	6.27e-05	***
trial_exp154	-0.170154	0.052114	7680.366019	-3.265	0.001099	**
trial_exp155	-0.178675	0.052460	7699.492192	-3.406	0.000663	***
trial_exp156	-0.165310	0.051798	7650.956358	-3.191	0.001421	**
trial_exp157	-0.195672	0.052456	7713.692080	-3.730	0.000193	***
trial_exp158	-0.177773	0.052455	7713.313409	-3.389	0.000705	***
trial_exp159	-0.182959	0.052460	7699.379537	-3.488	0.000490	***
trial_exp160	-0.178569	0.052123	7682.883854	-3.426	0.000616	***
trial_exp161	-0.093508	0.052572	8006.529218	-1.779	0.075333	.
trial_exp162	-0.215600	0.052823	7724.401361	-4.082	4.52e-05	***
trial_exp163	-0.145438	0.052823	7723.311231	-2.753	0.005913	**
trial_exp164	-0.152632	0.052469	7705.304586	-2.909	0.003636	**
trial_exp165	-0.199109	0.052469	7705.484512	-3.795	0.000149	***
trial_exp166	-0.151021	0.052823	7724.142772	-2.859	0.004261	**
trial_exp167	-0.189804	0.053216	7737.267452	-3.567	0.000364	***
trial_exp168	-0.240469	0.052823	7723.469474	-4.552	5.39e-06	***
trial_exp169	-0.140466	0.052469	7705.027134	-2.677	0.007441	**
trial_exp170	-0.212495	0.052823	7723.236267	-4.023	5.81e-05	***
trial_exp171	-0.171798	0.052823	7738.143604	-3.252	0.001149	**
trial_exp172	-0.190539	0.053201	7728.907788	-3.581	0.000344	***
trial_exp173	-0.161694	0.052469	7705.213844	-3.082	0.002065	**
trial_exp174	-0.158864	0.052469	7705.156534	-3.028	0.002472	**
trial_exp175	-0.198457	0.052469	7705.127167	-3.782	0.000157	***
trial_exp176	-0.152690	0.052833	7733.969041	-2.890	0.003863	**
trial_exp177	-0.049601	0.052967	8063.015751	-0.936	0.349068	
trial_exp178	-0.154056	0.052841	7704.507328	-2.915	0.003562	**
trial_exp179	-0.106461	0.052832	7701.084309	-2.015	0.043932	*
trial_exp180	-0.124861	0.052831	7700.760855	-2.363	0.018133	*
trial_exp181	-0.105858	0.052478	7684.805189	-2.017	0.043709	*
trial_exp182	-0.231986	0.052832	7701.513335	-4.391	1.14e-05	***
trial_exp183	-0.193973	0.052832	7701.559777	-3.671	0.000243	***
trial_exp184	-0.227532	0.052832	7701.459463	-4.307	1.68e-05	***
trial_exp185	-0.198201	0.052478	7684.942333	-3.777	0.000160	***
trial_exp186	-0.180127	0.052478	7685.167526	-3.432	0.000601	***
trial_exp187	-0.146182	0.052832	7700.901711	-2.767	0.005673	**
trial_exp188	-0.106739	0.052839	7704.785073	-2.020	0.043410	*
trial_exp189	-0.178008	0.052478	7684.961702	-3.392	0.000697	***
trial_exp190	-0.185151	0.052832	7701.424310	-3.505	0.000460	***
trial_exp191	-0.193757	0.052478	7685.077303	-3.692	0.000224	***
trial_exp192	-0.169097	0.052478	7684.701053	-3.222	0.001277	**
trial_exp193	-0.087503	0.051765	7198.236514	-1.690	0.090995	.
trial_exp194	-0.184846	0.052032	6924.751475	-3.553	0.000384	***
trial_exp195	-0.157291	0.052032	6924.952212	-3.023	0.002512	**
trial_exp196	-0.168923	0.052360	6978.349741	-3.226	0.001260	**
trial_exp197	-0.202634	0.052356	6963.862308	-3.870	0.000110	***
trial_exp198	-0.200304	0.052032	6925.231480	-3.850	0.000119	***
trial_exp199	-0.188199	0.052360	6978.191502	-3.594	0.000327	***
trial_exp200	-0.169150	0.052032	6924.764061	-3.251	0.001156	**
trial_exp201	-0.154631	0.052032	6925.305059	-2.972	0.002970	**
trial_exp202	-0.187587	0.052032	6925.162264	-3.605	0.000314	***

trial_exp203	-0.182330	0.052686	7034.329048	-3.461	0.000542	***
trial_exp204	-0.233511	0.052363	6968.787150	-4.459	8.35e-06	***
trial_exp205	-0.203378	0.052363	6969.088037	-3.884	0.000104	***
trial_exp206	-0.181179	0.052032	6925.030957	-3.482	0.000501	***
trial_exp207	-0.232009	0.052356	6963.681042	-4.431	9.51e-06	***
trial_exp208	-0.239259	0.052353	6949.381241	-4.570	4.96e-06	***
trial_exp209	-0.088164	0.053191	7327.023223	-1.657	0.097461	.
trial_exp210	-0.205266	0.052692	6978.629950	-3.896	9.89e-05	***
trial_exp211	-0.207344	0.053038	7023.831515	-3.909	9.34e-05	***
trial_exp212	-0.135698	0.052692	6977.985964	-2.575	0.010035	*
trial_exp213	-0.192405	0.052691	6978.806421	-3.652	0.000263	***
trial_exp214	-0.144457	0.053059	7024.089467	-2.723	0.006494	**
trial_exp215	-0.191212	0.053059	7023.712485	-3.604	0.000316	***
trial_exp216	-0.234801	0.053437	7068.628137	-4.394	1.13e-05	***
trial_exp217	-0.194780	0.053054	7036.795727	-3.671	0.000243	***
trial_exp218	-0.260601	0.052692	6978.664541	-4.946	7.76e-07	***
trial_exp219	-0.184584	0.052692	6978.090214	-3.503	0.000463	***
trial_exp220	-0.225166	0.052691	6978.782334	-4.273	1.95e-05	***
trial_exp221	-0.223751	0.053059	7024.379113	-4.217	2.51e-05	***
trial_exp222	-0.217872	0.053059	7024.907153	-4.106	4.07e-05	***
trial_exp223	-0.292537	0.053437	7068.801731	-5.474	4.54e-08	***
trial_exp224	-0.181455	0.053059	7023.969667	-3.420	0.000630	***
trial_exp225	-0.102219	0.052437	7290.929385	-1.949	0.051292	.
trial_exp226	-0.169906	0.052032	6924.058551	-3.265	0.001098	**
trial_exp227	-0.134357	0.052364	6965.303526	-2.566	0.010314	*
trial_exp228	-0.152683	0.052708	6982.184955	-2.897	0.003782	**
trial_exp229	-0.159987	0.052368	6956.150168	-3.055	0.002259	**
trial_exp230	-0.111103	0.052355	6948.437253	-2.122	0.033863	*
trial_exp231	-0.245141	0.052678	7000.580416	-4.654	3.32e-06	***
trial_exp232	-0.190392	0.052032	6924.463876	-3.659	0.000255	***
trial_exp233	-0.200366	0.052032	6924.014708	-3.851	0.000119	***
trial_exp234	-0.207027	0.052698	7029.701856	-3.929	8.63e-05	***
trial_exp235	-0.201341	0.052355	6948.695722	-3.846	0.000121	***
trial_exp236	-0.215750	0.052356	6962.499882	-4.121	3.82e-05	***
trial_exp237	-0.228020	0.052346	6967.518427	-4.356	1.34e-05	***
trial_exp238	-0.194336	0.052346	6966.174916	-3.713	0.000207	***
trial_exp239	-0.151427	0.052368	6955.874422	-2.892	0.003845	**
trial_exp240	-0.212117	0.052695	6988.789086	-4.025	5.75e-05	***
trial_exp241	-0.044838	0.052804	7300.252001	-0.849	0.395835	
trial_exp242	-0.196101	0.053828	7168.332059	-3.643	0.000271	***
trial_exp243	-0.125431	0.053425	7105.079714	-2.348	0.018912	*
trial_exp244	-0.168704	0.053054	7067.721632	-3.180	0.001480	**
trial_exp245	-0.238308	0.053054	7067.953799	-4.492	7.17e-06	***
trial_exp246	-0.204777	0.053441	7116.417679	-3.832	0.000128	***
trial_exp247	-0.102222	0.053832	7158.649475	-1.899	0.057616	.
