Cognition Seen Through the Eyes of Hearing Aid Users

Working Memory Resource Allocation
for Speech Perception and Recall

Andreea Micula

Cognition Seen Through the Eyes of Hearing Aid Users:

Working Memory Resource Allocation for Speech Perception and Recall

Andreea Micula

Linköping Studies in Arts and Sciences No. 826
Studies from the Swedish Institute for Disability Research No. 103
Linköping University,
Department of Behavioural Sciences and Learning
Linköping 2022

Linköping Studies in Arts and Sciences • No. 826

Studies from the Swedish Institute for Disability Research • No. 103

At the Faculty of Arts and Sciences at Linköping University, research and doctoral studies are carried out within broad problem areas. Research is organized in interdisciplinary research environments and doctoral studies mainly in graduate schools. Jointly, they publish the series Linköping Studies in Arts and Sciences. This thesis comes from the Swedish Institute for Disability Research at the Department of Behavioural Sciences and Learning.

Distributed by:

Department of Behavioural Sciences and Learning.

Linköping University

SE-581 83 Linköping

Andreea Micula

Cognition Through the Eyes of Hearing Aid Users:

Working Memory Resource Allocation for Speech Perception and Recall

Edition 1:1

ISBN 978-91-7929-137-2 (print)

ISBN 978-91-7929-138-9 (PDF)

ISSN 0282-9800

ISSN 1650-1128

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

<https://creativecommons.org/licenses/by-nc/4.0/>

©Andreea Micula

Department of Behavioural Sciences and Learning, 2022

Printed by: LiU-tryck, Linköping 2022

“Do the best you can until you know better.
Then when you know better, do better.”

Maya Angelou

ABSTRACT

This thesis investigates how hearing aid users allocate working (WM) memory resources under various task demands when listening to and storing speech in memory for later recall. This was done by combining an auditory recall task, the Sentence-final Word Identification and Recall (SWIR) test, with pupillometry. Different pupillary responses were used to obtain insights into momentary WM resource allocation and overall WM resource allocation over time. The task demands were manipulated by varying hearing aid noise reduction settings, as well as by varying the task difficulty of the SWIR test and the task difficulty predictability.

The findings from the first two studies showed that recall performance in competing speech was better, and baseline pupillary responses were higher when noise reduction was activated compared to when it was not. This indicates that attenuating background noise frees up WM resources to be used for storing speech in memory rather than speech processing. While unpredictable task difficulty elicited higher baseline pupillary responses than predictable task difficulty, it did not have any effect on recall performance. This finding suggests that task difficulty predictability does not affect WM resource allocation. Instead, unpredictable task difficulty may lead to increased alertness in anticipation of the end of the SWIR test list. The findings of the third study showed that increased transient task-evoked pupillary responses, which reflect the momentary intensity of attention during encoding, were associated with a higher likelihood of subsequent recall. Moreover, higher WM capacity was also linked to higher likelihood of subsequent recall, presumably due to the ability to allocate more attentional resources during encoding. Lastly, the findings from the fourth study indicated that the combination of the SWIR test and pupillometry is suitable for capturing WM resource allocation. Although arousal decreased over time, recall performance remained stable, suggesting that participants did not reach the point of disengagement.

Overall, a novel learning from this thesis is that increased pupillary responses may be a marker of “successful effort” when additional WM resources are allocated to achieve a better recall performance in the SWIR test. Furthermore, this thesis gives insights into which factors affect WM resource allocation and how to reduce the amount of processing resources required to understand speech, which may contribute to optimizing auditory rehabilitation in the future.

LIST OF PAPERS

Paper 1

Micula, A., Ng, E. H. N., El-Azm, F., & Rönnberg, J. (2020). The effects of noise reduction, background noise and task difficulty on recall. *International Journal of Audiology*, 59(10), 792–800. <https://doi.org/10.1080/14992027.2020.1771441>

Paper 2

Micula, A., Rönnberg, J., Fiedler, L., Wendt, D., Jørgensen, M. C., Larsen, D. K., & Ng, E. H. N. (2021). The Effects of Task Difficulty Predictability and Noise Reduction on Recall Performance and Pupil Dilation Responses. *Ear & Hearing*, 42(6), 1668–1679. <https://doi.org/10.1097/aud.0000000000001053>

Paper 3

Micula, A., Rönnberg, J., Książek, P., Nielsen, R. M., Wendt, D., Fiedler, L., Ng, E. H. N. A glimpse of memory through the eyes: Pupillary responses and working memory capacity reflect the likelihood of subsequent memory recall in an auditory free recall test. Submitted manuscript.

Paper 4

Micula, A., Rönnberg, J., Zhang, Y., Ng, E. H. N. A decrease in physiological arousal accompanied by stable behavioral performance reflects task habituation. Manuscript in preparation.

TABLE OF CONTENTS

ABBREVIATIONS	1
INTRODUCTION	2
BACKGROUND	4
Hearing impairment and hearing aid signal processing	4
Beyond the audiogram	5
Theoretical framework: working memory resource allocation in a communicative context	5
Working memory	6
Measures of working memory resource allocation	11
The Sentence-final Word Identification and Recall (SWIR) test	11
Pupillometry	12
Combining measures of working memory resource allocation	13
Targeting different processes via different pupillary responses	14
Task demand manipulations	16
THESIS AIMS	17
EMPIRICAL STUDIES	19
General methods	19
Participants	19
Administered measures	19
Summary of the papers	23
Paper 1	23
Paper 2	25
Paper 3	26
Paper 4	27

GENERAL DISCUSSION 29

 Main findings and conclusions 29

 The modified SWIR test combined with pupillometry 30

 Effects of noise and noise reduction 30

 Effects of task difficulty predictability 32

 An alternative method of investigating working memory resource allocation 34

 Taking advantage of both task-evoked and baseline pupillary responses 34

 Behavioral data disentangles the ambiguity of pupillary responses 38

 Theoretical implications 38

 Methodological considerations 41

 Design of studies combining the SWIR test and pupillometry 41

 Background noise 41

FUTURE DIRECTIONS 42

ACKNOWLEDGEMENTS 44

REFERENCES 46

ABBREVIATIONS

SWIR test Sentence-final Word Identification and Recall test

WM Working Memory

RS test Reading Span test

PPD Peak Pupil Dilation

4T babble Four-Talker babble

SNR Signal-to-Noise Ratio

INTRODUCTION

The most common treatment for hearing impairment is providing amplification via hearing aids. Advanced hearing aid signal processing has been designed to compensate for the consequences of auditory impairment by improving audibility, speech intelligibility, sound quality and listening comfort, especially in difficult listening environments (Anderson et al., 2018). Despite the advancements in hearing aid signal processing technologies, individuals with hearing impairment still report that listening is effortful, even when most of the speech can be understood. In clinical audiological praxis, assessment mainly consists of measuring thresholds for pure-tone audiometry or speech audiometry, which may not be sensitive to evaluating the complaint of effortful listening (Pichora-Fuller et al., 2016).

In recent years, researchers have worked towards developing methods that could assess the effects of hearing aid signal processing on working memory (WM) resource allocation. Ng et al. (2013, 2015) have designed an auditory recall test, the Sentence-final Word Identification and Recall (SWIR) test, to measure the effects of noise reduction in hearing aids on memory for highly intelligible speech heard in background noise. The task of the SWIR test is to listen to lists of sentences, to repeat the last word immediately after each sentence, and then to recall as many of the repeated words as possible at the end of the list. The first studies using the SWIR test showed that recall performance was better when noise reduction was activated compared to when it was not. This demonstrates that noise reduction frees up WM resources otherwise needed to process speech, allowing them to be allocated to storage of speech. However, Ng et al. also speculated that if the task difficulty exceeds individual WM capacity, the effect of noise reduction on recall performance cannot be captured. This raises the questions whether the task difficulty of the SWIR test should be adapted to individual WM capacity. Furthermore, the use of pupillometry as a measure of WM resource allocation during speech recognition tasks has become increasingly widespread. Several studies have shown that task-evoked pupillary responses are smaller when noise reduction is on compared to off, reflecting a decrease in allocation of WM resources for speech processing, even when speech understanding is nearly 100% (Ohlenforst et al., 2018; Wendt et al., 2017). The advantage of pupillometry is that it can provide temporal information, that is, it allows tracking WM resource allocation over time (Unsworth & Robison, 2014).

This thesis investigates how hearing aid users allocate WM resources under various task demands when speech not only needs to be understood, but also recalled. This was done by combining the SWIR test with pupillometry and using different pupillary responses to index momentary WM resource allocation and overall WM resource allocation over time. The aims of the first study were to investigate whether varying the task difficulty of the SWIR test had an effect on the cognitive benefit of noise reduction in terms of WM resource allocation indexed by recall performance. Additionally, the effect of task difficulty predictability on WM resource allocation was explored. The aim of the second study was to investigate the effects of noise reduction and task difficulty predictability on WM resource allocation indexed via recall performance and pupillary responses measured over the course of a SWIR test list. The aim of the third study was to explore a different application of the combination of the SWIR test with pupillometry, by investigating whether the magnitude of the pupillary responses measured while listening to the SWIR test sentences and preparing to encode the target items into memory reflect the likelihood of subsequent memory recall. Furthermore, the effects of WM capacity and noise reduction on subsequent memory recall were investigated. The aim of the fourth study was to investigate whether the combination of the SWIR test and pupillometry is a suitable method for measuring WM resource allocation, by analyzing changes in arousal over time and their effect in task engagement.

BACKGROUND

Hearing impairment and hearing aid signal processing

The number of people living with hearing impairment has increased drastically over recent years. It is estimated that 1.34 billion people have mild to severe hearing loss, accounting for 18.1 % of the global population. Furthermore, hearing impairment has been found to be the fourth leading cause for years lived with disability, accounting for 5.8% of the sum of years lived with disability due to all other causes. The World Health Organization estimates that the yearly cost of unaddressed hearing loss is approximately 750-790 billion international dollars (Tucci et al., 2017). Hearing loss can be conductive, stemming from issues in the outer or middle ear, or sensorineural, stemming from problems in the inner ear, auditory nerve, or central auditory pathways. Conductive hearing loss mainly affects audibility and can often be resolved via medication or surgery. Sensorineural hearing loss, however, is irreversible and it negatively affects abilities such as frequency selectivity, loudness perception, pitch perception and spatial sound localization (Moore, 1996). Besides the perceptual consequences, untreated hearing impairment may also lead to poorer life quality in general, associated with social isolation and depression (Arlinger, 2003; Keidser et al., 2015). Additionally, hearing impairment is associated with a decrease in cognitive function, having been identified as the largest potentially modifiable risk factor for dementia occurring in midlife (Livingston et al., 2017, 2020)

The most common intervention for hearing impairment is the use of amplification through hearing aids (Chisolm et al., 2007). Modern hearing aid signal processing includes technologies such as noise reduction, directional microphones, and wide-dynamic range compression, aimed at reducing the effects of hearing impairment by improving audibility and listening comfort, especially in noisy environments. While hearing aid signal processing may provide a benefit in terms of audibility, speech intelligibility and listening comfort, it may also have a negative effect on cognitive processing. Research has shown that hearing aid signal processing may introduce artefacts and signal distortion, to which people with lower cognitive abilities may be susceptible (Arehart et al., 2013, 2015; Lunner et al., 2009). However, individual factors other than the audiogram are usually not taken into account when prescribing amplification or fine-tuning signal processing features (Anderson et al., 2018). This results in individuals with similar hearing impairment in terms of audiometric configuration obtaining varying benefit from similar

hearing aid fittings (Lunner et al., 2009). There are currently no suitable clinical measures for assessing the effect of hearing aid fitting on cognitive processing.

Beyond the audiogram

Assessment in clinical praxis is mainly based on pure-tone audiometry and speech audiometry. Pure-tone audiometry measures the threshold of the softest audible tone, while speech audiometry usually measures the percentage of correct repetition of standardized test material. Although these measures are important, they are not designed to address complaints about listening being effortful even when it is loud enough to be comprehended. Supplementary measures are needed in order for clinicians to understand how cognitive resources are expended during communication and how auditory intervention can support this (Anderson et al., 2018; Pichora-Fuller et al., 2016; Rönnberg et al., 2021).

The field of cognitive hearing science has developed as a result of increasing awareness about the link between hearing and cognition, especially in complex daily listening situations, as well as the link between hearing and changes in the brain (Arlinger et al., 2009). Studies have found that individuals with poorer hearing have lower brain matter volume in the auditory cortex (F. R. Lin et al., 2014; Peelle et al., 2011; Rudner et al., 2019), as well as different neural activation patterns in response to speech of varying grammatical complexity compared to individuals with better hearing (Peelle et al., 2011). Furthermore, even mild to moderate degrees of hearing loss have been shown to cause cortical re-organization, where areas of the auditory cortex are recruited for other functions (Cardon & Sharma, 2018). However, recent evidence suggests that this re-organization may be reversible if appropriate amplification is provided (Glick & Sharma, 2020). Additionally, age-related hearing loss has been linked to cognitive decline (Lin & Albert, 2014) and has been identified as the highest modifiable risk factor in relation to the development of dementia (Livingston et al., 2017, 2020). These findings highlight the importance of continuing to explore the relationship between cognition and hearing or hearing aid use, as well as the importance of integrating cognitive assessments into clinical praxis. Clinically feasible measures aimed at assessing the effect of hearing aid signal processing on cognitive processing have not yet been developed. However, several such measures are currently being explored within the research field of cognitive hearing science (Lunner et al., 2020; Rönnberg et al., 2013, 2021).

Theoretical framework: working memory resource allocation in a communicative context

There are several cognitive functions that are related to listening and understanding speech, such as WM, phonological processing abilities, information processing speed and

lexical decision-making speed amongst other executive functions (Lunner, 2003; Pichora-Fuller et al., 2016; Rönnberg et al., 2008, 2013, 2019). WM in particular has been extensively studied as a predictor of speech understanding in challenging listening situations (Arehart et al., 2013, 2015; Gordon-Salant & Cole, 2016; Lunner et al., 2009; Ng & Rönnberg, 2020; Rönnberg et al., 2008, 2013, 2019).

