Motivation and fatigue effects in pupillometric measures of listening effort

Defne Alfandari Menase MSc

Thesis submitted to the University of Nottingham

for the degree of Doctor of Philosophy

December 2021

Revised in June 2022

Abstract

Listening effort and fatigue are common among individuals with hearing impairment, although within hearing sciences the relationship between effort and fatigue is not fully understood. The objective measurement of listening effort has commonly involved pupillometry in combination with the speech-in-noise paradigm, where the baseline pupil diameter (BPD) is used as a reference point, and the mean and peak pupil dilation (MPD and PPD) as indices of listening effort. Research to quantify listening effort to date has mostly investigated effort as a function of listening demands. Recent research has shown a correlation between PPD and daily-life fatigue. Furthermore, monetary incentives have been shown to increase the PPD, suggesting that listening demands alone are insufficient to explain variation in PPD. How motivation and fatigue influence the pupil metrics remains unclear.

Frameworks and models of listening effort and listening-related fatigue postulate that larger previous load should result in larger subsequent fatigue. Fatigue manifests as a reluctance to mobilize further effort, particularly when listening conditions are difficult, and when expected rewards are low. In this study we experimentally investigated the interactive influences of task-induced fatigue, motivation, and listening demands on the BPD, MPD, and PPD in the speech-in-noise task. Participants completed a speech-in-noise task of 30-40 minutes without any breaks (“load sequence”) in 3 experiments. Pre- and post- load sequence listening effort was assessed under varying signal-to-noise ratios (SNRs; listening demands), monetary incentives (motivation level; Experiment 1), and magnitudes of load sequence (fatigue level; Experiment 2)

in normal hearing adults. In adults with impaired hearing (Experiment 3), pre- and post-load sequence listening effort was investigated when wearing hearing aids that were described as novel technology (implied reward; high motivation) and conventional technology (no reward; low motivation). In addition to the pupil metrics, need for recovery, self-reported effort, estimated performance, and tendency to quit listening were assessed.

In Experiment 1 and 2, there was a consistent decline in BPD from pre- to post- load sequence, independent of listening demands, monetary incentives, and load sequence magnitude. These results indicate declines in arousal with time-on-task. Replicating earlier research, the MPD scaled with listening demand. Crucially, the pre- to post- load sequence change in PPD was larger with smaller monetary incentives, and larger with larger load sequence magnitude. This result is in line with the conceptualization that larger previous effort results in larger mental fatigue. Furthermore, in a state of fatigue, the mobilization of effort depends on the willingness to perform well. In Experiment 3, we have shown that in adults with impaired hearing, need for recovery and load sequence may interactively influence the BPD and PPD. In all experiments, there was no evidence for a change in speech reception performance, self-reported effort, estimated performance, or tendency to quit listening from pre- to post-load sequence, as a function of load sequence magnitude, monetary incentives, or hearing aid descriptions. This is in line with the understanding that the PPD reveals valuable information that is not captured through traditional speech recognition tests or self-report measures.

Overall, by demonstrating the direct influence of previous load sequence on the PPD, these studies suggest that listening-related fatigue may influence listening effort. Furthermore, this influence may be modulated by motivational factors. Thus, models that explain listening effort and listening-related fatigue need to consider the interactive influence of motivation and fatigue on listening effort.

Acknowledgements

Graham, Dorothea, Lorenz, Michael. Andrew, Pat, David, and all staff at the Hearing Sciences - Scottish section. Renskje, Sarah, Eva, Julia, Soffi, Lena, Antoine, Patrycja and all staff at the Eriksholm Research Centre, part of Oticon A/S. All Innovative Hearing Aid Research – Ecological Conditions and Outcome Measures (HEAR-ECO) Industrial Training Network (ITN) students and supervisors.

The HEAR-ECO 765329 project has received funding from the European Union’s Horizon 2020 research and innovation programme under the Marie-Sklodowska-Curie grant agreement No 765329.

Author's Declaration

This thesis is the result of the author's original research. It has been composed by the author and has not been previously submitted for any other academic qualification.

Defne Alfandari Menase

December 2021

Contents

Chapter 1: Relations between mental fatigue, motivation, and listening effort - a narrative overview

1.1. Introduction

Listening effort has been investigated within hearing sciences for more than 10 years (McGarrigle et al., 2014; Pichora-Fuller et al., 2016). A growing body of research suggests that in daily life individuals with hearing impairment (HI) experience more effort and fatigue as compared to their normal hearing (NH) peers (e.g. Alhanbali, Dawes, Lloyd, & Munro, 2017; Davis et al., 2020; Holman, Drummond, Hughes, & Naylor, 2019; Hornsby & Kipp, 2016). Such hearing-related mental distress can result in increased need for recovery and social isolation (Hétu et al., 1988; Nachtegaal et al., 2009). Adults with HI are known to take more sick leave from work (Kramer et al., 2006). School children with HI are shown to expend more effort in listening and report more fatigue (e.g. Bess & Hornsby, W., Y., 2014; Hicks & Tharpe, 2002). Addressing these issues requires a better understanding of the relations between listening effort and listening-related fatigue (Pichora-Fuller et al., 2016).

Research to quantify listening effort has utilized several self-report, behavioral, and physiological measurement techniques (see McGarrigle et al., 2014; Pichora-Fuller et al., 2016 for reviews). For example, rating scales with questions about effort have been administered after participants engaged in listening tasks (e.g. Luts et al., 2010; Rudner, Lunner, Behrens, Thorén, & Rönnberg, 2012). During the performance of cognitively demanding auditory tasks, increases in primary and secondary task reaction-times (RT; e.g. Gosselin & Gagné, 2010; Houben et al., 2013; Sarampalis et al., 2009) and larger amplitude in the task-evoked pupil dilation responses (e.g. Kuchinsky et al., 2013; Zekveld, Kramer, & Festen, 2010) have been attributed to effortful listening. Changes in alpha and theta band, and in the latency of N1 ERP component as measured by EEG (e.g. Alavash et al., 2018; Bernarding, Strauss, Hannemann, Seidler, & Corona-Strauss, 2013; Obleser, Wöstmann, Hellbernd, Wilsch, & Maess, 2012; Wöstmann, Herrmann, Wilsch, & Obleser, 2015); increased recruitment of attentional and control networks as observed in fMRI and fNIRS (e.g. Rovetti, Goy, Pichora-Fuller, & Russo, 2019; Wild et al.,

2012) have been suggested as markers of listening effort. Whereas traditional audiometry and speech reception tests measure the successful reception of sounds and speech, measures of listening effort are thought to capture the recruitment of resources (effort) that enable successful performance in challenging listening settings (Peelle, 2018).

The bulk of the studies that have quantified listening effort have focused on the relationship between listening demand and listening effort (see Francis & Love, 2020; Mattys, Davis, Bradlow, & Scott, 2012; McGarrigle, Rakusen, & Mattys, 2021 for reviews). Listening demand was manipulated through degradations in the speech sound, such as noise-vocoded speech that varied in envelope modulation depth (e.g. Lawrence, Wiggins, Anderson, Davies-Thompson, & Hartley, 2018), spectral resolution (e.g. Winn, Edwards, & Litovsky, 2016) or low-pass filtering frequency (Pals et al., 2019); different types of background noise that mask the target speech (Koelewijn et al., 2012; Zekveld & Kramer, 2014); varying signal-to-noise ratios (SNR) between the target speech and background noise (Kramer et al., 1997; Krueger et al., 2017); effects of hearing aid (HA) algorithms that are expected to reduce task demand (see Ohlenforst et al., 2017 for a review). Such studies have reported physiological and behavioral correlates of listening demand as indices of listening effort. However, psychological theories conceptualize effort not only as a function of task demand, but also as a function of the willingness of an individual to expend effort (e.g. see Pichora-Fuller et al., 2016; Richter, Gendolla, & Wright, 2016). In studies that measure the effect of task demand on listening effort, such motivation and fatigue effects are usually attempted to be minimized by excluding the initial trials of a session from the analyses, and by including regular breaks in an experimental session for rest.

A few recent studies that aimed to quantify the effect of task demand on listening effort in HI and NH adults refer to motivation and fatigue effects in interpreting their findings (Ohlenforst, Zekveld, Lunner, et al., 2017; Wendt et al., 2018; Wu et al., 2017). For example, Ohlenforst et al (2017) compared peak pupil dilation while HI adults performed a speech-in-noise task across a continuum of SNRs that ranged from +16 to -12 dB in 4-talker babble and in stationary noise. Peak pupil dilation was maximum at the SNRs that were in the middle of the continuum. At very low SNRs, the peak pupil dilation was relatively little (see Lawrence et al., 2018; Zekveld & Kramer, 2014b for similar results in NH listeners). Ohlenforst et al (2017) inferred that at the lower SNRs participants had “given up” listening. That is, participants would not believe that the task was achievable, and, consequently, were not willing to mobilize any effort. This interpretation suggests that in HI adults motivation for listening mediates how much effort is mobilized; however, the mechanisms through which individuals become motivated for listening are unclear.

The relationship between fatigue and listening effort has been suggested by correlational (regression) analyses (Wang et al., 2018). Wang et al. (2018) measured peak pupil dilation in HI and NH adults as they performed a speech reception task that targeted 50% correct sentence performance level. Participants reported their subjective daily-life fatigue through completing the Need for Recovery scale and Checklist Individual Strength questionnaire (Nachtegaal et al., 2009; Vercoulen et al., 1994). A multiple regression analysis showed that factors representing hearing acuity and self-reported daily-life fatigue were equally and independently predictive of the peak pupil dilation while listening. Better hearing acuity and smaller fatigue would predict larger pupil dilation (Wang et al., 2018). This result underlines the relevance of taking fatigue into account when measuring listening effort. However, the mechanisms through which fatigue impact listening effort are not completely clear.

Recent conceptualizations of listening effort acknowledge that the willingness of an individual to mobilize effort for listening plays a role in how much effort is invested (Eckert et al., 2016; Francis & Love, 2020; Matthen, 2016; Pichora-Fuller et al., 2016). For example, the Framework for Understanding Effortful Listening (FUEL) incorporates an adaptation of Kahneman’s (1973) Capacity Model of Attention. In this model, which is central in the framework, it is acknowledged that the available capacity that an individual has for exerting effort is not stable and fluctuates from moment to moment in parallel to fluctuations in arousal[1]. Fatigue and motivation are hypothesized to influence (1) available capacity, and (2) policy for the allocation of capacity. Fatigue is thought to reduce the available capacity for listening. In a fatigued state, an individual is expected to have low arousal and to perceive a listening situation as more challenging than in a rested or recovered state. In addition, fatigue is thought to influence one's own judgement of whether they have enough capacity to meet task demands. For example, when performing a speech-in-noise task in a fatigued state, one may not only be in a state of low arousal, but also judge that they do not have enough capacity to meet the listening demands, and therefore, deliberately give up listening. Motivation is thought to influence both available capacity, and the policy for the allocation of capacity. For example, if one has no interest in listening to a certain conversation (i.e. low motivation), then little capacity is allocated for listening. Conversely, if a conversation is interesting, then one may intentionally upregulate their available capacity and allocate effort for listening.

In the FUEL, recent neuroscientific and psychological frameworks and models of effort and fatigue are considered as complementary to the Capacity Model of Attention (Pichora-Fuller et al., 2016). According to these theories and models, although fundamentally different constructs, motivation and fatigue arise in overlapping brain networks and can mask or boost the observable influence that each alone has on behavior. For example, giving up listening during a speech-in-noise task could be both because of reduced interest (willingness) to perform or an increased state of fatigue. Therefore, arguably, considering both constructs together can prevent hearing scientists from misinterpretations of the influence that each construct has on listening effort. The FUEL paves the way for grounding listening effort in existing theories and

---

[1] In this model, following the observation that available mental capacity fluctuates in parallel to fluctuations of physiological arousal, the term “available capacity” is used to describe the maximum of the mental resources that one has at a given moment. Thus, in the model, as well as in the current review, the terms “exerting capacity”, and “mobilizing effort” are used interchangeably.

models to explain the influence of motivation and fatigue on listening effort, but it does not provide specific quantifiable predictions.

Throughout the last few years, several experimental studies within the hearing sciences have manipulated motivation and fatigue during listening tasks. The results of those experiments reveal the neurocognitive mechanisms by which listening effort is influenced by motivation and fatigue. The aim of this review is to provide an overview of those studies and to set the scene for the next chapters. In section 1.2 and 1.3 the conceptualization of motivation and effort in relevant (neuro)psychological models and theories are briefly and selectively described to give the reader an understanding of the experimental paradigms that are inspired by those. An exhaustive compare and contrast of the models are beyond the scope of this review. After the description of those models, research within hearing sciences on the influence of motivation on listening effort is described. These are interpreted from the perspective of FUEL, where motivation is hypothesized to influence "intentional attention" and the allocation policy of capacity. In section 1.4 prominent conceptualizations of fatigue, and research withing hearing sciences that is inspired by those models are described. The findings of the described studies are interpreted from the perspective of FUEL. That is, with the research questions of how fatigue influences (1) available capacity and (2) the allocation of capacity for listening.

1.2. Motivation as a moderator/mediator of listening effort

The FUEL draws upon the motivation intensity theory (MIT; Brehm & Self, 1989; Silvestrini & Gendolla, 2019) in predicting how much capacity will be allocated for listening at a given moment. The MIT postulates that the investment of effort is governed by the energy conservation principle. That is, individuals seek to avoid spending more resources than are necessary to accomplish a task. Where task demands are known to the individual, the harder the task, the more effort is mobilized in achieving the task. How important it is for an individual to be successful in the task, i.e. success importance, determines potential motivation. Potential motivation dictates the tipping point, where for larger levels of task demand, the outcome is considered to be not worth the extra effort. Thus, the MIT predicts that motivation for listening should influence the mobilization of effort, particularly when listening is challenging.

Within hearing sciences, several studies have experimentally investigated the moderating effect of motivation on the relation between listening demand and listening effort by manipulating the importance of success. Here seven studies that were published in peer-reviewed journals within the discipline of hearing sciences before 1 February 2021 (see Table 1.1.) are described. Picou and Rickets (2014) conducted two experiments using auditory-only and audiovisual stimuli. In both experiments, adults with NH performed a speech-in-noise task in fixed SNRs that would approximate 80% (Easy SNR), and 50% (Hard SNR) correct sentence recognition. In a within-participant design, participants were told to listen to passages carefully (low motivation), and to listen to passages that would be quizzed afterwards for comprehension (high motivation). In the audio-visual experiment, participants reported increased effort when listening to passages that were later to be quizzed as compared to passages that were not going to be quizzed. In the auditory-only experiment, in line with the predictions of the MIT, larger effort was reported with high motivation as compared to low motivation in the Hard SNR condition, but no such relationship was observed in the Easy SNR condition. To note is that participants in the study were not asked to repeat the passages. Hence, the effect of motivation on performance was not measured.

Table 1.1. Studies that experimentally manipulated motivation to investigate its influence on the effort invested for listening

Publication	**Sample**	**Experimental paradigm**	**Effort-related measures**	**Results related to influence of motivation**
Picou & Rickets (2014)	Two samples of 16 NH adults each (mean age = 24.87 and 23.44)	Speech-in-noise with… … 2 levels of SNR … 2 levels of motivation (listening carefully only, or listening carefully for a later comprehension quiz)	Subjective ratings	Motivation leads to increased effort and tiredness, particularly in lower SNR
Richter (2016)	16 (self-reported) NH adults (mean age = 23.88)	Auditory discrimination task with… … 2 levels of task difficulty … 2 levels of monetary reward	Cardiac pre-ejection period	Model with success importance and difficulty most likely predictor of PEP
Koelewijn et al. (2018)	24 adults with NH (median age = 21, range: 18-52)	Speech reception task… …2 levels of task demand (50% and 85% SRT) and quiet control condition … 2 levels of monetary reward	Pupil dilation response Subjective ratings Need for recovery	Larger peak and mean pupil dilation with higher reward, only when listening in noise
Plain et al. (2020)	32 adults with NH (mean age =22, range: 18-40)	Speech in noise test with… … 6 levels of fixed SNRs … 2 levels of monetary reward	Cardiac pre-ejection period Subjective scores Need for recovery	Better performance with reward, and no effect of reward on PEP
Neher et al. (2019)	16 NH and 15 with HL (mean age = 67 and 74, range: 62-75 and 63 – 88 years)	Speech comprehension in cafeteria noise… … 2 levels of task demand (with and without directional HA simulation) … 2 levels of reward (with and without monetary reward)	Electroencephalography impulse responses to speech	Better performance with reward in the low SNR condition only, and for HI participants only. No effect of reward on EEG impulses
Zhang et al. (2019)	35 NH adults (mean age = 22.43, range: 19-32)	Value-based strategic auditory comprehension effort allocation with… …5 levels of reward (motivation) …5 levels of demand (speech rate)	Pupillometry with wavelet analyses Speech comprehension	Interaction effect of reward and task demand on pupil dilation and comprehension accuracy

McLaughlin et al. (2021)	50 NH young adults (mean age = 19.9, range; 18-24) and 50 NH older adults (mean age = 70.9; range: 65 – 79)	Value-based decision-making paradigm with… … speech in speech-shaped noise (SNR: +20, +4,0, -4, -8, -12) … reward of $2 per trial of difficult choice.	Discounting behavior (i.e. the money that is “given up” to avoid effort by choosing the easy task, as an index of the subjective value of the difficult task)	Older adults discounted more than younger adults at difficult SNRs. Among older adults poorer hearing and lower working memory predict selection of lower reward values.

Complementing the results of Picou & Rickets (2014), Richter (2016) and Koelewijn et al. (2018) used physiological measurements to investigate the moderating effect of motivation on the listening demand – listening effort relationship. Richter (2016) measured the pre-ejection period (PEP; the time interval between the onset of ventricular depolarization and the opening of the aortic valve- valid and reliable indicator of beta-adrenergic influences on the heart) in a sample of normal hearing adults. In a within-participant design, participants performed an auditory tone discrimination task that was either easy or difficult. For both difficulty conditions, participants were promised either high or low monetary reward that was conditional upon their performance. PEP was the smallest, suggesting largest effort, in the difficult-high reward condition as compared to the other conditions. In explaining the data, a model that included both success importance and task demand was deemed more likely as compared to a model that only included task difficulty. In a similar within-participant experimental design of 2 (task difficulty) x 2 (monetary reward), Koelewijn (2018) measured the mean and peak pupil dilation response during a speech reception task. Similarly to the effects observed by Richter (2016), on average, mean and peak pupil dilation were largest in the high-difficulty-high-reward condition. Although the reward x difficulty interaction effect was not significant in any of the pupil metrics, there was a significant main effect of reward on the peak pupil dilation (Koelewijn et al., 2018). These results suggest that particularly in challenging listening conditions, the willingness to perform well determines the investment of listening effort.

Contrary to the findings of Koelewijn et al. (2018), where effects of reward on physiological outcome measures are observed in the absence of changes in speech reception performance, Neher (2019) and Plein (2020) report effects of monetary reward on speech reception and comprehension performance in the absence of physiological correlates. The discrepancies in the findings can be attributed to the differences in the study paradigms and associated performance strategies. Koelewijn et al. (2018) presented everyday speech sentences and measured pupil dilation during a 3- second retention interval, whereas Plein et al. (2020) used a 1-talker masker and included a retention interval of only 0.5 seconds. Whereas successful performance in Koelewijn et al. (2018) would require the use of working memory to sustain representations of heard sentences during the retention phase, successful performance in Plein et al (2020) may be driven by an increased reliance in guessing during the response stage. Neher (2019) recorded effort during listening and measured comprehension performance after listening. Participants with NH and HI listened to 1-minute-long audiobook recordings in cafeteria noise. In half of the trials, participants were promised monetary reward upon correctly answering comprehension questions. The cross-correlation pattern between the speech envelope and the neural oscillations (measured by EEG) that are known to synchronize to the attended speech was measured. Although reward led to increased performance, particularly in the low SNR condition, and particularly for the HI group, there was no reward-related change in the EEG responses. One explanation for this finding may be that the increase in effort occurred during the response stage of the task rather than during listening. Together, these studies illustrate the diversity in the paradigms that aimed at measuring the effect of reward on listening effort. The reviewed studies point at the importance of alignment between the interval where physiological measures are recorded and the interval where reward-related effort is expected.

In sum, although scarce, the reviewed evidence suggest that the effort invested in listening, particularly in challenging listening conditions, may be moderated by the importance of performing well. Future fMRI and fNIR studies could reveal which networks are implicated in increasing the effort for listening. To note is that except one (Picou & Ricketts, 2014), all reviewed studies have used extrinsic rewards to manipulate motivation. These reliably show that by wanting to perform better, listeners were able to deliberately invest additional effort. The extent to which intrinsic rewards (e.g. an interesting conversation) moderate listening effort remains unexplored.

1.3. Cost/benefit calculations as underpinnings of motivation

Recent neurocognitive research on cognitive control and motivation sheds light on the mechanisms underlying the energization of behavior by reward (see Botvinick & Braver, 2015 for a review). In the context of such research, cognitive control could be described as the set of higher cognitive functions that encode and maintain representations of goals, and assemble subordinate functions such as working memory, semantic memory, episodic memory; perceptual, motor, and attentional selection and inhibition. Motivation has been defined as the "biasing and energizing effect of prospective rewards on behavior and cognition" (Botvinick & Braver, 2015). Processes that demand cognitive control are thought to be costly and effortful. They are often distinguished from habitual and automatic processes that are considered to be effortless (e.g. Norman & Shallice, 1986; Posner & Snyder, 1975; Shriffin & Schneider, 1977).

Models of language comprehension describe listening to speech as an automatic process for people with NH in ideal conditions, and postulate that controlled processes are recruited when such automaticity fails, for example when listening with HI (e.g. Lunner et al., 2009; Stenfelt & Rönnberg, 2009; Rönnberg et al., 2021; Shinn-Cunningham & Best, 2008). Indeed, neuroimaging studies show increased recruitment of control networks in HI listeners during speech reception in challenging conditions (see Eckert et al., 2016 for a review). Cognitive control models postulate that the amount of control demanded by a task (i.e. cost), and the expected reward from the task (i.e. benefit) are integrated to determine which task to invest control in, and how much the investment should be. Effort is then conceptualized as the subjective cost of implementing control (e.g. Shenhav et al., 2013; see

Silvestrini, 2017 for an integration of the MIT with models of cognitive control). Thus, when speech reception is not automatic, the investment of effort should depend on the subjective value of the expected outcome. Note that value includes a combination of factors such as the probability of attaining the award, the cost of attaining the award, the context in which the reward is presented etc.

Two recent studies have used decision-making paradigms to investigate the subjective value of listening. In the study of Zhang et al (2019) a value-based strategic effort allocation paradigm was used together with pupillometry. Participants listened to sentences in combinations of 5 difficulty (in speech rate) and 5 reward levels (in points). The sentences included reasoning questions about the spatial relationship between two or more objects (e.g. "*The kite is in front of the stone, the stone is in front of the ball, where is the kite in relation to the stone?*"). After listening, participants could choose to either answer the question, or skip the trial. Crucially, participants were told that a minimum number of points need to be collected to attain the full monetary reward and the time they could spend on the experiment was limited. Before the beginning of each sentence, an auditory cue indicated the difficulty and reward level of the trial. All trials except wrongly answered ones and ones with no answer were included in the analyses. Zhang et al (2019) report an interaction effect of reward and task difficulty on both the pupil dilation during the interval where participants prepared to respond, and on the performance accuracy. For trials assigned to low rewards, pupil dilation did not differ depending on difficulty. For trials assigned to high rewards, pupil dilation decreased with increasing difficulty. Furthermore, reward led to better performance for difficult trials, but not for easy trials. This pattern of results provides evidence that young adults with normal hearing strategically allocate effort to information that is expected to have the largest value. It is also in line with the understanding that success importance determines the investment of effort, particularly at difficult listening conditions.