trial_exp248	-0.183130	0.053054	7068.143372	-3.452	0.000560	***
trial_exp249	-0.198104	0.053442	7118.834451	-3.707	0.000211	***
trial_exp250	-0.229745	0.053442	7116.272991	-4.299	1.74e-05	***
trial_exp251	-0.212695	0.053425	7105.059473	-3.981	6.93e-05	***
trial_exp252	-0.176952	0.053843	7161.800680	-3.286	0.001019	**
trial_exp253	-0.211203	0.053054	7067.990364	-3.981	6.93e-05	***
trial_exp254	-0.249186	0.053442	7118.726289	-4.663	3.18e-06	***
trial_exp255	-0.240082	0.053852	7168.968352	-4.458	8.39e-06	***
trial_exp256	-0.190294	0.054263	7191.912320	-3.507	0.000456	***
snr-6:lightlevel2	-0.021135	0.017451	8594.767498	-1.211	0.225886	
snr0:lightlevel2	0.026013	0.017329	8604.205069	1.501	0.133371	
snr3:lightlevel2	0.033720	0.017626	8497.361183	1.913	0.055773	.
snr-6:lightlevel3	-0.003191	0.017556	8594.528231	-0.182	0.855768	
snr0:lightlevel3	0.022434	0.017534	8304.439247	1.279	0.200765	
snr3:lightlevel3	0.070523	0.017812	6986.575269	3.959	7.59e-05	***

```
snr-6:lightlevel4    0.009735    0.017598 8630.255294   0.553 0.580161
snr0:lightlevel4     0.069330    0.017498 8631.537381   3.962 7.49e-05 ***
snr3:lightlevel4     0.102995    0.017610 8534.576399   5.849 5.14e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

18.8 Appendix 8: Output from Trial-Level Mean Dilation and Time-On-Task MEM: Final Model Results

```
Output captured by 'eatGet', Version 0.0.9, build 2018-05-15.
User: bald0106, computer: HXD4WF2, R version 4.0.1 (2020-06-06), Time: Tue
Oct 26 10:26:04 2021
Linear mixed model fit by maximum likelihood . t-tests use Satterthwaite's
method ['lmerModLmerTest']
Formula: meandilation ~ snr * lightlevel + trial_exp + (1 | participant) +
(1 | sentence)
   Data: pup_data

     AIC      BIC   logLik deviance df.resid
 -4720.8  -2781.8   2634.4  -5268.8     8475

Scaled residuals:
    Min      1Q  Median      3Q     Max
-6.4188 -0.5552  0.0045  0.5628  6.4695

Random effects:
 Groups      Name        Variance  Std.Dev.
 sentence    (Intercept) 0.0002105 0.01451
 participant (Intercept) 0.0039056 0.06249
 Residual                0.0314218 0.17726
Number of obs: 8749, groups:  sentence, 256; participant, 36

Fixed effects:
                     Estimate  Std. Error          df t value Pr(>|t|)
(Intercept)          0.373281    0.035509 3116.747548  10.512  < 2e-16 ***
snr-6               -0.010430    0.010777 8676.296800  -0.968 0.333187
snr0                -0.068541    0.010654 8711.683262  -6.433 1.32e-10 ***
snr3                -0.130124    0.010833 8543.498978 -12.012  < 2e-16 ***
lightlevel2         -0.054462    0.010716 7363.057429  -5.082 3.82e-07 ***
lightlevel3         -0.098055    0.010740 6565.968161  -9.130  < 2e-16 ***
lightlevel4         -0.123474    0.010684 7332.367135 -11.557  < 2e-16 ***
trial_exp2          -0.062715    0.046479 8239.599173  -1.349 0.177273
trial_exp3          -0.024753    0.046483 8223.270247  -0.533 0.594388
trial_exp4          -0.036201    0.046479 8222.126712  -0.779 0.436078
trial_exp5          -0.039231    0.045764 8192.231159  -0.857 0.391333
trial_exp6          -0.073400    0.045764 8192.235671  -1.604 0.108776
trial_exp7          -0.044606    0.046113 8207.263019  -0.967 0.333412
trial_exp8           0.001992    0.045764 8192.306530   0.044 0.965277
trial_exp9          -0.012020    0.045764 8192.214596  -0.263 0.792829
trial_exp10         -0.085614    0.045764 8192.235831  -1.871 0.061411 .
trial_exp11         -0.090887    0.046113 8207.352809  -1.971 0.048762 *
trial_exp12         -0.117557    0.045764 8192.137045  -2.569 0.010223 *
trial_exp13         -0.095564    0.045764 8192.184771  -2.088 0.036810 *
trial_exp14         -0.139878    0.046483 8223.074463  -3.009 0.002627 **
trial_exp15         -0.049641    0.046113 8206.961787  -1.076 0.281739
trial_exp16         -0.060416    0.045764 8191.898591  -1.320 0.186812
trial_exp17         -0.053672    0.044550 8515.225471  -1.205 0.228323
trial_exp18         -0.111073    0.045149 8122.166629  -2.460 0.013909 *
trial_exp19         -0.081376    0.045149 8122.217648  -1.802 0.071523 .
trial_exp20         -0.064834    0.045149 8122.505196  -1.436 0.151045
trial_exp21         -0.073789    0.045137 8168.100869  -1.635 0.102135
trial_exp22         -0.067520    0.044565 8101.141055  -1.515 0.129789
trial_exp23         -0.129867    0.044850 8111.555302  -2.896 0.003794 **
trial_exp24         -0.105510    0.044850 8111.239405  -2.353 0.018671 *
trial_exp25         -0.126382    0.044850 8110.810826  -2.818 0.004846 **
```

trial_exp26	-0.078309	0.045149	8122.473331	-1.734	0.082877	.
trial_exp27	-0.032697	0.045140	8151.537404	-0.724	0.468873	
trial_exp28	-0.101302	0.044850	8111.542876	-2.259	0.023929	*
trial_exp29	-0.090288	0.044565	8101.151408	-2.026	0.042800	*
trial_exp30	-0.178809	0.044850	8111.349000	-3.987	6.75e-05	***
trial_exp31	-0.021989	0.044565	8101.007220	-0.493	0.621732	
trial_exp32	-0.065985	0.044850	8110.994633	-1.471	0.141266	
trial_exp33	-0.007840	0.044266	8519.658505	-0.177	0.859419	
trial_exp34	-0.067087	0.044845	8125.851038	-1.496	0.134697	
trial_exp35	-0.022629	0.044845	8125.513756	-0.505	0.613848	
trial_exp36	-0.057534	0.045139	8152.451484	-1.275	0.202491	
trial_exp37	-0.010348	0.045144	8137.744212	-0.229	0.818701	
trial_exp38	-0.062431	0.044561	8115.014572	-1.401	0.161243	
trial_exp39	-0.053422	0.045137	8152.121080	-1.184	0.236628	
trial_exp40	-0.002689	0.045139	8136.381757	-0.060	0.952491	
trial_exp41	-0.131254	0.044845	8126.831842	-2.927	0.003434	**
trial_exp42	-0.117032	0.044561	8114.918410	-2.626	0.008647	**
trial_exp43	-0.123670	0.045144	8137.901161	-2.739	0.006168	**
trial_exp44	-0.142654	0.044561	8114.927544	-3.201	0.001373	**
trial_exp45	-0.097086	0.044561	8115.011726	-2.179	0.029380	*
trial_exp46	-0.133104	0.044561	8114.899130	-2.987	0.002826	**
trial_exp47	-0.094153	0.044561	8114.873370	-2.113	0.034639	*
trial_exp48	-0.089272	0.044841	8125.040151	-1.991	0.046530	*
trial_exp49	-0.042238	0.044557	8539.612352	-0.948	0.343172	
trial_exp50	-0.076676	0.044561	8114.426839	-1.721	0.085342	.