Working memory

WM is most commonly defined as a limited-capacity system that temporarily holds information needed to perform various cognitive tasks for simultaneous processing and storage (Baddeley, 2012; Oberauer, 2019; Wingfield, 2016). According to Baddeley's WM model, the system is split into various components: the central executive, the phonological loop, the visuo-spatial sketchpad, and the episodic buffer. The central executive is presumed to be an attentional system comprising several executive functions, such as decision-making, memory retrieval, attention maintenance and inference-making. The visuo-spatial sketchpad and phonological loop are domain specific, the former acting as a temporary store for visual and spatial input, while the latter as a temporary store where information is maintained by vocal or subvocal rehearsal. The episodic buffer allows input from different sources to be bound into episodes or chunks, which are held as multidimensional codes/representations. The episodic buffer not only links the WM components with each other, but also links WM to long-term memory.

WM plays an essential role for understanding speech, particularly in background noise, both for individuals with hearing impairment (Arehart et al., 2013, 2015; Davies-Venn & Souza, 2014; Lunner et al., 2009; Ng & Rönnberg, 2020; Souza et al., 2015) and individuals with normal hearing (Dryden et al., 2017; Gordon-Salant & Cole, 2016). The Ease of Language Understanding (ELU) model specifies the role of WM in supporting listening during adverse conditions (Rönnberg et al., 2008, 2013, 2019). Furthermore, the Framework for Understanding Effortful Listening (FUEL) (Pichora-Fuller et al., 2016) gives an overview of the factors that affect WM resource allocation during listening and which responses could be appropriate measures of listening effort.

Ease of Language Understanding (ELU) model

The ELU model describes how WM resources are allocated in a communicative context. According to the ELU model (figure 1), under optimal listening conditions, speech input is rapidly and automatically bound into phonological representations in a buffer termed RAMBPHO. If the information held in the RAMBPHO buffer matches the phonological representations stored in semantic long-term memory, access to the mental lexicon is successful with relatively little need for top-down processes. This automatic and rapid implicit processing continues as long as the RAMBPHO buffer output provides matching information. However, under sub-optimal listening conditions, a mismatch between the phonological representations held in the RAMBPHO buffer and those stored in semantic

long-term memory may occur. In order to remediate the mismatch, the explicit processing loop is activated, recruiting additional WM processing resources to retrieve information from semantic or episodic long-term memory. Such top-down processing may involve inference-making, inhibiting irrelevant information, switching attention and storing information.

Figure 1: Illustration of the Ease of Language Understanding model (Rönnberg et al., 2019)

Complex WM tasks, such as listening to speech in background noise, may require simultaneous processing and storage. Resources can be flexibly allocated to these functions within the capacity constraint of the WM system. Thus there is a trade-off between WM resources that can be allocated to processing and storage (Daneman & Carpenter, 1980). This is illustrated in figure 2. Consequently, the need for a higher degree of explicit processing as described in the ELU model, compromises the amount of WM resources that can be allocated for storage (Lunner et al., 2009).

Figure 2: Illustration of the trade-off between working memory resource allocation to processing (dark grey) and storage (light grey) and the effect on recall performance (modified from Lunner 2009).

Sub-optimal listening conditions include speech masked by background noise, but also degraded speech input due to the perceptual consequences of hearing impairment. Thus, individuals with hearing loss are more prone to experiencing mismatch and engage in explicit processing. Compensating for hearing loss as well as attenuating background noise via hearing aid signal processing may be essential for reducing the need for explicit processing in communicative situations.

Framework for Understanding Effortful Listening (FUEL)

The FUEL (Pichora-Fuller et al., 2016) is adapted from the Capacity Model of Attention by Daniel Kahneman (1973). The core assumptions of Kahneman's model are that the capacity that can be allocated to various activities is limited and that this limit is modulated by the level of arousal. The levels of arousal and effort allocated to a task will depend on the load imposed by the task demands. Figure 3 shows an illustration of the FUEL, which keeps the core components from Kahneman's model, but expands on them in relation to listening effort. Pichora-Fuller et al. (2016) define listening effort as "the deliberate allocation of mental resources to overcome obstacles in goal pursuit when carrying out a task".

The FUEL specifies that, while the evaluation of task demands determines the level of arousal and how available cognitive resources are allocated to a task, this in turn can be influenced by fatigue or displeasure with the task. This may explain why a task is abandoned even if the available capacity has not been exceeded. The allocation policy may be directly influenced by automatic attention, such as sudden or novel stimuli, or intentional attention, such as following a task instruction. The latter may be affected by factors including low motivation or displeasure. Furthermore, input-related demands, for instance background noise or hearing status, may indirectly influence the allocation policy via the cognitive capacity component. Lastly, responses have been grouped into two types, automatic arousal responses and attention-related responses, which may provide suitable measures of listening effort. In the present thesis, both types of responses are investigated. Recall performance in the SWIR test can be categorized under attention-related responses, while pupillary responses can be categorized under both attention-related and arousal responses.

Figure 3: Illustration of the Framework for Understanding Effortful Listening, based on Kahneman's (1973) Capacity Model of Attention (Pichora-Fuller et al., 2016).

An integral part of the FUEL is the concept of motivation and how motivation alongside task demands relate to effort, which is illustrated in the 3D plot shown in figure 4. The demands dimension would be determined by input-related demands, as well as task demands based on factors related to automatic and intentional attention. The motivation dimension would be affected by how arousal influences the listener’s evaluation of the value of spending the resources needed to meet the demands on capacity.

Figure 4: Three-dimensional illustration of how the effort required to perform a task may vary as a function of the demands on capacity and individual motivational arousal (Pichora-Fuller et al., 2016).

The arrows and time points on the 3D plot show an example of how effort may vary as a function of task demands and motivational arousal for a person attending a cocktail party. Initially, there is little change in effort as demands remain constant, but motivation increases as the person becomes more engaged in an interesting conversation (t0-t1). Next, the level of motivation remains relatively unchanged, but effort increases because of increasing task demands imposed by higher levels of background noise (t1-t3). Lastly, while the task demands remain constant, effort decreases due to an abrupt decrease in motivational arousal (t3-t4) as the topic of conversation becomes uninteresting, but the level of background noise remains high. This may lead to decreasing task engagement and ultimately the development of fatigue. It is essential to consider this, since it illustrates that the relationship between task demands and WM resource allocation is complex. Arousal and motivation may influence an individual’s choice to continue

engaging in a challenging listening task. As mentioned previously, the listener may disengage, even if the task does not exceed the individual WM capacity.

Measures of working memory resource allocation

The ELU model and FUEL describe how varying types and degrees of task demand may affect WM resource allocation. Measures of WM resource allocation could be essential to understand how challenges, such as background noise, degraded speech input and other external factors, may affect the allocation policy. Until relatively recently, the effect of hearing aids on WM resource allocation and cognitive performance had not been taken into account when designing signal processing algorithms (Edwards, 2016). However, the importance of investigating this effect was highlighted when studies using dual-task paradigms showed that, even when speech intelligibility was nearly 100% and signal processing could not further contribute to better speech intelligibility, it provided a cognitive benefit reflected in improved performance on the secondary task (Desjardins & Doherty, 2014; Sarampalis et al., 2009). Since these initial studies, various methods designed to investigate the effects of hearing aid signal processing on WM resource allocation have been used (Lunner et al., 2016, 2020; Neher et al., 2018; Ng et al., 2013, 2015; Ohlenforst et al., 2017, 2018; Wendt et al., 2017).

Measures of WM resource allocation could be valuable for guiding the development of signal processing technologies with the aim of not only improving audibility, but also reducing the amount of resources needed for speech processing (i.e. explicit processing). Moreover, such measures could provide insights into how individual factors, such as WM capacity, and hearing aid signal processing interact to modulate resource allocation or listening effort (Arehart et al., 2013, 2015; Souza et al., 2015). Consequently, this knowledge could be implemented when clinicians select and fine-tune various hearing aid features, resulting in better and individualized hearing aid fittings. The general view is that experiencing less listening effort may facilitate participation in social interactions for hearing aid users, thus decreasing the risk for the potential long-term consequences of hearing loss, such as cognitive decline.

The Sentence-final Word Identification and Recall (SWIR) test

The SWIR test is an auditory recall test developed by Ng et al. (2013) to measure the effect of noise reduction on memory for highly intelligible speech heard in background noise. The test material was composed of sentences from the Swedish version of the Hearing In Noise Test (HINT) (Hällgren et al., 2006) that end with a bi- or trisyllabic noun. The task of the SWIR test consists of listening to lists of sentences, repeating the last word immediately after each sentence and recalling as many of the repeated words as possible when the list is finished. Two scores are obtained: an identification score, based on the percentage of correctly repeated final words, and a final recall score, based on the

proportion of recalled words. Incorrectly repeated (i.e., misperceived) words are included in the latter score if they are correctly recalled.

The first studies using the SWIR test demonstrated that background noise has a negative effect on recall performance for individuals with hearing loss, especially four-talker (4T) babble compared to speech-shaped noise (SSN). Importantly, these studies have also shown that recall performance improved when noise reduction was on compared to off in 4T babble (Ng et al., 2013, 2015). These findings demonstrate that listening in 4T babble is cognitively more demanding than listening in SSN. Due to its lexical-semantic content, 4T babble competes with the target signal for attentional resources and the two signals are harder to segregate from each other (Mattys et al., 2009; Ng & Rönnberg, 2020; Sörqvist & Rönnberg, 2012). Attenuating background noise makes it easier to segregate the target speech, facilitating the function of the RAMBPHO buffer and lowering the degree of explicit processing. Although the ELU model does not make predictions about encoding speech into memory, it can be deduced that attenuating background noise frees up WM resources to be allocated for storage, which is reflected in the improved recall performance when noise reduction is on (light grey bar in figure 2).

Interestingly, the improvement in recall performance was only seen for participants with high WM capacity in the first study (Ng et al., 2013). A SWIR test version with lists of eight sentences was used in this study and the speech recognition performance decreased to approximately a mean of 85% when noise reduction was off in 4T babble. In a subsequent study (Ng et al., 2015), the list length was shortened to seven sentences and the SNR level was individualized using a procedure to estimate 95% speech intelligibility. These findings showed that recall performance was better with noise reduction on compared to off for both participants with high and low WM capacity. This was likely due to a lower task difficulty in terms of list length, which did not exceed the capacity of people with lower WM, as well as the more favorable SNR levels, which decreased the need for WM resources required for speech processing. Thus, the sensitivity of the SWIR test to capturing resource allocation may depend on individual differences in WM capacity.

Pupillometry

The use of pupillometry as an objective measure of WM resource allocation during tasks involving auditory stimuli has become wide-spread in the field of cognitive hearing science (Zekveld et al., 2018). Pupillary responses are an index of the noradrenergic function of the locus coeruleus, which is associated with physiological responses to arousal, stress, anxiety, emotion, engagement, attention and other cognitive functions. Pupillometry is considered an objective measure, since pupillary responses occur spontaneously and involuntarily (Aston-Jones & Cohen, 2005; Beatty, 1982; Laeng et al., 2012; Mathôt, 2018; Pichora-Fuller et al., 2016; Winn et al., 2018). Furthermore,

pupillometry is a time-series measurement, offering the possibility to track how resources are allocated at varying time points during a task (Winn et al., 2018). Pupillary responses have therefore been defined as an online measure of “attentional allocation” (Unsworth & Miller, 2020).

Two types of pupillary responses are typically measured in research studies investigating cognitive processes: the baseline pupillary responses, which reflect the overall brain state and its effects on arousal or attention over relatively long time intervals, and the task-evoked pupillary responses, which are transient responses triggered by specific task stimuli (Joshi & Gold, 2020). It is the magnitude of the task-evoked pupillary responses that is commonly used to quantify the allocation of processing resources to a task or listening effort. Pupillary dilations in response to cognitive tasks are relatively small, typically between 0.1 mm and 0.5 mm, compared to dilations in response to changes in luminance, which can be up to 4 mm. The task-evoked pupillary responses are usually calculated as either the maximum or mean dilation relative to the baseline prior to the stimulus (Książek et al., 2021; Winn et al., 2018; Zekveld et al., 2018). The baseline pupillary response has been used to a lesser degree in studies investigating the allocation of WM resources to auditory tasks. It is usually measured immediately prior to a stimulus (Joshi & Gold, 2020; Winn et al., 2018; Zekveld et al., 2018).

Studies combining speech recognition tests with pupillometry have demonstrated that task-evoked pupillary responses are lower when noise reduction is activated compared to when it is not, even when speech intelligibility close to 100% and cannot be improved (Ohlenforst et al., 2018; Wendt et al., 2017). The reduced pupillary responses in this case have been interpreted as an index of reduced listening effort. As mentioned previously, noise reduction facilitates the function of the RAMBPHO and decreases the need for explicit processing when listening to speech (dark grey bar in figure 2). The findings from the pupillometry studies by Wendt et al. (2017) and Ohlenforst et al. (2018) are in agreement with the SWIR test outcomes when it comes to the effect of noise reduction on WM resource allocation.

Combining measures of working memory resource allocation

Francis and Love (2020) argue that there is extensive overlap between the physiological systems linked to the amount of processing resources a listener expends, such as pupillary responses or heart rate and blood pressure, and the physiological systems linked to affective responses, that is, the emotional responses to the amount of effort that is being exerted (e.g., arousal related to motivation to continue exerting effort). Interestingly, measures of “listening effort”, as they are sometimes defined (Pichora-Fuller et al., 2016), do not always correlate with each other. The lack of correlation indicates that various

measures may tap into different overlapping dimensions of listening effort, such as WM resource allocation, attention, stress, arousal, and individual evaluation of task demands. This is consistent with the multidimensional outputs described in the FUEL (Alhanbali et al., 2019; Pichora-Fuller et al., 2016). The FUEL suggests that an interplay between these factors may modulate listening effort. For instance, arousal and individual task demands evaluation may have an effect on how WM resources are allocated. Combining different measures may give more in-depth insights into the multidimensional mechanisms of WM resource allocation and the phenomenon of listening effort.