The findings of McLaughlin et al (2021) support the understanding that listening demand and reward both contribute towards the decision for expending effort. Participants in the study of McLaughlin et al (2021) performed a value-based decision-making task, where they chose between listening to difficult trials that were paired with high monetary reward and easy trials that were paired with low monetary reward. After selecting their preferred option, participants listened to sentences in noise and repeated what they heard. Importantly, participants would get the monetary reward assigned to the trial independently of their speech recognition performance. The easy alternative was always presented at +12 dB SNR and would be paired with a different SNR (+4, 0, -4, -8) in each block. Within blocks, the monetary reward for choosing the difficult alternative was always the same, while the reward for choosing the easy alternative would change from trial to trial. The aim of this titration process was to determine how much money an individual was willing to give up to avoid effort. On average, participants chose the low-reward-Easy-SNR option more often as the difficulty of the alternative option increased. The decisions taken by the participants in the experiment reveal that the value of exerting effort diminishes with increasing task difficulty. Particularly when listening demands are high, individuals exert listening effort only when the expected rewards are sufficient.

1.4. Effects of task-induced fatigue on listening effort

Fatigue, the subjective mental state of exhaustion, is classically thought to arise with effortful exertion over time and to decrease after some rest or recovery (Boksem & Tops, 2008). Recent neurocognitive accounts consider fatigue as arising in the same brain networks that implement the selection of task-relevant stimuli, that motivate the pursuit of rewards, and integrate the outcomes of such behaviors with interoceptive information (Boksem & Tops, 2008; see Müller & Apps, 2018 for a review of neurocognitive underpinnings of motivational fatigue). In the context of adaptive control, mental fatigue has been considered as a cost signal that increases over time and that directs one towards the selection of less costly (automatic) behaviors (see Eckert et al., 2016 for a neurobiologically plausible explanation of listening-related fatigue; Hockey, 2011; Inzlicht et al., 2014; Schneider et al., 2019).

Psychological studies of fatigue traditionally have focused on decrements in the speed, accuracy, and force of behaviors over time as observable manifestations of fatigue (e.g. Boksem et al., 2005; Tanaka et al., 2014). These time-on-task effects have been shown to depend not only on the time spent on task, but also on the difficulty of the task. That is, in tasks with greater demand, performance declines are observed to occur more rapidly as compared to tasks with less demand (e.g. Nuechterlein et al., 1982; Earle et al., 2015). Furthermore, the observation that performing one cognitively demanding task can lead to performance decrements in another cognitively demanding task led to the argument that fatigue arises in domain-general cognitive control networks (e.g. Inzlicht and Schmeichel, 2012; Shenhav et al., 2017).

Within the hearing sciences numerous studies have induced listening-related fatigue in the laboratory and measured the resulting changes in available capacity and allocated effort through dual-task RTs (Hornsby et al., 2013) pupillometry (McGarrigle et al., 2017; McGarrigle et al., 2017b) and EEG (Key et al., 2017; Gustafson et al., 2018; Antons et al., 2012; Moore et al., 2017; see Table 1.2. for a summary). Participants in the study of Hornsby et al (2013) engaged in a dual-task speech-in-noise paradigm for approximately 1 hour. The task included word recognition (primary task), word recall (primary task), and visual RTs (secondary task). Experimental sessions were scheduled at the end of the working day of the participants. During their working day participants wore HAs in advanced settings, HAs in basic settings, or were unaided. The same task demand conditions were kept throughout the experimental sessions. Participants reported increased fatigue after the experiment as compared to before, independently of the condition, suggesting that the fatigue manipulation was successful. Word recognition and recall stayed stable over time in all conditions. Secondary task RTs stayed stable in the aided conditions, but showed a gradual increase over time in the unaided condition. No differences between the basic and advanced hearing aid conditions were observed in the changes of secondary task RT with time-on-task. This result could be attributed to the background noise in the experiment (fixed at 55dBA), which may not have been high enough to activate the advanced (i.e. digital noise reduction) settings of hearing aids. Overall, this study demonstrates that by reducing task demands hearing aids reduce the capacity-limiting effects of fatigue. While how the magnitude of fatigue relates to the magnitude of observable limitations remains to be further investigated by future research, the results are in line with the notion that fatigue limits the capacity to exert effort.

Table 1.2. Studies that experimentally investigated the influence of fatigue on the effort mobilized for listening

Publication	**Sample**	**Experimental paradigm**	**Effort and fatigue measures**	**Results related to motivation and/or fatigue**
Hornsby et al. (2013)	16 adults with HL (mean age =65.8, range: 47 – 69)	Dual-task involving word recognition in noise, word recall, and visual reaction time with… …3 levels of task demand (unaided, basic aided, and advanced aided)	Behavioral (speech recognition, recall, visual reaction time) Subjective ratings	Increased self-reported fatigue from pre- to post- in all conditions. Increased RT over time-on-task in unaided condition and stable RT over time-on-task in both aided conditions
McGarrigle et al. (2017a)	24 NH adults (age range: 18-30)	Speech-picture verification task with (i.e. continuous demanding block to induce fatigue) with… …2 levels of demand (15 dB and -8 dB SNR)	Pupillometry with growth curve analyses Response time	Steeper within-trial decrease in pupil size in the difficult condition, particularly in the second half of the session No change in RT with time-on-task; self-reported fatigue did not depend on task demand
McGarrigle et al. (2017b)	41 NH children (age range: 8-11)	Speech-picture verification task with… … 2 levels of task demand (+15dB SNR and -2dB SNR)	Pupillometry with growth curve analyses Response time	No significant effect of time-on task of self-reported fatigue, nor on pupil metrics
Key at al. (2017)	27 NH children (mean age= 9.28, range: 6-12.9)	Auditory oddball with speech syllables in noise… … pre- and post-sustained speech processing (i.e. pre- and post- fatigue)	Evet-related potentials (P1, N1, N2), performance accuracy, psychomotor vigilance, subjective ratings	Increased lapses of attention, longer reaction times, reduced P300 amplitudes to oddballs, and greater self-reported fatigue during post- as compared to pre-

Gustafson et al. (2018)	34 children with HI (mean age= 10.03, range: 6-12)	Auditory oddball with speech syllables in noise… … pre- and post-sustained speech processing (i.e. pre- and post- fatigue)	Event-related potentials (P1, N1, N2), performance, and subjective ratings	Increased lapses of attention, longer reaction times, reduced P300 amplitudes to oddballs post- as compared to pre-
Antons et al., 2012	18 NH adults (mean age = 25.56, range = 21-31)	Listening followed by comprehension questions with… … 2 levels of task demand (recordings in different bit rate)	EEG recordings Subjective ratings	Increase of theta band, and earlier increase of alpha band activity with time-on-task for the high demand as compared to low demand condition
Moore et al. (2017)	19 NH adults	Auditory choice paradigm (i.e. continuous demanding block to induce fatigue)	Event-related potentials (ERN, N1), electrical neuroimaging, response time, accuracy, subjective ratings	Increased subjective fatigue, decreased subjective motivation in second half; trending decrease in ERN (conflict monitoring); decrease in N1 amplitude (arousal) from first to second half

The operationalization of mental fatigue as the changes in effort and performance that arise with time-on-task does not explicitly demonstrate the domain-general nature of fatigue. An alternative approach, the so-called “indirect” study of fatigue, is the induction and measurement of fatigue over distinct tasks, where domain-general fatigue effects are expected to transfer from one task to another (Ackerman, 2011; Hornsby et al., 2016). Another advantage of the indirect approach is that it allows ruling out boredom and habituation effects, which are likely to arise with time spent on the same task (Ackerman, 2011). Two studies have investigated the effect of fatigue on subsequent effort mobilized for listening using the indirect approach (Key et al., 2017; Gustafson et al., 2017). To induce listening-related fatigue in the lab, Key et al. (2017) asked NH children to engage in a series of three demanding speech-processing tasks concurrently with only short breaks. Both before and after engaging in these tasks, participants completed a self-report fatigue scale, a visual vigilance test, and an auditory oddball paradigm. The oddball paradigm consisted of speech syllables that were presented in noise. Participants were instructed to listen and notice the deviant syllable. Participants reported increased fatigue after completing the speech-processing tasks, compared to before them. In addition, there was a decrease in the amplitude of the parietal P300 to targets, suggesting reduced cognitive processing of the target (e.g. Murata et al., 2005). Furthermore, lower ERP amplitudes of parietal P300 were associated with more attentional lapses and longer reaction times in the vigilance task. An experiment with a similar methodology was conducted with children with hearing loss (Gustafson et al., 2017). Consistent with the reports of the effects of fatigue in NH children, Gustafson et al. (2017) report reduced parietal P300, increased RT, and more attentional lapses from pre- to post-fatigue in HI children. The pre- to post-fatigue difference in self-reported fatigue, however, did not reach statistical significance in HI children. The discrepancy in the results may be because for children with hearing loss, the experience of fatigue may be more ubiquitous than for children without hearing loss. Another reason may be that children with hearing loss experienced fatigue earlier in the fatigue-inducing task, and switched to alternative strategies, or withdrew cognitive effort during the induction task itself.

McGarrigle et al. (2017a) investigated listening-related fatigue using behavioral, self-report, and physiological measures in a sample of NH adults. Participants listened to ~12-second long passages of speech in multitalker babble-noise and judged whether pictures matched the content of the passages. Passages were presented in blocks of +15 and -8 dB SNR. Although participants reported more effort in the hard condition as compared to the easy, there was no effect of task demand or block order on self-reported fatigue. RTs were longer with more task demand but did not depend on block order. Growth curve analysis revealed a steeper within-trial decline in pupil size in the second block of the session with larger task demand. This is attributed to decreased arousal, an indicator of fatigue. A similar experiment investigated listening-related fatigue in children with NH (McGarrigle et al., 2017b). Children performed a speech-picture verification task in hard and easy conditions (-2 and +15 dB SNR, respectively) that were distributed in blocked fashion. The authors hypothesized a steeper decline in pupil response over the course of the experiment, as well as larger increase in RTs to the visual images with time on task in the difficult condition as compared to the easy condition. Furthermore, larger self-report of fatigue was hypothesized in the difficult condition as compared to the easy one. However, analyses showed no differences between the easy and difficult conditions in terms of RT, pupil response, or self-report. The discrepancy between the results of McGarrigle et al 2017a and 2017b may be because of shorter task duration and more positive SNR in the Hard condition 2017b as compared to 2017a, which may have limited the induction of fatigue.

Two studies used electroencephalogram (EEG) recordings to investigate the effect of sustained auditory processing on subsequent fatigue (Moore et al., 2017; Antons et al., 2012). Moore et al (2017) asked normal hearing adults to determine whether trains of three pure tones contained a specific stimulus combination, continually for around 50 minutes. Participants responded by pressing keyboard buttons to answer. No feedback was given to their responses. During the second half of the experiment, participants expressed increased fatigue and decreased motivation to continue the task. Although not statistically significant, visual inspection showed a decrease in error-related negativity (ERN) amplitude. The amplitude of N1 to the first stimulus of the train showed a significantly decreased amplitude in the second block of the experiment as compared to the first block. Furthermore, electrical neuroimaging analyses with EEG recordings showed a significantly decreased response strength in auditory and dorsal attention areas. Participants made more mistakes in the second half of the experiment as compared to the first. Together these results suggest that participants were aware of their errors, but were not willing or not able to increase effort for corrective adjustments. Antons et al. (2012) showed higher self-report of mental fatigue after listening to degraded audio compared with intact audio in NH adults. Participants listened to degraded audio passages and later answered comprehension questions. The theta alpha band activity (8 – 10 Hz) of the EEG was greater during the last 10 minutes of listening as compared with the first 10 minutes.

In sum, these findings are in line with the notion that fatigue limits the capacity to exert further effort, although it is less clear under which conditions the capacity limitations are overcome by more effort, and under which conditions listeners give up. Thus, the conditions under which fatigue effects manifest differently remain to be investigated.

1.5. Subjective evaluation of demands on capacity

In the FUEL, fatigue is thought to influence the evaluation of demands on capacity, which in turn is thought to influence both capacity and the allocation of capacity for possible activities (Pichora-Fuller et al., 2016). Central in many models of motivation is the flexible adaptation of effort (capacity) depending on the judgements of one's ability to reach goals and on the outcomes of one's pursuit to attain goals (e.g. Silvestrini et al., 2017; Shenhav et al., 2013; Kahneman, 1973). Such conceptualization is supported by neuroscientific evidence of brain structures that integrate such information to adaptively (flexibly) drive the implementation of control and its concomitant effort-related autonomic activity (Eckert et al., 2016; Silvestrini et al., 2017; Gianaros et al., 2005). Thus, adjusting the capacity allocated for listening should be dependent on one's subjective evaluation of listening demands and listening goals.

Within hearing sciences two studies show effects of subjective capacity judgements on effort allocated for listening. Ayasse and Wingfield (2020) measured the baseline pupil dilation throughout 160 trials of an auditory sentence comprehension task. The participants included adults with NH and with mild-to-moderate HI. Throughout the experiment baseline pupil size declined, whereas performance increased. Furthermore, participants with larger HI had larger baseline pupil size at the beginning of the experiment, and a larger decline in pupil size during the initial trials of the experiment. This pattern of results was attributed to anticipatory recruitment of capacity at the beginning of the session, perhaps related to a lack of confidence or anxiety over the task.

Although participants performing a speech-in-noise task are known to have accurate judgements of their performance (e.g. Koelewijn et al., 2018), receiving explicit feedback from the experimenter has been shown to affect the capacity allocated for listening (Zekveld et al., 2019). Zekveld et al. (2019) provided NH and HI participants with explicit feedback throughout a speech-in-noise task. The feedback included a written information after each trial, in addition to a target performance level expected from the participant, and verbal nudges that aimed to pressure participants to try harder. Regardless of hearing status, such extensive feedback was shown to increase the peak pupil dilation and increase speech reception performance. These results are in line with the theorizing that performance outcomes (e.g. success in speech reception, or rewards) and capacity judgements are integrated to flexibly determine the allocation of capacity during listening (Pichora-Fuller et al., 2016; Muller & Apps 2019; Scheider et al., 2018)

1.6. Conclusions

In sum, the research discussed here suggests that the effort mobilized for listening with HI is influenced by fatigue and motivation through multiple mechanisms. Particularly in demanding listening conditions, the importance of success and subjective cost/benefit evaluations are shown to influence whether and how much effort is allocated for listening. Fatigue is suggested to reduce the capacity to expend effort on listening, and the allocation of effort for listening. Lastly, the evidence, although scarce, suggests that effort is flexibly adjusted depending on one’s subjective evaluation of their capacity to meet their target listening goals.

1.7. Remaining questions

The research reviewed above suggests that listening-related fatigue should reduce the capacity allocated for listening, however it is unclear whether the reduction in the allocated capacity is related to the increased perception of task demands or decreased motivation to achieve listening goals, or both.

Some of the remaining questions that arise out of the review are as follows.

1. How does listening-related fatigue impact available capacity and the allocation of capacity for listening?
2. How does listening-related fatigue manifest in varying listening demands and varying motivational states?

In Experiments 1 and 2 (Chapters 2 and 3) the interactive effects of mental fatigue and motivation on listening effort are investigated in NH adults. In Experiment 1, listening demands, listening-related fatigue, and motivation are manipulated. Their interactive influence on available capacity and allocation of capacity is investigated. In Experiment 2, the influence of different levels of mental fatigue and motivation on available capacity and allocation of capacity are examined. In Experiment 3, the same research question is investigated in a sample of adults with HI. Here motivation is manipulated by describing hearing aids as novel technology (intrinsic reward).

1.8. Pupil size and dilation as indices of capacity and effort

In all experiments described in the following chapters, pupillometric measures are used as indices of available capacity and the allocation of capacity for listening. Pupillometry, the continuous recording of the pupil size, is perhaps one of the most thoroughly used measures of listening effort (Naylor et al., 2018). As compared to other physiological measures, modern eye tracking glasses offer a cheaper alternative for professionals and are fairly comfortable to wear for patients. Pupil size has long been considered as an index of arousal. Arousal, which can be defined as “a period of heightened sensory responsiveness and perception that involve autonomic and endocrine activation” (Larsen & Waters, 2018) is thought to primarily be mediated by the activity of subcortical brain structures (see Larsen & Waters, 2018; van den Brink, Pfeffer, & Donner, 2019 for reviews). Changes in neuromodulation are thought to drive the cortical neuron responsiveness and the signal-to-noise ratio of responses to sensory stimuli (i.e. neural gain). Intermediate levels of arousal, where the pupil is dilated 40-60 % of its maximum, are commonly observed to elicit the best performance in auditory perceptual tasks (McGinley et al., 2015; Larsen & Waters, 2018). Although the locus coerulus-noreprinephine (LC-NE) system is most classically thought to modulate pupil-linked arousal (pupillary dilation), dopaminergic neural circuits have recently been shown to mediate pupil dilation in anticipation of rewards (e.g. Drew et al., 2020; Muhammed et al., 2016; Schneider et al., 2020).

The task-evoked pupil dilation has widely been used as an index of cognitive effort and an indirect measure of brain state (Beatty, 1982; Hess & Polt, 1964; van der Wel & Steenbergen, 2018; Larsen & Waters, 2018). In hearing sciences, pupil size has commonly been combined with the speech-in-noise paradigm, where participants listen to everyday sentences in noise and are asked to repeat the sentences to the best of their ability. Whereas the baseline pupil diameter (BPD), which is commonly measured during the 1 second before the presentation of the sentences, is thought to index anticipatory arousal, the baseline-corrected task-evoked pupil dilation (TEPD) that is coupled to the onset of the sentences is thought to index the effort invested in listening (Kramer et al., 1997; Winn et al., 2018). In all experiments in the next chapters, in addition to pupillometric measures, self-report measures of effort, perceived performance and tendency to quit were collected. Correlations between pupillometric and self-report outcomes were computed to investigate how pupillometric measures relate to self-report outcomes.

Chapter 2: Impact of listening-related fatigue and motivation on the relationship between signal-to-noise ratio and listening effort as indexed by the baseline and mean pupil dilation in adults with normal hearing

2.1. Introduction

As described in Chapter 1, most previous research to quantify listening effort has focused on the relationship between listening demand and listening effort (see McGarrigle et al., 2014). The studies that have investigated the moderating effect of motivation and fatigue on listening effort focused either only on motivation (e.g. Koelewijn, Zekveld, Lunner, & Kramer, 2018), or only on fatigue (e.g. Hornsby, 2013), thereby neglecting the interactive effects of both on listening effort (see Chapter 1 for a review). Although fundamentally different constructs, fatigue and motivation are known to arise in overlapping brain networks and to influence the observable influence that each has alone on behavior (Boksem & Tops, 2008; Hockey, 1997a, 2010; Hopstaken et al., 2015a; Müller & Apps, 2019). In this chapter, pupillometry is used to investigate how fatigue and motivation interactively influence the capacity for and investment of effort for listening.

In auditory perception tasks, monetary incentives have previously been shown to elicit larger TEPD, particularly when task demands are high (see Zekveld, Koelewijn, & Kramer, 2018 for a review; Bijleveld, Custers, & Aarts, 2009; Knapen et al., 2016; Koelewijn, Zekveld, Lunner, & Kramer, 2018; although see Stanners, Coulter, Sweet, & Murphy, 1979 for no effect of reward on the TEPD). For example, Koelewijn et al. (2018) recorded pupil size while normal-hearing adults engaged in an adaptive speech-in-noise task tracking 50% and 85% correct sentence repetition. In addition, all participants completed a speech reception task in quiet. Either 5 euros (high reward) or 0.20 euro (low reward) was offered per block of 25 trials to the participants on the condition that they repeated 70% or more of the sentences correctly. The peak of the task-evoked pupil dilation (i.e. peak pupil dilation; PPD) was larger for high than for low reward when perceiving speech in noise across the 50% and 85 % SRTs, but not in quiet.

In auditory tasks time-on-task has previously been associated with decreases in baseline and TEPRs (Ayasse & Wingfield, 2020; see Zekveld, Koelewijn, & Kramer, 2018 for a review). For example, Zekveld, Kramer, & Festen, 2010 report declines in both baseline and mean pupil diameter throughout the first 5-to-45 trials of a speech-in-noise task. Given that this decline was mainly in the first block in the session Zekveld et al. (2010) attributed this to habituation. Recently, over the course of an approximately 40 min. long speech-picture verification task, the TEPR has been shown to have a larger decline during the second half of the experiment, particularly when listening in lower SNRs (McGarrigle et al., 2017b). In parallel to the pupil results, self-rated fatigue is reported to increase in the second half of the experiment, particularly in the difficult SNRs (McGarrigle et al., 2017b). Thus, in line with the view that fatigue reduces allocation of capacity (Kahneman, 1973; Pichora-Fuller et al., 2016) the decline in the TEPR has been interpreted to reflect fatigue (McGarrigle et al., 2017b). Whether and how fatigue changes the listening effort remains unknown.

A few studies have used pupillometry to investigate the interaction between fatigue and motivation in demanding visual tasks (Gergelyfi et al., 2015; Herlambang et al., 2019, 2021; Hopstaken et al., 2015b, 2015a). These mostly (although inconsistently) show decreasing TEPR with time-on-task. When rewards are introduced that are conditional upon performance, TEPR has been shown to recover to the same level as in the beginning of the sessions (e.g. Hopstaken, van der Linden, Bakker, & Kompier, 2015). In sum, existing pupillometry research suggests that fatigue and motivation may interactively determine the TEPR in demanding tasks.

A previous study investigated the effect of daily life fatigue as measured by the Need for Recovery Scale, an 11 item scale that is an operationalization of early symptoms of work-related fatigue, (NFR; Van Veldhoven & Broersen, 2003) on the TEPR in a speech-in-noise task with HI and NH adults (Wang et al., 2018). The authors report an association between larger TEPR with smaller NFR. However, this relationship hasn’t been observed consistently (Koelewijn, Zekveld, Lunner, & Kramer, 2018). Thus, further investigation of the relationship between NFR and TEPR is warranted.

Although lacks of relations between self-reported measures and pupil dilation measures of listening effort have been commonly reported (e.g. Koelewijn, Zekveld, Festen, & Kramer, 2012; McGarrigle, Dawes, Stewart, Kuchinsky, & Munro, 2017; Pichora-Fuller et al., 2016; Strand, Brown, Merchant, Brown, & Smith, 2018), a recent study reports a relationship between pupil and self-report measures of tiredness (McGarrigle, Rakusen, & Mattys, 2021). While most research investigated between-subject correlations in the outcome measures, McGarrigle et al. (2021) report a correlation using a within-subject repeated measures of effort. In the current experiment thus the relationship between objective and subjective outcomes was investigated in a within-individual approach.

2.1.1. Aims and hypotheses

The current experiment aimed to investigate the interactive effects of listening-related fatigue and motivation on listening effort. To this end, listening-related fatigue was induced in a sample of NH adults. Both pre- and post- fatigue induction, participants performed a speech-in-noise task with Easy and Hard conditions, where the SNR was individually pre-determined and fixed at approximately 80% and 50% sentence comprehension, respectively. To manipulate motivation, participants were offered High and Low monetary incentives for correct sentence recognition. Following theories of motivation and previous research on listening effort (Pichora-Fuller et al., 2016) larger pupil dilation in the Hard SNR condition as compared to the Easy SNR condition was hypothesized (Hypothesis 1). Larger effect of the monetary incentives on the pupil dilation was hypothesized in the Hard SNR condition as compared to the Easy SNR condition (Hypothesis 2). In line with previous research (Zekveld et al., 2018) diminishing BPD after the fatigue induction was hypothesized (Hypothesis 3). Lastly, according to the motivational models of (listening-related) fatigue (Hockey, 2010; Müller & Apps, 2019; Pichora-Fuller et al., 2016; Schneider, Bernarding, Francis, Hornsby, & Strauss, 2019), after the fatigue induction and when task demands are high, participants should only mobilize effort when it feels worth doing so. That is, the difference in pupil dilation between High and Low monetary incentive conditions in the Hard SNR condition should be larger post-fatigue as compared to pre-fatigue (Hypothesis 4).