trial_exp51	-0.073367	0.044561	8114.389906	-1.646	0.099714	.
trial_exp52	-0.084657	0.044561	8114.246215	-1.900	0.057496	.
trial_exp53	-0.052935	0.044561	8114.508889	-1.188	0.234899	
trial_exp54	-0.101562	0.044845	8125.937640	-2.265	0.023556	*
trial_exp55	-0.093848	0.044561	8114.255871	-2.106	0.035229	*
trial_exp56	-0.107057	0.045142	8136.569100	-2.372	0.017735	*
trial_exp57	-0.143127	0.044845	8124.955400	-3.192	0.001421	**
trial_exp58	-0.109668	0.044845	8125.042207	-2.445	0.014488	*
trial_exp59	-0.127024	0.045146	8136.403427	-2.814	0.004910	**
trial_exp60	-0.154235	0.044561	8114.470188	-3.461	0.000541	***
trial_exp61	-0.138622	0.044561	8114.476024	-3.111	0.001872	**
trial_exp62	-0.154695	0.044561	8114.450180	-3.472	0.000520	***
trial_exp63	-0.146895	0.045460	8150.326382	-3.231	0.001237	**
trial_exp64	-0.136336	0.045143	8136.694059	-3.020	0.002535	**
trial_exp65	-0.025088	0.044554	8539.823639	-0.563	0.573387	
trial_exp66	-0.088475	0.044843	8124.197320	-1.973	0.048530	*
trial_exp67	-0.126603	0.044843	8124.144809	-2.823	0.004766	**
trial_exp68	-0.107379	0.044843	8124.127117	-2.395	0.016663	*
trial_exp69	-0.072460	0.045137	8151.601715	-1.605	0.108460	
trial_exp70	-0.081328	0.045453	8163.814371	-1.789	0.073609	.
trial_exp71	-0.114475	0.045141	8135.308693	-2.536	0.011233	*
trial_exp72	-0.119914	0.045141	8135.654365	-2.656	0.007912	**
trial_exp73	-0.139728	0.045142	8136.019623	-3.095	0.001973	**
trial_exp74	-0.122484	0.045141	8135.448117	-2.713	0.006674	**
trial_exp75	-0.190416	0.044843	8124.420010	-4.246	2.20e-05	***
trial_exp76	-0.166591	0.044843	8124.321324	-3.715	0.000205	***
trial_exp77	-0.137449	0.044843	8124.370505	-3.065	0.002183	**
trial_exp78	-0.159551	0.044843	8124.288337	-3.558	0.000376	***
trial_exp79	-0.146296	0.044843	8124.376248	-3.262	0.001109	**
trial_exp80	-0.128333	0.044843	8124.424137	-2.862	0.004223	**
trial_exp81	-0.067031	0.044555	8540.882686	-1.504	0.132502	
trial_exp82	-0.114837	0.045143	8134.920266	-2.544	0.010982	*
trial_exp83	-0.112975	0.044844	8123.902687	-2.519	0.011779	*
trial_exp84	-0.161582	0.045143	8134.568683	-3.579	0.000346	***
trial_exp85	-0.092345	0.044844	8123.879712	-2.059	0.039503	*

trial_exp86	-0.192087	0.044844	8123.951566	-4.283	1.86e-05	***
trial_exp87	-0.106134	0.044844	8123.997311	-2.367	0.017969	*
trial_exp88	-0.116253	0.044844	8123.981687	-2.592	0.009549	**
trial_exp89	-0.165295	0.044844	8123.730014	-3.686	0.000229	***
trial_exp90	-0.096464	0.044844	8123.840933	-2.151	0.031499	*
trial_exp91	-0.177838	0.044844	8123.936996	-3.966	7.38e-05	***
trial_exp92	-0.156710	0.045142	8134.614965	-3.471	0.000520	***
trial_exp93	-0.125171	0.045139	8150.545536	-2.773	0.005566	**
trial_exp94	-0.170300	0.044844	8123.809956	-3.798	0.000147	***
trial_exp95	-0.138762	0.044844	8123.611959	-3.094	0.001979	**
trial_exp96	-0.152483	0.045453	8162.244185	-3.355	0.000798	***
trial_exp97	0.009125	0.044548	8519.656505	0.205	0.837703	
trial_exp98	-0.096634	0.044841	8139.942119	-2.155	0.031189	*
trial_exp99	-0.128815	0.044841	8139.872964	-2.873	0.004080	**
trial_exp100	-0.106831	0.045771	8209.766057	-2.334	0.019618	*
trial_exp101	-0.063256	0.044841	8139.902345	-1.411	0.158378	
trial_exp102	-0.080550	0.044841	8139.992862	-1.796	0.072476	.
trial_exp103	-0.131363	0.045452	8164.259604	-2.890	0.003861	**
trial_exp104	-0.119069	0.045450	8180.132871	-2.620	0.008814	**
trial_exp105	-0.094130	0.045139	8151.713184	-2.085	0.037069	*
trial_exp106	-0.186815	0.045447	8179.319189	-4.111	3.99e-05	***
trial_exp107	-0.109414	0.044841	8139.843821	-2.440	0.014707	*
trial_exp108	-0.202853	0.045138	8151.961535	-4.494	7.09e-06	***
trial_exp109	-0.145829	0.045136	8167.396975	-3.231	0.001239	**
trial_exp110	-0.156495	0.044841	8139.985185	-3.490	0.000486	***
trial_exp111	-0.157821	0.045453	8165.055930	-3.472	0.000519	***
trial_exp112	-0.095961	0.044841	8140.044358	-2.140	0.032382	*
trial_exp113	-0.074291	0.044550	8529.164381	-1.668	0.095434	.
trial_exp114	-0.098955	0.044560	8114.020422	-2.221	0.026397	*
trial_exp115	-0.132122	0.044560	8114.034959	-2.965	0.003035	**
trial_exp116	-0.106292	0.044560	8113.862555	-2.385	0.017084	*
trial_exp117	-0.071431	0.044560	8113.967477	-1.603	0.108967	
trial_exp118	-0.096713	0.045137	8152.077082	-2.143	0.032170	*
trial_exp119	-0.156071	0.044839	8138.937481	-3.481	0.000503	***
trial_exp120	-0.117173	0.044560	8113.930006	-2.630	0.008566	**
trial_exp121	-0.186628	0.044560	8114.022522	-4.188	2.84e-05	***
trial_exp122	-0.123939	0.044560	8114.004180	-2.781	0.005425	**
trial_exp123	-0.118864	0.044560	8113.974400	-2.667	0.007657	**
trial_exp124	-0.042291	0.044844	8124.212430	-0.943	0.345680	
trial_exp125	-0.162448	0.045140	8151.010519	-3.599	0.000322	***
trial_exp126	-0.168204	0.044843	8124.248949	-3.751	0.000177	***
trial_exp127	-0.138103	0.044560	8113.887789	-3.099	0.001947	**
trial_exp128	-0.148794	0.045776	8192.312694	-3.250	0.001157	**
trial_exp129	-0.037040	0.044264	8520.045890	-0.837	0.402734	
trial_exp130	-0.126218	0.044844	8125.375292	-2.815	0.004895	**
trial_exp131	-0.119806	0.044844	8125.358846	-2.672	0.007564	**
trial_exp132	-0.148818	0.044844	8125.188970	-3.319	0.000909	***
trial_exp133	-0.087611	0.044844	8125.526953	-1.954	0.050774	.