The advantage of the SWIR test is that it targets more complex cognitive processes compared to the often-used speech recognition tests, thus being more representative of the types of cognitive processes that are involved in communicative contexts. Although both the SWIR test and pupillometry have been used as measures of WM resource allocation or listening effort, they may not necessarily tap into the same underlying phenomena. The SWIR test was designed to test the hypothesis presented in figure 2, that is, the trade-off between allocation of WM resources to processing and storage of speech. In three of the papers included in this thesis the SWIR test was combined with pupillometry in order to track WM resource allocation at different time points while listening to and encoding speech for later recall. However, pupillary responses may also capture some of the factors that affect WM resource allocation, such as arousal. While WM resource allocation is the main focus of this thesis, it is considered to be only one of the aspects that affects listening effort as defined by the authors of the FUEL. As illustrated in the FUEL, cognitive-behavioral responses, such as the recall performance in the SWIR test, are categorized under attention-related responses. However, changes in pupillary responses are categorized under both attention-related and automatic arousal responses, thereby presumably capturing other processes besides WM resource allocation (Pichora-Fuller et al., 2016). Thus, combining the SWIR test with pupillometry offers the possibility to observe the effects of various task demand manipulations on the behavioral and physiological responses. This combination may contribute to a more thorough understanding of the influence of such factors on WM resource allocation.

Targeting different processes via different pupillary responses

Pupillary responses can be calculated on different time scales and using different methods (see section on Pupillometry). In this thesis, different pupillary responses are used to target different processes related to WM resource allocation when listening to and maintaining speech in memory for subsequent recall. Zekveld et al. (2018) suggest that in a concurrent speech recognition and memory task, the pupillary responses are more sensitive to the memory load rather than perceptual factors. Thus, even though pupillary responses have been shown to reflect the amount of WM resources allocated to speech

processing in speech recognition tasks (Ohlenforst et al., 2018; Wendt et al., 2017), they are expected to index the amount of WM resources allocated for storage of the target words in the SWIR test.

Figure 5 illustrates the different pupillary responses and the time intervals of the SWIR test in which they were calculated. The transient task-evoked pupillary responses are considered to reflect momentary intensity of attention when listening to a SWIR test sentence and preparing to encode a particular target word (Miller et al., 2019). Thus, it is hypothesized that target words of sentences that receive a higher degree of attentional processing, indexed by larger task-evoked pupillary responses, are more likely to be subsequently recalled (Bergt et al., 2018; Kucewicz et al., 2018; Miller et al., 2019; Papesch et al., 2012). The task-evoked pupillary responses were calculated as the maximum dilation in response to a SWIR test sentence. This is referred to as the peak pupil dilation (PPD). The sentence baseline pupillary responses, which were measured immediately prior to each sentence onset, are considered to reflect WM resource allocation over the course of a SWIR test list (Bönitz et al., 2021). Increased sentence baseline pupillary responses over the course of a list are expected to be accompanied by better recall performance, reflecting the amount of WM resources allocated for storage of speech over time.

Figure 5: Illustration of the different types of pupillary responses used in the current thesis and the SWIR test time intervals during which they were measured.

Although recall performance in the SWIR test might be limited by an individual's WM capacity, this limitation is not imposed on pupillary responses. The pupil can dilate up to 4 mm in response to light, but the maximum dilation in response to cognitive tasks is typically up to 0.5 mm (Winn et al., 2018). Taking this into consideration, pupillary responses do not reach their absolute maximum, even if recall performance does. Pupillary responses are hypothesized to reflect WM resource allocation to storage of

speech, thereby being linked to recall performance. Lastly, the pupil baseline, which is calculated as the mean of all sentence baseline pupillary responses in a list, is used to investigate potential changes in arousal over the course of the entire SWIR test session. This is of interest as arousal may modulate task engagement, which in turn may affect the listeners' willingness to continue allocating the necessary WM resources to meet task demands (Alhanbali et al., 2020; Ayasse & Wingfield, 2020; Pichora-Fuller et al., 2016).

Task demand manipulations

Task demands are manipulated in various ways in the current thesis in order to obtain a better understanding of which factors may affect WM resource allocation when listening to and storing speech in memory.

1. Noise reduction: The test was conducted with noise reduction activated and deactivated, as well as in different types of background noise. These types of demand manipulations have been previously used when investigating listening effort and have been shown to affect the degree of explicit processing and thereby WM resource allocation (Bönitz et al., 2021; Lunner et al., 2016; Ng et al., 2013, 2015; Ohlenforst et al., 2018; Wendt et al., 2017).
2. Task difficulty: Ng et al. speculated that if task demands exceed an individual's cognitive capacity, this may interfere with capturing the effect of noise reduction on WM resource allocation. Therefore, the task demands are also manipulated by varying the task difficulty of the SWIR test, that is the number of sentences per list. Although it is expected that more resources are free to be used for storage of target items when noise reduction is on, which would result in increased recall performance, this may not be the case when the task difficulty surpasses the individual capacity limit.
3. Task difficulty predictability: This manipulation refers to knowing or not knowing how many sentences will be presented in an upcoming list. Since this is a novel manipulation for the SWIR test, it is not known whether it has an impact on WM resource allocation or on other factors, for instance arousal as highlighted in the FUEL. It is hypothesized that task difficulty predictability will have an effect on both recall performance and pupillary responses if it is linked to WM resource allocation. Conversely, if an effect is only found on pupillary responses, or on neither pupillary responses nor recall performance, task difficulty predictability is speculated to be linked to other processes, such as arousal or task engagement. Consequently, it is important to investigate whether there is a difference between predictable and unpredictable task difficulty, and which processes the potential difference may be related to. This may guide how studies using this methodology are designed.

THESIS AIMS

This thesis investigated how hearing aid users allocate WM resources when listening to and encoding speech for subsequent recall under various task demands. A behavioral auditory recall test, the Sentence-Final Word Identification and Recall (SWIR) test, and a physiological measure, pupillometry, were combined in order to reach a more in-depth understanding of WM resource allocation. Both the momentary WM resource allocation and overall WM resource allocation over time were examined, indexed by different pupillary responses. Specifically, the aims of the thesis were to investigate 1) whether the SWIR test of varying task difficulty was sensitive to capturing the effect of noise reduction on WM resource allocation; 2) the effects of task difficulty predictability on WM resource allocation; 3) whether the magnitude of pupillary responses measured during encoding were linked to subsequent memory recall; 4) whether WM capacity and noise reduction have an effect on subsequent memory recall; and 5) whether the combination of the SWIR test and pupillometry is an appropriate method to assess WM resources allocation by tracking potential changes in arousal and their effect on task engagement over time.

The first two aims were investigated in Paper 1 using the SWIR test. Based on findings by Ng et al. (2013, 2015), it was expected that recall performance would be better when noise reduction was activated, reflecting the ability to allocate more WM resources to storing speech in memory. The aims of Paper 2 were similar to those of Paper 1, but the SWIR test was supplemented with pupillometry. Based on the findings from Paper 1, which showed that task difficulty predictability did not have an effect on recall performance, the prediction in Paper 2 was that task difficulty predictability would not have an effect on WM resource allocation. Any changes in pupillary responses were hypothesized to be a reflection of the effect of task difficulty predictability on other cognitive processes. Recall performance was expected to be better with noise reduction on compared to off. Moreover, the task-evoked PPD was expected to be lower (Ohlenforst et al., 2018; Wendt et al., 2017), and sentence baseline pupillary responses higher (Bönitz et al., 2021; Zhang et al., 2021) with noise reduction on.

The third and fourth aims were investigated in Paper 3. Previous studies have shown that the magnitude of the task-evoked pupillary responses measured during item encoding is an index of the likelihood of subsequent memory recall (Bergt et al., 2018; Kucewicz et al., 2018; Miller et al., 2019; Papesh et al., 2012). However, this has not been investigated in individuals with hearing loss and in conditions with background noise. It was

hypothesized that the larger the task-evoked PPD during encoding, the higher the likelihood of subsequently recalling a word. Moreover, it was expected that the higher the WM capacity, the higher the likelihood of subsequent memory recall. Lastly, consistent with Paper 2, it was expected that noise reduction would also be associated with higher likelihood of subsequent memory recall.

The last aim was investigated in Paper 4. Since the SWIR test is lengthy and cognitively demanding, one concern is that listeners may gradually disengage from the task. This could interfere with capturing the effect of various manipulations on WM resource allocation. Consequently, the combination of SWIR test and pupillometry is only considered to be a suitable method if participants are able to maintain the necessary task engagement level to meet task demands throughout the test session. This is particularly relevant in studies including individuals with hearing loss, who may be more susceptible to increased listening effort and are thereby more likely to disengage (Pichora-Fuller et al., 2016). Based on previous studies, a decrease in pupil baseline over the course of the SWIR test was expected. It was hypothesized that if recall performance also decreased over the course of the SWIR test, the pupil baseline would reflect decreasing task engagement (Alhanbali et al., 2020; Hopstaken et al., 2015). However, if recall performance did not change or improved over the course of the SWIR test, it was hypothesized that the pupil baseline would reflect task habituation (Ayasse & Wingfield, 2020).

EMPIRICAL STUDIES

General methods

Participants

The participants who took part in the study reported in Paper 1 were native Swedish speakers recruited from the audiology clinic at the University Hospital of Linköping (Sweden). The participants who took part in the study reported in Paper 2 were native Danish speakers recruited from the database at Oticon A/S (Denmark). The studies reported in Papers 3 and 4 share the same pool of participants, who were native Danish speakers recruited from Eriksholm Research Centre (Denmark). All participants had been using hearing aids for approximately one year at the time of testing. Table 1 gives an overview over the participants in each paper. None of the participants reported otologic or psychological disorders. Moreover, the participants included in the studies reported in Papers 2-4 had normal or corrected to normal vision and had no history of eye diseases or surgeries. All participants signed written informed consent forms. The study presented in Paper 1 was approved by the Regional Ethics Review Board in Linköping and the studies presented in Papers 2-4 have been exempted from ethical application by the Science Ethics Committee for the Capital Region of Denmark.

Administered measures

Table 1 gives an overview over the measures included in each paper.

Table 1: Overview of participants and measures reported in each paper.

Paper	Participants			Measures			
	N	Age (years)	Pure-tone average at 0.5, 1, 2 and 4 kHz dB HL	Sentence-final Word Identification and Recall (SWIR) test	Pupillometry	Montreal Cognitive Assessment	Reading Span test
1	32 (F: 13, M:19)	M = 65, SD = 6.4, range: 46 – 74	Moderate symmetrical sensorineural hearing loss M = 49.8, SD = 5.7	<ul style="list-style-type: none">• Language: Swedish• Lists of 3, 5, 7 and 9 sentences (total = 48)• Estimated SNR for 95% word recognition in 4T babble and noise reduction off• Noise reduction: off/on• Background noise: 4T babble/SSN• Predictable vs. unpredictable task difficulty	---	Swedish, Version 7.0	Swedish 24-item version
2	24 (F: 11, M: 13)	M = 65, SD = 8.0, range: 43 – 75	Moderate to moderately severe symmetrical sensorineural hearing loss M = 53.2, SD = 7.7	<ul style="list-style-type: none">• Language: Danish• Lists of 3, 5 and 7 sentences (total = 42)• Estimated SNR for 95% word recognition with noise reduction off• Noise reduction: off/on• Background noise: 4T babble• Predictable vs. unpredictable task difficulty	<ul style="list-style-type: none">• Peak pupil dilation: maximum dilation in a 4s interval from sentence onset• Sentence baseline dilation: mean dilation in a 1s interval prior to sentence onset	Danish, Version 8.1	Danish 54-item version
3	21 (F: 8, M: 13)	M = 58, SD = 11.3, range: 22 – 73	Mild to moderately severe symmetrical sensorineural hearing loss M = 49.3, SD = 11.4	<ul style="list-style-type: none">• Language: Danish• Lists of 7 sentences (total = 28)• Estimated SNR for 95% word recognition with noise reduction off• Background noise: 16T babble	<ul style="list-style-type: none">• Peak pupil dilation: same as Paper 2• Pupil baseline: mean of all sentence baseline pupillary responses in a list	---	Danish 24-item version
4							

Modified versions of the SWIR test

Various versions of the SWIR test, with different task demand manipulations, were included in all the papers. The SWIR test was administered at an SNR level estimated to yield 95% correct word recognition in multi-talker babble in the respective language in which each study was conducted. For the purpose of estimating this level, the HINT was administered. For the study reported in Paper 1, a Swedish version of the HINT (Hällgren et al., 2006) was used. For the studies reported in Papers 2-4, a Danish version of the HINT (Nielsen & Dau, 2011) was used. For all papers, a modified procedure was followed in order to estimate the 80% speech intelligibility in the HINT, which roughly corresponds to 95% correct word recognition in the SWIR test. In order to obtain the 80% speech intelligibility level in the HINT, the SNR was decreased by 0.8 dB after correct repetition and increased by 3.2 dB after incorrect repetition. For the first five sentences the step size was twice as large. The SNR obtained from the HINT was used as a starting point for the SWIR test training. Four training lists were administered in order to adjust the SNR if needed. If six or seven (86-100%) of the last words in a training list were repeated correctly, the SNR obtained from the HINT was left unchanged after the respective list. If four or five of the last words in a list were repeated correctly, the SNR was increased by 1 dB after the respective training list. If fewer words were repeated correctly, the SNR was increased by 2 dB. This procedure for estimating 95% correct word repetition in the SWIR test has been used in several studies (Bönitz et al., 2021; Lunner et al., 2016).