2.2. Methods

2.2.1. Experimental design

A pre-/post- fatigue experiment with a within-subject design was used to investigate the interactive effects of listening-related fatigue and motivation on listening effort. Figure 2.1 illustrates the experimental design. A challenging sustained speech reception task (i.e. “load sequence”) was used to induce fatigue. Listening effort was evaluated pre- and post-load sequence using pupillometry and self-report. Both pre- and post- load sequence, 4 blocks (i.e. 2 x monetary incentive [high and low amount] x task demand [easy and hard SNR] of 20 trials each were administered. Per participant, the order of the pre-load sequence blocks was the same as the post-load sequence blocks. Between participants, the order of the pre- and post- load sequence blocks was counterbalanced using a Latin square method. After the last pre-load sequence block, participants took 5 minutes break to recover from the possibly fatiguing effects of the pre-load sequence blocks.

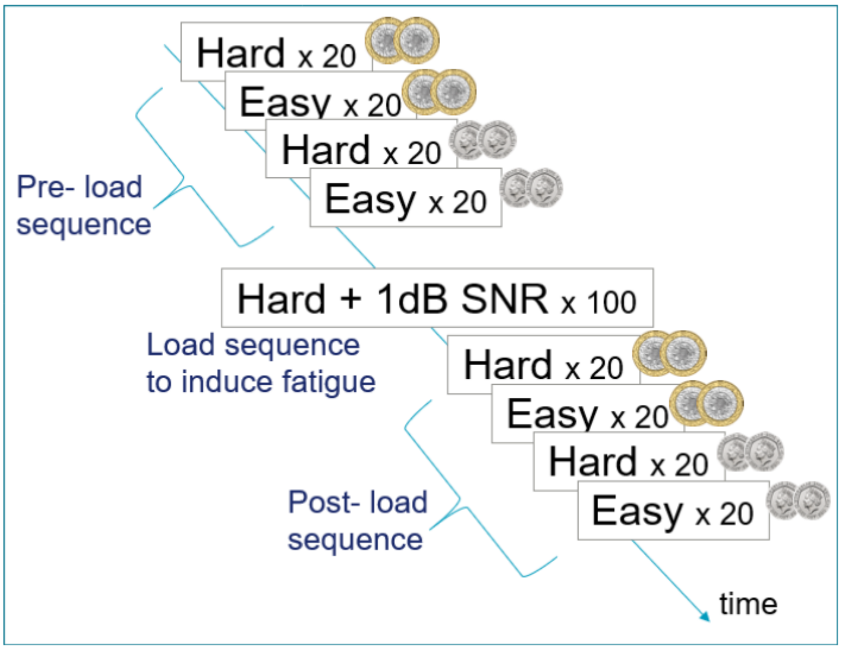


Figure2.1. Experiment design. To induce fatigue, participants were asked to engage in a 100-trial long speech-in-noise task ("load sequence"). Both before and after the load sequence, participants completed 4 blocks of speech-in-noise tasks in 2 levels of task-demand and 2 levels of monetary incentives.

Previous research shows that at the SNR where sentence perception performance is around 80%, speech reception is reported as easy and elicits relatively small pupil dilation (Wendt et al., 2018). At the SNR where sentence perception accuracy is around 50% correct, speech reception is reported as challenging and elicits the largest pupil dilation as compared to the SNR where performance is poorer or better (Wendt et al., 2018). To investigate the influence of fatigue and motivation on easy and difficult listening conditions, the SNR during the pre- and post-load sequence blocks were pre-determined and fixed at the SNRs that would correspond to 50% and 80% correct speech reception thresholds of each participant (i.e. SRT50 and SRT80).

The SRT50 and SRT80 were calculated based on a sentence-based adaptive estimation procedure (described in detail below). In both conditions the intensity of the 4-talker background babble noise masker was fixed at 70dB while the intensity of the target sentences was fixed at the estimated dB.

Hereafter the SNRs that are individually fixed at SRT80 and SRT50 will be referred to as the Easy SNR and Hard SNR conditions, respectively.

The SNR during the load sequence (fatigue induction) was individually set at 1 dB above the SNR of the Hard SNR condition. This was to set the load sequence to be both demanding enough to lead participants to experience fatigue and achievable enough to prevent participants from giving up.

The monetary incentives during the pre- and post-load sequence blocks were approximately based on the amounts that successfully motivated participants in a previous study (Koelewijn et al., 2018). Per block of 20 trials, participants could earn either £4 or 40 pence by repeating 14 (i.e. 70%) or more of the presented sentences correctly (cf. Koelewijn et al., 2018). Ethical guidelines, which suggest that participation in the study should not be incentivized by money, prevented the offer of larger monetary amounts. The incentive were offered per block (that is, not per trial) to keep participants motivated for at least 14 trials, as approximately 20 trials are known to be required for observing effort-related effects in pupil dilation responses (Winn et al., 2018). The threshold for earning the reward wasn’t set for a larger number of trials than 14, because earning the reward needed to be perceived as achievable by the participants. Hereafter the conditions with £4 and 40 pence reward offers will be referred to as the High and Low monetary incentive conditions, respectively.

In this study the load sequence was only used to induce fatigue. Investigation of the pupil size during the load sequence (see Appendices 1 and 2 for plots of BPD and MPD during the load sequence) was not part of the research questions. The load sequence was assumed to have induced mental fatigue and was thus not further investigated.

2.2.2. Participants

The data of 30 volunteers (age between 28 – 72 years, *M* = 54.67, *SD* = 10.91) are reported here. All participants had normal hearing (i.e. hearing thresholds were less than or equal to 25 dB HL over the frequencies from 0.25 to 4 kHz for both ears). Participants reported no neurological or psychiatric disorders and normal or corrected-to normal vision. All participants were native speakers of the language. Participants were naïve to the fatigue-related aim of the experiment. This was to avoid fatigue-related interoceptive awareness interfering with the natural motivation to perform during the experiment. Later they were verbally debriefed about the purpose of the experiment. The study was approved by the ethics committee of the University of Nottingham, Faculty of Medicine and Health Sciences. All participants provided written informed consent.

To the best of our knowledge this is the first study to experimentally investigate fatigue x motivation x task demand interaction in a speech-in-noise task. Based on a small-to-medium effect size of the motivation x listening demand x fatigue interaction ($d = 0.35$) on the pupil dilation, we calculated that a sample size of 30 would be needed using the G*Power software (version 3.1.9.7). During our pilot testing, we noticed that the quality of our pupillometry recordings differed widely across participants. Thus, taking possible exclusions of participants based on data quality and errors in recording into account, an additional 16 participants were recruited. In total 46 participants were recruited from the database of the Hearing Sciences – Scottish Section, School of Medicine, University of Nottingham. Later, the data of 6 participants were excluded due to errors in the presented SNRs, and the data of 10 participants were excluded due to excessive missing samples (see a detailed description of data exclusion criteria below in section “pre-processing”).

2.2.3. Stimuli and trial sequence

The stimuli in the pre- and post- load sequence blocks consisted of everyday speech sentences from the Bamford-Kowal-Bench corpus (Bench et al., 1979) whereas those in the load sequence block consisted of sentences from the Institute of Hearing Research corpus (MacLeod & Summerfield, 1990). The sentences were read by the same male speaker. The 4-talker babble masker noise consisted of speech segments from 2 female and 2 male talkers. The long-term average frequency spectrum of the masker noise was identical to that of the target speech signal.

Figure 2.2 illustrates the trial sequence. Trials started 2.2 seconds after the experimenter pressed a button. A 4-talker babble noise (at 70 dB SPL for all trials) was presented. The target sentences (length: *M* = 1.62 seconds, *SD* = 0.40 seconds) started 3 seconds after the start of the babble noise. Due to a technical error that prevented the babble noise to continue for longer, the babble noise continued for only 0.5 seconds after the ending of the target sentences. Thereafter participants were prompted to repeat the sentences that they heard. The experimenter scored the sentence as correct when all the keywords were repeated correctly (sentence-based scoring) by a button-press. The button-press of the experimenter would start the next trial.

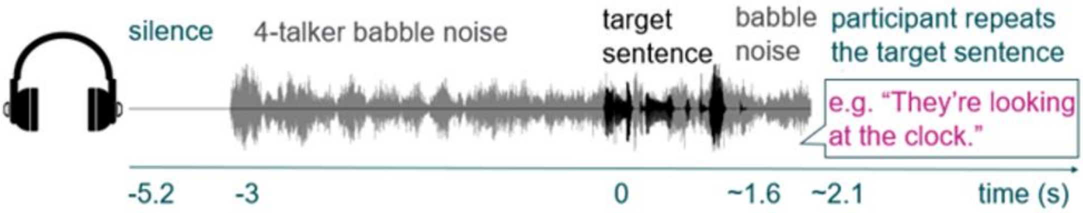


Figure 2.2. Trial sequence. All trials started with 2.2 seconds of silence. This was followed by 3 seconds of 4-talker babble noise. Then the target sentence was presented (mean duration = 1.6 seconds). After 0.5 seconds of babble noise participants were asked to repeat the target sentence.

2.2.4. Subjective report scales

After each of the of 20-sentence blocks during pre- and post- load sequence, participants were asked to report how much effort they mobilized for listening, how well they thought they performed, and how much they felt an inclination for quitting listening (cf. Koelewijn et al., 2018). Three visual analogue scales were printed on a A4 paper with the words "How much effort did understanding the sentences in the last block require?" (0=no effort, 10 = very much effort), "How would you estimate the amount of sentences that you repeated correctly" (0 = none of the sentences, 10 = all of the sentences), "How would you rate your tendency for quitting listening because the sentence was too difficult?" (0 = This happened for none of the sentences, 10= this happened for all of the sentences). To prevent participants from switching between the eye-tracker glasses to their reading glasses, which would be a break and a potential recovery from fatigue, the experimenter read the sentences aloud to the participants. Verbal responses from the participants were noted by the experimenter. Answers in fractional numbers (e.g. 3.5) were allowed.

2.2.4. The need for recovery scale

The Need for Recovery Scale (NFR; Van Veldhoven & Broersen, 2003) aims to assess how much an individual typically feels fatigued after a working day. It has been shown to have good internal consistency (alpha =0.91) and favourable construct validity (De Vries et al., 2003). The scale consists of 11 statements about feelings and attitudes at the end of a working day (e.g. "*When I get home from work, I need to be left in peace for a while*."). Listeners respond to each statement with a yes/no answer. The total NFR percentage score is calculated by dividing the number of yes responses by 11 and multiplying the outcome by 100. To accommodate for the possibility that not all participants would be employed, before filling in the scale, participants were first asked whether they were working full time. In case they didn't, they were asked to complete the scale by considering a busy day wherever a statement in the scale referred to a working day.

2.2.5. Apparatus

The sentences were presented through a software running on MATLAB (MATLAB, 2016a) using the SoundMexPro tool. The audio output from the software was amplified using RME Babyface Pro audio interface and presented to the participants through circumaural AKG K-702 Harman High End Reference headphones. The speech reception test software was integrated with the Tobii Pro 2 Eye Tracker Glasses. The glasses were set to a sampling frequency of 100 Hz for the first 15 participants. The sampling frequency was later set to 50 Hz for the remaining participants to improve data quality. The pupil diameter of both eyes was recorded. Sessions took place in a sound-attenuated booth. The booth was illuminated by LED strip lights. Illumination was adjusted for each participant using a dimmer switch, such that the illumination was set to the level that would correspond to the middle of the dynamic range of the pupils of each participant (cf. Winn et al., 2018). The experimenter monitored the experimental stimuli, gaze position, and pupil recordings through two screens in an adjacent room. The experimenter listened to and scored the responses of the participants through a graphical interface created in MATLAB.

2.2.6. Procedure

Prior to the experiment, participants received information about the study. The information focused on listening effort and motivation, but not fatigue. Participants visited our facilities twice within the course of 3 weeks. The first visit included air-conduction audiometry and ear examination, followed by the administration of the NFR scale, the estimation of SRT levels to be used in the experiment, and a pilot eye-tracker calibration as a preparation for the second visit. The pre- /post- fatigue experiment was scheduled at the second visit. The first visit took approximately 40 minutes whereas the second visit was 2-2.5 hours long.

The SNRs corresponding to SRT50 and SRT80 were estimated in two adaptive speech-in-noise tracks. The adaptive procedure for both SRT50 and SRT80 was run twice. The first run was recorded as practice whereas the resulting SNRs from the second run were noted to be used in the experiment in the second visit. During the adaptive tracks 30 sentences from the BKB corpus were presented (Bench et al., 1979) Throughout the whole procedure the intensity of the background masker was 70dB (as averaged across 30 seconds). During the estimation of SRT50 the first sentence was presented at 70 dB and increased by 4dB until the participants repeated all keywords correctly. The intensity of the following three sentences were decreased or increased by 4 dB, depending on whether the participants repeated all the keywords in the sentence correctly or not, respectively. The intensity of the fifth sentence was calculated by taking the average of the intensity levels of the first four sentences and the intensity level of what the fifth sentence would have been with a 4dB stepsize. For the rest of the sentences, the level of the target sentence was decreased or increased by 2 dB depending on correct or incorrect (incomplete) repetition of the keywords, respectively. The estimated SRT50 was calculated as the average SNR of the last five sentences and what an extra sentence would have been presented at. The SRT80 was estimated with a similar procedure to that of SRT50, where the stepsizes for the first four sentences were -1.6 dB in correct and +6.4 dB in incorrect recognition. After the 5th sentence, the stepsizes were -0.8 and +3.2 dB, respectively.

The second visit started with the eye-tracker calibration. Participants wore the glasses and looked at a fixation cross card that was placed approximately 1 meter away from them. They were asked to keep their gaze on the card throughout the experiment and to schedule their blinks after the end of the target sentences. The illumination in the booth was adjusted for the dynamic range of the pupils of each participant (cf. Zekveld, Kramer, & Festen, 2010). For this, pupil size was recorded for 30 seconds in dark and 30 seconds in maximum illumination (i.e. 750 lux as measured when the light meter faced the wall). The light was adjusted with a dimmer switch in 30-second long recordings until the elicited pupil size was approximately half-way between those in dark and maximum illumination. After this light adjustment, participants practiced the speech-in-noise task with 10 sentences in the Easy SNR and 10 sentences in the Hard SNR. The aim of the practice round was to firstly remind participants of the speech-in-noise procedure and, secondly, to give the participants an understanding of the task demands in the Easy and Hard SNR conditions. To this end, during the practice round participants received feedback on how many sentences they repeated correctly. Participants were told that they could earn additional monetary reward by correctly repeating minimum 14 of the sentences out of blocks of 20 sentences. Before each of the 20-trial blocks, written reminders about the task-demand and monetary reward (e.g. “-Difficult- -4£-” in font-size 35) were placed at the visual periphery of the participants. Participants were informed beforehand that this session would take approximately 2 hours, but were not informed about how many blocks of trials the experiment had in total. After completing the pre-load sequence blocks, participants took a 5-minute long break, where they could take the eye-tracker glasses off. The breaks were mostly spent conversing with the experimenter in quiet. At the beginning of the load sequence, participants were told that the coming block did not entail any rewards. The block would be slightly easier than the Hard SNR blocks and somewhat longer than the previous blocks. Immediately after the load sequence, participants continued with the post-load sequence blocks. Throughout the experiment the shades on the glass window of the sound-proof booth were half open, so that participants could see the experimenter who sat outside of the booth in the periphery of their vision.

2.2.7. Pre-processing and calculation of the pupil dilation indices

Pupil data were processed using MATLAB R2018b (MathWorks, Natick, MA). Pupil traces starting 1 second before the presentation of the target sentence until the end of the retention interval sentence were selected for analyses. All traces were re-sampled to 50 Hz to eliminate regular artifacts that were observed in some of the traces by visual inspection. Traces with more than 30 % of missing data were considered invalid and excluded from the analyses (cf. Wendt et al., 2018). The data of a given participant was included in the analyses if 10 or more of the trials per block of 20 trials of each experimental condition was valid.

Pupil diameter values more than 2 standard deviations away from the median of the trial were coded as blinks. Blinks were then interpolated in a linear fashion. The interpolation started for the samples within 35 ms before and ended for the samples within 100 ms after the blink to account for blink-related changes in pupil size. Later a moving average filter with a window of the samples within 15 ms length was used to smooth the de-blinked trials and to remove any high-frequency artefacts.

For each trial, the mean value of the pupil trace corresponding to the 1-second period before the onset of the target sentence was taken as BPD. The average pupil size during the presentation of the target sentence minus the baseline pupil size for each trial was taken as the mean pupil dilation (MPD).

2.2.8. Statistical analyses

The dependent measures of interest (performance accuracy, BPD, MPD, self-rated effort, performance, and quitting tendency) were each entered in a 3-way repeated-measures analysis of variance (ANOVA) in SPSS with probe time (pre- vs post-load sequence), SNR (easy vs hard), and monetary incentives (high vs low) as independent variables. Furthermore, linear mixed models were used to explore the relationship between subjective and objective measures of effort.

2.3. Results

2.3.1. SNR levels

Figure 2.3 shows the estimated SRT50 and SRT80 values in dB SNR for each participant. These were later used in the Hard SNR and Easy SNR conditions of the experiment, respectively. The estimated SNR for SRT50 was statistically significantly lower than that of SRT80 [$t_{paired}(29) = 18.210$, $p < 0.001$].

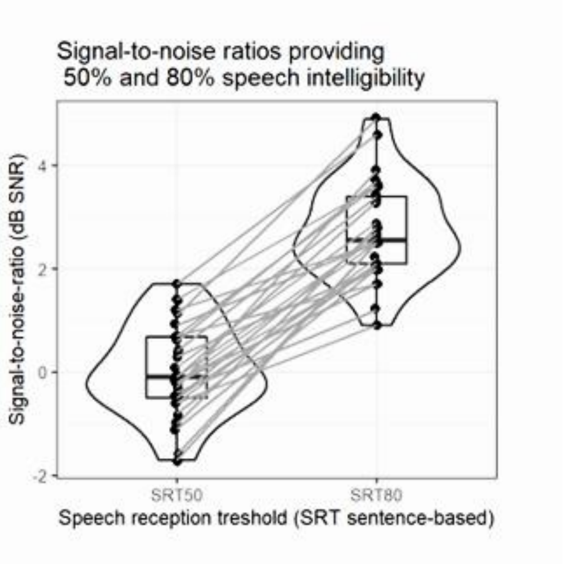


Figure 2.3. Estimated signal-to-noise ratios corresponding to 50 % and 80 % speech reception thresholds (SRTs) in boxplots, violin plots and scattered

values. For each participant the estimated SRTs were used in the experiment in the Hard SNR and Easy SNR

2.3.2. Performance

Figure 2.4 shows sentence recognition performance as a function of probe time (pre-/post- load sequence), monetary incentive amount (High and Low), and SNR (Easy and Hard). The repeated-measures ANOVA showed a main effect of SNR on speech recognition performance [$F(1, 29) = 101.706$; $p <.001$; $\eta_p^2 = 0.778$]; speech recognition performance was better in the Easy SNR condition as compared to that in the Hard SNR. There were no significant evidence for the effects of probe time [$F(1, 29) = 0.513$; $p =.480$; $\eta_p^2 = 0.017$] or monetary incentive amount [$F(1, 29) = 1.629$; $p = .212$; $\eta_p^2 = 0.053$]. Nor was there a significant evidence for the probe time x monetary incentive amount interaction effect [$F(1, 29) = 0.003$; $p =.956$; $\eta_p^2 < 0.001$] on performance. There was no significant evidence for the SNR x probe time interaction [$F(1, 29) = 0.719$; $p =.403$; $\eta_p^2 = 0.024$], or SNR x monetary incentive [$F(1, 29) = 0.271$; $p =.607$; $\eta_p^2 = 0.009$], or on the probe time x SNR x monetary incentive interaction [$F(1, 29) = 0.115$; $p =.737$; $\eta_p^2 = 0.004$]. In sum, there was only significant evidence for the effect of SNR on performance.

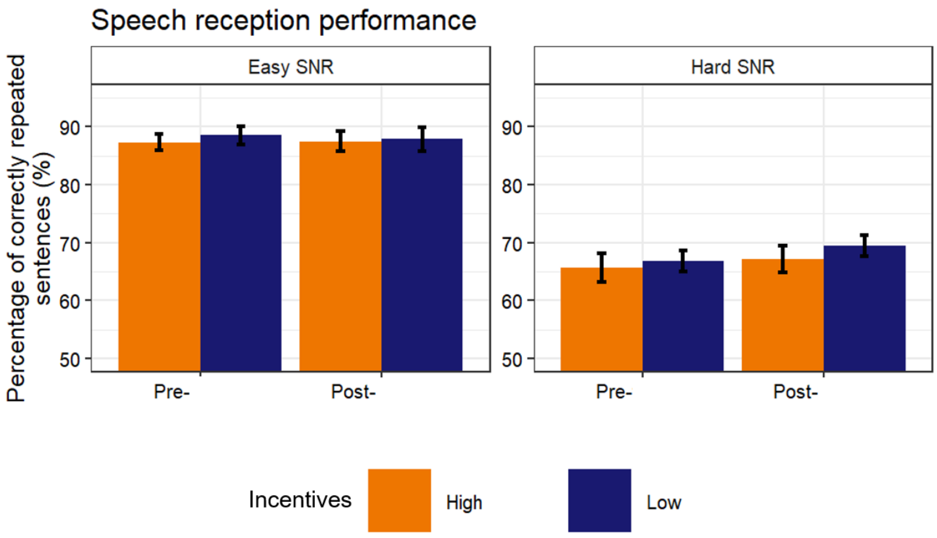


Figure 2.4. Speech reception performance as a function of fatigue, reward, and task-demand. Error bars represent the within-subject standard error of the mean (Cousineau, 2005).

2.3.4. Pupil traces

Figure 2.5 shows averaged, baseline-corrected pupil traces for the SNR, monetary incentive, and probe time conditions. The traces are plot starting from 1 second before until 3 seconds after the start of trials. Visual inspection of the interval between the seconds 2 and 2.5 shows on average larger dilation (> 0.16 mm) in the Hard SNR condition as compared to Easy SNR (< 0.16 mm). Inspection of the same interval for the Easy SNR condition shows a decline in pupil diameter from pre- to post-load sequence. In the same interval for the Hard SNR condition there is an increase in pupil diameter from pre- to post- load sequence.

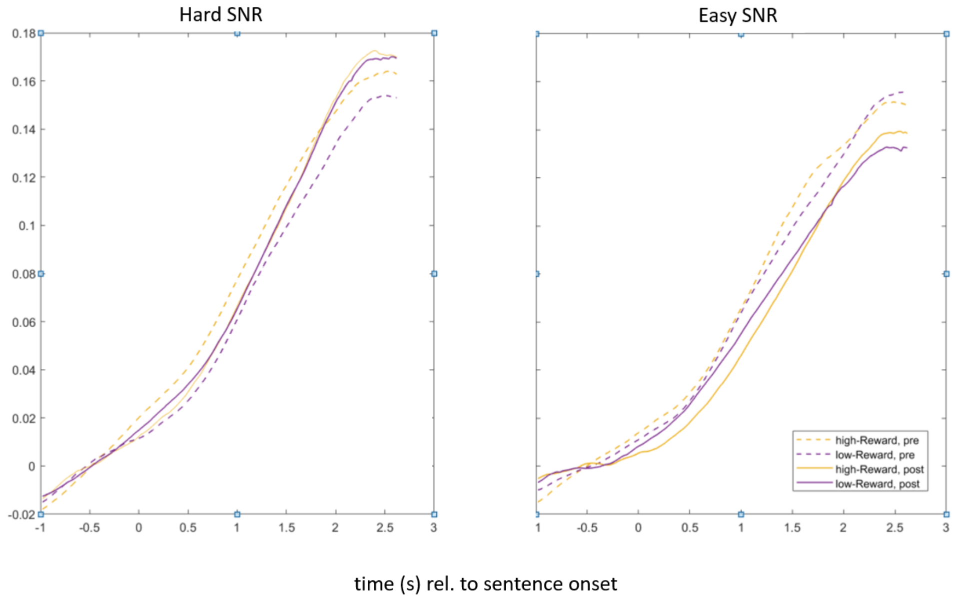


Figure 2.5. Baseline pupil diameter as a function of SNR, reward, and fatigue. Error bars represent the within-subject standard error of the mean (Cousineau, 2005).