trial_exp134	-0.138462	0.044844	8125.382997	-3.088	0.002024	**
trial_exp135	-0.205303	0.044560	8114.244576	-4.607	4.14e-06	***
trial_exp136	-0.111371	0.044560	8114.003781	-2.499	0.012461	*
trial_exp137	-0.180441	0.044844	8125.258663	-4.024	5.78e-05	***
trial_exp138	-0.192292	0.044844	8125.371300	-4.288	1.82e-05	***
trial_exp139	-0.122031	0.044560	8114.089329	-2.739	0.006184	**
trial_exp140	-0.181745	0.044844	8125.256131	-4.053	5.11e-05	***
trial_exp141	-0.170176	0.044560	8114.315026	-3.819	0.000135	***
trial_exp142	-0.193649	0.045143	8137.280770	-4.290	1.81e-05	***
trial_exp143	-0.166045	0.045142	8135.878715	-3.678	0.000236	***
trial_exp144	-0.198170	0.044560	8113.987933	-4.447	8.81e-06	***
trial_exp145	-0.141117	0.044544	8506.565249	-3.168	0.001540	**

trial_exp146	-0.138739	0.044844	8124.869023	-3.094	0.001983	**
trial_exp147	-0.140984	0.044840	8139.274485	-3.144	0.001672	**
trial_exp148	-0.165393	0.044560	8114.015505	-3.712	0.000207	***
trial_exp149	-0.083245	0.044839	8139.336754	-1.857	0.063416	.
trial_exp150	-0.161225	0.044560	8114.021682	-3.618	0.000298	***
trial_exp151	-0.121585	0.044840	8139.813337	-2.712	0.006712	**
trial_exp152	-0.175654	0.045142	8136.202493	-3.891	0.000101	***
trial_exp153	-0.166705	0.045450	8164.468159	-3.668	0.000246	***
trial_exp154	-0.131172	0.044840	8139.236286	-2.925	0.003450	**
trial_exp155	-0.139001	0.045139	8152.229213	-3.079	0.002081	**
trial_exp156	-0.161222	0.044560	8113.926410	-3.618	0.000299	***
trial_exp157	-0.162470	0.045134	8166.503257	-3.600	0.000320	***
trial_exp158	-0.125309	0.045134	8166.747305	-2.776	0.005510	**
trial_exp159	-0.150510	0.045139	8152.122561	-3.334	0.000859	***
trial_exp160	-0.164461	0.044840	8140.055089	-3.668	0.000246	***
trial_exp161	-0.086726	0.045153	8503.667909	-1.921	0.054800	.
trial_exp162	-0.179242	0.045453	8163.478613	-3.943	8.10e-05	***
trial_exp163	-0.112985	0.045453	8164.545161	-2.486	0.012947	*
trial_exp164	-0.126122	0.045138	8151.205482	-2.794	0.005216	**
trial_exp165	-0.148049	0.045138	8151.568773	-3.280	0.001043	**
trial_exp166	-0.139327	0.045453	8163.768368	-3.065	0.002181	**
trial_exp167	-0.146270	0.045785	8160.738753	-3.195	0.001405	**
trial_exp168	-0.183133	0.045452	8164.642794	-4.029	5.65e-05	***
trial_exp169	-0.128143	0.045138	8151.589999	-2.839	0.004538	**
trial_exp170	-0.170242	0.045452	8164.639619	-3.746	0.000181	***
trial_exp171	-0.153893	0.045445	8180.020393	-3.386	0.000712	***
trial_exp172	-0.166367	0.045789	8161.404695	-3.633	0.000281	***
trial_exp173	-0.147048	0.045138	8151.481589	-3.258	0.001128	**
trial_exp174	-0.138829	0.045138	8151.641095	-3.076	0.002107	**
trial_exp175	-0.168001	0.045138	8151.461798	-3.722	0.000199	***
trial_exp176	-0.135932	0.045451	8163.803856	-2.991	0.002792	**
trial_exp177	-0.059212	0.045509	8564.480914	-1.301	0.193251	
trial_exp178	-0.141985	0.045460	8133.891083	-3.123	0.001795	**
trial_exp179	-0.099103	0.045459	8132.904668	-2.180	0.029284	*
trial_exp180	-0.110733	0.045458	8132.537371	-2.436	0.014874	*
trial_exp181	-0.090036	0.045144	8121.579757	-1.994	0.046141	*
trial_exp182	-0.199088	0.045459	8133.193616	-4.379	1.20e-05	***
trial_exp183	-0.165515	0.045459	8133.413911	-3.641	0.000273	***
trial_exp184	-0.169974	0.045459	8133.353928	-3.739	0.000186	***
trial_exp185	-0.157429	0.045144	8121.789430	-3.487	0.000491	***
trial_exp186	-0.169445	0.045144	8121.763169	-3.753	0.000176	***
trial_exp187	-0.133709	0.045459	8132.973456	-2.941	0.003278	**
trial_exp188	-0.107246	0.045459	8133.359197	-2.359	0.018339	*
trial_exp189	-0.172452	0.045144	8121.440949	-3.820	0.000134	***
trial_exp190	-0.150534	0.045459	8133.300583	-3.311	0.000932	***
trial_exp191	-0.181943	0.045144	8121.407643	-4.030	5.62e-05	***
trial_exp192	-0.149631	0.045144	8121.116037	-3.314	0.000922	***
trial_exp193	-0.078182	0.044262	8504.124315	-1.766	0.077372	.
trial_exp194	-0.129359	0.044568	8086.121329	-2.902	0.003712	**
trial_exp195	-0.112655	0.044568	8086.610326	-2.528	0.011500	*
trial_exp196	-0.131489	0.044849	8111.546222	-2.932	0.003379	**
trial_exp197	-0.149433	0.044853	8097.554694	-3.332	0.000867	***
trial_exp198	-0.159780	0.044568	8086.541717	-3.585	0.000339	***
trial_exp199	-0.132411	0.044849	8111.382004	-2.952	0.003162	**
trial_exp200	-0.130271	0.044568	8086.353362	-2.923	0.003477	**
trial_exp201	-0.117017	0.044568	8086.729755	-2.626	0.008667	**
trial_exp202	-0.141622	0.044568	8086.542001	-3.178	0.001490	**
trial_exp203	-0.141979	0.045140	8152.380457	-3.145	0.001665	**
trial_exp204	-0.191162	0.044853	8097.101779	-4.262	2.05e-05	***
trial_exp205	-0.170463	0.044853	8096.688977	-3.801	0.000145	***

trial_exp206	-0.150461	0.044568	8086.862513	-3.376	0.000739	***
trial_exp207	-0.180813	0.044853	8097.058973	-4.031	5.60e-05	***
trial_exp208	-0.202793	0.044848	8111.022574	-4.522	6.22e-06	***
trial_exp209	-0.071546	0.045496	8533.470059	-1.573	0.115853	
trial_exp210	-0.160607	0.045142	8137.385337	-3.558	0.000376	***
trial_exp211	-0.169826	0.045448	8164.378713	-3.737	0.000188	***
trial_exp212	-0.091814	0.045142	8137.021954	-2.034	0.041995	*
trial_exp213	-0.142293	0.045142	8137.600690	-3.152	0.001627	**
trial_exp214	-0.127734	0.045457	8149.447026	-2.810	0.004966	**
trial_exp215	-0.141161	0.045457	8149.320992	-3.105	0.001907	**
trial_exp216	-0.198188	0.045789	8163.228001	-4.328	1.52e-05	***
trial_exp217	-0.155210	0.045452	8165.109341	-3.415	0.000641	***
trial_exp218	-0.209470	0.045142	8137.290501	-4.640	3.53e-06	***
trial_exp219	-0.149611	0.045142	8137.215822	-3.314	0.000923	***
trial_exp220	-0.172589	0.045142	8137.585050	-3.823	0.000133	***
trial_exp221	-0.207784	0.045457	8149.725083	-4.571	4.93e-06	***
trial_exp222	-0.173759	0.045457	8150.144833	-3.823	0.000133	***
trial_exp223	-0.236436	0.045789	8163.050874	-5.164	2.48e-07	***
trial_exp224	-0.164987	0.045457	8149.116014	-3.629	0.000286	***
trial_exp225	-0.083496	0.044850	8537.600760	-1.862	0.062683	.