The version of the SWIR test reported in Paper 1 consisted of the same test material as the Swedish SWIR test version used by Ng et al. (2013, 2015). However, the task difficulty was varied by restructuring the material to build lists containing three, five, seven and nine sentences. Since Ng et al. (2015) did not find any effect of last word repetition, the participants were not instructed to repeat the last word. Moreover, task difficulty predictability was manipulated by having a group of participants who were informed about how many sentences the upcoming list would contain (predictable task difficulty), and a group who was not informed about the list length (unpredictable task difficulty).

In Paper 2, a Danish version of the SWIR test was used, which was composed of Danish HINT sentences, arranged in lists of three, five and seven sentences. The study was conducted over two test visits with the same participants. During one visit the task difficulty was predictable and during the other visit the task difficulty was unpredictable. In the study reported in Papers 3 and 4 a Danish version of the SWIR test was administered as well. However, the task difficulty was not manipulated, so all the lists contained seven sentences, similarly to the version used in the study by Ng et al. (2015). Since the SWIR test was accompanied by pupillometry in Papers 2-4, the participants were instructed to repeat the target word. This provides a time interval between sentences

during which the pupillary responses can return to the baseline size after the dilation in response to the stimulus.

Hearing aid signal processing and background noise

In Paper 1, only amplification based on the Voice Aligned Compression (VAC+) rationale (Buus & Florentine, 2001) was provided when noise reduction was off using a hearing aid simulator. Compensation for hearing loss was provided based on the individual audiometric thresholds in the frequency range of 250 to 8000 Hz. The VAC+ rationale is quasilinear and has lower compression kneepoints to provide more compression at low input levels and less compression at high input levels. The noise reduction system used a smoothed SNR estimate, which had been mapped to a continuous gain for each time-frequency unit (Boldt et al., 2008). Attenuation was applied in time-frequency units with low SNR. The noise reduction was pre-processed. The SWIR test was conducted in Swedish 4T babble and SSN. The 4T babble consisted of recordings of two female and two male native Swedish speakers reading different newspaper paragraphs. The SSN was stationary speech-shaped noise. Both 4T babble and SSN had long-term average spectra resembling the spectrum of the HINT sentences.

In Paper 2, all test participants were fitted with Oticon OpnS1™ (mini-Receiver-in-the-ear) and amplification based on the VAC+ rationale was provided. A microphone setting that simulates the acoustic effect of the pinna was used when noise reduction was off. When noise reduction was on, a fast-acting version of a minimum-variance distortionless response beam-former was applied, which uses spatial filtering to attenuate background noise coming from behind the listener (Kjems & Jensen, 2012), and a single-channel Wiener postfilter was applied to further attenuate background noise (Jensen & Pedersen, 2015). The SWIR test was conducted in Danish 4T babble, which is similar to the Swedish 4T babble.

The data used for Papers 3 and 4 was part of a study conducted at Eriksholm Research Centre. The SWIR test was conducted in Danish 16-talker (16T) babble, which is similar to the 4T babble, but the same four talkers are presented from four different loudspeakers rather than one talker in each loudspeaker. This background noise was chosen in order to maximize the effect of the noise reduction system. Since 4T babble may be identified as speech, it is less likely to be attenuated by the noise reduction system compared to 16T babble, which more closely resembles SSN. The test was conducted using two models of hearing aids (Oticon OpnS1™ and Oticon More™ mini-Receiver-in-the-ear) with noise reduction activated and deactivated. Since it was not within the scope of Paper 3 to evaluate the differences between the two hearing aids, these were pooled and only the noise reduction was further explored. Hearing aid noise reduction was not analyzed in paper 4, as this was not within the scope of the study.

Reading Span (RS) test

The RS test is often used as a measure of WM capacity, as it involves simultaneous processing and storage of information (Daneman & Carpenter, 1980; Rönnberg et al., 1989). The task of the RS test is to read lists of three-word sentences, verbally indicate whether the sentence is sensible or absurd immediately after each sentence and, when the list is finished, recall either all the first or last words of each sentence. The RS test was scored as the percentage of correctly recalled words out of the total, regardless of the recall order (Petersen et al., 2016).

The Swedish version of the short RS test (Ng et al., 2013, 2015; Rönnberg et al., 1989) was reported in Paper 1. This version of the test is composed of three-, four- and five-sentence lists. The lists are visually presented on a laptop screen. Two lists of each length are presented in ascending order. The long 54-item version of the Danish RS test was administered in Paper 2. The RS test was not reported in Paper 2, as the outcome did not contribute with additional knowledge in comparison to Paper 1. A Danish version of the RS test (Petersen et al., 2016) was reported in Papers 3 and 4, which had been shortened to have the same structure as the version included in Paper 1.

Montreal Cognitive Assessment (MoCA)

The MoCA is a short screening tool for mild cognitive dysfunction which includes major cognitive domains, such as visuospatial/executive functions, naming, attention, language, abstraction, delayed recall and orientation (Nasreddine et al., 2005). Nasreddine et al. (2005) recommend using a cut-off score of 26 out of 30 points as an indication of normal cognitive function, although another study recommends a less conservative cut-off score of 23 (Carson et al., 2018). In Paper 1, the validated Swedish version of the MoCA test was reported (Borland et al., 2017). In Paper 2, a version that has been translated to Danish, but not validated, was reported.

Summary of the papers

Paper 1

Introduction

This study investigates the effects of various task demands manipulations on WM resource allocation indexed by recall performance. The first aim was to investigate whether the task difficulty of the SWIR test, determined by varying the list length, had an effect on the benefit that noise reduction can give in terms of recall performance. Moreover, the effect of background noise on recall performance was examined. The correlation between WM capacity and recall performance at each task difficulty level was

investigated. Lastly, the effect of task difficulty predictability, i.e. knowing compared to not knowing list length in advance, on recall performance was investigated.

Method

The RS test was administered to measure individual WM capacity. The SWIR test with lists of three, five, seven and nine sentences was administered in 4T babble and SSN with noise reduction off or on. The participants received amplification based on the VAC+ rationale using a hearing aid simulator. The SWIR test sentences were presented at 65 dB SPL and the background noise level was individualized to yield 95% word recognition in 4T babble with noise reduction off. The SWIR test score was calculated as the percentage of recalled words out of the total number of sentences in a list. The participants were divided in two groups: in the predictable task difficulty group, the participants were informed before each list how many sentences it would contain. In the unpredictable task difficulty group, the participants were not informed about list length, although they knew it varied.

Results and discussion

A repeated-measures analysis of variance (ANOVA) with task difficulty predictability group as a between-subjects factor was performed. The results showed that recall performance decreased significantly with increasing task difficulty. However, there was no interaction between list length and noise reduction, indicating that task difficulty does not have an effect on the recall benefit of noise reduction. The correlation analysis between the SWIR test and the RS test revealed a significant positive relationship between the two measures for lists of five, seven and nine sentences. The recall performance for lists of three sentences was nearly 100%, which likely accounts for the correlation with the RS test not being significant. Thus, the SWIR test with varying task difficulty is sensitive to capturing the effect of noise reduction on recall performance, as long as the performance is not at ceiling. This suggests that the SWIR test could be adapted to individual WM capacity.

We replicated previous findings showing that recall performance was significantly better with noise reduction on compared to off in 4T babble, but not in SSN (Ng et al., 2013, 2015). From the perspective of the ELU model, this finding demonstrates that noise reduction decreases the need for explicit processing when listening to speech in noise (Rönnerberg et al., 2013, 2019). This interaction between background noise and noise reduction was more prominent for the unpredictable task difficulty group, although there was no significant difference between the groups in terms of overall recall performance. In order to examine this finding, a second ANOVA was performed focusing on the unpredictable task difficulty group in 4T babble. The outcome showed that there was a higher tendency to start recall from the first item when noise reduction was off compared to on. This suggests that listeners employ different encoding/retrieval strategies with noise reduction on and off when task difficulty is unpredictable.

Paper 2

Introduction

This study expands on the findings of Paper 1 by combining the SWIR test with pupillometry. The SWIR test is an offline measure of WM resource allocation, while pupillometry is an online measure, as it allows tracking WM resource allocation changes over the course of the SWIR test list. The aims were to investigate the effects of task difficulty predictability and noise reduction on WM resource allocation. Moreover, the outcomes of both measures were compared in order to investigate whether they reflect the same processes related to WM resource allocation.

Method

Pupillometry was measured while the participants listened to the SWIR test sentences in a background noise of 4T babble. The target sentences were presented at 65 dB SPL and the level of the 4T babble was individually adjusted to yield 95% word recognition when noise reduction was off. In order to manipulate task difficulty predictability, the SWIR test included lists of three, five and seven sentences. The test consisted of two visits. During one visit the participants were informed before each list how many sentences it would contain, hence the task difficulty was predictable. During the other visit, the task difficulty was unpredictable, since the participants did not know how many sentences the upcoming list would contain. Amplification was provided based on the VAC+ rationale using hearing aids (Oticon OpnS1™ mini-Receiver-in-the-ear). In each visit, half of the SWIR test lists were administered with noise reduction off and half with noise reduction on. Two pupillary responses were obtained. The sentence baseline pupillary response was calculated as the mean pupil dilation during a time interval of one second before each sentence. The task-evoked PPD was defined as the maximum pupil dilation that occurs in a four-second interval starting from sentence onset.

Results and discussion

One ANOVA was conducted for recall performance and one for identification performance in the SWIR test. Three sets of ANOVA were performed for sentence baseline pupillary responses and three sets for PPD, one for each list length. The Benjamini–Hochberg method was used to correct for false discovery rate. The results showed that significantly more words were identified correctly when lists contained three sentences compared to five or seven, as well as when noise reduction was on compared to off. As the list length increased, significantly fewer words were recalled. Furthermore, recall performance was significantly better with noise reduction on compared to off. Task difficulty predictability did not have an effect on either identification nor recall performance. Sentence baseline pupillary responses for lists of five and seven sentences were higher when task difficulty was unpredictable compared to predictable, except for the first sentence in the list. The PPD was significantly higher when task difficulty was unpredictable compared to predictable only for the first sentence of seven-sentence lists.

Sentence baseline pupillary responses increased significantly more towards the end of lists with seven sentences when noise reduction was on compared to off. There was no effect of noise reduction on PPD.

Task difficulty predictability does not seem to influence WM resource allocation, since it did not have an effect on identification or recall performance. Thus, the increased sentence baseline pupillary responses when task difficulty is unpredictable likely indicate a higher degree of task engagement or arousal in anticipation of the end of the list. The increased sentence baseline pupillary responses towards the end of longer lists, when memory load is highest, in conjunction with better recall performance are an index of the increased capacity to allocate resources to storage of speech. The effect of task difficulty predictability was only captured by pupillometry, while the effect of noise reduction was captured by both pupillometry and the SWIR test. Task difficulty predictability presumably involves cognitive processes which cannot be captured by the SWIR test. Since the SWIR test is an offline measure and pupillometry is an online measure, they reflect different temporal aspects of the effect of noise reduction on WM resource allocation.

Paper 3

Introduction

In this study a different application of the combined SWIR test and pupillometry methodology was explored. The aim was to investigate whether the magnitude of the task-evoked PPD measured while listening to the SWIR test sentences and preparing to encode the target items is an index of the likelihood of subsequent memory recall. Furthermore, the effect of WM capacity and noise reduction on subsequent memory recall was investigated.

Method

The SWIR test with lists of seven sentences was administered in a background noise composed of 16T babble. Pupillometry was measured simultaneously. The background noise was presented at 70 dB SPL and the level of the target sentences was individually adjusted to yield 95% speech recognition when noise reduction was off. Amplification was provided using Oticon OpnS1™ and Oticon More™ (mini-Receiver-in-the-ear), but the data from the two hearing aid models was pooled and only the effect of noise reduction was investigated. The baseline-corrected PPD was measured during an interval of four seconds starting at sentence onset. The individual WM capacity was measured using the RS test.

Results and discussion

A binary logistic mixed effects model was constructed including PPD, RS test score, noise reduction, age and PTA as fixed effects. Test participant and the serial position of a sentence in a list were included as crossed random effects. The outcomes demonstrated that only the magnitude of the PPD and RS test scores were significantly linked to subsequent memory recall. The larger the PPD in response to a SWIR test sentence or the higher the RS test score, the higher the likelihood of subsequent memory recall. A post hoc Pearson's correlation analysis between PPD and RS test score did not yield significant outcomes.

The magnitude of the task-evoked PPD is considered to reflect the intensity of attentional processing allocated during encoding (Miller et al., 2019). Interestingly, this is the case even when attentional processing may be increased due to hearing loss and background noise. Individuals with higher WM capacity are presumably able to devote more attentional processing resources during encoding, which accounts for the higher RS test scores being associated with a higher likelihood of subsequent memory recall. While RS test score reflects an individual's maximum WM capacity, the PPD is sensitive to temporal fluctuations in attentional processing. This may account for the non-significant correlation between the RS test and WM capacity. The lack of association between noise reduction and subsequent memory recall may be attributed to the use of 16T babble, which may be less cognitively demanding than 4T babble due to its resemblance to SSN. This type of background noise is presumably less distracting, especially at SNR levels where speech intelligibility is nearly perfect.

Paper 4

Introduction

The aim of this study was to investigate whether the combination of the SWIR test and pupillometry is a suitable method to investigate WM resource allocation. This was done by examining how potential changes in arousal, indexed by the pupil baseline, affect task engagement over the course of the entire SWIR test session. Moreover, it was investigated whether individual factors, namely the WM capacity and the SNR level at which the SWIR test was performed, have an effect on the change in pupil baseline over the course of the task.

Method

The test set-up and procedure were the same as in Paper 3. The pupil baseline was calculated as the mean of all sentence baseline pupillary responses in a list. This yielded one pupil baseline value for each of the 28 administered SWIR test lists, which are referred to as blocks. The RS test was used as a measure of WM capacity.