2.3.3. Baseline and mean pupil dilation

Figure 2. 6 shows baseline pupil dilation (BPD) as a function of SNR (Easy and Hard), monetary incentive (High and Low), and probe time (pre- and post-). A repeated-measures ANOVA showed a main effect of SNR, as BPD was larger in

anticipation of hard trials as compared to that in anticipation of easy trials, [$F(1, 29) = 13.790$; $p < .01$; $\eta_p^2 = 0.332$]. While there was no main effect of monetary incentive on the BPD [$F(1, 29) = 0.002$; $p = .964$; $\eta_p^2 < 0.001$], there was a main effect of probe time on baseline pupil size, as baseline pupil size was larger in the pre-load sequence condition [$F(1, 29) = 25.573$; $p < .001$; $\eta_p^2 = 0.469$]. The effect of probe time on BPD did not depend on monetary incentive [$F(1, 29) = 0.281$; $p = .600$; $\eta_p^2 = 0.010$, or on SNR [$F(1, 29) = 0.645$; $p = .429$; $\eta_p^2 = 0.022$]. We observed a monetary incentive x SNR interaction effect [$F(1, 29) = 5.312$; $p < .05$; $\eta_p^2 = 0.155$]. Post-hoc analyses showed that there was no evidence for an effect of SNR on BPD in the low monetary incentive condition [$F(1, 29) = 0.865$; $p = .360$; $\eta_p^2 = 0.029$], but in the high monetary incentive condition, BPD was larger in the hard SNR [$F(1, 29) = 14.015$; $p < .05$; $\eta_p^2 = 0.326$] as compared to low SNR. The three-way SNR x probe time x monetary incentive interaction effect on BDP did not reach significance [$F(1, 29) = 1.807$; $p = .189$; $\eta_p^2 = 0.059$].

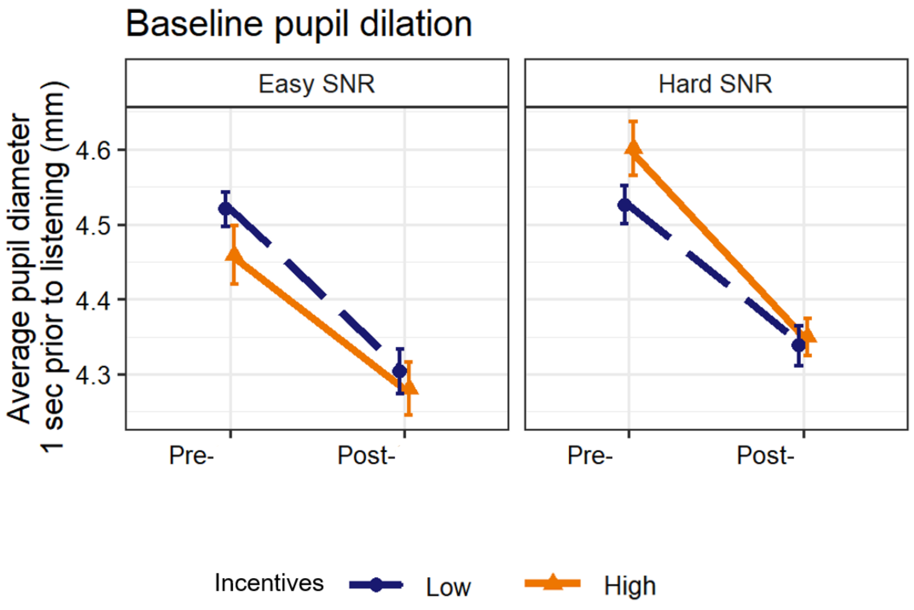


Figure 2.6 Baseline pupil diameter as a function of SNR, reward, and fatigue. Error bars represent the within-subject standard error of the mean (Cousineau, 2005).

Figure 2.7 shows baseline-corrected MPD as a function of SNR, probe time, and monetary incentive amount. A 2x2x2 repeated measures ANOVA showed a significant main effect of SNR on the MPD, as the MPD was larger in the Hard SNR condition in comparison to that in the Easy SNR condition [$F(1, 29) = 10.481$; $p < .01$; $\eta_p^2 = 0.265$]. There was no main effect of monetary incentive [$F(1, 29) <0.001$; $p = 0.992$; $\eta_p^2 < 0.001$], or probe time [$F(1, 29) = 0.581$; $p = 0.452$; $\eta_p^2 < 0.020$] on the MPD. The effect of SNR on the MPD did not depend on monetary incentive [$F(1, 29) = 1.118$; $p = .299$; $\eta_p^2 < 0.037$] or on probe time [$F(1, 29) = 0.001$; $p = .970$; $\eta_p^2 < 0.001$]. The monetary incentive x probe time interaction effect on the MPD was not significant [$F(1, 29) = 1.118$; $p = .229$; $\eta_p^2 < 0.037$]. Nor was there a significant 3-way SNR x monetary incentive x probe time interaction effect on MPD [$F(1, 29) = 0.597$; $p = .446$; $\eta_p^2 < 0.020$].

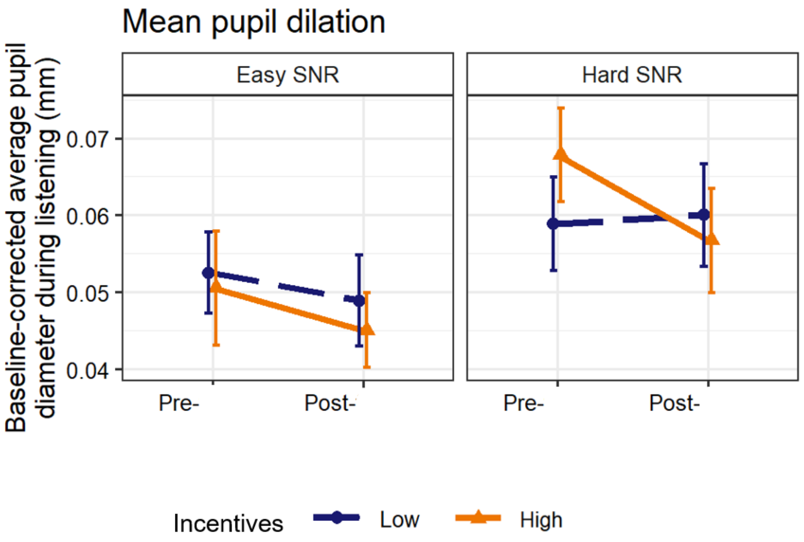


Figure 2.7 Mean pupil dilation (relative to baseline) as a function of SNR, Reward, and fatigue. Error bars represent the within-subject standard error of the mean (Cousineau, 2005).

2.3.4. Subjective ratings of effort, performance, and quitting

Table 2.1. shows self-rated effort, performance, and tendency for giving up listening as a function of SNR, monetary incentive, and probe time. Shapiro

Wilk-tests of normality confirmed that all the rating scores were normally distributed (all p’s < 0.05).

Table 2.1.

Average self-rated effort, performance, and tendency to give up

		SNR			
		Easy		Hard	
	Fatigue Reward	Pre-	Post-	Pre-	Post-
Self-rated effort	High	6.258 (0.250)	6.879 (0.256)	7.931 (0.196)	8.189 (0.278)
	Low	6.120 (0.387)	6.241 (0.301)	7.672 (0.238)	7.965 (0.256)
Self-rated performance	High	6.913 (0.256)	6.965 (0.294)	5.741 (0.196)	5.551 (0.221)
	Low	7.206 (0.224)	6.827 (0.178)	5.568 (0.248)	5.948 (0.243)
Self-rated tendency to give up	High	1.189 (0.294)	1.482 (0.181)	1.862 (0.205)	2.103 (0.205)
	Low	1.413 (0.167)	1.517 (0.196)	1.035 (0.192)	2.344 (0.202)

Table 2.1. Mean self-reported effort, performance, and tendency to give up. In parentheses are the within-subject standard error of the mean (Cousineau, 2005).

2.3.4.1. Self-rated effort

The repeated measures ANOVA with self-rated effort as the dependent variable showed a main effect of SNR [F(1, 29) = 37.384; p < .001; η_p^2 = .572], as participants reported putting more effort in the Hard SNR condition. There was a significant main effect of monetary incentive on self-rated effort [F(1, 29) = 1.793; p = 0.191; η_p^2 = 0.060] as participants reported larger effort after the High monetary incentive blocks. Although on average participants appeared to rate greater effort in the post-load sequence blocks as compared to the pre-load sequence blocks, the effect of probe time on self-rated effort did not reach significance, F(1, 29) = 1.793; p = .191; η_p^2 = .060. The effect of SNR on self-rated effort did not depend on monetary incentive [F(1, 29) = 0.155; p = .697; η_p^2 = .005], or on probe time [F(1, 29) = 0.050; p = 0.825; η_p^2 = 0.002]. The monetary incentive x probe time interaction effect was not significant [F(1, 29) = 0.602; p = 0.444; η_p^2 = 0.021]. Nor was there an SNR x monetary incentive x probe time interaction, $F(1, 29) = 1.077$; $p = 0.308$; $\eta_p^2 = 0.037$.

2.3.4.2. Self-rated performance

The repeated measures ANOVA with performance as the dependent variable showed a main effect of SNR, [$F(1, 29) = 27.282$; $p < .001$; $\eta_p^2 = .494$] as participants rated their performance as poorer in the Hard SNR condition. The main effects of monetary incentive [$F(1, 29) = 0.389$; $p = .538$; $\eta_p^2 = .014$] and probe time [$F(1, 29) = 0.055$; $p = .817$; $\eta_p^2 = .002$] were not significant. The effect of SNR did not depend on monetary incentive [$F(1, 29) = 0.014$; $p = .907$; $\eta_p^2 < .001$] or probe time [$F(1, 29) = 0.541$; $p = .468$; $\eta_p^2 = .019$]. The monetary incentive x probe time interaction effect was not significant [$F(1, 29) = 0.090$; $p = 0.766$; $\eta_p^2 = 0.003$]. Nor was the SNR x probe time x monetary incentive interaction significant [$F(1, 29) = 2.598$; $p = .118$; $\eta_p^2 = .085$].

2.3.4.2. Self-rated tendency to quit

The repeated measures ANOVA with self-rated tendency for quitting as the dependent variable showed a main effect of SNR [$F(1, 29) = 12.689$; $p <.001$; $\eta_p^2 = .312$], as on average, self-rated quitting was greater in the Hard SNR condition. There was no significant main effect of monetary incentive [$F(1, 29) = 0.646$; $p = 0.428$; $\eta_p^2 = 0.023$]. Although the average tendency for quitting was larger at all post-load sequence blocks as compared to the corresponding pre-load sequence blocks, the main effect of probe time did not reach significance [$F(1, 29) = 2.358$; $p = .136$; $\eta_p^2 = .0789$]. The effect of SNR did not depend on monetary incentive [$F(1, 29) = 0.005$; $p = .945$; $\eta_p^2 < .001$] or on probe time [$F(1, 29) = 0.289$; $p = .595$; $\eta_p^2 = .010$]. There was no monetary incentive x probe time interaction [$F(1, 29) = 0.028$; $p = .869$; $\eta_p^2 = .001$]. The average increase in self-reported tendency for quitting from pre- to post- load sequence was largest in the Hard SNR – low monetary incentive block, but the 3-way SNR x monetary incentive x probe time interaction did not reach significance [$F(1, 29) = 2.048$; $p = .163$; $\eta_p^2 = .068$].

2.3.5 Relations between objective and subjective measures

To explore the relationships between the subjective and objective measurements of effort, four repeated measures correlations were computed, where baseline and baseline-corrected mean pupil size were the dependent variables. Self-rated effort, and self-rated tendency for quitting were independent variables. The analyses showed a significant positive association between MPD and subjective effort scores [*F*(12, 222.885) = 2.019; p < .05; see Figure 2.8.], and subjective tendency for quitting [*F*(12, 186.269) = 1.806; *p* < .05; see Figure 2.9.]. There was no significant association between BPD and subjective effort scores [*F*(12, 204,705) = 0.436; *p* =.947; see Figure 2.10.] nor between BPD and subjective tendency for quitting [*F*(12, 208,446) = 1.223; *p* =.265; see Figure 2.11.].

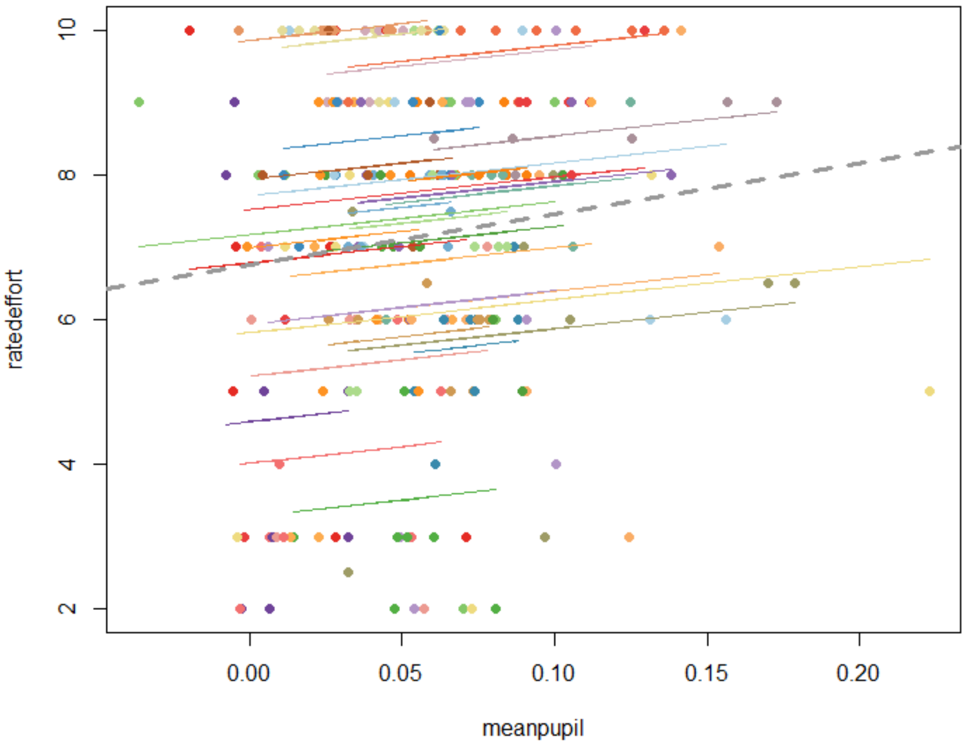


Figure 2.8 Scatterplot of mean pupil dilation and self-rated effort. Each dot represents the averaged value per monetary incentive, SNR and probe time condition per participant. Observations from the same participant are given the same color and the corresponding lines show the rmcorr fit for each

participant.

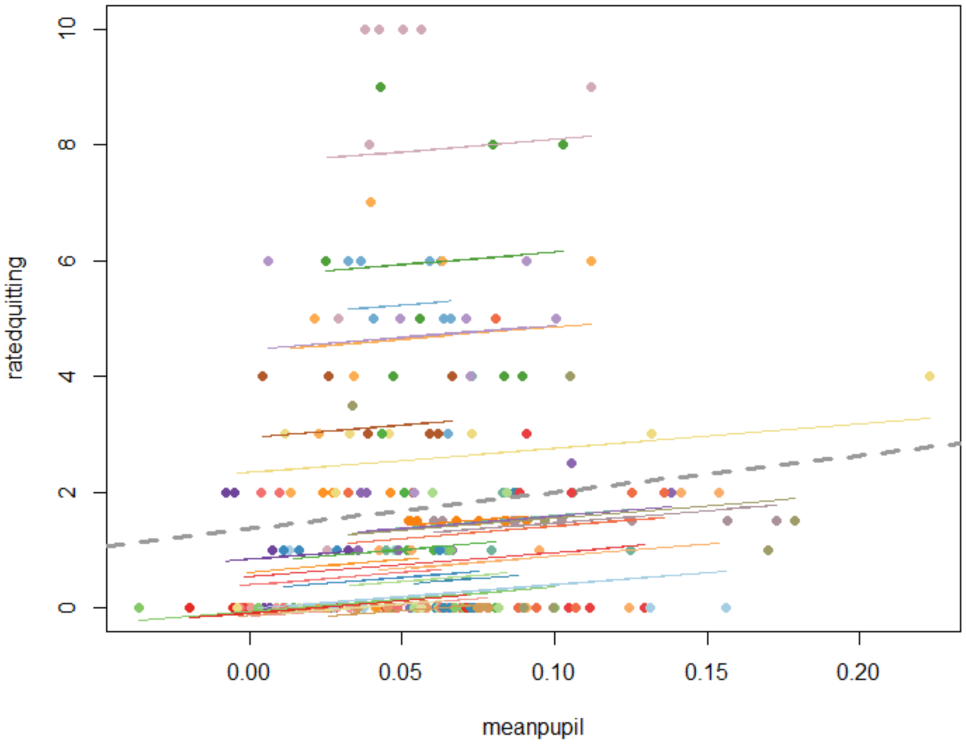


Figure 2.9. Scatterplot of mean pupil dilation and self-rated effort. Each dot represents the averaged value per monetary incentive, SNR and probe time condition per participant. Observations from the same participant are given the same color and the corresponding lines show the rmcorr fit for each participant.

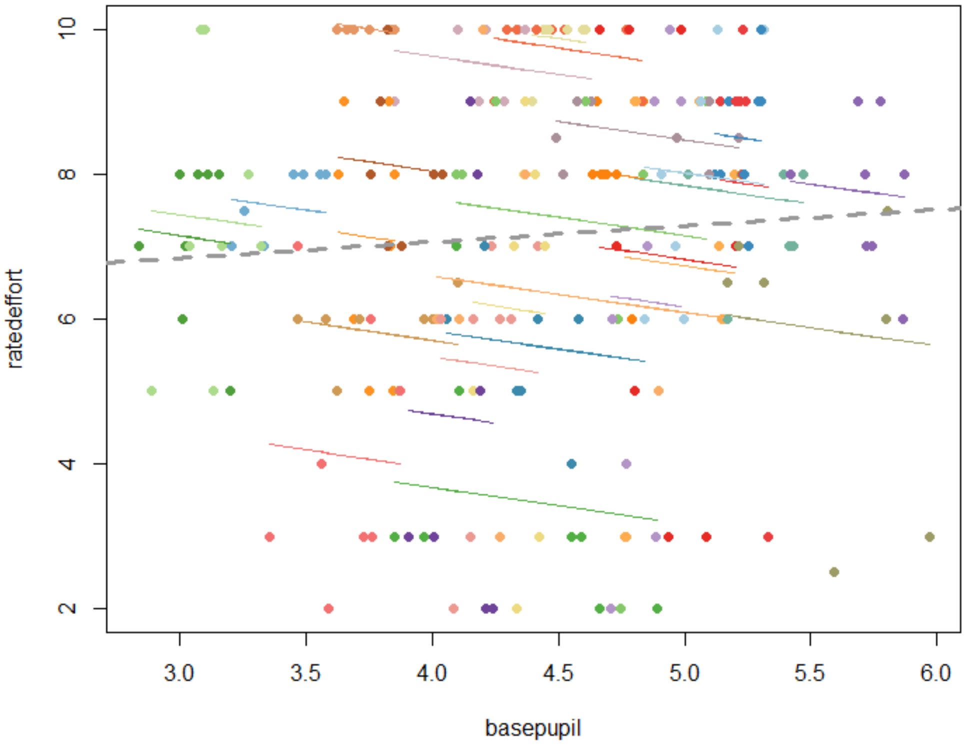


Figure 2.10. Scatterplot of baseline pupil diameter and self-rated quitting. Each dot represents the averaged value per monetary incentive, SNR and probe time condition per participant. Observations from the same participant are given the same color and the corresponding lines show the rmcorr fit for each participant.

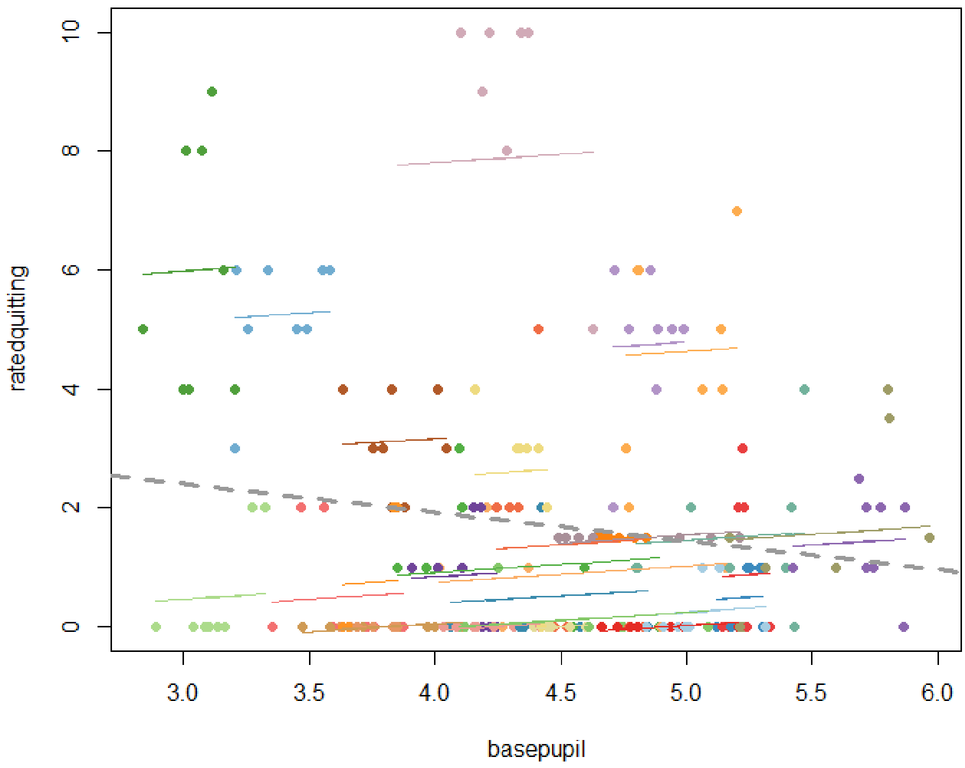


Figure 2.11. Scatterplot of baseline pupil diameter and self-rated quitting. Each dot represents the averaged value per monetary incentive, SNR and probe time condition per participant. Observations from the same participant are given the same color and the corresponding lines show the rmcorr fit for each participant.

2.3.6. Need for recovery

The mean score for the need for recovery in our sample was 17.83 % [*Median* = 18 %, *SD* = 2.550 %]. A Shapiro-Wilk test conducted with the Need for Recovery scores showed that the scores were not normally distributed [*W* = 0.917; *p* < 0.026]. To explore the association of Need for Recovery with the baseline-corrected mean pupil dilation, therefore, a 2-tailed Spearman's correlation was conducted. The results did not show any evidence for a significant correlation between the Need for Recovery scores and average baseline-corrected mean pupil dilation in the experiment [Spearman's *rho* = 0.120; *p* = .535].

2.4. Discussion

The main aim of this study was to investigate the effects of task-induced fatigue and motivation on the relationship between listening demand and listening effort. To induce fatigue, a sample of adults with NH engaged in a challenging sustained speech reception task (“load sequence”). Listening effort was probed during a speech-in-noise task through pupillometry both pre- and post-load sequence in 2 SNR and 2 monetary incentive conditions. In line with existing conceptualizations of fatigue (e.g. Hockey, 1997; Kahneman, 1973) and previous reports of on time-on-task effects on the pupil metrics (e.g. Herlambang, Cnossen, & Taatgen, 2021; Pichora-Fuller et al., 2016; Zekveld, Koelewijn, & Kramer, 2018), diminished BPD from pre- to post- load sequence was hypothesized. Furthermore, in line with the view that fatigue diminishes the perception of how much effort is worth investing (e.g. Müller & Apps, 2019), particularly in demanding tasks, larger difference in listening effort between the monetary incentive conditions during the post-load sequence probe as compared to pre-load sequence was hypothesized, particularly when listening in the Hard SNR condition. The hypotheses were partially supported.