trial_exp226	-0.131984	0.044568	8085.923554	-2.961	0.003071	**
trial_exp227	-0.109897	0.044853	8095.517915	-2.450	0.014300	*
trial_exp228	-0.107320	0.045153	8104.784774	-2.377	0.017487	*
trial_exp229	-0.120451	0.044857	8080.376410	-2.685	0.007263	**
trial_exp230	-0.091407	0.044849	8110.304185	-2.038	0.041573	*
trial_exp231	-0.181581	0.045143	8121.793622	-4.022	5.81e-05	***
trial_exp232	-0.141387	0.044568	8086.039020	-3.172	0.001518	**
trial_exp233	-0.159424	0.044568	8085.567488	-3.577	0.000349	***
trial_exp234	-0.182201	0.045144	8136.849572	-4.036	5.49e-05	***
trial_exp235	-0.169667	0.044849	8110.345798	-3.783	0.000156	***
trial_exp236	-0.188793	0.044853	8096.150637	-4.209	2.59e-05	***
trial_exp237	-0.198612	0.044848	8111.276666	-4.429	9.61e-06	***
trial_exp238	-0.175695	0.044846	8110.853657	-3.918	9.01e-05	***
trial_exp239	-0.123236	0.044857	8079.712328	-2.747	0.006022	**
trial_exp240	-0.182211	0.045148	8121.678123	-4.036	5.49e-05	***
trial_exp241	-0.018861	0.045165	8526.490562	-0.418	0.676239	
trial_exp242	-0.171475	0.046125	8225.809090	-3.718	0.000202	***
trial_exp243	-0.109279	0.045781	8194.358265	-2.387	0.017008	*
trial_exp244	-0.119417	0.045452	8180.407392	-2.627	0.008622	**
trial_exp245	-0.206424	0.045452	8180.571821	-4.542	5.66e-06	***
trial_exp246	-0.175447	0.045784	8194.711914	-3.832	0.000128	***
trial_exp247	-0.082133	0.046135	8210.378798	-1.780	0.075068	.
trial_exp248	-0.144095	0.045452	8180.625906	-3.170	0.001529	**
trial_exp249	-0.183212	0.045785	8195.086272	-4.002	6.35e-05	***
trial_exp250	-0.202150	0.045784	8194.559320	-4.415	1.02e-05	***
trial_exp251	-0.178873	0.045781	8194.444304	-3.907	9.41e-05	***
trial_exp252	-0.186283	0.046136	8209.818393	-4.038	5.45e-05	***
trial_exp253	-0.156328	0.045452	8180.571744	-3.439	0.000586	***
trial_exp254	-0.245662	0.045785	8195.067167	-5.366	8.29e-08	***
trial_exp255	-0.201975	0.046136	8209.881759	-4.378	1.21e-05	***
trial_exp256	-0.171798	0.046508	8206.660517	-3.694	0.000222	***
snr-6:lightlevel2	-0.024474	0.015118	8690.579218	-1.619	0.105521	
snr0:lightlevel2	0.026633	0.015042	8712.805423	1.771	0.076653	.
snr3:lightlevel2	0.028347	0.015267	8621.922053	1.857	0.063378	.
snr-6:lightlevel3	-0.018027	0.015209	8685.741007	-1.185	0.235930	
snr0:lightlevel3	0.025518	0.015213	8475.793728	1.677	0.093500	.
snr3:lightlevel3	0.059756	0.015442	7268.277355	3.870	0.000110	***
snr-6:lightlevel4	-0.004882	0.015237	8698.209348	-0.320	0.748684	
snr0:lightlevel4	0.057384	0.015175	8701.738520	3.781	0.000157	***
snr3:lightlevel4	0.087620	0.015256	8644.406178	5.743	9.60e-09	***

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

18.9 Appendix 9: Output from Trial-Level Baseline Diameter and Time-On-Task MEM: Final Model Results

```
Output captured by 'eatGet', Version 0.0.9, build 2018-05-15.
User: bald0106, computer: HXD4WF2, R version 4.0.1 (2020-06-06), Time: Tue
Oct 26 10:36:13 2021
Linear mixed model fit by maximum likelihood . t-tests use Satterthwaite's
method ['lmerModLmerTest']
Formula: baseline ~ lightlevel * snr + trial_exp + (1 + lightlevel + snr |
participant)
   Data: pup_data
Control: lmerControl(optimizer = "bobyqa")

     AIC      BIC   logLik deviance df.resid
  2291.2   4414.2   -845.6   1691.2     8449

Scaled residuals:
    Min      1Q  Median      3Q     Max
-5.0204 -0.5015  0.0106  0.5021  7.8207

Random effects:
 Groups      Name        Variance Std.Dev. Corr
 participant (Intercept) 0.393986 0.62768
             lightlevel2 0.078464 0.28012  -0.87
             lightlevel3 0.144167 0.37969  -0.89  0.97
             lightlevel4 0.161405 0.40175  -0.91  0.96  0.98
             snr-6       0.005092 0.07136   0.04 -0.16 -0.24 -0.24
             snr0        0.009721 0.09860  -0.40  0.33  0.35  0.47 -0.09
             snr3       0.013443 0.11594  -0.37  0.20  0.23  0.33  0.19  0.83
 Residual                0.066413 0.25771
Number of obs: 8749, groups:  participant, 36

Fixed effects:
                     Estimate  Std. Error          df t value Pr(>|t|)
(Intercept)          4.937176    0.116102   54.042223  42.524  < 2e-16 ***
lightlevel2         -0.830922    0.049170   41.863101 -16.899  < 2e-16 ***
lightlevel3         -1.172974    0.065134   38.997380 -18.009  < 2e-16 ***
lightlevel4         -1.375501    0.068703   38.742214 -20.021  < 2e-16 ***
snr-6                0.110360    0.019757  126.316239   5.586 1.36e-07 ***
snr0                -0.041936    0.022557   73.444303  -1.859 0.067020 .