Results and discussion

Two linear mixed effects model were constructed. They were identical, except for the outcome variables, which were pupil baseline and recall performance respectively. Block, SNR level and RS test score were included as continues fixed effects, as well as their three-way interaction. Participants were included as a random effect and by-participant random intercepts and random slopes for all fixed effects were included. After backward elimination, the main effects of all fixed effects as well as the two-way interaction between SNR level and WM capacity remained in the model with pupil baseline as the outcome variable. In the model with recall performance as the outcome variable, only the RS test score remained. The random intercept and slope for block remained in both models.

The findings showed that the pupil baseline decreased over the course of the trials. However, recall performance did not significantly change over the course of the blocks. Consequently, the decrease in arousal was believed to be attributed to habituation to the task (Ayasse & Wingfield, 2020). According to the FUEL, the individual's evaluation of task demands may affect arousal. Thus, the listeners may have re-evaluated the SWIR test as less challenging after becoming familiar with the test procedure. The recall performance indicated that the participants continued allocating the necessary WM resources to perform the task. The individual WM capacity and SNR level did not have an effect on the decrease in pupil baseline. However, individuals with higher WM capacity exhibited overall a larger pupil baseline at low SNR levels than individuals with lower WM capacity. This may reflect the mechanisms that allow individuals with higher WM capacity to maintain a stable performance in challenging listening situations.

GENERAL DISCUSSION

Main findings and conclusions

Papers 1 and 2 showed that recall performance in the SWIR test was better when noise reduction was activated compared to when it is not in 4T babble, replicating the findings from Ng et al. (2013, 2015). This corroborates evidence that noise reduction facilitates segregation of the target speech from the competing speech, decreasing the need for explicit processing (Rönnberg et al., 2021). Furthermore, in Paper 1 it was shown that the SWIR test with varying list length is sensitive to measuring the cognitive benefit of noise reduction. Although there was a significant correlation between recall performance and WM capacity for lists of five, seven or nine sentences the correlation was not significant for lists of three sentences. This suggests that with three sentences per list the task is easy for individuals with normal cognitive abilities, as recall performance is at ceiling. In Paper 2 the SWIR test outcomes were supplemented with pupillometry outcomes, showing that sentence baseline pupillary responses were higher towards the end of the SWIR test list when noise reduction was on compared to off, especially for longer lists. Since the increased sentence baseline pupillary responses were accompanied by improved recall performance, this combination of responses was considered to reflect the amount of WM resources allocated for storage of the target words over the course of a list.

Task difficulty predictability did not seem to have an effect on WM resource allocation, as recall performance was not affected by this manipulation in Papers 1 and 2. However, the pupillometry findings from Paper 2 showed that the sentence baseline pupillary responses were higher over the course of the SWIR test list when task difficulty was unpredictable compared to predictable. This was assumed to be attributed to higher levels of task engagement or arousal in anticipation of the end of the list when task difficulty was unpredictable. Hence, task difficulty predictability seems to have an effect on cognitive processes other than WM resource allocation to speech storage. The findings from Paper 2 also highlight how behavioral responses can facilitate the interpretation of pupillary responses, since similar dilation patterns were observed for both the effects of noise reduction and task difficulty predictability.

An alternative application of the combined SWIR test and pupillometry method was investigated in Paper 3. The findings showed that the magnitude of the task-evoked PPD measured during encoding on a sentence-by-sentence basis was associated with the likelihood of subsequent memory recall during the SWIR test recall phase. The higher the

task-evoked PPD when listening to a SWIR test sentence, reflecting increased intensity of attentional processing, the higher the likelihood of subsequently recalling the target word. Similar findings have been obtained in previous studies, but not including participants with hearing loss and conducting the testing in background noise. These conditions presumably entail a higher degree of explicit processing. The findings also showed that higher WM capacity was associated with higher likelihood of subsequent memory recall. This is likely due to individuals with higher WM capacity being able to allocate more attentional processing resources to the task (Miller et al., 2019). The effect of noise reduction on subsequent memory recall was not significant, presumably due to the lower degree of lexical interference caused by 16T babble compared to the 4T babble used in the previous studies.

In Paper 4 the effect of changes in arousal over the course of the SWIR test, indexed by the pupil baseline, on task engagement was investigated. Moreover, the effects of WM capacity and individual SNR level on changes in arousal were investigated. The findings demonstrated that the pupil baseline decreased over the course of the task, but recall performance remained relatively stable. Hence, the decrease in arousal was attributed to task habituation (Ayasse & Wingfield, 2020) rather than lower task engagement (Alhanbali et al., 2020; Hopstaken et al., 2015). The individual WM capacity and SNR level did not have an effect on the decrease in pupil baseline. Individuals with higher WM capacity exhibited overall larger pupil baseline at low SNR levels than individuals with lower WM capacity, which may reflect the mechanisms that allow individuals with higher WM to maintain a good level of performance in challenging listening conditions.

The modified SWIR test combined with pupillometry

Effects of noise and noise reduction

Paper 1 replicates the findings from the study by Ng et al. (2013) demonstrating that competing speech impairs recall performance to a higher degree than speech-shaped noise. Background noise results in target signal degradation, which impairs lexical access caused by mismatch of heard input to phonological representations in semantic long-term memory (Rönnberg et al., 2013). Due to its lexical-semantic content, background noise composed of speech may compete with the target speech for the listener's attention to a higher degree than stationary noise (Mattys et al., 2009; Rosen et al., 2013). Rosen et al. (2013) demonstrated that speech recognition performance was lower in competing speech compared to stationary noise. The decrease in performance was most noticeable with two or four competing talkers, where opportunities for glimpsing the target signal are minimized, but it was even observed with a single competing speaker. Consequently, a higher degree of explicit processing is required to segregate target speech from 4T babble compared to SSN, increasing dependency on WM (Rönnberg et al., 2013). In fact, it has

been shown that the correlation between WM capacity and speech recognition performance is stronger in 4T babble than in SSN. Interestingly, this correlation weakens over time with increasing hearing aid experience for SSN, but not for 4T babble (Ng & Rönnberg, 2020). Taking these considerations into account, only 4T babble was used in Paper 2.

The SWIR test outcomes from Papers 1 and 2 corroborate the findings from Ng et al. (2013, 2015), demonstrating that recall performance is better when noise reduction is on compared to when it is off in 4T babble. Although the ELU model does not address speech encoding for subsequent recall, these findings align with the predictions made regarding understanding speech in challenging listening environments. When noise reduction is off, background noise is more likely to interfere with the matching process between the phonological representations held in the RAMBPHO buffer and the phonological representations stored in semantic long-term memory. Hence, explicit processing resources are needed to remediate the instances of mismatch. When more explicit processing resources are needed for speech understanding, it can be assumed that fewer resources are available for storage of speech, reflected in the lower recall performance. Papers 1 and 2 alongside the studies by Ng et al. provide evidence that noise reduction decreases the need for explicit processing required to understand speech in cognitively demanding listening situations, freeing up resources to be allocated for storage. From the perspective of the FUEL, the improvement in recall performance with noise reduction on can be attributed to a decrease in WM resources required to achieve successful speech understanding, thus reducing task demands and thereby listening effort. Consequently, more attentional resources can be allocated to maintaining items in memory for subsequent recall when noise reduction is on.

In Paper 2 the investigation of WM resource allocation was expanded by including pupillometry as a physiological measure alongside the behavioral performance in the SWIR test. Previous research has mainly combined pupillometry with speech recognition tests in order to quantify the effect of noise reduction on the allocation of resources to speech processing. These studies found the magnitude of the task-evoked PPD to be a reliable index of WM resources allocated to speech processing with noise reduction on or off (Ohlenforst et al., 2018; Wendt et al., 2017). The findings from Paper 2 on the other hand do not show an effect of noise reduction on the task-evoked PPD over the course of the SWIR test list, but on sentence baseline pupillary responses instead. This is in line with evidence suggesting that adding a memory load to a speech recognition task overshadows the effect of perceptual factors on task-evoked pupillary responses (Zekveld et al., 2018). A study by Bönitz et al. (2021) is the only one to have combined the SWIR test with pupillometry in a similar way to Paper 2 thus far, but including a group of individuals with normal hearing. The authors interpret the change of the task-evoked pupillary responses over the course of a list as an index of listening effort, which is in line with the interpretation of the FUEL. The change in sentence baseline pupillary responses

is considered to reflect a modulation of “memory effort” (Bönitz et al., 2021). In other words, the sentence baseline pupillary responses reflect the number of items held in memory for later recall. This is in agreement with the findings from Paper 2, where the higher increase in sentence baseline pupillary responses towards the end of longer SWIR test lists with noise reduction on also coincided with better recall performance compared to when noise reduction was off. Simply put, the increased allocation of WM resources to storage seemed to pay off in terms of the ability recall more words, which is reflected in the increasing sentence baseline pupillary responses over the course of a SWIR test list.

Effects of task difficulty predictability

Ng et al. (2013, 2015) administered the SWIR test with a fixed number of sentences per list (eight and seven respectively). Since the participants were aware of the number of sentences in a list, the task difficulty can be seen as being predictable. In their studies, there was a significant effect of noise reduction on recall performance. In contrast, the findings from Paper 1 suggest that the cognitive benefit of noise reduction was less prominent when task difficulty was predictable. This may have been due to the varying list length. When list length is constant, participants may use the same encoding strategy throughout the test or they may settle on one relatively fast (Unsworth et al., 2019). In Paper 1, the participants who were informed about the list length in advance, may have switched encoding or retrieval strategies between lists based on task difficulty. Although this was not analyzed statistically, visual inspection of table 2 suggests that this may have been the case. Table 2 is summarized from table 3 in Paper 1 and it shows the probability of starting recall from the first or the last item in the SWIR test list for each list length when task difficulty is predictable or unpredictable in 4T babble. Note that the noise reduction conditions have been pooled. The tendency of starting recall from the first item in the shorter lists and the last item in the longer lists is more pronounced when task difficulty is predictable. This suggests that test participants may adapt their encoding or retrieval strategy based on task difficulty when they know the list length in advance (Grenfell-Essam & Ward, 2012).

The task difficulty predictability did not have any effect on overall recall performance in neither Paper 1 nor Paper 2. Additionally, there was no effect of task difficulty predictability on word recognition in Paper 2. Therefore, this manipulation did not seem to affect how WM resources are allocated to neither processing nor storage of speech. The additional analysis in Paper 1 suggests that task difficulty predictability may have an effect on encoding or retrieval strategies. Investigating the effect of task difficulty predictability using pupillometry in Paper 2 has given further insights into how this manipulation could affect other processes. The findings demonstrated that sentence baseline pupillary responses were higher over the course of a SWIR test list when task difficulty was unpredictable compared to predictable. In this case, the sentence baseline

pupillary responses were interpreted as an index of arousal associated with anticipation of the end of the list or increased alertness or task engagement throughout the list (Ayasse & Wingfield, 2020; Unsworth & Robison, 2014). Consequently, the encoding or recall strategy used depending on task difficulty predictability may modulate arousal levels.

Table 2: The probability of first recall for the first and last input position at each task difficulty level in 4T babble for unpredictable and predictable task difficulty (table summarized from Paper 1, noise reduction conditions are pooled).

		Unpredictable task difficulty		Predictable task difficulty	
		<i>Input position</i>			
Sentences/list		First	Last	First	Last
Probability of first recall	3	0.51	0.51	0.77	0.26
	5	0.35	0.50	0.46	0.55
	7	0.43	0.48	0.35	0.59
	9	0.39	0.50	0.39	0.65

Regardless of the strategy used in each task difficulty predictability condition and its effect on arousal, it seems that the test participants allocate the necessary WM capacity to accomplish the task, which accounts for the similar recall performance levels. These phenomena may be encompassed by the motivation aspect of the FUEL, as motivation and arousal are closely related (Aston-Jones & Cohen, 2005). In the explanation of the 3D plot (figure 4), a scenario is described where the listener’s motivation increases as they become more engaged, while task demands remain constant. Thus, WM resource allocation does not change. Similarly, the findings from Paper 2 show that the sentence baseline pupillary responses reflect increased arousal or task engagement when task difficulty is unpredictable compared to predictable, but the recall performance suggests that WM resource allocation, and thereby listening effort, is relatively equal in both conditions.

An alternative method of investigating working memory resource allocation

For Paper 3 the same combination of measures was used as in Paper 2 in order to investigate WM resource allocation from a different perspective. Rather than analyzing overall recall performance and pupillary responses averaged over several lists, the magnitude of the task-evoked PPD measured during encoding on a sentence-by-sentence basis and the likelihood of subsequently recalling the corresponding target word was investigated. The PPD in this case was seen as an index of the momentary intensity of attention (Kahneman, 1973; Miller et al., 2019). According to Kahneman's Capacity Model of Attention (1973), which the FUEL is based on, the available cognitive capacity that can be allocated to a task is limited. Having to divide attention during encoding lowers the chance of retrieving an item (Miller et al., 2019). In Paper 3, this may be due to the challenges of listening to speech in competing babble noise, while trying to maintain the target SWIR test items in memory for later recall.

The findings of Paper 3 demonstrated that a larger PPD measured when listening to a SWIR test sentence and encoding the target word was associated with higher likelihood of recalling that particular item, which corroborates previous findings on individuals with normal hearing (Bergt et al., 2018; Kucewicz et al., 2018; Miller et al., 2019; Papesh et al., 2012). This may seem to contradict the interpretation that the PPD measured during speech recognition tasks reflects increased allocation of WM resources to speech processing (Ohlenforst et al., 2018; Wendt et al., 2017). When more WM resources are required for speech processing, less WM resources are expected to be available for encoding words into memory (Lunner et al., 2009). Interestingly, the findings from Paper 3 suggest that the PPD measured on a sentence-by-sentence basis during the SWIR test reflects the amount of attentional resources devoted to encoding items into memory for subsequent recall. Consequently, this provides evidence that pupillary responses are sensitive to the type of task, indicating that the effect of memory load on pupillary responses is more dominant than the effect of perceptual factors (Zekveld et al., 2018; Zhang et al., 2021).