The results showed a smaller BPD in anticipation of listening post-load sequence as compared to pre- load sequence. Although this effect could partially be attributed to habituation (Metalis & Hess, 1981; Verbaten et al., 1986), diminished alertness over time-on-task (Hopstaken et al., 2015a) and reduced preparation for effortful trials (Ackerman, 2010; Boksem, Meijman, & Lorist, 2006; Lorist, 2008; Lorist et al., 2000; McGarrigle, Dawes, Stewart, Kuchinsky, & Munro, 2017b, 2017a) have typically been considered important characteristics of fatigue. A blending of habituation, learning, and fatigue effects over time-on-task is commonly reported in previous research (Faber et al., 2012; Hopstaken et al., 2015a; van der Linden et al., 2003). Overall, in the current study the decline in BPD, although in line with the hypotheses, is perhaps not to be attributed to a single source.

In addition to the effect of probe time on BPD, there was an effect of SNR on BPD, as BPD was larger in anticipation of hard trials, suggesting larger recruitment of capacity in anticipation of larger demands. The effect of SNR depended on incentives. Whereas in the low monetary incentive condition there was no effect of SNR on BPD, in the high monetary incentive condition BPD was larger with more difficult SNR. Together, these results suggest that the BPD is sensitive to listening—related fatigue and motivation as these seem to affect the pro-active (pre-emptive) allocation of capacity in anticipation of listening in plausible ways. The results support the view that BPD does not reflect a unitary construct (Ayasse & Wingfield, 2020).
In line with previous research on the effect of SNR on the MPD (e.g. Wendt, Koelewijn, Książek, Kramer, & Lunner, 2018; Zekveld, Kramer, & Festen, 2010) we observed larger MPD in the hard SNR condition as compared to the easy SNR condition. This result shows that our analyses were able to reproduce the effects of task demand on listening effort. We thereby were able to confirm the validity of our pupillometry set-up and analyses.

There was no main effect of monetary incentives on the MPD. Previously Koelewijn et al. (2018) reported no effect of monetary incentives on the BPD or MPD, but a significant effect on the PPD. The observed pattern of results is similar to the results of Stanners et al., 1979, where incentives led to increased anticipatory dilation and no effect on the task-evoked dilation in an auditory digit memorization task. The lack of incentive effect on the MPD in our study may be due to the earlier dilation of the pupil in anticipation of trials, which may have nullified the effects of incentives on the MPD after baseline correction.

There was no significant effect of probe time on the MPD. This may be related to the nature of the fatigue manipulation, and the timing of the trial sequences. Due to time constraints and ethical considerations the fatigue manipulation was not strong; participants may have been able to exert effort even after the manipulation. Furthermore, the short rest periods in between the trials may have allowed participants to regulate their arousal and thereby exert enough capacity even after the fatigue manipulation.

Lastly, contrary to hypothesized, there was no significant evidence for an SNR x monetary incentive x probe time interaction effect on the MPD. Whereas we expected a stronger decline from pre- to post- load sequence in the High incentive condition, as compared to Low, particularly in the High SNR condition, this effect was not significant. The lack of the interaction effect may be due to the incentives that may have been insufficient to keep participants motivated in the post-load sequence condition. The amount of monetary incentives were based on those previously reported by Koelewijn et al., 2018; however, the current participant sample was much older than those of Koelewijn et al. As older age is known to reduce sensitivity to rewards (Dhingra et al., 2020), future studies with similar age groups could use larger or more intrinsic incentives.

Performance was on average better in the Easy SNR condition as compared to the Hard SNR condition. In both conditions, average performance, *M* = 67 (*SD* = 1) % and *M* = 88 (*SD* = 2) %, respectively, was better than what the SRT estimations had targeted, 50 % and 80 %, respectively. This suggests that the Hard SNR condition was in fact on average easier than how we expected it to be. This may explain lack of SNR x probe time interactions on the BPD and MPD. Even after the fatigue manipulation, the Hard SNR condition may have been perceived as achievable and worth exerting effort. Although participants practiced with 30 trials before the adaptive SRT estimation track, this practice may have been insufficient, leading to the improvement effects that were observed. Future studies may opt for longer practice periods before running the adaptive estimation procedure.

There were no effects of monetary incentive, probe time, or the interaction of the two, on the speech reception performance. The lack of monetary incentive effect on performance is in line with that reported by Koelewijn et al. (2018) and support the view that speech-in-noise tasks by nature offer little room for improvement. On the other hand, evaluative feedback and encouraging nudges have recently been shown to increase performance in a speech-in-noise task (Zekveld et al., 2019). Together these suggest that motivation manipulations in other forms may be stronger than that through monetary incentives. Conversely, the SNR x monetary incentive interaction on the BPD and the lack thereof on performance support the view that pupil metrics reveal information about the state of the listeners that is not captured by speech reception tests alone.

Self-rated effort and self-rated tendency to quit listening both separately predicted MPD, whereas no such relationship existed with the self-report measures and BPD. This pattern of results are similar to those reported by Hopstaken et al. (2015) in a visual n-back task with multiple blocks. They suggest that there may be shared within-subject variance between the objective and subjective measures of effort.

Self-rated effort was significantly larger in the Hard SNR condition as compared to the Easy SNR condition. This finding confirms that the Easy SNR was indeed perceived as easier than the Hard SNR. Furthermore, self-rated effort was significantly larger with higher incentives. Note that the wording of the effort scale tapped more into the demanded effort rather than the exerted one; thus one interpretation of this finding may be the thinking in listeners that they were asked by the experimenter to exert more effort in the high incentive condition. Although not statistically significant, self-rated effort showed an increasing trend from pre- to post- fatigue for all the incentive and SNR conditions, suggesting that participants may have felt the need for compensatory effort in the post-load sequence block to keep up. This compensatory effort may be another reason for the lack of pre- to post load sequence decline in the mean pupil dilation. Self-reported performance was statistically significantly better in the Easy SNR condition as compared to the Hard SNR condition, whereas there were no effects of probe time or incentive on the self-reported performance. Interestingly, in the Hard SNR condition, on average, the estimated performance was poorer than 60%, even before the load sequence. This suggests an underestimation of performance, which may be another explanation for the pattern of the observed results in the pupil metrics. That is, for the Hard SNR condition, the reward offer may not have had the intended motivating effect as participants may have perceived the 70 % performance condition as too hard to achieve.

The self-reported tendency for quitting was larger for the Hard SNR condition as compared to the Easy SNR. This supports the understanding that listening effort involves a deliberate component for overcoming obstacles as defined in the FUEL (Pichora-Fuller et al., 2016). Although the SNR x monetary incentive x probe time interaction on the self-reported tendency to quit was not statistically significant, the size of this effect was medium; on average participants rated largest tendency to quit in the post-fatigue-Hard SNR-low Incentive block. This result suggests that in our sample of participants, on average, previous load sequence and monetary incentives may have contributed to the willingness to exert listening effort.

In the current experiment the MPD did not correlate with the Need for Recovery scores. Previously Koelewijn et al. (2018) reported no relation between the PPD and NFR, whereas Wang et al. (2018) reported that individuals larger NFR had smaller PPD. Whereas both Koelewijn et al (2018) and Wang et al (2018) investigated relations with PPD, we investigated the relationship of NFR with the MPD, thus our results are not directly comparable. Nevertheless, it is to note that the mean NFR score reported in Koelewijn et al (2018) and the current study (17.9 % and 17.8 %), where the samples were NH, were both lower as compared to the score in the Wang et al (2018) study (39.4 %), where the sample was HI. The relation between the pupil dilation response and NFR may be stronger for the individuals who report larger NFR. Another reason for the lack of relation between NFR and MPD in the current results may be the relative small variance in the NFR in the current sample.

Finally, we observed a within-subject relationship between self-rated and pupillometric outcomes of listening effort. Whereas McGarrigle et al. (2021) report a relationship between tiredness and TEPR, we observed a relationship between the MPD and self-rated effort and tendency for quitting. This result supports the view that within-subject measurement with self-rated and pupillometry measures may capture overlapping variance that reflects moment-to-moment changes in the state of the listener.

There are several limitations of the present study to be reported. The trial sequence in the speech-in-noise task only included half a second of retention interval (i.e. a break after the presentation of the sentence and before participants repeated what they heard). This short retention interval did not allow us to observe the full pupil dilation response and therefore the PPD was not included in our pupil metrics. Thus, we were unable to observe the influence of previous load sequence and monetary incentives on the PPD. In addition, monetary incentives may have been insufficient to induce motivation in a sample that was middle aged on average. Future studies could use other forms of incentives (e.g. intrinsic pleasure), which may be more effective in manipulating motivation. Lastly, the current study presents association of pupil metrics with task-induced fatigue. Thus, the conclusions that can be drawn based on our results about any effect of daily-life fatigue on the pupil metrics are limited. Future pupillometry studies with longitudinal designs are thus needed to describe the development of daily-life fatigue in in adults with HI.

2.5. Summary and conclusions

To conclude, in this study the interactive effects of motivation (monetary incentives) and fatigue (load sequence) on listening effort were measured using pupillometry in a speech-in-noise task with easy and hard SNR conditions. Although easier SNR elicited smaller MPD, monetary incentive and previous load sequence effects on the MPD were not present. Instead, monetary incentives interacted with SNR in influencing BPD, and previous load sequence was shown to decrease the BPD. These results suggest that motivation and fatigue may influence the capacity exerted in anticipation of listening. In addition, larger MPD was shown to predict larger self-reported effort and larger self-reported tendency for quitting. If the "tendency to quit" measure was an accurate indication of how often listeners gave up, one might expect larger ratings to predict smaller dilation. Instead, this result is in line with the understanding that larger effort is accompanied by larger tendency of quitting but no actual quitting. No relation was found between need for recovery and average pupil dilation. These results support the view that objective and subjective measures of effort may capture shared within-subject variation (McGarrigle et al., 2021).

In the current experiment we explored task- induced hearing-related fatigue by comparing pre- to post- load sequence probes of listening effort. How much of the pre- to post- changes in the pupil metrics can be attributed to fatigue, habituation, or the mere passage of time, remain unclear. Thus, in Experiment 2 (Chapter 3) the effect of fatigue on listening effort was investigated with a load sequence that included a heavy vs. light contrast. If the decline in BPD is due to task-induced fatigue, then heavier load sequence should result in larger decline in BPD.

Chapter 3: Influence of motivation and of the magnitude of previous task demand on subsequent listening effort as measured by the baseline, mean, and peak pupil dilation in adults with normal hearing

3.1. Introduction

As mentioned in Chapter 1, mental fatigue is well known to arise with time spent on effortful tasks (Boksem et al., 2006; Boksem & Tops, 2008; Hockey, 1997a; Müller & Apps, 2019). Although fatigue is subjectively experienced as an inability to exert further effort, previous research has shown that in a fatigued state individuals can and do exert effort, particularly in real world tasks, as long as their motivation to perform well is high (Hockey, 2010). Thus, it has been acknowledged that whether a state of fatigue leads to a suppression of effort depends on the willingness to exert effort (Boksem et al., 2006; Boksem & Tops, 2008; Hockey, 1997a; Müller & Apps, 2019). How listening effort is influenced by the interactions between motivation and fatigue remains largely unexplored.

Previous studies investigating the effect of time on task on the pupil dilation response to auditory stimuli mostly report declines over time in the baseline pupil diameter (BPD) and in the task-evoked pupil responses (TEPR; Beatty, 1982; see Zekveld, Koelewijn, & Kramer, 2018 for a review). Previous research on the effect of monetary incentives in auditory tasks show larger TEPRs when participants are offered larger amounts (e.g. Koelewijn, Zekveld, Lunner, & Kramer, 2018; see Zekveld et al., 2018 for a review). In Experiment 1 (Chapter 2) we used pupillometry to investigate the interactive effects of listening-related fatigue and motivation on listening effort in a pre-/post-fatigue experiment, where pre- and post-load sequence listening effort was measured in 2 SNRs and 2 monetary incentive levels. Results showed a decline in BPD from pre- to post-load sequence. The decline in BPD was attributed to having performed the load sequence. However, given no change in performance, TEPR (in this case MPD), or self-reported tendency for quitting from pre- to post- load sequence, the relative contributions of habituation, learning, and fatigue to the decline in BPD are hard to disentangle. Hence, in this chapter, we investigate how much the *magnitude* of load sequence influences the change in the pupil metrics from the pre- to post- sequence probes. If changes in the pupil metrics from pre- to post- are related to having performed the load sequence, then a heavier load sequence should elicit larger changes in the pupil metrics.

Previous research in psychology shows that heavier task load over time leads to larger suppression of subsequent effort (e.g. Earle, Hockey, Earle, & Clough, 2015; van der Linden, Frese, & Meijman, 2003). Previous pupillometry research with auditory stimuli has shown larger decline in the TEPR when performing more difficult tasks as compared to easier ones (Ambler et al., 1976). For example, in a dichotic listening task with shadowing, i.e. repeating one of the messages concurrently aloud, the decline in the TEPR over time was larger when the listening task was more difficult (Ambler et al., 1976). It remains unknown whether a heavier task load sequence leads to larger decline in TEPR in a subsequent speech-in-noise task, and to what extent the decline depends on the willingness of individuals to perform well.
Although lack of between-subject correlations between self-reported measures and pupil dilation measures of listening effort is commonly reported (e.g. Koelewijn, Zekveld, Festen, & Kramer, 2012; McGarrigle, Dawes, Stewart, Kuchinsky, & Munro, 2017; Pichora-Fuller et al., 2016; Strand, Brown, Merchant, Brown, & Smith, 2018), recent research suggests an association between within-individual variability of subjective and objective measurement outcomes (McGarrigle, Rakusen, & Mattys, 2021; Experiment 1 [Chapter 2]). Hence in the current experiment, the relationship between objective and subjective outcomes was investigated in a within-individual approach (cf. Experiment 1).

The PPD has previously been shown to be negatively correlated with daily-life fatigue (Wang et al., 2018); however, this relationship hasn’t been consistently observed (Koelewijn, Zekveld, Lunner, & Kramer, 2018; Experiment 1 [Chapter 2]). To seek further clarification on the relationship between daily life fatigue and the PPD, a Need for Recovery scale was used to measure daily life fatigue (Van Veldhoven & Broersen, 2003a).

3.1.1 Aims and hypotheses

In the present experiment, we sought to investigate the interactive effects of listening-related fatigue and motivation on listening effort. This was achieved through using pupillometry in a pre-/ post- fatigue experiment that utilized a speech-in-noise task to both induce and probe fatigue. The fatigue induction was through either a heavy or light “load sequence”. Listening effort was probed pre- and post-load sequence, where high and low monetary incentives for correct speech recognition were used to manipulate motivation. Parallel to previous effects of time-on-task on the BPD, we expected a general decline in BPD from pre- to post- load sequence (hypothesis 1). If a larger load sequence leads to larger fatigue, than any decline in pre- to post-load sequence in MPD, PPD and BPD should be larger in the heavy load sequence condition (hypothesis 2). Assuming that in a fatigued state the mobilization of effort depends on motivation, then the pre- to post- load sequence decline in MDP and PPD should be larger when monetary incentives are low (hypothesis 3). This effect should be particularly pronounced in the heavy load sequence condition (hypothesis 4).

3.2. Methods

3.2.1 Experimental design

Figure 3.1 illustrates the experimental design. Fatigue and motivation were both manipulated in two levels. Listening effort was measured using pupillometry and self-report.

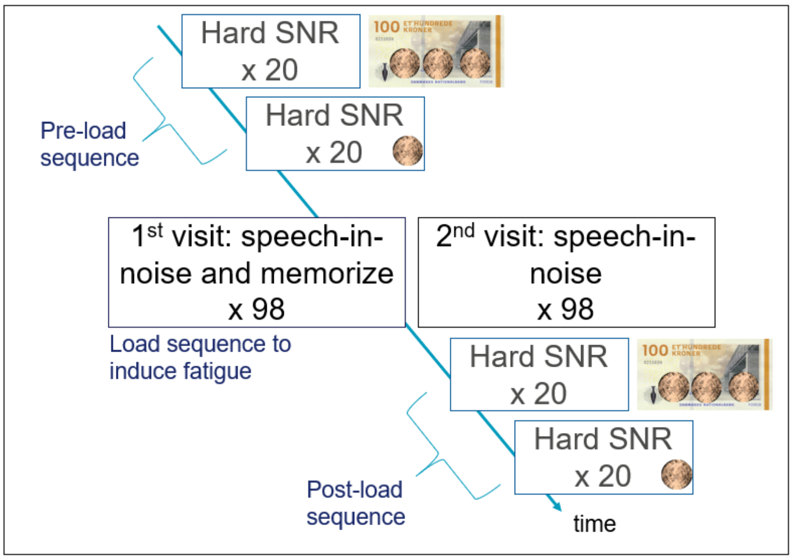


Figure 3.1. Experimental design. To induce fatigue, a long speech-in-noise task of 98 trials with or without concurrent memory task (i.e. "heavy and light load sequence") was used. Shorter blocks of 20 trials were administered before and after the load sequence to probe the effect of previous task demand on listening effort. To manipulate motivation, participants were offered two levels of monetary incentives (20 DK and 160 DK) during the pre- and post-load sequence probes. Whereas the SNR during the load sequence was fixed across participants, the SNR in the pre- and post-test blocks was individually pre-determined for each participant through an adaptive track prior to the experiment.

To induce fatigue, a speech-in-noise task with 98 sentences was used (i.e. 'load sequence'; approx. 40 minutes long). Each participant completed the experiment in 2 sessions. The speech-in-noise task during the load sequence either excluded or included a concurrent memory task (i.e. "light load sequence" and "heavy load sequence", conditions respectively). The order of the sessions were counterbalanced over participants. In the light load sequence condition, participants only repeated the final words of the sentences that were presented. In the heavy load sequence condition, in addition to repeating the final words of the sentences, participants were also asked to memorize the final words of blocks of 7 sentences. At the end of each 7th sentence, participants were asked to recall the 7 words that were to be memorized. The SNR within 7-sentence blocks was constant at either +1 or -4 dB, and changed across 7-sentence blocks within sessions in a pre-determined sequence. The sequence was +1, +1, -4, -4, +1, -4, +1, +1, -4, -4, -4, +1, -4, and +1 dB. The sequence was created to prevent participants from predicting the upcoming SNR during the session and to make the load sequence relatively homogenously taxing across time. Per participant the SNR order was the same for the light load sequence and heavy load sequence conditions (i.e. in both sessions of the experiment). Across participants, the SNR was cycled through the sequence with a different starting point for each participant.

To assess the influence of load sequence on subsequent listening effort, both before and after the load sequence, participants performed two speech-in-noise tasks (i.e. pre- and post-load sequence 'probes'). The pre- and post-load sequence probes of 2 blocks of 20 trials.

To manipulate motivation in the pre- and post-load sequence blocks, participants were offered either 20 or 160 Danish Krones per block (high and low monetary incentive conditions, respectively) on the condition that they correctly repeated at least 12 of the 20 sentences (i.e. 60% correct). Note that this percentage was smaller than that in Experiment 1 (%70; chapter 2) to increase the perceived achievability of the task. The reasoning was that this condition should be perceived as achievable even in a state of mild fatigue.

Per participant, the order of the monetary incentive conditions was the same in the pre- and post-load sequence blocks, and across both experimental sessions. The SNR in the pre- and post-test blocks was individually pre-determined prior to the experiment. We reasoned that fatigue would have the strongest effect on listening effort when listening is the most effortful. Previous research has shown that the SNR at SRT50 elicited the largest pupil dilation as compared to the other SRT levels, suggesting maximal effort at SRT50 (e.g. Wendt, Koelewijn, Książek, Kramer, & Lunner, 2018; Ohlenforst et al., 2018). Therefore, the SNR that corresponds to SRT50 was chosen for the probe speech-in-noise task.

The effects of load sequence (light and heavy) and monetary incentives (high and low) on the change in listening effort were investigated by calculating the change in the pupil dilation metrics and self-report scores from the pre- to the post-load sequence blocks (i.e. Δpost-pre).

In this study the load sequence was only used to induce fatigue. The investigation of the pupil size during the load sequence (see Appendices 3,4,5, and 6 for plots of BPD and PPD during the light and heavy load sequences, respectively) was not part of the research questions. The load sequence was assumed to have induced fatigue, and was not investigated further.

3.2.2. Participants

Participants were 18 adults with normal hearing (i.e. hearing thresholds were less than or equal to 25 dB HL over the frequencies from 0.25 to 4 kHz for both ears). They reported no neurological or psychiatric disorders, normal or corrected-to normal vision, and Danish as native language. Participants were naïve to the fatigue-related aim of the experiment. This was to avoid interoceptive awareness interfering with the natural motivation to perform during the experiment. Later they were verbally debriefed about the purpose of the experiment. The study was approved by the local ethics committee and all participants provided written informed consent.

Power calculations for a small to medium effect size of d = 0.25 for the monetary incentive amount x load sequence magnitude interaction using GPower 3.1.9.7. software showed that at least 30 participants would be needed to observe the expected interaction effect. 30 participants were recruited for the study from the database of the Eriksholm Research Centre, part of Oticon A/S. Among those recruited, the data of 18 participants were clean to use for analyses (see pre-processing section below for inclusion criteria). This meant that the counterbalancing of the session order did not hold. There were more participants (11) who took the light load sequence condition as first session.

3.2.3. Stimuli and trial sequence

Target stimuli in the load sequence consisted of everyday sentences from the Danish Hearing in Noise Test (HINT; Nielsen & Dau, 2011), whereas target stimuli in the pre- and post-load sequence blocks consisted of sentences from the Danish Dagmar-Asta-Tine corpus (DAT; Nielsen, Dau, & Neher, 2014). The 16-talker babble masker noise consisted of speech segments from 2 female and 2 male talkers. The long-term average frequency spectrum of the masker noise was identical to that of the target speech signal of the corresponding speech corpus, separately for the DAT and HINT material. In this study listening effort was only investigated during the pre- and post-load sequence block. The purpose of the load sequence was to merely induce fatigue. Therefore, only the sequence of trials in the probe blocks are described here. All stimuli were presented through loudspeakers in a sound attenuated booth (see apparatus section for details)

Figure 3.2. illustrates the trial sequence in the pre- and post-load sequence blocks. Trials started 2.2 seconds after the experimenter pressed a button. A 16-talker babble noise (at 70 dB SPL for all trials) was presented throughout the trials. The target sentences started 3 seconds after the start of the babble noise. All sentences were 2 seconds long and had the same carrier words "Dagmar tænkte på (~550 ms) … og en (~166 ms) … I går (~333 ms)" (English: Dagmar thought of … and … yesterday"); and 2 keywords (~450 ms and ~500 ms, respectively). The babble noise continued for 3 seconds after the ending of the target sentences. Thereafter participants were asked to repeat the keywords. Answers were considered correct if both keywords had been repeated correctly The experimenter scored the answers by a button-press ("correct or incorrect"). The button-press of the experimenter would start the next trial.

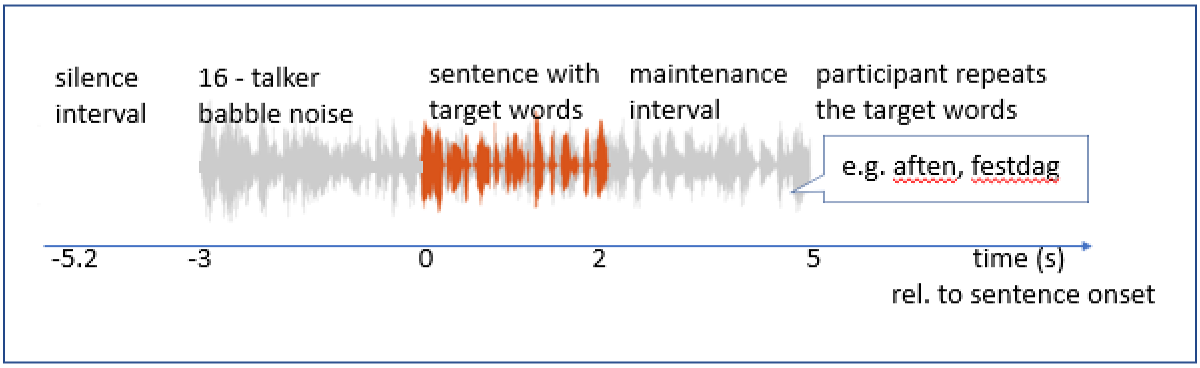


Figure 3.2. Example trial sequence during the pre-and post-load sequence blocks. All trials started with 2.2 seconds silence, followed by 16-talker babble noise. Then participants were presented with a sentence (e.g. Dagmar tænkte på en aften o gen festdag I går). The noise continued for 3 seconds after the end of the sentence. Participants were asked to repeat the keywords as accurately and as quickly as possible (e.g. "en aften, festdag").