snr3                -0.088657    0.024962   65.781392  -3.552 0.000713 ***
trial_exp2          -0.135957    0.067177 8505.336067  -2.024 0.043015 *
trial_exp3          -0.228866    0.067176 8505.124363  -3.407 0.000660 ***
trial_exp4          -0.288934    0.067151 8501.408608  -4.303 1.71e-05 ***
trial_exp5          -0.301910    0.066109 8500.699803  -4.567 5.02e-06 ***
trial_exp6          -0.296635    0.066109 8500.699804  -4.487 7.32e-06 ***
trial_exp7          -0.391409    0.066629 8502.727951  -5.874 4.40e-09 ***
trial_exp8          -0.461160    0.066109 8500.699803  -6.976 3.27e-12 ***
trial_exp9          -0.468029    0.066109 8500.699803  -7.080 1.56e-12 ***
trial_exp10         -0.416710    0.066109 8500.699800  -6.303 3.06e-10 ***
trial_exp11         -0.503092    0.066627 8503.564983  -7.551 4.77e-14 ***
trial_exp12         -0.482108    0.066109 8500.699801  -7.293 3.31e-13 ***
trial_exp13         -0.540960    0.066109 8500.699802  -8.183 3.17e-16 ***
trial_exp14         -0.501819    0.067178 8505.114618  -7.470 8.82e-14 ***
trial_exp15         -0.575000    0.066630 8502.233097  -8.630  < 2e-16 ***
trial_exp16         -0.614504    0.066109 8500.699801  -9.295  < 2e-16 ***
trial_exp17         -0.285904    0.065789 7848.941192  -4.346 1.41e-05 ***
trial_exp18         -0.409763    0.066223 7878.874438  -6.188 6.41e-10 ***
trial_exp19         -0.543570    0.066228 7873.363475  -8.208 2.62e-16 ***
```

trial_exp20	-0.641605	0.066228	7873.363472	-9.688	< 2e-16	***
trial_exp21	-0.624391	0.066221	7892.010936	-9.429	< 2e-16	***
trial_exp22	-0.614598	0.065369	7843.354455	-9.402	< 2e-16	***
trial_exp23	-0.541860	0.065782	7867.761073	-8.237	< 2e-16	***
trial_exp24	-0.606596	0.065787	7867.468873	-9.221	< 2e-16	***
trial_exp25	-0.622169	0.065772	7859.312758	-9.459	< 2e-16	***
trial_exp26	-0.656358	0.066212	7881.968424	-9.913	< 2e-16	***
trial_exp27	-0.727123	0.066224	7872.060922	-10.980	< 2e-16	***
trial_exp28	-0.653613	0.065787	7867.468877	-9.935	< 2e-16	***
trial_exp29	-0.699248	0.065369	7843.354470	-10.697	< 2e-16	***
trial_exp30	-0.615326	0.065784	7854.577868	-9.354	< 2e-16	***
trial_exp31	-0.676549	0.065369	7843.354453	-10.350	< 2e-16	***
trial_exp32	-0.638681	0.065789	7848.941193	-9.708	< 2e-16	***
trial_exp33	-0.174936	0.065386	7846.382488	-2.675	0.007479	**
trial_exp34	-0.316568	0.065803	7870.887969	-4.811	1.53e-06	***
trial_exp35	-0.448653	0.065806	7855.903979	-6.818	9.92e-12	***
trial_exp36	-0.546836	0.066233	7869.813425	-8.256	< 2e-16	***
trial_exp37	-0.542653	0.066251	7873.324536	-8.191	3.00e-16	***
trial_exp38	-0.569910	0.065386	7846.382500	-8.716	< 2e-16	***
trial_exp39	-0.586649	0.066231	7881.406495	-8.858	< 2e-16	***
trial_exp40	-0.664660	0.066218	7890.238865	-10.037	< 2e-16	***
trial_exp41	-0.564537	0.065797	7866.225680	-8.580	< 2e-16	***
trial_exp42	-0.510242	0.065386	7846.382497	-7.804	6.80e-15	***
trial_exp43	-0.594657	0.066248	7874.243507	-8.976	< 2e-16	***
trial_exp44	-0.554070	0.065386	7846.382498	-8.474	< 2e-16	***
trial_exp45	-0.643560	0.065386	7846.382499	-9.842	< 2e-16	***
trial_exp46	-0.583414	0.065386	7846.382493	-8.923	< 2e-16	***
trial_exp47	-0.605608	0.065386	7846.382493	-9.262	< 2e-16	***
trial_exp48	-0.617555	0.065782	7864.864457	-9.388	< 2e-16	***
trial_exp49	-0.143548	0.066336	6349.314976	-2.164	0.030506	*
trial_exp50	-0.270281	0.065932	6297.512231	-4.099	4.19e-05	***
trial_exp51	-0.442783	0.065932	6297.512231	-6.716	2.04e-11	***
trial_exp52	-0.425789	0.065932	6297.512235	-6.458	1.14e-10	***
trial_exp53	-0.470017	0.065932	6297.512233	-7.129	1.13e-12	***
trial_exp54	-0.485360	0.066348	6348.043032	-7.315	2.88e-13	***
trial_exp55	-0.496813	0.065932	6297.512232	-7.535	5.56e-14	***
trial_exp56	-0.573789	0.066773	6386.234765	-8.593	< 2e-16	***
trial_exp57	-0.558803	0.066336	6349.314966	-8.424	< 2e-16	***
trial_exp58	-0.562526	0.066336	6349.314984	-8.480	< 2e-16	***
trial_exp59	-0.594598	0.066777	6402.219487	-8.904	< 2e-16	***
trial_exp60	-0.631867	0.065932	6297.512244	-9.584	< 2e-16	***
trial_exp61	-0.678256	0.065932	6297.512229	-10.287	< 2e-16	***
trial_exp62	-0.594461	0.065932	6297.512228	-9.016	< 2e-16	***
trial_exp63	-0.594262	0.067239	6437.364407	-8.838	< 2e-16	***
trial_exp64	-0.640282	0.066772	6393.408515	-9.589	< 2e-16	***
trial_exp65	-0.190661	0.067241	6148.301328	-2.836	0.004590	**
trial_exp66	-0.301315	0.067241	6148.301359	-4.481	7.56e-06	***
trial_exp67	-0.358558	0.067241	6148.301340	-5.332	1.00e-07	***
trial_exp68	-0.419890	0.067241	6148.301366	-6.245	4.53e-10	***
trial_exp69	-0.436045	0.067659	6191.426184	-6.445	1.25e-10	***
trial_exp70	-0.492000	0.068102	6257.376287	-7.224	5.63e-13	***
trial_exp71	-0.435169	0.067671	6202.454666	-6.431	1.37e-10	***
trial_exp72	-0.474386	0.067671	6202.454671	-7.010	2.63e-12	***
trial_exp73	-0.516124	0.067672	6201.096234	-7.627	2.77e-14	***
trial_exp74	-0.551280	0.067671	6202.454672	-8.146	4.49e-16	***
trial_exp75	-0.514387	0.067241	6148.301363	-7.650	2.32e-14	***
trial_exp76	-0.586878	0.067241	6148.301368	-8.728	< 2e-16	***
trial_exp77	-0.634255	0.067241	6148.301332	-9.433	< 2e-16	***
trial_exp78	-0.597878	0.067241	6148.301338	-8.892	< 2e-16	***
trial_exp79	-0.599074	0.067241	6148.301362	-8.909	< 2e-16	***

trial_exp80	-0.656313	0.067241	6148.301352	-9.761	< 2e-16	***
trial_exp81	-0.292065	0.067749	5157.971129	-4.311	1.66e-05	***
trial_exp82	-0.386875	0.068188	5225.712856	-5.674	1.47e-08	***
trial_exp83	-0.464814	0.067749	5157.971125	-6.861	7.65e-12	***
trial_exp84	-0.464823	0.068164	5208.510801	-6.819	1.02e-11	***
trial_exp85	-0.499222	0.067749	5157.971131	-7.369	2.00e-13	***
trial_exp86	-0.431894	0.067749	5157.971122	-6.375	1.99e-10	***
trial_exp87	-0.500328	0.067749	5157.971123	-7.385	1.77e-13	***