Taking advantage of both task-evoked and baseline pupillary responses

Research has mainly focused on the task-evoked pupillary responses as an index of WM resource allocation or listening effort. Baseline pupillary responses have mainly been used as baseline correction for task-evoked pupillary responses. However, it is important not to overlook the potential value of baseline pupillary responses (Winn et al., 2018). Being measured over longer time intervals, baseline pupillary responses may be a suitable index of overall WM resource allocation. As demonstrated in Paper 2, sentence baseline

pupillary responses tracked over the course of a SWIR test list may reflect the amount of WM resources devoted to maintaining an increasing number of items in memory for subsequent recall. This is in agreement with the findings reported by Bönitz et al. (2021) in their study combining the SWIR test with pupillometry. The PPD may be less sensitive to capturing overall WM resource allocation, as it is a transient response to a specific stimulus. Thus, the PPD may be more sensitive to capturing the momentary attentional processing (Miller et al., 2019) when listening to a sentence and encoding the target word. Paper 3 provides evidence in support of this, demonstrating that the magnitude of the PPD was significantly associated with the likelihood of subsequent memory recall.

Figure 6: Hypothetical illustration of the task-evoked peak pupil dilation (vertical continuous grey lines) and sentence baseline pupillary response (dotted black line) as indices of different aspects of working memory resource allocation. The encircled serial positions represent the items in a SWIR test list that are more likely to be recalled based on the intensity of attentional processing devoted to them.

Both the task-evoked PPD and the sentence baseline pupillary responses were analyzed in this thesis in order to investigate different aspects of WM resource allocation to

processing and storage. The cognitive processes captured by each of these responses are illustrated in figure 6. The figure suggests that sentence baseline pupillary responses (dotted black line) reflect an accumulation of WM resources allocated for storage over time, while the magnitude of the PPD (continuous grey line) reflects the momentary attentional processing during each SWIR test sentence. The encircled serial positions represent items that are more likely to be recalled based on the amount of attentional processing devoted to them. It should be noted that this is a hypothetical illustration and not based on real data. Moreover, although it is speculated that the sentence baseline pupillary response increases as the number of words maintained in memory accumulate, this increase may not be linear.

Koelewijn et al. (2012) found that individuals with better cognitive abilities were more successful in a speech recognition in noise task, but also exhibited a larger PPD than individuals with poorer cognitive abilities. The authors argue that individuals with higher cognitive capacity are able to expend more processing resources to understand speech in challenging listening situations, which is reflected in the increased PPD. This suggests that the magnitude of the PPD is associated with cognitive capacity. The data from the studies in this thesis did not show any significant relationships between WM capacity and PPD nor WM capacity and sentence baseline pupillary responses. Consequently, these pupillary responses are considered to reflect the momentary intensity of attention and overall WM resource allocation during the SWIR test regardless of WM capacity. It is therefore speculated that the dilation of the pupillary responses exemplified in figure 6 is modulated by the allocation of WM resources, rather than limited based on WM capacity.

Investigating the pupil baseline over the course of an entire test session can provide additional insights into cognitive processing during the SWIR test. It has been shown that the baseline pupillary responses are lower towards the end of a test block, which also coincides with decreased behavioral performance, suggesting lapses in attentional engagement (Van Den Brink et al., 2016). Hopstaken et al. (2015) and Alhanbali et al. (2020) have interpreted the decreased baseline pupillary responses over the course of a test session accompanied by decreased behavioral performance as an index of fatigue. Alternatively, it has also been shown that although baseline pupillary responses decrease over the course of a task, behavioral performance does not decrease, suggesting a reduction in the participants' anxiety related to own task performance rather than low task engagement or fatigue (Ayasse & Wingfield, 2020).

The SWIR test is assumed to be a cognitively demanding test and decreasing task engagement, as highlighted by the FUEL, may be of concern. Task engagement has been defined as the "readiness to invest resources to accomplish a task goal" (Lemke & Besser, 2016). Pichora-Fuller et al. (2016) have also adopted this definition. If listeners disengage from the SWIR test and cease to invest the necessary WM resources, even when the task demands do not exceed their individual maximum capacity, it may interfere with

capturing the effect of various task manipulations on WM resource allocation. However, the evidence from Paper 4 suggests that this is not the case. Although the decrease in pupil baseline suggests a decrease in arousal over the course of the test session, recall performance remains relatively constant. This indicates that the participants continue allocating the necessary WM resources to perform the task. Thus, the findings suggest that participants habituate to the SWIR test, but do not reach the point of disengagement.

Figure 7 presents speculative hypotheses on engagement and disengagement based on the findings from Paper 4 and a hypothetical scenario based on previous studies (Alhanbali et al., 2020; Hopstaken et al., 2015). The figure proposes how task (dis)engagement may be reflected in behavioral and pupillary responses as well as how it may relate to WM capacity and resource allocation.

Figure 7: Hypothetical illustration of task engagement and its effect on behavioral and pupillary responses, as well as on overall working memory resource allocation.

The pupil baseline over the course of a test session has been typically shown to decrease (Alhanbali et al., 2020; Ayasse & Wingfield, 2020; Hopstaken et al., 2015; Zekveld, Koelewijn, et al., 2018), which is indicated by the downward pointing arrow in both engagement scenarios. When accompanied by the stable recall performance (represented by the double hyphens), as was the case in Paper 4, the decreasing pupil baseline is considered to reflect a decrease in arousal due to task habituation. Based on the FUEL, the task habituation is speculated to be a result of the SWIR test being evaluated as less challenging once the participants get used to the test procedure. This suggests that the necessary resources are invested to perform the SWIR test, regardless of how resources may be divided between speech processing and storage (represented by the full bar). In this case the listener maintains a relatively constant level of engagement throughout the entire test session. In the hypothetical situation it is speculated that a decrease in pupil baseline accompanied by a decrease in recall performance over the test session would

index disengagement from the task. This could be due to the listener not allocating sufficient WM resources to meet task demands, despite the task not exceeding their individual capacity. The decrease in investment of WM resources relative to the necessary amount to successfully perform the task is shown in the dotted bar. This is in line with the FUEL, which indicates that arousal may affect a listener's choice to engage in a task.

Behavioral data disentangles the ambiguity of pupillary responses

Pupillary responses may be difficult to interpret on their own. In Paper 2 it was found that the effects of noise reduction and task difficulty predictability resulted in similar pupillary response patterns. The accompanying behavioral performance data was essential for the interpretation of the different effects. Furthermore, in Paper 4, the decreasing pupil baseline over the course of the test session could have been interpreted as disengagement from the task if it was not for the behavioral performance data suggesting otherwise. These findings highlight the advantages of combining behavioral and physiological measures when it comes to the interpretation of results.

In general, within the field of cognitive hearing science, increased pupillary responses have been viewed somewhat negatively, being closely linked with increased listening effort (Ohlenforst et al., 2018; Pichora-Fuller et al., 2016; Wendt et al., 2017; Zekveld, Koelewijn, et al., 2018). However, Winn et al. (2018) encourage caution about equating pupil dilation magnitude with effort, since certain individuals, such as those who have hearing loss or those who are older, may exhibit smaller pupil sizes, despite reports of increased listening effort. Additionally, increased effort may not always be negative. In Papers 2 and 3 increased pupillary responses were accompanied by better recall performance, suggesting that more WM resources were allocated for storing speech (figure 2). This is another instance that highlights the usefulness of behavioral data for a more accurate interpretation of the pupillary responses. Here, the concept of “successful effort” is proposed, referring to the deliberate allocation of additional WM resources to achieve a better performance level. In this case, allocating more resources or exerting more effort is considered successful, since it leads to a better outcome.

Theoretical implications

The ELU model focuses on speech understanding during communication and the dependency on WM in sub-optimal listening environments. However, it does not make predictions about processes that may be involved when speech needs to be stored for later recall. When listening to a talker, the recipient attempts to remember the gist and take turns during the dialogue (Rönnberg et al., 2021). In order to simulate such communicative situations, speech recognition tests have been used in pupillometry

studies investigating WM resource allocation. Typically, these are conducted in different levels and types of background noise and/or with various hearing aid settings. These factors have been shown to affect the trade-off between implicit and explicit speech processing (Koelewijn et al., 2012; Ohlenforst et al., 2018; Wendt et al., 2017; Zekveld et al., 2011). Listening during speech recognition tests may be more passive than listening during real-life dialogue, since preparation for turn-taking is not needed. Furthermore, each sentence is repeated immediately after being heard, thus eliminating the need to keep attention on the gist for longer time intervals. The SWIR test on the other hand requires that participants maintain parts of the speech in memory for later recall, reflecting some of the processes that occur during real-life dialogue. It is important to note that the SWIR test demands more than remembering the gist, as specific words must be recalled accurately. Thus, the memory component may be over-estimated in relation to typical real-life dialogue. Since the memory component is such an essential part of the task, it is speculated that memory-related processes, which are not within the original scope of the ELU model, are involved.

When only the gist needs to be understood, the WM resources required for storing speech are low and the balance between implicit and explicit speech processing is mainly determined by the mismatch mechanisms. Previous studies have shown that the task-evoked pupillary responses reflect the amount of WM resources allocated for speech understanding, thereby being associated with the trade-off between implicit and explicit processing (Koelewijn et al., 2012; Ohlenforst et al., 2018; Wendt et al., 2017; Zekveld et al., 2011). In contrast, the evidence from the present thesis indicates that, when speech needs to be recalled, both baseline and task-evoked pupillary responses reflect different mechanisms related to storing speech (figure 6) rather than speech processing. Overall, increased pupillary responses are linked to a higher likelihood of successful recall. Furthermore, the pupillary responses are considered to reflect the allocation of WM resources to encoding, regardless of WM capacity. Figure 8 shows the ELU model illustration with speculative additions, adapting it to a speech understanding and recall scenario.

During continuous speech, the loop shown in the ELU model illustration repeats multiple times to unlock the lexical meaning of each word. Two additional processes are proposed when particular words need to be stored for later recall (figure 8):

- a) Speech encoding: Regardless of whether speech processing is implicit or explicit, once lexical access has been obtained, the explicit speech encoding process is then activated within the loop. Speech encoding is described as “explicit”, given that recalling specific words accurately may involve more conscious processing in comparison to the processes required to remember the gist. The probability of successful speech encoding is reflected in the magnitude of the task-evoked PPD, which is an index of the intensity of attention devoted when preparing to encode

- a target word (Paper 3). Successful encoding may be dependent on the degree to which speech processing is implicit or explicit, based on the assumption that the more WM resources are occupied for speech processing, the fewer are available for storage.
- b) Maintenance for recall: The target words need to be kept in memory until the free recall phase of the SWIR test. Thus, an ongoing process outside of the loop ensures the maintenance of the target words over a longer time interval. The amount of WM resources allocated to maintaining the items in memory for later recall are reflected in the sentence baseline pupillary responses (Paper 2). The maintenance process is presumably dependent on the outcome of the speech encoding processes and is therefore indirectly impacted by the balance between implicit and explicit processing.

Figure 8: Illustration of the Ease of Language Understanding model adapted to include the processes involved in speech encoding for subsequent recall as well as the corresponding pupillary responses.

In summary, it is speculated that the ELU model may be supplemented with two processes that are relevant for an auditory recall task: the momentary explicit speech encoding reflected in the transient task-evoked PPD and the ongoing maintenance of words for subsequent recall reflected in the sentence baseline pupillary responses.

Methodological considerations

Design of studies combining the SWIR test and pupillometry

The implications of manipulating task difficulty predictability are important to consider in terms of the design and methodology of the SWIR test in combination with pupillometry. Unpredictable task difficulty may be considered to have higher ecological validity, as utterance length is unpredictable in real life communication. However, Papers 1 and 2 suggest that task difficulty predictability may affect cognitive processes other than WM resource allocation, such as encoding/retrieval strategies or arousal/task engagement, which the SWIR test is not designed to capture. This may influence the ability to capture the effects of signal processing on WM resource allocation. Although overall recall performance was not affected, pupillary responses were sensitive to this manipulation. Consequently, this should be taken into account when designing a study that combines the SWIR test with pupillometry, especially when interpreting of the pupillometry outcomes.

Background noise

The outcomes from Paper 1 are in agreement with evidence demonstrating that 4T babble increases dependency on WM compared to SSN (Ng & Rönnerberg, 2020), which is why noise reduction provides a higher cognitive benefit in competing speech (Ng et al., 2013). However, this may only be the case for background composed of a limited number of speakers. In Papers 3 and 4, which are based on the same study sample, 16T babble was used as background noise. No effect of hearing aid signal processing was found in these data. One possible reason for this may have been the choice of background noise. As the number of talkers increases, the competing speech becomes less intelligible. This results in decreased competition with the target speech for the listener's attentional resources (Rosen et al., 2013). The 16T babble may be perceived more similarly to the way SSN is perceived, thereby being less cognitively demanding than 4T babble. Given the high speech intelligibility level at which the SWIR test is conducted, attenuating the 16T babble further may not provide a cognitive benefit, as was the case for SSN in Paper 1 and in the study by Ng et al. (2013). The selection of background noise may therefore be an important consideration if the aim is to evaluate the effect of noise reduction algorithms on WM resource allocation using the SWIR test.

FUTURE DIRECTIONS

The combination of the SWIR test with pupillometry is a relatively novel, but promising method of investigating various aspects of WM resource allocation. Thus, there are additional research and potential clinical applications to be explored.