3.2.4. Subjective report scales

After each of the pre- and post- load sequence blocks, participants were asked to report how much effort they mobilized for listening, how well they think they performed, and how much they felt an inclination to quit listening (Koelewijn et al., 2018). Three 100-point visual analog scales were printed in 18 font size on an A4 paper. The questions were (1) "*Hvor meget anstrengte du dig for at høre sætningerne? – EN: How much effort did you put into hearing the sentences?*" (0 = Ingen, 25 = Lav, 50 = Moderat, 75 = Høj, 100 = Meget høj – EN: 0 = None, 25 = Low, 50 = Moderate, 75 = High, 100 = Very High), (2) "Hvor mange af ordene tror du, at du forstod korrekt?" – EN: *How many of the words do you think you understood correctly*?" (0 = Ingen, 25 = Mindre end halvdelen, 50 = Halvdelen, 75 = Mere end halvdelen, 100 = Alle – EN: 0 = None, 25 = Less than half, 50 = Half, 75 = More than half, 100 = All), and (3) "*Hvor ofte måtte du opgive at forstå sætningen? – EN: How often did you have to give up understanding the phrase?* (0 = Aldrig, 25 = Mindre end halvdelen af tiden, 50 = Halvdelen af tiden, 75 = Mere end halvdelen af tiden, 100 = Altid – EN: 0 = Never, 25 = Less than half the time, 50 = Half the time, 75 = More than half the time, 100 = all the time). Participants marked their answers on paper.

3.2.5. The need for recovery scale

The need for recovery scale measures the early symptoms of fatigue at work, particularly the extent an effortful task induces a need to recuperate (Van Veldhoven & Broersen, 2003b). The Danish translation of the scale was used to assess the daily life fatigue of the participants.

3.2.6. Procedure and apparatus

Prior to the experiment participants were informed that the study was about listening effort and motivation. All participants visited the laboratory twice within the course of 3 weeks. Sessions began at either 9:00 or 13:00, and per participant, both visits were scheduled at the same time of the day to minimize variability of diurnal changes in cortisol (Kirschbaum & Hellhammer, 1989). Participants were advised to not take any alcohol 24 hours prior to, sleep well the night before, and not drink any caffeinated beverages 2 hours prior to the experiment. The luminance in the booth was dim (approx. 100lux) for all participants during all experimental sessions to avoid floor or ceiling effects in the pupil diameter measures.

All sessions took place in a sound-proof booth. Pupil size was recorded through the Tobii Pro Spectrum eye tracker at a sampling frequency of 1200 Hz. All auditory stimuli were presented via five loudspeakers. The loudspeakers (Genelec 8040A; Genelec Oy, Iisalmi, Finland) were distributed around the participant along an imaginary circle with 1 meter radius at the side and back of the listener (+/- 90 and +/- 270). The target stimuli were presented from the loudspeaker in the front of the participants while the 16-talker background noise was presented through the remaining loudspeakers.

The procedure included an audiometry test and an ear examination at the beginning of either the first or the second sessions of the experiment. During the audiometry test pure tones were presented at 250, 500, 1000, 2000, and 4000 Hz. The intensity of the tones was decreased in 10 dB intervals until the participant did not detect the tone and increased in 5 dB intervals when participants did detect the tone. The lowest intensity at which the tone was detected at least twice out of 3 times was recorded.

Thereafter, the speech reception threshold corresponding to 50% correct sentence recognition (SRT50) was measured using a software running on MATLAB 2018b (MathWorks, Natick, MA). This included a practice round and a test round with 20 DAT (Nielsen et al., 2014) sentences each. Throughout the whole procedure the intensity of the background masker was 70dB (as averaged across 30 seconds). During the estimation of SRT50 the first sentence was presented at 70 dB and increased by 4dB until the participants repeated all keywords correctly. The intensity of the following sentences were decreased or increased by 4 dB, depending on whether the participants repeated all the keywords in the sentence correctly or not, respectively. The intensity of the fifth sentence was calculated by taking the average of the intensity levels of the first four sentences and the intensity level of what the fifth sentence would have been with a 4dB stepsize. For the rest of the sentences, the level of the target sentence was decreased or increased by 2 dB depending on correct or incorrect (incomplete) repetition of the keywords, respectively. The estimated SRT50 was calculated as the average of the last five sentences and what an extra sentence would have been presented at. Prior to the start of the experiment, to provide participants with an understanding of the task demands in the pre- and post- load sequence blocks, participants practiced with 20 sentences in blocks of 10, and received feedback from the experimenter per block on how many sentences were repeated correctly.

The experiment started with a 5-point eye-tracker calibration procedure. Thereafter, participants were explained that they could earn additional money by correctly repeating minimum 12 of the sentences out of blocks of 20 sentences. Participants were instructed to fixate on a cross on a black screen that was approx. 60 cm away from them and to schedule their blinks at the end of their response period. Before each of the 20-trial blocks, participants were informed about the amount of monetary incentives though a verbal message from the experimenter (e.g. “Now we are proceeding with a block of 20 sentences. If you repeat all keywords in more than 12 sentences correctly, you will receive 160 Danish krones”). In addition, visual reminders of the amount were placed at approximately 20 cm away from the fixation cross on the computer screen. These reminders stayed on the screen for the duration of each block.

Participants were informed beforehand that this session would take approximately 2 hours, but were not informed about how many blocks the experiment had in total. After completing the pre-load sequence blocks, participants took a 5-minute break. At the beginning of the load sequence, participants received written and verbal information on the load sequence. Immediately after the load sequence, participants continued with the post-load sequence blocks. After the experiment, participants completed the Need for Recovery scale (Van Veldhoven & Broersen, 2003a).

3.2.7. Pre-processing and calculation of the pupil dilation indices

Pupil data were processed using MATLAB R2018b (MathWorks, Natick, MA), similarly to the procedure reported in Wendt et al. (2018). Pupil traces starting 1 s preceding sentence onset until the end of the retention interval were selected for analysis. Traces with more than 30 % of missing data were considered invalid. Per participant, the data from the eye with most valid traces was selected for analysis. The data of a given participant were included in the analyses if 10 or more of the trials per block of 20 trials of each experimental condition were valid. Pupil diameter values that were more than 2 standard deviations away from the trial median were coded as blinks. Blinks were later interpolated in a linear fashion. The interpolation started for the samples within 35 ms before and ended for the samples within 100 ms after the blink to account for blink-related changes in pupil size. Later a moving average filter with a window of samples that would fall within 15 ms length was used to smooth the de-blinked trials and to remove any high-frequency artefacts. The trials were averaged per condition, and per participant. That is, there were 8 averaged traces per participant (2 probe time x 2 load sequence magnitude x 2 monetary incentive amount). For each averaged trace, the mean value of the trace corresponding to the 1-second interval before the onset of the target sentence was taken as the BPD. The mean pupil size during the interval from the onset of the sentence until the end of the presentation of the background babble was saved as MPD, and the maximum pupil size within the same interval was noted as the PPD.

3.2.8. Statistical analyses

To assess whether there was an overall change from pre- to post- load sequence, t-tests were conducted, where the dependent measures of interest were speech recognition performance, BPD, MPD, PPD, self-rated effort, self-rated performance, and self-rated tendency to quit, and the independent variable was pre- vs. post-load sequence (i.e. probe time). To rule out any effect of fatigue that is not related to task performance, for example existing fatigue that diminish the baseline pupil size or the task-evoked pupil response, pre- and post- load sequence difference scores were calculated. Later repeated measures ANOVA's were conducted, where the change from pre- to post- load sequence in the same pupil and self-report measures were the dependent variables, and load sequence magnitude (light and heavy) and monetary incentive amount (high and low) were the independent variables. Finally, repeated measures correlations were used to explore the relationship between subjective and objective measures of effort.

3.3. Results

3.3.1. SNRs used in the probe blocks and performance

The average SNR corresponding to the SRT50 of the participants in a background noise of 70 dB was -4.8 dB (range: -7.4 - -1.6; *SD* = 1.5 dB). Figure 3.3 shows pre- and post-load sequence speech recognition performance as a function of load sequence magnitude and monetary incentive amount. The average performance was 51.7 % (*Median* = 52.5, *SD* = 10.9) and did not change from pre- to post- load sequence, [$t_{paired}(17) = 0.695$, $p = .497$]. The repeated measures ANOVA with Δpost-/pre- performance score as the dependent variable, and load sequence magnitude (light vs. heavy) and monetary incentive amount (high vs. low) as independent variables showed no main effect of load sequence magnitude, [$F(1,17) = 0.716$, $p = .409$, $\eta_p^2 = .040$], no main effect of monetary incentive amount, [$F(1,17) = 0.250$, $p =$

.624, $\eta_p^2 = .014$], and no load sequence magnitude x monetary incentive amount interaction effect, [$F(1,17) < 0.001$, $p = 1$, $\eta_p^2 < .001$], on the Δpost-/pre-performance score.

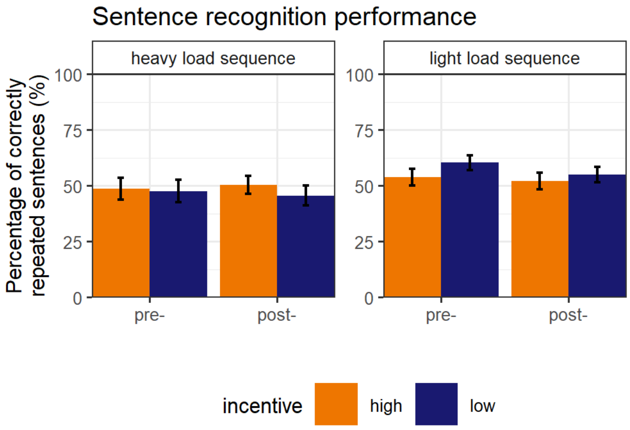


Figure 3.3. Pre- and post-load sequence sentence recognition scores, as a function of load sequence magnitude and the monetary incentive amount. The error bars represent the standard error of the mean (SEM) for within-subject variables (Cousineau, 2005; Morey, 2008).

3.3.2. Pupil indices

3.3.2.1. Pupil traces

Figure 3.4 shows the baseline-corrected pre- and post-load sequence pupil traces starting from the beginning of the baseline interval, until the end of the retention interval, averaged across the participants, and grouped together according to the load sequence magnitude and monetary incentive amount conditions. By visual inspection, the traces show a general decline in pupil dilation that was most evident after the heavy load sequence in the low monetary incentive condition.

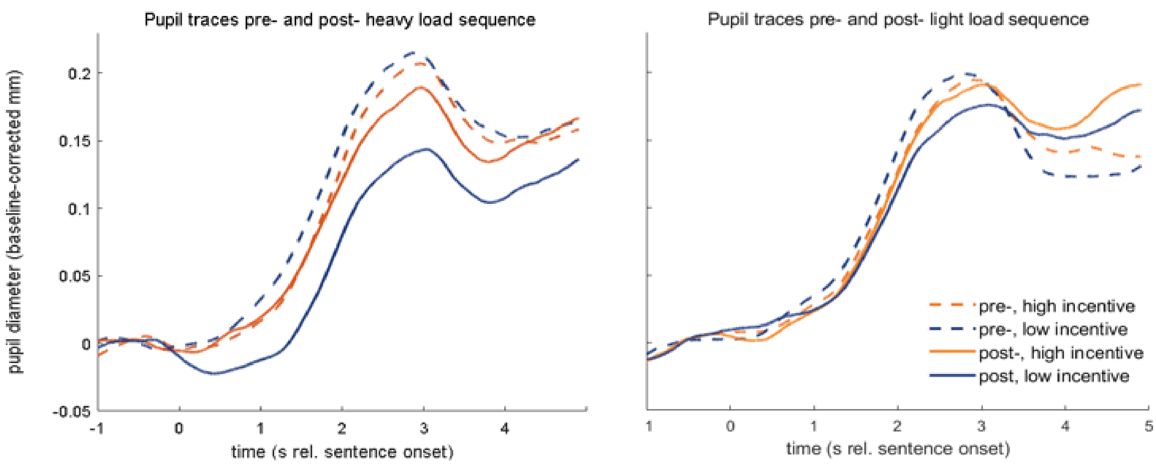


Figure 3.4. Pre- and post-load sequence pupil traces averaged across participants for the heavy (left) and light load sequence (right), and for high (orange) and low (blue) monetary incentive conditions.

3.3.2.2. Baseline pupil diameter

Figure 3.5. shows the BPD. On average pre-load sequence BPD was larger than post- load sequence BPD, [t_{paired} (17) = 4.599, $p < .001$]. The repeated measures ANOVA with Δpost-/pre- of BPD as dependent variable, and monetary incentive amount and load sequence magnitude as the independent variables showed no effect of monetary incentive amount, [F (1,17) = 0.639, $p = .435$, $\eta_p^2 < .036$], load sequence magnitude, [F (1,17) = 2.187, $p = .157$, $\eta_p^2 < .114$], or monetary incentive amount x load sequence magnitude interaction effect, [F (1,17) = 1.2, $p = .289$, $\eta_p^2 < .066$].

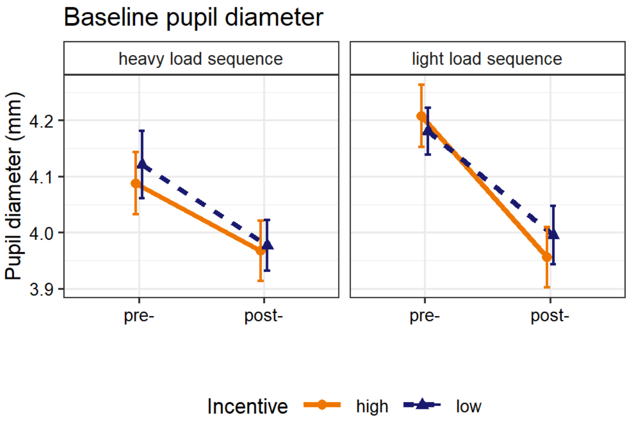


Figure 3.5. Pre- and post-load sequence baseline pupil diameter (BPD) as a function of load sequence and monetary incentives. The error bars represent

the standard error of the mean (SEM) for within-subject variables (Cousineau, 2005; Morey, 2008).

3.3.2.3. Peak pupil diameter

Figure 3.6. shows the extracted PPD (baseline-corrected). Averaged across the monetary incentive amount and load sequence magnitude conditions, overall, the difference in PPD between the pre- and post- load sequence conditions was not significant, [$t_{paired}(17) = 1.195$, $p = .248$]. The repeated measures ANOVA with Δpost-/pre- of PPD as dependent variable, and monetary incentive amount and load sequence magnitude as the independent variables showed a main effect of load sequence magnitude, [$F(1,17) = 4.899$, $p = .041$, $\eta_p^2 < .224$], as the decline in PPD from pre- to post-load sequence probe was larger after the heavy load sequence as compared to the light load sequence. In addition, there was a main effect of monetary incentive, [$F(1,17) = 11.886$, $p = .003$, $\eta_p^2 < .411$], as the decline from pre- to post- was larger when monetary incentives were low. Although the mean difference between the monetary incentive conditions in the decline from pre- to post load sequence was larger after the heavy load sequence as compared to the light load sequence, the monetary incentive amount x load sequence magnitude interaction effect on the Δpost-/pre- of PPD was not significant, [$F(1,17) = 1.286$, $p = .273$, $\eta_p^2 < .070$].

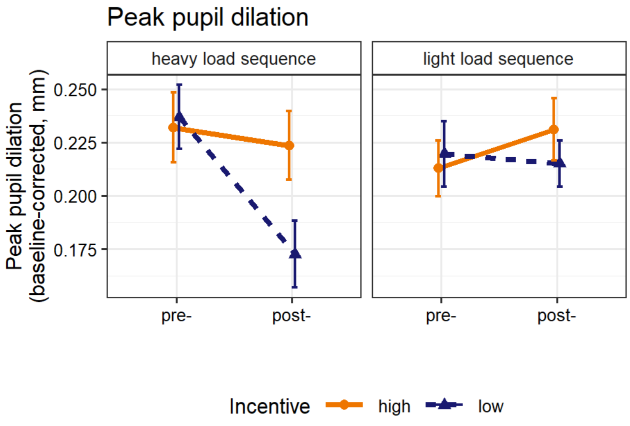


Figure 3.6. Pre- and post-load sequence peak pupil diameter (PPD) relative to baseline, as a function of load sequence and monetary incentives. The error

bars represent the standard error of the mean (SEM) for within-subject variables (Cousineau, 2005; Morey, 2008).

3.3.2.4. Mean pupil diameter

Figure 3.7. shows the MPD during the pre- and post-load sequence trials as a function of load sequence magnitude and monetary incentive amount. Visual inspection suggests overall a similar pattern in MPD to PPD. On average pre-load sequence MPD was larger than post-load sequence MPD, [$t_{paired}(17) = 2.336$, $p < .05$]. The repeated measures ANOVA with Δpost-/pre- of MPD as dependent variable, and load sequence and monetary incentive as the independent variables showed an effect of monetary incentive, [$F(1,17) = 16.094$, $p < .001$, $\eta_p^2 < .486$], as on average the pre- to post- decline in MPD was larger when the incentives were low. Although the mean decline in MPD from pre- to post- load sequence was larger after the heavy load sequence as compared to light, the main effect of load sequence magnitude was not significant, [$F(1,17) = 3.723$, $p = .071$, $\eta_p^2 = .180$]. The effect of monetary incentive amount on the Δpost-/pre- MPD was larger in the high load sequence condition, but the monetary incentive amount x load sequence magnitude interaction did not reach significance, [$F(1,17) = 2.533$, $p = .130$, $\eta_p^2 = .130$].

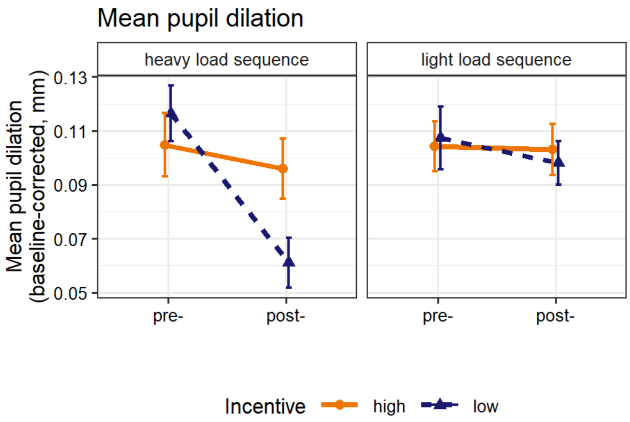


Figure 3.7. Pre- and post-load sequence mean pupil diameter (MPD) relative to baseline, as a function of load sequence and monetary incentives. The error bars represent the standard error of the mean (SEM) for within-subject variables (Cousineau, 2005; Morey, 2008).

3.3.3. Subjective ratings of effort, performance, and quitting

Table 3.1. shows pre- and post-load sequence self-rated effort, performance, and tendency to quit listening as a function of load sequence and monetary incentive. Shapiro-Wilk tests of normality confirmed that all the rating scores were normally distributed [all *p*'s > 0.05].

Table 3.1.

Self-rated effort, performance, and tendency to quit listening. Mean (s.d.) across participants

		Fatigue manipulation			
		heavy load sequence		light load sequence	
Ratings (range 0-100)	Probe time / Incentives	Pre-	Post-	Pre-	Post-
Effort	High	80 (3.11)	78.06 (3.34)	74.72 (3.39)	75.00 (3.84)
	Low	79.44 (3.08)	78.33 (3.11)	74.44 (3.64)	75.83 (2.84)
Performance	High	52.50 (4.70)	49.16 (3.98)	53.61 (3.63)	52.55 (4.52)
	Low	52.22 (4.61)	51.38 (4.86)	60.55 (3.78)	47.50 (4.04)
Tendency to quit listening	High	27.22 (4.85)	28.06 (3.73)	25.83 (4.66)	29.44 (5.21)
	Low	28.61 (3.81)	28.33 (4.53)	23.61 (5.06)	31.67 (5.53)

3.3.3.1. Self-rated effort

On average there was no evidence for a change in self-rated effort from pre- to post- load sequence, [$t_{paired}(17) = 0.200$; p =.844]. The repeated measures ANOVA with Δpost-/pre- of self-rated effort as dependent variable, and monetary incentive amount and load sequence magnitude as the independent variables showed no effect of load sequence magnitude, [F (1,17) = 1.253, p = .280, η_p^2 = .073], or monetary incentive amount [F (1,17) = 1.916, p = .185, η_p^2 = .107]. Nor was there a load sequence magnitude x monetary incentive amount interaction effect, [F (1,17) = 0.427, p = .523, η_p^2 = .026].

3.3.3.2. Self-rated performance

On average there was no evidence for a change in self-rated performance from pre- to post- load sequence, [$t_{paired}(17) = 1.797$; p =.090]. The repeated measures ANOVA with Δpost-/pre- of self-rated performance as dependent variable, and load sequence magnitude and monetary incentive amount as the independent variables showed no effect of load sequence magnitude, [F (1,17) = 0.773, p = .392, η_p^2 = .073], or monetary incentive amount [F (1,17) =

1.909, p = .185, η_p^2 = .101]. Nor was the load sequence magnitude x monetary incentive amount interaction effect significant, [$F(1,17) = 4.360$, p = .052, η_p^2 = .204].

3.3.3.3. Self-rated tendency to quit listening

On average there was no evidence for a change in self-reported tendency for quitting from pre- to post- fatigue, [$t_{paired}(17) = 0.200$; p =.844]. The repeated measures ANOVA with Δpost-/pre- of self-rated tendency for quitting as dependent variable, and monetary incentive amount and load sequence magnitude as the independent variables showed no effect of load sequence magnitude, [$F(1,17) = 0.961$, p = .341, η_p^2 = .053], monetary incentive amount [$F(1,17) = 1.156$, p = .697, η_p^2 = .009], or a load sequence magnitude x monetary incentive amount interaction effect, [$F(1,17) = 0.836$, p = .373, η_p^2 = .047].

3.3.4. Relations between subjective and objective measures

To explore the relationships between subjective and objective measurements of effort, repeated measures correlations between PPD and self-rated effort, and PPD and self-rated tendency for quitting were computed ("rmcorr"; Bakdash & Marusich, 2017). Figure 3.8. and 3.9. illustrate the PP and ratings. The analyses showed no significant association between PPD and self-rated effort scores, [*"rmcorr" r* (124) = 0.103, *CI 95%* = [-0.075 – 0.274], p = 0.250]. Nor was there a significant association between PPD and self-rated tendency to quit, ["rmcorr" r (125) = -0.052, CI 95% = [-0.225 – 0.124], p = 0.561].