trial_exp88	-0.545848	0.067749	5157.971131	-8.057	9.64e-16	***
trial_exp89	-0.493311	0.067749	5157.971121	-7.281	3.80e-13	***
trial_exp90	-0.578865	0.067749	5157.971129	-8.544	< 2e-16	***
trial_exp91	-0.584762	0.067749	5157.971128	-8.631	< 2e-16	***
trial_exp92	-0.607358	0.068174	5211.770715	-8.909	< 2e-16	***
trial_exp93	-0.644357	0.068172	5210.214663	-9.452	< 2e-16	***
trial_exp94	-0.582094	0.067749	5157.971107	-8.592	< 2e-16	***
trial_exp95	-0.674059	0.067749	5157.971118	-9.949	< 2e-16	***
trial_exp96	-0.554503	0.068610	5261.500245	-8.082	7.84e-16	***
trial_exp97	-0.303645	0.066294	6733.532546	-4.580	4.73e-06	***
trial_exp98	-0.250565	0.066294	6733.532563	-3.780	0.000158	***
trial_exp99	-0.415182	0.066294	6733.532547	-6.263	4.02e-10	***
trial_exp100	-0.457992	0.067659	6860.886140	-6.769	1.40e-11	***
trial_exp101	-0.496085	0.066294	6733.532556	-7.483	8.18e-14	***
trial_exp102	-0.517308	0.066294	6733.532548	-7.803	6.95e-15	***
trial_exp103	-0.466926	0.067190	6822.753766	-6.949	4.01e-12	***
trial_exp104	-0.489077	0.067182	6821.141070	-7.280	3.71e-13	***
trial_exp105	-0.553384	0.066727	6776.335892	-8.293	< 2e-16	***
trial_exp106	-0.483750	0.067181	6821.061773	-7.201	6.63e-13	***
trial_exp107	-0.624731	0.066294	6733.532550	-9.424	< 2e-16	***
trial_exp108	-0.560942	0.066724	6776.529001	-8.407	< 2e-16	***
trial_exp109	-0.601373	0.066731	6773.346045	-9.012	< 2e-16	***
trial_exp110	-0.573458	0.066294	6733.532563	-8.650	< 2e-16	***
trial_exp111	-0.600119	0.067182	6813.289182	-8.933	< 2e-16	***
trial_exp112	-0.648399	0.066294	6733.532559	-9.781	< 2e-16	***
trial_exp113	-0.287074	0.067241	6194.552506	-4.269	1.99e-05	***
trial_exp114	-0.345631	0.066832	6136.642911	-5.172	2.39e-07	***
trial_exp115	-0.463835	0.066832	6136.642905	-6.940	4.31e-12	***
trial_exp116	-0.558079	0.066832	6136.642900	-8.351	< 2e-16	***
trial_exp117	-0.548591	0.066832	6136.642903	-8.209	2.71e-16	***
trial_exp118	-0.477379	0.067671	6238.809203	-7.054	1.92e-12	***
trial_exp119	-0.510314	0.067233	6184.733439	-7.590	3.66e-14	***
trial_exp120	-0.579902	0.066832	6136.642914	-8.677	< 2e-16	***
trial_exp121	-0.530996	0.066832	6136.642913	-7.945	2.29e-15	***
trial_exp122	-0.567536	0.066832	6136.642902	-8.492	< 2e-16	***
trial_exp123	-0.618524	0.066832	6136.642898	-9.255	< 2e-16	***
trial_exp124	-0.725804	0.067235	6174.192147	-10.795	< 2e-16	***
trial_exp125	-0.599954	0.067667	6230.682890	-8.866	< 2e-16	***
trial_exp126	-0.553880	0.067238	6186.853339	-8.238	< 2e-16	***
trial_exp127	-0.635518	0.066832	6136.642910	-9.509	< 2e-16	***
trial_exp128	-0.615196	0.068574	6324.689522	-8.971	< 2e-16	***
trial_exp129	-0.351567	0.067339	5053.750862	-5.221	1.85e-07	***
trial_exp130	-0.352758	0.067734	5100.076051	-5.208	1.98e-07	***
trial_exp131	-0.473283	0.067734	5100.076052	-6.987	3.16e-12	***
trial_exp132	-0.511263	0.067729	5102.939148	-7.549	5.18e-14	***
trial_exp133	-0.550986	0.067734	5100.076034	-8.135	5.15e-16	***
trial_exp134	-0.516992	0.067734	5100.076035	-7.633	2.73e-14	***
trial_exp135	-0.469802	0.067339	5053.750842	-6.977	3.41e-12	***
trial_exp136	-0.564124	0.067339	5053.750834	-8.377	< 2e-16	***
trial_exp137	-0.512480	0.067729	5102.939164	-7.567	4.52e-14	***
trial_exp138	-0.505763	0.067734	5100.076044	-7.467	9.61e-14	***
trial_exp139	-0.542562	0.067339	5053.750846	-8.057	9.66e-16	***

trial_exp140	-0.554972	0.067729	5102.939160	-8.194	3.17e-16	***
trial_exp141	-0.554341	0.067339	5053.750825	-8.232	2.32e-16	***
trial_exp142	-0.558478	0.068160	5149.543318	-8.194	3.17e-16	***
trial_exp143	-0.577410	0.068164	5172.683555	-8.471	< 2e-16	***
trial_exp144	-0.574991	0.067339	5053.750839	-8.539	< 2e-16	***
trial_exp145	-0.236872	0.067228	6095.181537	-3.523	0.000429	***
trial_exp146	-0.378195	0.067222	6094.379467	-5.626	1.93e-08	***
trial_exp147	-0.516550	0.067217	6100.527562	-7.685	1.77e-14	***
trial_exp148	-0.468167	0.066826	6053.740734	-7.006	2.72e-12	***
trial_exp149	-0.564745	0.067223	6108.115928	-8.401	< 2e-16	***
trial_exp150	-0.539577	0.066826	6053.740726	-8.074	8.12e-16	***
trial_exp151	-0.504804	0.067241	6100.277614	-7.507	6.90e-14	***
trial_exp152	-0.543612	0.067638	6140.248040	-8.037	1.10e-15	***
trial_exp153	-0.549504	0.068077	6181.870587	-8.072	8.26e-16	***
trial_exp154	-0.616949	0.067217	6100.527553	-9.178	< 2e-16	***
trial_exp155	-0.654640	0.067657	6142.695829	-9.676	< 2e-16	***
trial_exp156	-0.636434	0.066826	6053.740728	-9.524	< 2e-16	***
trial_exp157	-0.596834	0.067652	6149.132212	-8.822	< 2e-16	***
trial_exp158	-0.663058	0.067657	6154.822203	-9.800	< 2e-16	***
trial_exp159	-0.562397	0.067657	6142.695823	-8.312	< 2e-16	***
trial_exp160	-0.653890	0.067241	6100.277598	-9.725	< 2e-16	***
trial_exp161	-0.298190	0.068090	6180.247856	-4.379	1.21e-05	***
trial_exp162	-0.275114	0.068070	6175.673292	-4.042	5.37e-05	***
trial_exp163	-0.476261	0.068085	6178.748991	-6.995	2.93e-12	***
trial_exp164	-0.491649	0.067642	6130.818251	-7.268	4.09e-13	***
trial_exp165	-0.489584	0.067642	6130.818244	-7.238	5.11e-13	***
trial_exp166	-0.566538	0.068070	6175.673291	-8.323	< 2e-16	***
trial_exp167	-0.588636	0.068575	6243.105940	-8.584	< 2e-16	***
trial_exp168	-0.514982	0.068085	6178.748985	-7.564	4.48e-14	***
trial_exp169	-0.593943	0.067642	6130.818238	-8.781	< 2e-16	***
trial_exp170	-0.595203	0.068085	6178.748984	-8.742	< 2e-16	***
trial_exp171	-0.600013	0.068095	6181.869420	-8.811	< 2e-16	***
trial_exp172	-0.652708	0.068553	6229.259588	-9.521	< 2e-16	***
trial_exp173	-0.638496	0.067642	6130.818229	-9.439	< 2e-16	***
trial_exp174	-0.714418	0.067642	6130.818250	-10.562	< 2e-16	***
trial_exp175	-0.624837	0.067642	6130.818237	-9.237	< 2e-16	***
trial_exp176	-0.676279	0.068103	6196.572556	-9.930	< 2e-16	***
trial_exp177	-0.346974	0.067702	6766.221451	-5.125	3.06e-07	***