- Investigating WM resource allocation in individuals with mild cognitive impairment:

Even when audiometric thresholds fall within the area of normal hearing, individuals with mild cognitive impairment exhibit poorer speech recognition performance in noise than individuals with normal cognitive function (Jalaei et al., 2019). The combined SWIR test and pupillometry methodology could be used to investigate differences in WM resources allocation under various task demands between individuals with and without normal cognitive function, as well as whether noise reduction can provide a cognitive benefit for people who do not have normal cognitive function. This is of interest based on a study demonstrating that pupillary responses may be a biomarker of early risk for mild cognitive impairment (Granholm et al., 2017). This study suggests that increased pupil dilation in people with normal cognitive function may be an indicator of the risk for developing mild cognitive impairment.

- Guiding hearing aid signal processing development:

The SWIR test and pupillometry have been successfully used to evaluate the effects of hearing aid signal processing on WM allocation and listening effort (Ng et al., 2013, 2015; Ohlenforst et al., 2018; Wendt et al., 2017). The advantage of these measures is that they target cognitive functions and listening conditions that are relevant for real-life communication. Incorporating these measures to evaluate new hearing aid signal processing technologies would be beneficial for ensuring both improved audibility and optimizing WM resource allocation to ease communication in challenging listening environments. This could ultimately facilitate social participation, leading to improved life quality (Pichora-Fuller et al., 2016).

- Investigating WM resource allocation at an individual level:

Pupillometry has become a well-established measure of listening effort at a group level. However, recent ongoing research aims to investigate the reliability of pupillary responses as an index of individual listening effort (Neagu et al., 2019). If suitable features of pupillary responses are found, it may be relevant to investigate whether pupillary responses measured during the SWIR test could be

used to investigate WM resource allocation or listening effort at an individual level. Reliable indices at an individual level would bring this methodology a step closer to potential clinical implementation.

- Clinical feasibility:

No clinical measures of the effect of hearing aid signal processing on WM resource allocation or listening effort have been developed so far. Yet, this may be essential for providing individualized and optimal auditory rehabilitation. The current applications of the SWIR test for research purposes are too time-consuming to be clinically feasible. However, this thesis provides evidence that the SWIR test with varying list length is sensitive to capturing the cognitive benefit of noise reduction in 4T babble. Thus, the testing duration could be reduced by shortening the list length. Furthermore, considering the relationship between WM capacity and performance on the SWIR test, an adaptive procedure could be developed such that the list length increases only if all target words are correctly recalled, in order to ensure that task demands do not exceed individual WM capacity. Assuming that reliable pupillary indices of individual WM resource allocation are found, and a clinically feasible SWIR test version is developed, these measures could be incorporated in clinical praxis to guide the choice of hearing aid settings so as to find an optimal balance between audibility and minimal listening effort.

ACKNOWLEDGEMENTS

There are many people I would like to thank for being part of my PhD journey. I would like to start by thanking all the participants who bravely volunteered to take part in these studies. I hope that draining your working memory resources helped generate knowledge that will benefit people with hearing loss.

I am grateful to my main supervisor, **Jerker Rönnberg**, for his guidance and for sharing his valuable expertise with me. This project would not have been possible without my co-supervisor, **Elaine Hoi Ning Ng**. Thank you for your patience, keeping me on track, bringing fika into my life, and above all for encouraging me. I have learned a lot from both of you and I hope I will continue to do so in the future.

I would like to express my gratitude to all my co-authors for their support and valuable feedback. I would like to especially thank **Lorenz Fiedler** and **Dorothea Wendt** for sharing their knowledge with me and for their input at various crucial phases of my project. A heart-felt thank you to **Patrycja Książek** for being my PhD buddy, I appreciate your thoroughness, honesty and patience. I would like to thank **Reena Murmu Nielsen** for stepping in to help and being a great role model for women in academia. I am also grateful to **Yue Zhang** for her ideas and interest in my work. A very special thanks to **Fares El-Azm**, not only for the technical help, but also for the fun and laughs. Your good vibes have gotten me through some of the harder parts.

I would like to thank **Mary Rudner** and **Adriana Zekveld** for acting as opponents at my half-time and final seminars respectively. Your input has provided invaluable guidance along the way.

I would like to express a warm thank you to all my colleagues at the Disability Research Division. Especially, I would like to thank the Super Happy Team, **Karin Nilsson** (supreme leader) and **Daniel Schöld**. There are too many things I am grateful for to list here, but I will highlight bubble evenings and numerous priceless episodes such as falling-from-chair-in-middle-of-lecture. **Elisabeth Ingo**, thank you for taking care of me and always suggesting fika when I needed it most. **Lisa Palmqvist**, hearing your laughter from the other room can cheer up the worst moods! Thank you, **Michaela Socher**, for entertaining me during the foundation course and for diminishing my fear of programming. A big thanks to **Josefine Andin** for the warm smiles and encouraging words. **Carine Signoret**, thank you for making me feel welcome! Thank you, **Henrik**

Danielsson, for your friendliness and introduction to the Linköping tango scene! **Maria Hugo-Lindén**, I am grateful for the administrative help and patiently answering my questions.

A special thanks to my colleagues at Oticon who have supported my project in different ways. I would like to recognize **Jens-Christian Britze Kijne**, **Christian Stender Simonsen** and **Michael Syskind Pedersen** for their valuable technical support. Merci, **Pierre-Yves Hasan**, for your pupillometry advice. **Lu Xia**, thank you for your help and Confucius quotes. I would also like to thank all my colleagues in the **CAAR team** for cheering me on. I am deeply grateful to my inspiring friend, **Husmita Ratanjee-Vanmali**, for the advice and support.

Thank you, **Morten Ammekilde Nielsen**, **Soffi Skovlund Jensen** and **Tomas Bjuvmar**, for the help with participant recruitment.

A warm thank you to my dear friends **Pernille**, **Viktorija**, **Brian** and **Tirdad** for both professional and unprofessional help, your encouragement has been priceless. A special thank you to my best friend in the world **Vero**, for always being there.

Mami and **Tati**, you are the reason why I am where I am, thank you from my heart for believing in me!

Thank you to the best friend-sister **Claudia** (a.k.a. Prof. Garibaldi), for spoiling me and understanding me without words. Thank you, my love **Egil**, for your unconditional love and support. You two are the bestest, my most fierce supporters, and you deserve informal co-authorships on this thesis. And of course, thank you **Mango**, **Chili** and **Pringles** for being the best listeners.

REFERENCES

- 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*, 1–11. <https://doi.org/10.1080/14992027.2020.1853262>
- Anderson, M. C., Arehart, K. H., & Souza, P. E. (2018). Survey of current practice in the fitting and fine-tuning of common signal-processing features in hearing AIDS for adults. *Journal of the American Academy of Audiology*, 29(2), 118–124. <https://doi.org/10.3766/jaaa.16107>
- Arehart, K., Souza, P., Baca, R., & Kates, J. M. (2013). Working Memory, Age, and Hearing Loss: Susceptibility to Hearing Aid Distortion. *Ear and Hearing*, 34, 251–260.
- Arehart, K., Souza, P., Kates, J., Lunner, T., & Pedersen, M. S. (2015). Relationship between signal fidelity, hearing loss and working memory for digital noise suppression. *Ear and Hearing*, 36(5), 505–516. <https://doi.org/10.1097/AUD.0000000000000173>
- Arlinger, S. (2003). Negative consequences of uncorrected hearing loss—a review. *International Journal of Audiology*, 42(sup2), 17–20. <https://doi.org/10.3109/14992020309074639>
- Arlinger, S., Lunner, T., Lyxell, B., & Kathleen Pichora-Fuller, M. (2009). The emergence of cognitive hearing science. *Scandinavian Journal of Psychology*, 50(5), 371–384. <https://doi.org/10.1111/j.1467-9450.2009.00753.x>
- 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, 403–450. <https://doi.org/10.1146/annurev.neuro.28.061604.135709>
- Ayasse, N. D., & Wingfield, A. (2020). Anticipatory Baseline Pupil Diameter Is Sensitive to Differences in Hearing Thresholds. *Frontiers in Psychology*, 10, 1–7. <https://doi.org/10.3389/fpsyg.2019.02947>

- Baddeley, A. (2012). Working Memory: Theories, Models, and Controversies. *Annual Review of Psychology*, 63, 1–29. <https://doi.org/10.1146/annurev-psych-120710-100422>
- Beatty, J. (1982). 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>
- Bergt, A., Urai, A. E., Donner, T. H., & Schwabe, L. (2018). Reading memory formation from the eyes. *European Journal of Neuroscience*, 47, 1525–1533. <https://doi.org/10.1111/ejn.13984>
- Boldt, J. B., Kjems, U., Pedersen, M. S., Lunner, T., & Wang, D. (2008). Estimation of the Ideal Binary Mask using Directional Systems. In *Proceedings of the 11th International Workshop on Acoustic Echo and Noise Control International Workshop on Acoustic Echo and Noise Control, University of Washington Campus in Seattle*.
- 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*, Advance online publication. <https://doi.org/10.1097/AUD.0000000000001002>
- Borland, E., Nägga, K., Nilsson, P. M., Minthon, L., Nilsson, E. D., & Palmqvist, S. (2017). The Montreal Cognitive Assessment: Normative Data from a Large Swedish Population-Based Cohort. *Journal of Alzheimer's Disease*, 59(3), 893–901. <https://doi.org/10.3233/JAD-170203>
- Buus, S., & Florentine, M. (2001). Growth of Loudness in Listeners with Cochlear Hearing Losses: Recruitment Reconsidered. *Journal of the Association for Research in Otolaryngology*, 03, 120–139. <https://doi.org/10.1007/s101620010084>
- Cardon, G., & Sharma, A. (2018). Somatosensory cross-modal reorganization in adults with age-related, early-stage hearing loss. *Frontiers in Human Neuroscience*, 12, 1–11. <https://doi.org/10.3389/fnhum.2018.00172>
- Carson, N., Leach, L., & Murphy, K. J. (2018). A re-examination of Montreal Cognitive Assessment (MoCA) cutoff scores. *International Journal of Geriatric Psychiatry*, 33, 379–388. <https://doi.org/10.1002/gps.4756>
- Chisolm, T. H., Johnson, C. E., Danhauer, J. L., Portz, L. J. P., Abrams, H. B., Lesner, S., McCarthy, P. A., & Newman, C. W. (2007). A Systematic Review of Health-Related Quality of Life and Hearing Aids: Final Report of the American Academy of Audiology Task Force on the Health-Related Quality of Life Benefits of Amplification in Adults. *Journal of the American Academy of Audiology*, 18(2), 151–183.
- Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory

- during reading. *Journal Of Verbal Learning And Verbal Behavior*, 19(4), 450–466. [https://doi.org/10.1016/S0022-5371\(80\)90312-6](https://doi.org/10.1016/S0022-5371(80)90312-6)
- Davies-Venn, E., & Souza, P. (2014). The Role of Spectral Resolution, Working Memory, and Audibility in Explaining Variance in Susceptibility to Temporal Envelope Distortion. *Journal of the American Academy of Audiology*, 25, 592–604. <https://doi.org/10.3766/jaaa.25.6.9>
- Desjardins, J. L., & Doherty, K. A. (2014). The Effect of Hearing Aid Noise Reduction on Listening Effort in Hearing-Impaired Adults. *Ear and Hearing*, 35, 600–610.
- Dryden, A., Allen, H. A., Henshaw, H., & Heinrich, A. (2017). The Association Between Cognitive Performance and Speech-in-Noise Perception for Adult Listeners: A Systematic Literature Review and Meta-Analysis. *Trends in Hearing*, 21, 1–21. <https://doi.org/10.1177/2331216517744675>
- Edwards, B. (2016). A Model of Auditory-Cognitive Processing and Relevance to Clinical Applicability. *Ear and Hearing*, 37, 85S-91S. <https://doi.org/doi:10.1097/AUD.0000000000000308>.
- Glick, H. A., & Sharma, A. (2020). Cortical Neuroplasticity and Cognitive Function in Early-Stage, Mild-Moderate Hearing Loss: Evidence of Neurocognitive Benefit From Hearing Aid Use. *Frontiers in Neuroscience*, 14, 1–22. <https://doi.org/10.3389/fnins.2020.00093>
- Gordon-Salant, S., & Cole, S. S. (2016). Effects of age and working memory capacity on speech recognition performance in noise among listeners with normal hearing. *Ear and Hearing*, 37(5), 593–602. <https://doi.org/10.1097/AUD.0000000000000316>
- Granholm, E. L., Panizzon, M. S., Elman, J. A., Jak, A. J., Hauger, R. L., Bondi, M. W., Lyons, M. J., Franz, C. E., & Kremen, W. S. (2017). Pupillary Responses as a Biomarker of Early Risk for Alzheimer’s Disease. *Journal of Alzheimer’s Disease*, 56(4), 1419–1428. <https://doi.org/10.3233/JAD-161078>
- Grenfell-Essam, R., & Ward, G. (2012). Examining the relationship between free recall and immediate serial recall: The role of list length, strategy use, and test expectancy. *Journal of Memory and Language*, 67, 106–148. <https://doi.org/10.1016/j.jml.2012.04.004>
- Hällgren, M., Larsby, B., & Arlinger, S. (2006). A Swedish version of the Hearing In Noise Test (HINT) for measurement of speech recognition. *International Journal of Audiology*, 45(4), 227–237. <https://doi.org/10.1080/14992020500429583>
- Hopstaken, J. F., van der Linden, D., Bakker, A. B., & Kompier, M. A. J. (2015). A multifaceted investigation of the link between mental fatigue and task disengagement. *Psychophysiology*, 52(3), 305–315. <https://doi.org/10.1111/psyp.12339>
- Jalaei, B., Valadbeigi, A., Panahi, R., Nahrani, M. H., Arefi, H. N., Zia, M., & Ranjbar,