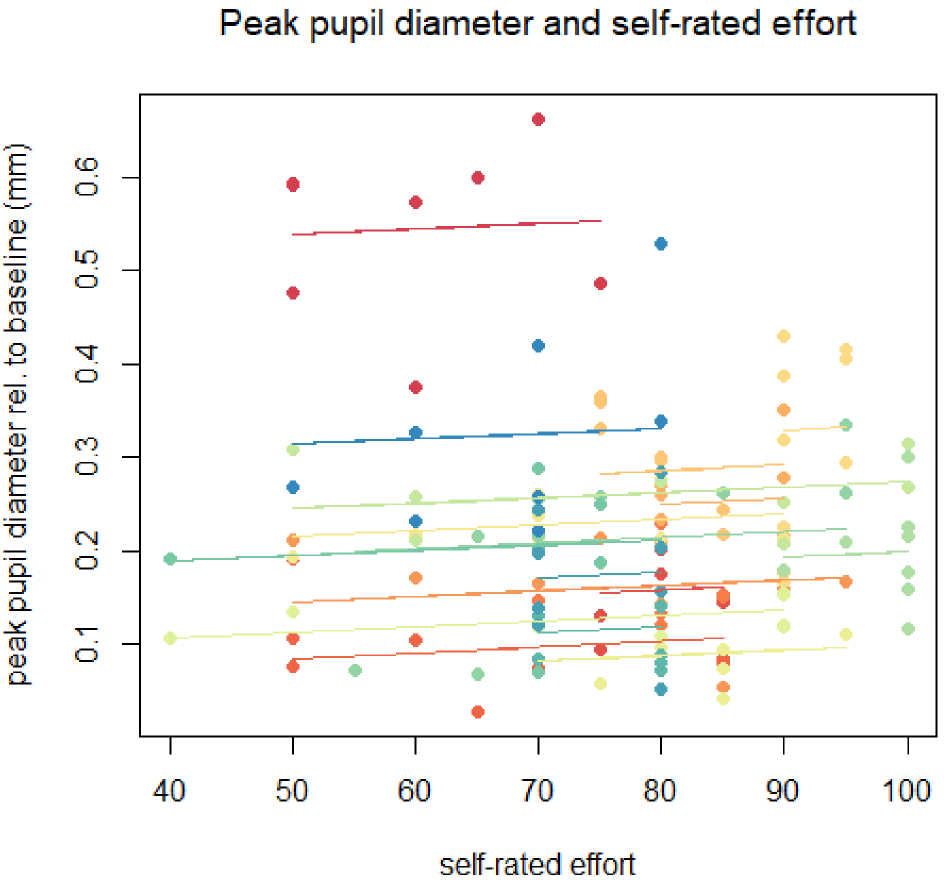


Figure 3.8. Scatterplot of peak pupil dilation and self-rated effort. Each dot represents the averaged value per monetary incentive, load sequence magnitude, and probe time condition per participant. Observations from the same participant are given the same color and the corresponding lines show the rmcorr fit for each participant.

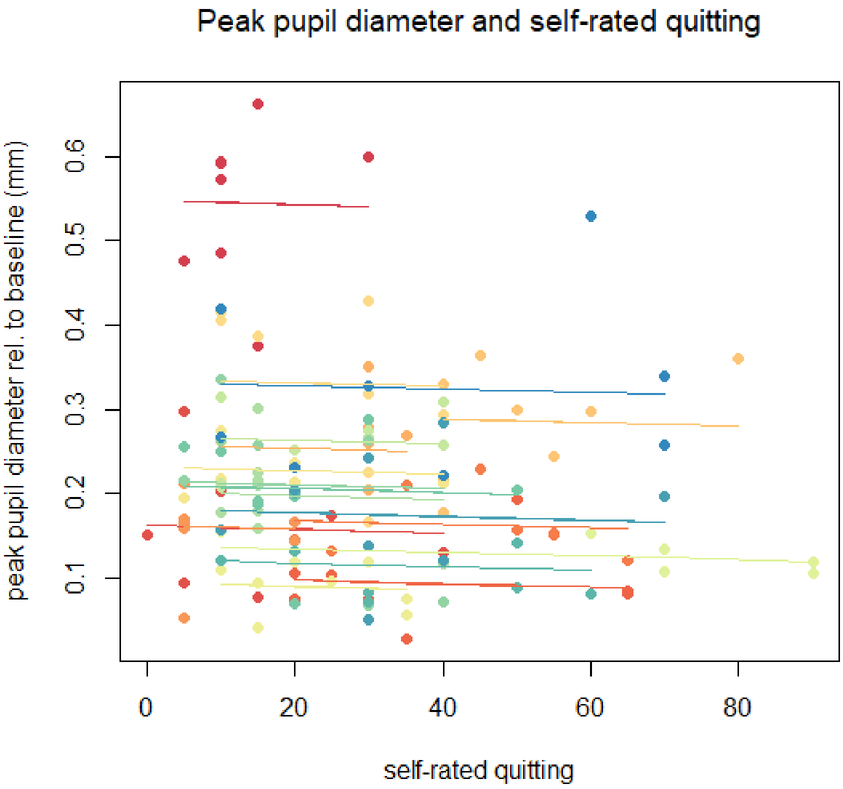


Figure 3.9. Scatterplot of peak pupil dilation and self-rated effort. Each dot represents the averaged value per monetary incentive, load sequence magnitude, and probe time condition per participant. Observations from the same participant are given the same color and the corresponding lines show the rmcorr fit for each participant.

3.3.5. Need for recovery

The Need for Recovery score of one participant was missing; the data of 17 participants was included in the analyses. The mean Need for Recovery score in our sample was 34.17 (*Median* = 27.00; *SD* = 22). A Shapiro-Wilk test showed that the scores were normally distributed, [*W* = 0.917; *p* = 0.396]. Thus, a two-tailed Pearson's correlation between Need for Recovery and PPD score was calculated. The results did not show an association between Need for Recovery and PPD, [Pearson's *r* = -.277, *p* = .282].

3.4. Discussion

The main aim of the current study was to investigate with pupillometry how listening effort is influenced by listening-related fatigue and motivation. This was realized through the assessment of listening effort in a pre-/post fatigue experiment, where two levels of load sequence (light and heavy) were used to induce fatigue. Pre- and post-load sequence listening effort was evaluated in two levels of monetary incentives (high and low). We hypothesized a decline in BPD from pre- to post- load sequence (hypothesis 1). Furthermore, following the view that the willingness to invest effort in a fatigued state depends on the perceived value of investing effort, we hypothesized that any decline in MPD and PPD should be larger after the heavier load sequence (hypothesis 2), and smaller with higher incentives (hypothesis 3). The difference in decline between the heavy and light conditions should be particularly sharper when incentives are low (hypothesis 4). To the best of my knowledge this was the first study to induce varying levels of hearing-related fatigue on the TEPR in varying levels of motivation in a speech-in-noise task. Therefore, the magnitudes of the main effects of fatigue and its interactions with motivation were not possible to estimate a priori. To observe the contributions of all effects to the variance in the TEPR we took the approach of testing the mentioned hypotheses with Type III ANOVA's instead of singling out specific contrasts. The results of the ANOVA's show that our hypotheses were partially supported.

As expected, there was a decline in BPD from pre- to post- load sequence. This is in line with previous literature showing time-on-task effects on BPD while performing a speech-in-noise task (Alhanbali et al., 2020; Ayasse & Wingfield, 2020; Koch & Janse, 2016; Zekveld et al., 2010). It is also in line with reports of declining BPD in studies that investigated fatigue with effortful visual tasks (e.g. Hopstaken, van der Linden, Bakker, & Kompier, 2015). The decline in BPD did not depend on whether the load sequence was heavy or light, or whether the incentives were high and low. Previously Hopstaken et al., (2015) report a steeper decline over the course of a 3-back task as compared to 2-back and 1-back tasks. This effect was accompanied by performance improvement in the 3-back task and performance declines in the 2-back and 1-back tasks. The steeper decline in BPD in the 3-back task was therefore attributed to learning effects. Similarly, Ayasse & Wingfield (2020) report a decline in BPD over the course of a speech-in-noise task that was steeper in individuals with poorer hearing thresholds. Inconsistent with the fatigue interpretation, there was an improvement in performance. The authors concluded that the BPD reflects a combination of different factors. In our experiment there was on average no change in performance from pre- to post- load sequence, independently of whether the load sequence was heavy or light. Self-reported tendency to quit did not change from the pre- to the post-load sequence trials either. This suggests that there is no evidence of any subjective experience of fatigue. On the other hand, self-report scale of tendency to quit may have been insensitive to capture the experience of fatigue due to a response bias. That is, participants may have been unwilling to admit their tendency to give up as part of social politeness (Grimm, 2010). Although it is difficult to completely rule out or pinpoint contributions of fatigue, habituation, or learning on the decrease in BPD (alertness) that developed over the experiment, a decline in baseline is consistent with the decline of capacity that is viewed as a hallmark of fatigue (Kahneman, 1973). Future studies may employ different (e.g. implicit) subjective measures or cognitive control measures to better assess the development of the subjective experience of fatigue.

The decline in PPD was larger after the heavy load sequence as compared to that after the light load sequence. This suggests that participants exerted less effort after the heavy load sequence, and demonstrates that the decline in PPD is not solely a result of the mere passage of time. Given no change in performance or self-reported tendency to quit, one possible interpretation for the larger decline in PPD from pre- to post-load sequence in the heavier load sequence may be larger fatigue after the heavy load sequence than after the light load sequence. After experiencing larger fatigue, participants may have been less willing to exert effort for listening. Given no statistically significant change in performance, one could argue for a learning effect. Namely, that the decline in PPD may reflect learning of the minimum effort needed to sustain the same performance as pre-load sequence (this argument is elaborated further below). However, this argument is insufficient to explain the interaction of the probe-time x motivation effect (this counter argument is further elaborated below). The interpretation of the decline in PPD from pre- to post-load sequence as fatigue effects would be in line with frameworks and models of listening effort and fatigue

Another explanation for the larger decline in PPD after the heavy load sequence could be the carry-over of learning to invest minimum effort for listening in the heavy load sequence. The additional working memory demands in the heavy load sequence may have led participants to perform the speech-in-noise task with a different strategy that puts less burden on working memory. For example, we could speculate that instead of listening attentively to all the words in the sentence, participants may have learned to focus only on the last word, as these were the only ones to be repeated. After completing the heavy load sequence, participants may have continued using the strategy that requires the least effort.

The decline in PPD from pre- to post- load sequence was larger when the incentives were low. This result suggests that motivational factors play a role in determining how much listeners exert effort after a load sequence. Although habituation/learning effects on the PPD should not be ruled out, habituation effects alone are insufficient to explain the effect of monetary incentives on the change in PPD. The moderating effect of monetary incentives can be interpreted in consideration of the FUEL and complementary models of motivational fatigue (Hockey, 2010; Müller & Apps, 2019; Pichora-Fuller et al., 2016). That is, after performing the load sequence, investing effort may have been perceived as worthwhile on trials with high monetary incentives and less worthwhile on trials with lower monetary incentives.

The MPD mirrored the PPD in the decline from pre- to post- load sequence probe, supporting the view that previous effortful exertion may result in a decline in subsequent listening effort. Furthermore, the decline was larger when the incentives were low. This supports the view that motivation contributes to how much effort is invested for listening (Koelewijn et al., 2018; Pichora-Fuller et al., 2016; Plain et al., 2020). There was no evidence for an effect of the load sequence on the decline of MPD from pre- to post- load sequence. The MPD may be less sensitive than the PPD to reflect the changes in listening effort that are dependent on how heavy the previous effortful exertion was. These results support the understanding that the PPD may be a better candidate to assess listening effort than MPD (Winn et al., 2018)

Whereas in Experiment 1 (Chapter 2) we reported an interaction of monetary incentive and SNR on the BPD, here in Experiment 2 we did not observe any effect of monetary incentive on BPD, but on the PPD and MPD. The discrepancy may be attributed to the difference in the timing of the trials in the experiments. Whereas in Experiment 1 participants were given 0.5 second to prepare their response, in Experiment 2 they were given 3 seconds. The speeded nature of the trials in Experiment 1 may have prompted participants to increase their arousal in anticipation of the trials instead of during the trials. Furthermore, in Experiment 2 the incentives may have been more effective. Not only was a larger amount offered, but the threshold to receive the offer was also lower (70% in Experiment 1 vs. 60 % in Experiment 2). Whereas in Experiment 1 participants underestimated their performance, and hence arguably were not expecting to perform well enough to receive the extra money, participants in Experiment 2 gave more accurate performance estimations, and thus may have believed that they had a chance to earn the money. Although another plausible reason for better performance estimation in Experiment 2 might also have to do with that there was only one difficulty level (60%) in Experiment 2 compared to two difficulty levels (50% and 80%) in Experiment 1, which might have had an influence on performance estimation.Together, the larger amount of monetary incentives and the belief that it is possible to achieve the money may have better motivated participants in Experiment 2 to mobilize effort during listening.

To investigate the relationship between pupil size and daily-life fatigue, we conducted a correlation analysis with PPD and need for recovery scores. Whereas Wang et al (2018) observed a negative relationship between PPD and daily life fatigue, this was not observed in the current study. This is similar to the results of Experiment 1 and of Koelewijn et al (2018), where no relationship was observed. Whereas the sample of Wang et al (2018) consisted of HI individuals with relatively larger need for recovery, the sample in the current study as well as in Experiment 1 (chapter 2) and Koelewijn et al (2018) was NH individuals with much lower self-reported need for recovery. Thus, our result supports the idea that such relationship may be more pronounced in samples who report larger daily-life fatigue.

To investigate whether within-subject changes in the pupil metrics of effort relate to changes in self-reported effort, two repeated measures correlations were computed between (1) PPD and self-reported effort and (2) PPD and tendency for quitting. There was no association between PPD and self-reported effort. This result is in line with the results of McGarrigle et al (2020), where no within-subject association between self-reported effort and PPD was found. Whereas McGarrigle (2020) explored changes over time, our study included motivation and fatigue manipulations which may have elicited a larger range of effort responses than observed by McGarrigle (2020). We did not find an association between the PPD and self-reported tendency for quitting. This finding could be considered in line with the view that there may be a mismatch between the subjective fatigue state and the investment of effort (Hockey, 1997a, 2010).

3.4.1. Limitations

The study has some limitations that should be mentioned. Technical difficulties in the collection of pupil data led to the exclusion of 12 participants from the study, potentially limiting the power of our statistical analyses. In addition, the self-report measures that were used in the study may have not been adequate to capture all dimensions and components of motivation and fatigue, such as sensitivity to monetary incentives, sleepiness and tiredness. Such measures may have been useful in evaluating the motivation that is manipulated by the monetary incentives, and the mental state that the load sequence task has elicited. To better assess the motivation and fatigue manipulations, future studies could utilize sub-scales from validated questionnaires (e.g. Carver & White, 1994; Smets, 1995). Lastly, as this study explored the effects of task-induced fatigue on listening effort, the applicability of the mechanisms explored here to listening-related daily life fatigue is left unclear. Longitudinal studies utilizing physiological methods are needed to shed light on the development of daily-life fatigue.

3.4.2. Conclusions

To conclude, the present study demonstrates with pupillometry in a pre-/post-fatigue experiment that listening-related fatigue and motivation both influence the capacity mobilized for listening. The decline in PPD from pre- to post- load sequence was larger when monetary incentives were low. This suggests that after sustained effortful listening, whether individuals continue to exert capacity for listening depends on the listening goals. Furthermore, pre- to post- load sequence decline in PPD was larger after performing a more taxing load sequence, suggesting that the magnitude of previous task demands affects subsequent capacity invested in listening. Future research could examine whether and how previous listening demand and listening goals (for example, gaining monetary rewards) influence listening effort in adults with hearing impairment.

Chapter 4: Influence of mental fatigue and motivation on listening effort as measured by the baseline and peak pupil dilation in adults with hearing impairment

4.1. Introduction

As referred to in the introduction (Chapter 1), mental fatigue is well known to develop over time spent on effortful tasks (Boksem & Tops, 2008; Müller & Apps, 2019). To date one study investigated the development of fatigue during an effortful auditory task in adults with HI (Hornsby, 2013). Hornsby (2013) studied task-induced listening-related fatigue in a 2-hour long speech-in-noise task with concurrent memory load and visual reaction time (RT) tasks, in aided (with and without noise reduction) and unaided conditions. Participants reported larger fatigue in the post-task assessment as compared to that in the pre-task. This confirmed that the task was indeed subjectively experienced as fatiguing. Hornsby (2013) reported that there was a larger increase in the visual RTs over time-on-task in the unaided condition as compared to the aided conditions, but there was no difference in the change in RT between the aided conditions. Hornsby (2013) reported large variance in the development of subjective fatigue and RT that may have occluded the expected difference in subjective fatigue and RT between the aided and unaided conditions. The author acknowledges that this variance may partly be due to individual differences in task-related motivation (Hornsby, 2013).

It is well established that in a state of fatigue, whether individuals withdraw or spend further effort depends on their motivation (Hockey, 1997, 2010; Hopstaken, van der Linden, Bakker, & Kompier, 2015; Boksem & Tops, 2008; Massar, Csathó, & Van der Linden, 2018; Müller & Apps, 2019; although see Gergelyfi, Jacob, Olivier, & Zénon, 2015). For example, Hopstaken et al. (2015) has shown that in a 2-hour long visual N-back task NH adults exhibited reduced performance and declining BPD over time (Hopstaken, van der Linden, Bakker, & Kompier, 2015). In the last block of the session, when participants were offered incentives that depended on their performance, both BPD and performance recovered to initial levels (Hopstaken, van der Linden, Bakker, & Kompier, 2015). Recently Koelewijn and colleagues (2018) showed in NH adults that the offer of monetary incentives increased the PPD in a speech-in-noise task (although see Koelewijn, Zekveld, Lunner, & Kramer, 2021). The results of Experiment 1 and 2 (Chapter 2, 3) suggested that fatigue and motivation may influence BPD and PPD in NH adults. How fatigue and motivation interact in a speech-in-noise task in influencing BPD and PPD in adults with HI is to date not investigated.

One way to motivate adults with HI in during the performance of a speech-in-noise task could be to describe the HAs that they are wearing as new technology. Although monetary incentives are well known to be successful motivators in studies with young participants (Koelewijn et al., 2021), previous research within hearing sciences using monetary incentives have shown inconsistent results (McLaughlin et al., 2020). It could be argued that for the regular participants of HA test studies, experiencing the sound through cutting-edge technology is intrinsically rewarding. The description of a HA as novel technology is known to influence the perception of users on the benefit that they get from the HA (Bentler et al., 2003; Dawes et al., 2011, 2013; Naylor et al., 2015). Such placebo effects with products are thought to arise as users are prone to remember the rewarding experiences that are in line with their beliefs about the products (Wegman et al., 2018). Those rewarding experiences are thought to reinforce the existing beliefs that users have on the products (Wegman et al., 2018). Therefore, it could be argued that adults with HI may find listening with new HAs intrinsically rewarding.

In addition to the fatigue and motivation states that can be temporarily induced by the conditions of a task, existing daily life fatigue or trait-like motivation can arguably be related to the effort mobilized for listening (Carolan, Heinrich, Munro, & Millman, 2021). For example, Wang et al. (2018) report that in adults with HI smaller Need for Recovery (NRS) predicted larger PPD during a speech-in-noise task. Need for Cognition (NCS) is conceptualized as a stable individual difference in the tendency to engage in, and enjoy, effortful cognitive activities (Cacioppo et al., 1996; Cacioppo & Petty, 1982). The effort put in adopting novel (superior technology) products has been shown to depend on the cognitive motivation of individuals (Wood & Swait, 2002). Individual differences in NCS have been argued to be related to individual differences in listening effort, although this hypothesis has to date not been tested (Carolan, Heinrich, Munro, & Millman, 2021a; Schneider, Bernarding, Francis, Hornsby, & Strauss, 2019).

In experiment 1 and 2 (Chapters 2 and 3) no statistically significant change was observed from pre- to post- load sequence in self-reported tendency to quit. Although self-report measures of mental fatigue, such as the checklist individual strength (Vercoulen et al., 1994) could be argued to have larger face validity in measuring fatigue than the tendency to quit scale, In Experiment 1 and these were not used. Direct reference to fatigue was not used not to change participant’s natural motivation to perform. Furthermore, it is well established that gauging the subjective experience of states and emotions through self-report is not easy due to limited introspective ability of humans (Hofmann et al., 2005). The implicit measurement of mental states has been argued to have relatively larger validity for capturing unverbalized mental fatigue (Rund, 2018). One of the established implicit measure of mental states is the lexical decision task (Moreno & van Orden, 2001), where humans have been shown to respond more quickly to words that correspond to their mental states (Olafson & Ferraro, 2001)

The aim of the current study was to explore the influences of task-induced fatigue and motivation on listening effort in adults with HI in a speech-in-noise task. A pre-/post- fatigue experimental design was used (cf. Experiment 1 and 2; Chapters 2 and 3), where a speech-in-noise task with concurrent memory load (“load sequence”) was used to induce fatigue. To manipulate motivation for listening, HAs were described as novel technology. Pre- and post- load sequence listening effort was measured using pupillometry while participants listened with HAs that were described as either conventional or novel. In addition to the pupil metrics, self-reported effort, estimated performance, and tendency to quit listening were collected to gauge the subjective experience of effort. As the effects of state fatigue and motivation arguably depend on existing NRS and existing trait-like NCS, information on these attributes were also collected. In addition to investigating objective and self-reported fatigue, to investigate any subjective experience of fatigue that is not verbalized, a lexical decision task was used.

We hypothesized that in line with earlier research (Experiment 1 and 2; Chapter 2 and 3) BPD would decline from pre- to post- load sequence (Hypothesis 1). Furthermore, given that the effects of fatigue on effort manifest more consistently when motivation is low, any decline in PPD should be smaller with HA descriptions that motivate to listen (hypothesis 2). We reasoned that NRS should influence PPD, such that larger NRS should be associated with smaller PPD (hypothesis 3); and moderate the influence of probe time on PPD, such that individuals with moderate-to-large levels of NRS should get more affected by the load sequence (hypothesis 4). The NCS could influence the level of PPD, as individuals with larger NCS would be expected to invest larger effort in the experiment (Hypothesis 5). The NCS should moderate the influence of the HA descriptions on PPD, such that individuals with larger NCS should invest larger effort when listening with the HAs described as test-prototype (Hypothesis 6). The NCS and NRS, therefore, may interactively influence the impact of load sequence, such that the moderating effect of NRS could be dampened by individuals with a high NCS (Hypothesis 7). To explore the relationship between pupillometric and self-report measures of effort, repeated measures correlations were used (Bakdash & Marusich, 2017; Bland & Altman, 1995). On the one hand, they would be expected to correlate because they measure a similar construct (Hypothesis 8). On the other hand, physiological measures may capture variance that is too small to be consciously accessible, thus, restricting the correlation between the measures (Hypothesis 9). Therefore, the relationship between objective and subjective measures of effort was an open question. Lastly, we hypothesized a decrease in RT from pre-to post-load sequence in the lexical categorization of fatigue-related words that is consistent with the induction of fatigue (Hypothesis 10).

4.2. Methods

4.2.1. Experimental design

Figure 4.1 depicts the experimental design. To investigate how fatigue and motivation interactively affect listening effort, a pre-/post- fatigue experiment was designed. Pre- and post- load sequence speech-in-noise tasks were performed in two HA description conditions (“test-prototype” vs. “own-conventional”) in a within-subject design. In these ‘probe’ blocks, listening effort was measured using pupillometry and self-report.

The load sequence block consisted of a speech-in-noise task with concurrent memory load, the sentence-final word identification and recall (SWIR) task that is known to be cognitively taxing (Ng, Rudner, Lunner, & Rönnberg, 2013, cf "heavy load sequence" in Experiment 2, Chapter 3) . The SWIR task consisted of 105 trials (approx. 40 minutes long). The sentences were sampled from the Danish HINT corpus (Nielsen & Dau, 2011). In each trial participants were asked to repeat the last word of the sentence that was presented to them. In addition, they were asked to memorize the last words of the last 7 sentences, and to recall these 7 words at the end of each 7th sentence. The SNR in the load sequence was fixed at individually pre-determined 80% speech reception threshold (SRT80). In case a participant could not recall any of the words during the first 2 7-sentence blocks, then the SNR was increased by 2 dB for the remaining of the blocks. This was to ensure that participants could hear well enough to understand and keep words in memory (Ng et al., 2013).