trial_exp178	-0.312697	0.067232	6720.492623	-4.651	3.37e-06	***
trial_exp179	-0.517462	0.067216	6720.887107	-7.698	1.58e-14	***
trial_exp180	-0.540443	0.067223	6721.286434	-8.040	1.06e-15	***
trial_exp181	-0.591357	0.066771	6676.107394	-8.856	< 2e-16	***
trial_exp182	-0.492469	0.067216	6720.887090	-7.327	2.64e-13	***
trial_exp183	-0.540166	0.067216	6720.887089	-8.036	1.09e-15	***
trial_exp184	-0.572987	0.067216	6720.887083	-8.525	< 2e-16	***
trial_exp185	-0.601823	0.066771	6676.107417	-9.013	< 2e-16	***
trial_exp186	-0.635112	0.066771	6676.107416	-9.512	< 2e-16	***
trial_exp187	-0.604590	0.067216	6720.887092	-8.995	< 2e-16	***
trial_exp188	-0.663973	0.067236	6723.792233	-9.875	< 2e-16	***
trial_exp189	-0.630533	0.066771	6676.107407	-9.443	< 2e-16	***
trial_exp190	-0.668814	0.067216	6720.887101	-9.950	< 2e-16	***
trial_exp191	-0.641082	0.066771	6676.107404	-9.601	< 2e-16	***
trial_exp192	-0.636912	0.066771	6676.107412	-9.539	< 2e-16	***
trial_exp193	-0.337019	0.066907	5068.194749	-5.037	4.89e-07	***
trial_exp194	-0.399969	0.066907	5068.194756	-5.978	2.41e-09	***
trial_exp195	-0.482369	0.066907	5068.194759	-7.210	6.44e-13	***
trial_exp196	-0.540238	0.067315	5135.253312	-8.026	1.24e-15	***
trial_exp197	-0.517987	0.067297	5113.748283	-7.697	1.66e-14	***
trial_exp198	-0.543591	0.066907	5068.194755	-8.125	5.59e-16	***
trial_exp199	-0.531164	0.067315	5135.253309	-7.891	3.65e-15	***

trial_exp200	-0.567446	0.066907	5068.194754	-8.481	< 2e-16	***
trial_exp201	-0.677938	0.066907	5068.194760	-10.133	< 2e-16	***
trial_exp202	-0.616380	0.066907	5068.194752	-9.213	< 2e-16	***
trial_exp203	-0.635910	0.067721	5199.017441	-9.390	< 2e-16	***
trial_exp204	-0.622893	0.067329	5148.928234	-9.251	< 2e-16	***
trial_exp205	-0.645245	0.067304	5117.835078	-9.587	< 2e-16	***
trial_exp206	-0.746119	0.066907	5068.194756	-11.152	< 2e-16	***
trial_exp207	-0.570718	0.067297	5113.748277	-8.481	< 2e-16	***
trial_exp208	-0.610854	0.067302	5097.796756	-9.076	< 2e-16	***
trial_exp209	-0.326661	0.067785	5667.487957	-4.819	1.48e-06	***
trial_exp210	-0.318361	0.066846	5536.694078	-4.763	1.96e-06	***
trial_exp211	-0.477817	0.067275	5595.659017	-7.102	1.38e-12	***
trial_exp212	-0.562679	0.066846	5536.694076	-8.418	< 2e-16	***
trial_exp213	-0.576167	0.066846	5536.694075	-8.619	< 2e-16	***
trial_exp214	-0.656196	0.067306	5600.860966	-9.749	< 2e-16	***
trial_exp215	-0.579487	0.067306	5600.860964	-8.610	< 2e-16	***
trial_exp216	-0.643135	0.067784	5662.866576	-9.488	< 2e-16	***
trial_exp217	-0.630872	0.067300	5600.366234	-9.374	< 2e-16	***
trial_exp218	-0.613955	0.066846	5536.694075	-9.185	< 2e-16	***
trial_exp219	-0.670461	0.066846	5536.694075	-10.030	< 2e-16	***
trial_exp220	-0.671067	0.066846	5536.694070	-10.039	< 2e-16	***
trial_exp221	-0.680244	0.067306	5600.860959	-10.107	< 2e-16	***
trial_exp222	-0.698949	0.067319	5603.751444	-10.383	< 2e-16	***
trial_exp223	-0.680520	0.067792	5664.820500	-10.038	< 2e-16	***
trial_exp224	-0.730843	0.067306	5600.860971	-10.859	< 2e-16	***
trial_exp225	-0.320184	0.068219	4449.947040	-4.693	2.77e-06	***
trial_exp226	-0.358777	0.067386	4303.618724	-5.324	1.07e-07	***
trial_exp227	-0.485860	0.067796	4366.946656	-7.166	8.99e-13	***
trial_exp228	-0.548992	0.068189	4380.901546	-8.051	1.05e-15	***
trial_exp229	-0.597649	0.067779	4346.679029	-8.818	< 2e-16	***
trial_exp230	-0.654599	0.067775	4336.275215	-9.658	< 2e-16	***
trial_exp231	-0.544764	0.068184	4419.349311	-7.990	1.71e-15	***
trial_exp232	-0.649325	0.067386	4303.618733	-9.636	< 2e-16	***
trial_exp233	-0.681254	0.067386	4303.618730	-10.110	< 2e-16	***
trial_exp234	-0.611865	0.068242	4448.395771	-8.966	< 2e-16	***
trial_exp235	-0.712033	0.067775	4336.275223	-10.506	< 2e-16	***
trial_exp236	-0.647615	0.067787	4353.481971	-9.554	< 2e-16	***
trial_exp237	-0.694787	0.067774	4358.378512	-10.252	< 2e-16	***
trial_exp238	-0.687892	0.067780	4353.531001	-10.149	< 2e-16	***
trial_exp239	-0.690483	0.067779	4346.679022	-10.187	< 2e-16	***
trial_exp240	-0.721807	0.068197	4387.880706	-10.584	< 2e-16	***
trial_exp241	-0.426195	0.068185	5193.205826	-6.251	4.42e-10	***
trial_exp242	-0.398475	0.069161	5340.923157	-5.762	8.80e-09	***
trial_exp243	-0.598973	0.068649	5252.578399	-8.725	< 2e-16	***
trial_exp244	-0.555741	0.068185	5193.205835	-8.150	4.50e-16	***
trial_exp245	-0.515135	0.068185	5193.205830	-7.555	4.92e-14	***
trial_exp246	-0.570689	0.068658	5251.676838	-8.312	< 2e-16	***
trial_exp247	-0.650493	0.069168	5331.010855	-9.405	< 2e-16	***
trial_exp248	-0.525056	0.068185	5193.205835	-7.700	1.61e-14	***
trial_exp249	-0.586253	0.068647	5250.744228	-8.540	< 2e-16	***
trial_exp250	-0.533220	0.068658	5251.676833	-7.766	9.65e-15	***
trial_exp251	-0.573209	0.068649	5252.578395	-8.350	< 2e-16	***
trial_exp252	-0.556990	0.069176	5334.745193	-8.052	9.98e-16	***
trial_exp253	-0.675771	0.068185	5193.205831	-9.911	< 2e-16	***
trial_exp254	-0.646081	0.068647	5250.744235	-9.412	< 2e-16	***
trial_exp255	-0.582376	0.069146	5312.041143	-8.422	< 2e-16	***
trial_exp256	-0.654944	0.069680	5400.239862	-9.399	< 2e-16	***
lightlevel2:snr-6	-0.060049	0.022028	8526.236270	-2.726	0.006422	**
lightlevel3:snr-6	-0.092791	0.022153	8540.473600	-4.189	2.83e-05	***
lightlevel4:snr-6	-0.085591	0.022256	8513.666292	-3.846	0.000121	***

```
lightlevel2:snr0    -0.009621    0.021805 8503.770265  -0.441 0.659047
lightlevel3:snr0     0.014265    0.022058 8535.203979   0.647 0.517845
lightlevel4:snr0    -0.009959    0.022126 8537.677576  -0.450 0.652632
lightlevel2:snr3     0.047435    0.022225 8546.166467   2.134 0.032844 *
lightlevel3:snr3     0.049718    0.022424 8523.205429   2.217 0.026638 *
lightlevel4:snr3     0.043993    0.022259 8552.388596   1.976 0.048136 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```