- N. (2019). Central auditory processing tests as diagnostic tools for the early identification of elderly individuals with mild cognitive impairment. *Journal of Audiology and Otology*, 23(2), 83–88. <https://doi.org/10.7874/jao.2018.00283>
- Jensen, J., & Pedersen, M. S. (2015). Analysis of beamformer directed single-channel noise reduction system for hearing aid applications. *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, Australia*, 5728–5732. <https://doi.org/10.1109/ICASSP.2015.7179069>
- 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>
- Kahneman, D. (1973). *Attention and Effort*. Prentice-Hall. <https://doi.org/10.2307/1421603>
- Keidser, G., Seeto, M., Rudner, M., Hygge, S., & Rönnberg, J. (2015). On the relationship between functional hearing and depression. *International Journal of Audiology*, 54(10), 653–664. <https://doi.org/10.3109/14992027.2015.1046503>
- Kjems, U., & Jensen, J. (2012). Maximum likelihood based noise covariance matrix estimation for multi-microphone speech enhancement. *Proceedings of the European Signal Processing Conference, Romania*, 295–299.
- 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*, 1–11. <https://doi.org/10.1155/2012/865731>
- 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>
- Kucewicz, M. T., Dolezal, J., Kremen, V., Berry, B. M., Miller, L. R., Magee, A. L., Fabian, V., & Worrell, G. A. (2018). Pupil size reflects successful encoding and recall of memory in humans. *Scientific Reports*, 8(1), 1–7. <https://doi.org/10.1038/s41598-018-23197-6>
- Laeng, B., Sirois, S., & Gredebäck, G. (2012). Pupillometry: A window to the preconscious? *Perspectives on Psychological Science*, 7(1), 18–27. <https://doi.org/10.1177/1745691611427305>
- Lemke, U., & Besser, J. (2016). Cognitive load and listening effort: Concepts and age-related considerations. *Ear and Hearing*, 37, 77S–84S. <https://doi.org/10.1097/AUD.0000000000000304>
- Lin, F. R., Ferrucci, L., An, Y., Goh, J. O., Doshi, J., Metter, E. J., Davatzikos, C., Kraut, M. A., & Resnick, S. M. (2014). Association of hearing impairment with brain

- volume changes in older adults. *NeuroImage*, 90, 84–92. <https://doi.org/10.1016/j.neuroimage.2013.12.059>
- Lin, Frank R., & Albert, M. (2014). Hearing loss and dementia - Who is listening? *Aging and Mental Health*, 18(6), 671–673. <https://doi.org/10.1080/13607863.2014.915924>
- Livingston, G., Huntley, J., Sommerlad, A., Ames, D., Ballard, C., Banerjee, S., Brayne, C., Burns, A., Cohen-Mansfield, J., Cooper, C., Costafreda, S. G., Dias, A., Fox, N., Gitlin, L. N., Howard, R., Kales, H. C., Kivimäki, M., Larson, E. B., Ogunniyi, A., ... Mukadam, N. (2020). Dementia prevention, intervention, and care: 2020 report of the Lancet Commission. *The Lancet*, 396(10248), 413–446. [https://doi.org/10.1016/S0140-6736\(20\)30367-6](https://doi.org/10.1016/S0140-6736(20)30367-6)
- Livingston, G., Sommerlad, A., Orgeta, V., Costafreda, S. G., Huntley, J., Ames, D., Ballard, C., Banerjee, S., Burns, A., Cohen-Mansfield, J., Cooper, C., Fox, N., Gitlin, L. N., Howard, R., Kales, H. C., Larson, E. B., Ritchie, K., Rockwood, K., Sampson, E. L., ... Mukadam, N. (2017). Dementia prevention, intervention, and care. *The Lancet*, 390, 2673–2734. [https://doi.org/10.1016/S0140-6736\(17\)31363-6](https://doi.org/10.1016/S0140-6736(17)31363-6)
- Lunner, T. (2003). Cognitive function in relation to hearing aid use. *International Journal of Audiology*, 42(Sup1), 49–58. <https://doi.org/10.3109/14992020309074624>
- Lunner, T., Alickovic, E., Graversen, C., Ng, E. H. N., Wendt, D., & Keidser, G. (2020). Three New Outcome Measures That Tap Into Cognitive Processes Required for Real-Life Communication. *Ear and Hearing*, 41, 39S-47S. <https://doi.org/10.1097/AUD.0000000000000941>
- Lunner, T., Rudner, M., & Rönnerberg, J. (2009). Cognition and hearing aids. *Scandinavian Journal of Psychology*, 50, 395–403. <https://doi.org/10.1111/j.1467-9450.2009.00742.x>
- Lunner, T., Rudner, M., Rosenbom, T., Ågren, J., & Ng, E. H. N. (2016). Using Speech Recall in Hearing Aid Fitting and Outcome Evaluation Under Ecological Test Conditions. *Ear and Hearing*, 37, 145S-154S. <https://doi.org/10.1097/AUD.0000000000000294>
- Mathôt, S. (2018). Pupillometry: Psychology, Physiology, and Function. *Journal of Cognition*, 1(1), 1–23. <https://doi.org/https://doi.org/10.5334/joc.18>
- 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>
- Miller, A. L., Gross, M. P., & Unsworth, N. (2019). Individual differences in working memory capacity and long-term memory: The influence of intensity of attention to items at encoding as measured by pupil dilation. *Journal of Memory and Language*, 104, 25–42. <https://doi.org/10.1016/j.jml.2018.09.005>
- Moore, B. C. J. (1996). Perceptual Consequences of Cochlear Hearing Loss and their

- Implications for the Design of Hearing Aids. *Ear and Hearing*, 17(2), 133–160.
- Nasreddine, Z. S., Phillips, N. a., Bedirian, V., Charbonneau, S., Whitehead, V., Collin, I., Cummings, J. L., & Chertkow, H. (2005). The Montreal Cognitive Assessment, MoCA: A Brief Screening Tool for Mild Cognitive Impairment. *Journal of the American Geriatrics Society*, 53(4), 695–699. <https://doi.org/10.1111/j.1532-5415.2005.53221.x>
- Neagu, M. B., Dau, T., Hyvärinen, P., Bækgaard, P., Lunner, T., & Wendt, D. (2019). Investigating pupillometry as a reliable measure of individual's listening effort. *Proceedings of the International Symposium on Auditory and Audiological Research*, 7, 365–372.
- Neher, T., Wagener, K. C., & Fischer, R. (2018). Hearing aid noise suppression and working memory function. *International Journal of Audiology*, 57(5), 335–344. <https://doi.org/10.1080/14992027.2017.1423118>
- Ng, E. H. N., & Rönnerberg, J. (2020). Hearing aid experience and background noise affect the robust relationship between working memory and speech recognition in noise. *International Journal of Audiology*, 59(3), 208–218. <https://doi.org/10.1080/14992027.2019.1677951>
- Ng, E. H. N., Rudner, M., Lunner, T., Pedersen, M. S., & Rönnerberg, J. (2013). Effects of noise and working memory capacity on memory processing of speech for hearing-aid users. *International Journal of Audiology*, 52(7), 433–441. <https://doi.org/10.3109/14992027.2013.776181>
- Ng, E. H. N., Rudner, M., Lunner, T., & Rönnerberg, J. (2015). Noise Reduction Improves Memory for Target Language Speech in Competing Native but Not Foreign Language Speech. *Ear and Hearing*, 36(1), 82–91. <https://doi.org/10.1097/AUD.0000000000000080>
- Nielsen, J. B., & Dau, T. (2011). The Danish hearing in noise test. *International Journal of Audiology*, 50(3), 202–208. <https://doi.org/10.3109/14992027.2010.524254>
- Oberauer, K. (2019). Working Memory and Attention – A Conceptual Analysis and Review. *Journal of Cognition*, 2(1), 1–23. <https://doi.org/10.5334/joc.58>
- 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., 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>

- Papesh, M. H., Goldinger, S. D., & Hout, M. C. (2012). Memory strength and specificity revealed by pupillometry. *International Journal of Psychophysiology*, 83(1), 56–64. <https://doi.org/10.1016/j.ijpsycho.2011.10.002>
- Peelle, J. E., Troiani, V., Grossman, M., & Wingfield, A. (2011). Hearing loss in older adults affects neural systems supporting speech comprehension. *Journal of Neuroscience*, 31(35), 12638–12643. <https://doi.org/10.1523/JNEUROSCI.2559-11.2011>
- Petersen, E. B., Lunner, T., Vestergaard, M. D., & Sundewall Thorén, E. (2016). Danish reading span data from 283 hearing-aid users, including a sub-group analysis of their relationship to speech-in-noise performance. *International Journal of Audiology*, 55(4), 254–261. <https://doi.org/10.3109/14992027.2015.1125533>
- Pichora-Fuller, M. K., Kramer, S. E., Eckert, M. A., Edwards, B., Hornsby, B. W. Y., 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, 5S-27S. <https://doi.org/10.1097/AUD.0000000000000312>
- Rönnerberg, 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>
- Rönnerberg, J., Holmer, E., & Rudner, M. (2021). Cognitive Hearing Science: Three Memory Systems, Two Approaches, and the Ease of Language Understanding Model. *Journal of Speech, Language, and Hearing Research*, January. <https://doi.org/10.1044/2020>
- Rönnerberg, 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. *Frontiers in Systems Neuroscience*, 7, 1–17. <https://doi.org/10.3389/fnsys.2013.00031>
- Rönnerberg, J., Lyxell, B., Arlinger, S., & Kinnefors, C. (1989). Visual evoked potentials: Relation to adult speech reading. *Journal of Speech and Hearing Research*, 32, 725–735.
- Rönnerberg, 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, 99S-105S. <https://doi.org/10.1080/14992020802301167>
- Rosen, S., Souza, P., Ekelund, C., & Majeed, A. A. (2013). Listening to speech in a background of other talkers: Effects of talker number and noise vocoding. *The Journal of the Acoustical Society of America*, 133(4), 2431–2443. <https://doi.org/10.1121/1.4794379>

- Rudner, M., Seeto, M., Keidser, G., Johnson, B., & Rönnerberg, J. (2019). Poorer speech reception threshold in noise is associated with lower brain volume in auditory and cognitive processing regions. *Journal of Speech, Language, and Hearing Research*, 62(4S), 1117–1130. https://doi.org/10.1044/2018_JSLHR-H-ASCC7-18-0142
- 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. [https://doi.org/10.1044/1092-4388\(2009/08-0111\)](https://doi.org/10.1044/1092-4388(2009/08-0111))
- Sörqvist, P., & Rönnerberg, J. (2012). Episodic Long-Term Memory of Spoken Discourse Masked by Speech: What Is the Role for Working Memory Capacity? *Journal Of Speech, Language, and Hearing Research*, 55(February), 210–218. [https://doi.org/10.1044/1092-4388\(2011/10-0353\)a](https://doi.org/10.1044/1092-4388(2011/10-0353)a)
- Souza, P., Arehart, K. H., & Neher, T. (2015). Working Memory and Hearing Aid Processing: Literature Findings, Future Directions, and Clinical Applications. *Frontiers in Psychology*, 6, 1–12. <https://doi.org/10.3389/fpsyg.2015.01894>
- Tucci, D. L., Wilson, B. S., & O'Donoghue, G. M. (2017). The growing-and now alarming-burden of hearing loss worldwide. *Otology and Neurotology*, 38(10), 1387–1388. <https://doi.org/10.1097/MAO.0000000000001593>
- Unsworth, N., & Miller, A. L. (2020). Encoding dynamics in free recall: Examining attention allocation with pupillometry. *Memory and Cognition*. <https://doi.org/10.3758/s13421-020-01077-7>
- Unsworth, N., Miller, A. L., & Robison, M. K. (2019). Individual differences in encoding strategies and free recall dynamics. *Quarterly Journal of Experimental Psychology*, 1–14. <https://doi.org/10.1177/1747021819847441>
- Unsworth, N., & Robison, M. K. (2014). Individual differences in the allocation of attention to items in working memory: Evidence from pupillometry. *Psychonomic Bulletin and Review*, 22(3), 757–765. <https://doi.org/10.3758/s13423-014-0747-6>
- Van Den Brink, R. L., Murphy, P. R., & Nieuwenhuis, S. (2016). Pupil diameter tracks lapses of attention. *PLoS ONE*, 11(10), 1–16. <https://doi.org/10.1371/journal.pone.0165274>
- 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>
- Wingfield, A. (2016). Evolution of models of working memory and cognitive resources. *Ear and Hearing*, 37, 35S–43S. <https://doi.org/10.1097/AUD.0000000000000310>
- 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>
- Zekveld, A. A., Koelewijn, T., & Kramer, S. E. (2018). The Pupil Dilation Response to Auditory Stimuli: Current State of Knowledge. *Trends in Hearing*, 22, 1–25. <https://doi.org/10.1177/2331216518777174>
- 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., Kramer, S. E., Rönnerberg, J., & Rudner, M. (2018). In a Concurrent Memory and Auditory Perception Task , the Pupil Dilation Response Is More Sensitive to Memory Load Than to Auditory Stimulus Characteristics. *Ear and Hearing*, 40(2), 272–286. <https://doi.org/10.1097/AUD.0000000000000612>
- Zhang, Y., Lehmann, A., & Deroche, M. (2021). Disentangling listening effort and memory load beyond behavioural evidence: Pupillary response to listening effort during a concurrent memory task. *PLoS ONE*, 16(3), 1–24. <https://doi.org/https://doi.org/10.1371/journal.pone.0233251>

Papers

The papers associated with this thesis have been removed for copyright reasons. For more details about these see:

<https://doi.org/10.3384/9789179291389>

FACULTY OF ARTS AND SCIENCES

Linköping Studies in Arts and Sciences No. 826
Studies from the Swedish Institute for Disability Research No. 103
Department of Behavioural Sciences and Learning (IBL)

Linköping University
SE-581 83 Linköping, Sweden

www.liu.se

SWEDISH INSTITUTE FOR DISABILITY RESEARCH
LINNAEUS CENTRE HEAD GRADUATE SCHOOL