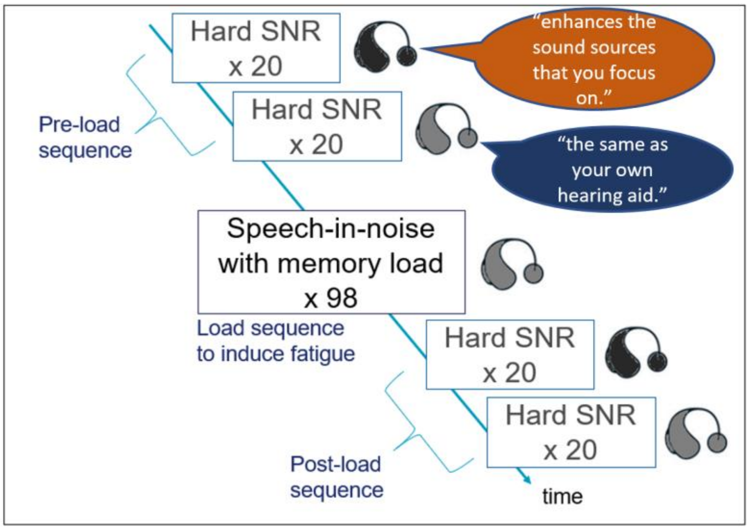


Figure 4.1. Experimental design. Participants engaged in a speech-in-noise task of 98 trials with concurrent memory load (i.e. "load sequence") without any breaks. Blocks of 20 trials were administered pre- and post- load sequence, where the SNR was individually pre-determined and fixed at sentence-based 50% speech reception threshold (SRT50). In the 20-trial blocks, participants wore HAs that were the same as their own. The HAs were the same in technology but differed in color. Participants were told that the gray HAs were the same as their own whereas black HAs were a prototype in test.

The pre- and post- load sequence probe blocks consisted of 2 blocks of 20 trials each. The SNR in these blocks was fixed at individually pre-determined 50% speech reception threshold (SRT50). The sentences were sampled from the Danish Dagmar-Asta-Tine corpus (Nielsen, Dau, & Neher, 2014). In one of the blocks participants wore a HA that was the same as their own ("own-conventional" condition). This was in gray color for all participants, because the HAs owned by many participants were also in gray color. In the other pre- and post- load sequence probe blocks participants wore a HA that was in reality the same as their own; but, was described as a cutting-edge HA in development ("test-prototype" condition), which would enhance the sounds in the focus of the user (English translation): "We are working towards getting HAs to help users hear exactly what they want without having to do anything. It's still a long way off, but the test devices you'll get on today will help you to amplify the sound source that you're focusing on." (In Danish: "Vi arbejder hen imod at få høreapparater til at hjælpe brugerne med at høre præcis det, de gerne vil, uden at de skal foretage sig noget som helst. Det er et stykke vej endnu, men de test-apparater, du vil få på i dag, vil hjælpe dig med at forstærke den lydkilde, du retter din opmærksomhed mod). Given that the regular participants of the Eriksholm Research Centre are known to be motivated to explore and experience new technology, it was assumed that this explanation would motivate participants to focus on the target source to experience the consequence of their focus. The color of the the hearing aid in the test-prototype condition was black.

For each participant the order of the HA description conditions was the same in the pre- and post- load sequence blocks. Across participants this order was counterbalanced.

In this study the load sequence was only used to induce fatigue. The investigation of the pupil size during the load sequence (see Appendices 7 and 8 for plots of BPD and PPD during the load sequence) was not part of the research questions. The load sequence was assumed to have induced fatigue and was not investigated further.

4.2.2. Participants

Eighteen adults (age range = 28-75, *M* = 65.778 years, *SD* = 12.539) with HI (defined as mean PTA between 35 - 65 dB HL at the octave frequencies between 250 and 4000 Hz, with left-right difference no greater than 20 dB) participated. All participants were native speakers of Danish. They had been wearing Oticon Opn S hearing aids for at least 3 months. This inclusion criteria was to ensure that all participants had acclimatized to the HAs that were to be used in the experiment, such that the contrast in the perceived novelty between the own-conventional and new-prototype conditions would be as large as possible. Participants reported normal or corrected-to-normal vision, no eye conditions, and no psychiatric or neurological conditions. All participants signed a standard informed consent that had been approved by the local ethics committee.

Power calculations for a small to medium effect size of d = 0.25 for the HA description x probe time interaction using GPower 3.1.9.7. software showed that at least 30 participants would be needed to observe the expected interaction effect. 26 participants were recruited for the study from the database of the Eriksholm Research Centre, part of Oticon A/S. Among those recruited, the data of 18 participants were clean to use for analyses (see pre-processing section below for inclusion criteria).

4.2.3. Trial sequence

Figure 4.2 illustrates the trial sequence in the pre- and post-load sequence probe blocks. Trials started 2.2 seconds after the experimenter pressed a button. A 16-talker babble noise (at 70 dB for all trials) was presented for 8 seconds. The 16-talker babble masker had speech segments of 2 female and 2 male talkers. The long-term average frequency spectrum of the masker was identical to that of the target speech signal of the target sentences. The target sentences started 3 seconds after the start of the babble noise. All sentences were 2 seconds long and had the same carrier words: “Dagmar tænkte på (~550 ms) … og en (~166 ms) … i går (~333 ms)” English: “Dagmar thought of … and … yesterday”); and 2 keywords embedded (~450 ms and ~500 ms, respectively). The babble noise continued for 3 seconds after the ending of the target sentences. Thereafter participants were asked to repeat back the keywords. The trial would be scored as correct if both keywords had been repeated correctly. The experimenter scored the answers by a button-press. The button-press of the experimenter would start the next trial.

Figure 4.2. Example trial sequence during the pre-and post-load sequence blocks. All trials started with 2.2 seconds silence, followed by 16-talker babble

noise. Then participants were presented with a sentence (e.g. Dagmar tænkte på en aften o gen festdag I går). The noise continued for 3 seconds after the end of the sentence. Participants were asked to repeat the keywords as accurately and as quickly as possible (e.g. "en aften, festdag").

4.2.4. Subjective report scales

After each of the 20-trial probe blocks, participants reported how much effort they mobilized for listening, how well they thought they performed, and how much they felt an inclination to quit listening on three 100-point visual analog scales (cf. Koelewijn, Zekveld, Lunner, & Kramer, 2018). The questions and anchors were (1) "*Hvor meget anstrengte du dig for at høre sætningerne? – EN: How much effort did you put into hearing the sentences?*" (0 = Ingen, 25 = Lav, 50 = Moderat, 75 = Høj, 100 = Meget høj – EN: 0 = None, 25 = Low, 50 = Moderate, 75 = High, 100 = Very High), (2) "Hvor mange af ordene tror du, at du forstod korrekt?" – EN: *How many of the words do you think you understood correctly*?" (0 = Ingen, 25 = Mindre end halvdelen, 50 = Halvdelen, 75 = Mere end halvdelen, 100 = Alle – EN: 0 = None, 25 = Less than half, 50 = Half, 75 = More than half, 100 = All), and (3) "*Hvor ofte måtte du opgive at forstå sætningen? – EN: How often did you have to give up understanding the phrase?* (0 = Aldrig, 25 = Mindre end halvdelen af tiden, 50 = Halvdelen af tiden, 75 = Mere end halvdelen af tiden, 100 = Altid – EN: 0 = Never, 25 = Less than half the time, 50 = Half the time, 75 = More than half the time, 100 = all the time).

4.2.5. Lexical decision implicit fatigue measure

A lexical decision task was used to measure implicit mental fatigue. In this task participants categorized visually presented words and non-words as either a word or non-word by a button press. The RT to the fatigue-related words (e.g. tired, sleepy, overworked) was compared to the RT to the neutral words (e.g. apple, umbrella, garage). The task was designed for the current experiment and piloted.

Participants in the pilot were 11 adults and native speakers of Danish. They had self-reported (corrected to) normal vision and no neurological or psychiatric conditions. A within-subject 2 x 3 design was used, where probe time (morning vs. afternoon) and word category (non-word, neutral-word, fatigue-word) were the independent variables, and RT was the dependent variable. Each trial in the experiment started with a fixation dot in the center of the screen (650 ms) to catch the gaze of the participant into the location where the words would be presented. Thereafter, a (non-) word was presented (max 2000 ms) in the center of the screen, until the button press of the participant. The button press of the participant led to the beginning of the next trial.

Pilot experiments started with a practice block that consisted of 49 trials (20 neutral-words and 29 non-words). After the end of the practice block, participants received feedback on their performance as their average RT and accuracy was presented to them on the screen. In case they did not have any questions or reflections about the task, they continued with the experimental block. The trials in the experimental block were the same as in the practice block. The experimental block consisted of 240 trials. There was no feedback after the experimental block. Open Sesame 3.3.3 software running on Python 3.7.6 on a Lenovo L480 ThinkPad laptop was used to present the stimuli and record responses.

Each participant took part in sessions in the mornings (between 7 and 10 am) and afternoon (between 2.30 and 4.30 pm) on the same day. Participants were naïve to the fatigue-related research question. They were told to press the “m” key if the word did and “z” key if the word did not exist in the Danish lexicon. They were told that responding correctly and quickly were both equally important. Whereas the practice block took around 2 minutes long, the experimental block took approximately 8 minutes to complete. After completing both sessions, participants were debriefed on the aims of the experiment. The pilot study showed that the RT to fatigue-related words was shorter in the afternoon sessions as compared to the morning sessions [$p < .05$], suggesting that the lexical decision task may successfully gauge experienced fatigue.

4.2.6. The Need for Recovery Scale

The Danish translation (unpublished) of the 11-item NRS (Van Veldhoven & Broersen, 2003) was used to measure daily-life fatigue. The concept of NRS describes the extent to which a task produces a need to recuperate from work induced fatigue, characterized by temporary feelings of overload, irritability, social withdrawal, lack of energy. When daily recuperation is not sufficient, individuals start the next day with residual need for recovery, which is thought to lead to lead to more adverse health outcomes in the long-term (Graham et al., 2020). The NRS thus is a concept that is in between short-term and long-term fatigue.

4.2.7. The Need for Cognition scale

The Danish translation (unpublished) of the 18-item (short) NCS scale (Cacioppo & Petty, 1982; Sadowsky, 1993) was used to measure (*Cronbach's alpha* = .86) the dispositional NCS .

4.2.8. Procedure and apparatus

Sessions were planned at either 9 am or 1 pm to minimize variability of diurnal changes in cortisol, as morning and mid-day are known to correspond to peak levels of cortisol (Lovallo & Buchanan, 2016). Participants were instructed to sleep well the night before the experiment, and not to drink any caffeinated beverages on the day of the experiment. They received the NRS through email and completed it before coming to the laboratory for the experiment.

All sessions started with the estimation of the SRT50 and SRT80 through an adaptive procedure in MATLAB. Throughout the estimation procedure participants wore their own HAs. Both estimation procedures included a practice round and a test round of 25 DAT (Nielsen et al., 2014) sentences each. The intensity of the background masker was kept constant at 70dB (as averaged across 30 seconds), whereas that of the target was adapted throughout the procedure. The first sentence was presented at 70 dB and increased by 4dB until participants repeated all keywords correctly. The intensity of the following sentences was decreased or increased by a stepsize of 4 dB, after incorrect and correct responses, respectively. The average of the intensity levels of the first four sentences and the intensity level of what the fifth sentence would have been with a 4dB stepsize was calculated as the intensity level of the fifth sentence. For the remaining sentences, the level of the target sentence was decreased or increased by 2 dB depending on correct or incorrect (incomplete) repetition of the keywords, respectively. The estimated SRT50 was calculated as the average SNR of the last five sentences and what an extra sentence would have been presented at. The SRT80 was estimated with the same procedure to that of SRT50, except that the stepsizes for the first four sentences were -1.6 dB in correct and +6.4 dB in incorrect repetition. After the 5th sentence, the stepsizes were -0.8 and +3.2 dB, respectively. All stimuli were presented through loudspeakers in a sound attenuated booth.

After the SRT50 estimation participants switched to wearing the test HAs. When wearing the gray-colored HAs they were told that the HAs were programmed identically to their own (i.e.“own-conventional” condition). When wearing the black colored HAs they were explained that this was a prototype HA (i.e. “test-prototype” condition). In fact, both pairs of HAs were programmed the same, and identically to the individual participant’s own HAs.

Following a 5-point eye-tracker calibration with the Tobii spectrum pro screen, participants were instructed to keep their gaze at a white- colored fixation cross on a dark screen that was approx. 1 meter distance. The pupil diameter was tracked at a sampling rate of 1200 Hz. The illumination in the booth was fixed at dim light (100 lux) to elicit a mid-size pupil diameter and thus prevent floor and ceiling effects in the measurement of task-evoked pupil responses.

The test HA’s to be used in the probe blocks were programmed before the sessions using the Oticon Genie 2 software in identical manner to the ones owned by the participants in all settings, including anti-feedback system, gain, transient and spatial noise management, noise reduction, directionality, openSound navigation, and speech rescue configuration. The OpenS engraving on the HAs were hidden with gray and black stickers.

The pre-load sequence block included 25 trials with the black HA, and 25 trials with the gray HA in the SNR fixed at the pre-determined SRT50. After each 25-trial block, participants completed the visual-analog self-report ratings using pen and paper. The scales were in 18 font size on A4 paper. After the second pre- load sequence block, participants were encouraged to take a break, visit the bathroom if needed, and go outside for fresh air. Throughout the load sequence participants wore their own HAs. After the load sequence participants followed the same procedure as in the pre-load sequence. A debriefing letter was sent to the participants at the end of the data-collection period, where participants were informed that both HAs in the probe blocks were the same in technology and programmed identically to their own, but differed only with respect to their paint color.

4.2.9. Pre-processing and calculation of the pupil dilation indices

Pupil data were processed using MATLAB R2018b (MathWorks, Natick, MA), similarly to the procedure reported in Wendt et al. (2018). Pupil traces starting 1 s preceding sentence onset until 3 seconds after the end of the target sentence were selected for analyses. Traces with more than 30 % of missing data were considered invalid (invalid trial). Per participant, the data from the eye with most valid traces (trials) was selected for analysis. The data of a given participant were included in the analyses if 10 or more of the trials per block of 25 trials of each experimental condition were valid. Pupil diameter values more than 2 standard deviations away from the trial median were coded as blinks. Blinks were later interpolated in a linear fashion. The interpolation started for the samples within 35 ms before and ended for the samples within 100 ms after the blink to account for blink-related changes in pupil size. Later a moving average filter with a window of 15 ms length was used to smooth the de-blinked trials and to remove any high-frequency artefacts. The trials were averaged per condition, per participant. That is, there were 4 averaged traces per participant (i.e. 2 probe time x 2 HA description). For each averaged trace, the mean value of the trace corresponding to the 1-second interval before the onset of the target sentence was taken as the BPD. The maximum pupil size during the interval from the onset of the sentence until the end of the background babble was saved as the PPD.

4.2.10. Statistical analyses

To explore the influence of probe time (pre- vs. post-load sequence), HA description (own-conventional vs. test-prototype), NRS, and NCS, and their interactions on the BPD and PPD, two linear mixed models were constructed, where BPD and PPD were the dependent variables; probe time (pre- vs. post-load sequence), HA description (own-conventional vs. test-prototype), NRS, and NCS, and their interactions were modelled as the independent variables using the lme4 package with default optimizers in R language (Bates et al., 2015). The “summary” and “anova” functions within the same package were used to obtain p-values for the t-tests performed with the Satterthwaite’s method for approximating degrees of freedom (Kuznetsova et al., 2017) as Satterthwaite’s approximation is thought to yield relatively few Type-I error rates in mixed effect models (Luke, 2017). To check whether the assumptions of the linear mixed modelling were met, all variables were inspected for linearity and homogeneity. Significance of the effects and interactions between them were assessed using a Type III Analysis of Variance (ANOVA) together with Satterthwaite’s approximation for degrees of freedom. This test was performed using the lmerTest package’s functionality (Kuznetsova et al. 2017).

To explore the influence of probe time (pre- vs. post-load sequence), HA description (own-conventional vs. test-prototype) on self- reported effort, performance, and tendency to quit listening, 3 repeated measures ANOVAs were conducted, where self-rated effort, performance, and tendency were the dependent variables; probe time (pre- vs. post-load sequence), HA description (own-conventional vs. test-prototype) were the independent variables.

To explore the relationships between subjective and objective measurements of effort, 2 repeated-measures correlational analyses (i.e. Bland Altman correlation) were conducted using the “rmcorr” package in R in default settings (i.e. without scaling; Bakdash & Marusich, 2017; Bland & Altman, 1995) , where PPD and BPD were the dependent variables; HA description (own-conventional vs. test-prototype), probe time (pre- and post-load sequence), were coded as repeated measure variables with equal variance; self-rated effort, and self-rated tendency for quitting were fixed variables with random intercepts but without random slopes.

4.3. Results

4.3.1. Speech reception performance

Figure 4.3 shows average speech reception performance (%) as a function of probe time (pre- vs. post- load sequence), and HA descriptions (own-conventional vs. test-prototype). Performance measures are plot in bar graph and not as point graph because the column from zero up until the mean point conveys that the participants had to achieve scores starting from zero until the mean to reach the level of performance. To investigate whether performance depends on probe time, HA description, Need for Recovery, and NCS, the fit of the linear mixed model with the notation [Performance ~ probe + description + NRS + NCS + probe : description + probe : NRS + description : NRS + description : NCS+ probe : description : NRS + probe : NRS : NCS + (1 | subject)] was constructed. The variance in performance explained by probe time, HA description, NCS and NRS was evaluated using a maximum likelihood test with Satterthwaite’s Type III ANOVA. The results showed that only the random effect of subject could improve model fit [$\chi 2$ (1) = 37.847, $p < 0.001$], while the fixed effects did not significantly improve the model [all p’s > 0.05]. The type III ANOVA showed that none of the fixed effects could predict the data [all p’s > 0.05].

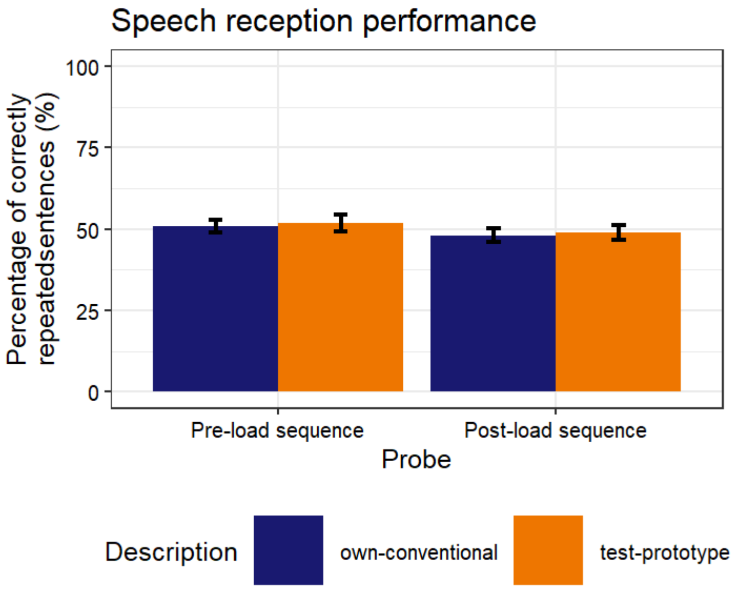


Figure 4.3. Speech reception performance during the pre- and post-load sequence probe blocks for the own-conventional and test-prototype HA description conditions. The error bars show the within-subject standard error of the mean (Cousineau, 2005; Morey, 2008).

4.3.2. Pupil dilation

Figure 4.4 shows pupil dilation during the probe blocks. The pupil dilation relative to baseline is plotted over time, starting from 1 second prior to the onset of the target sentence until 3 seconds after the end of the target sentence (in total approximately 6 seconds). The plot shows the averaged pupil traces during the pre- and post-load sequence blocks for the own-conventional and test-prototype conditions. Visual inspection of the averaged pupil traces shows that on average, the diameter peaked at approximately 1 second after the offset of the target sentence and declined thereafter for all conditions.

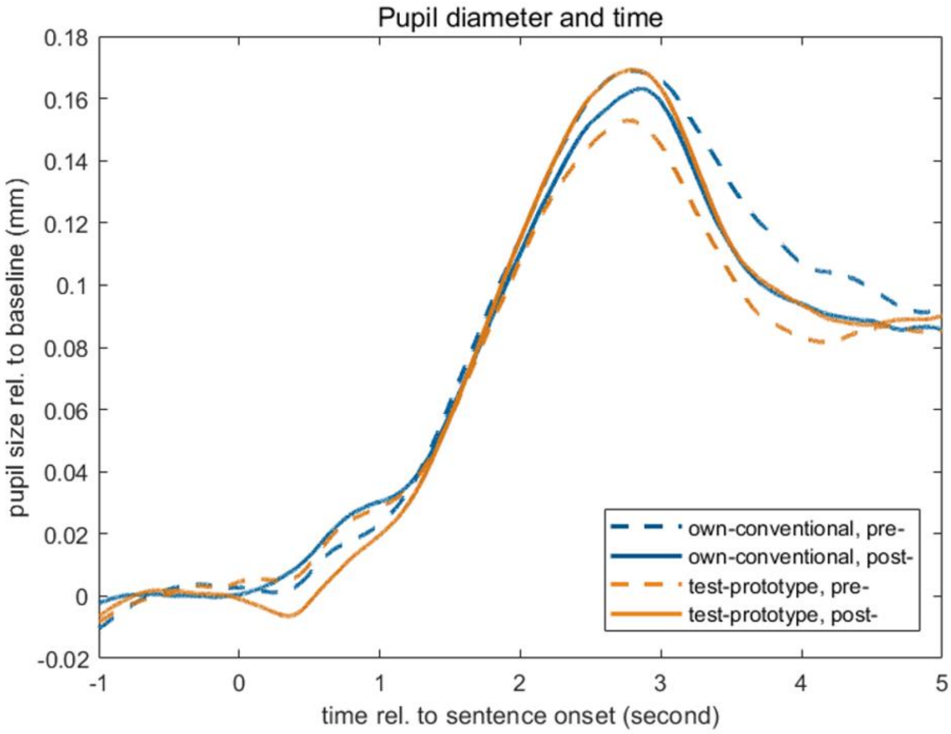


Figure 4.4. Pupil traces during the pre- and post-load sequence probe blocks for the own-conventional and test-prototype HA description conditions.

4.3.3. Baseline pupil diameter (BPD)

Figure 4.5 shows the baseline pupil diameter that is averaged across participants for the probe time and HA description conditions.

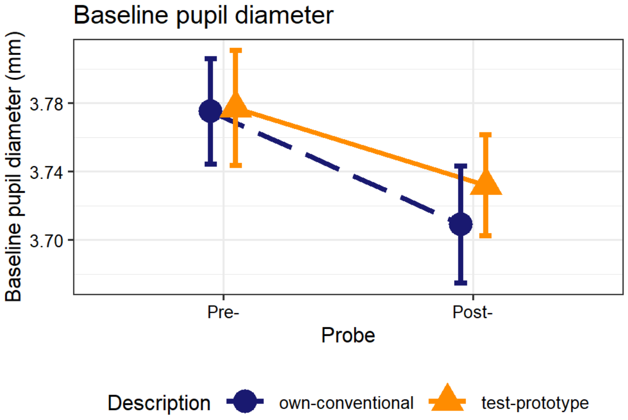


Figure 4.5. Baseline pupil size pre- and post- load sequence for the own-conventional and test-prototype HA description conditions. Error bars indicate the within-subject standard error of the mean (Cousineau, 2005; Morey, 2008).

The linear mixed model [notation: “BPD ~ probe + description + NRS + NCS + probe : description + probe : NRS + description : NRS + description : NCS+ probe : description : NRS + probe : NRS : NCS + (1 | subject)] was constructed to explain the variance in BPD with probe time, HA description, NCS and NRS (*AIC* = -181.5; See Table 4.1.); Overall, none of the fixed effects were statistically significant.

Table 4.1.

	β	*t* (54)	*Pr(>|t|)*
Probe time	-0.006	-0.106	0.916
HA description	-0.017	0.232	0.817
NRS	0.003	0.339	0.738
NCS	0.008	0.674	0.509
probe time x HA description	0.064	0.765	0.447
probe time x NRS	0.003	-1.904	0.062
HA description x NRS	<0.001	0.755	0.454
HA description x NCS	<-0.0005	-0.322	0.7490
Probe x HA description x NRS	<-0.0012	-0.708	0.482
Probe time 1 x NRS x NCS	<-0.0003	-1.406	0.177
probe time 2 x NRS x NCS	<-0.0003	-1.196	0.247

Figure 4.6. shows the BPD and the NRS of the participants. The Type III Analysis of Variance Table with Satterthwaite's method showed that the probe time x NRS interaction [F (1, 54) = 7.529, $p < 0.01$] contributed significantly to the variance in BPD.

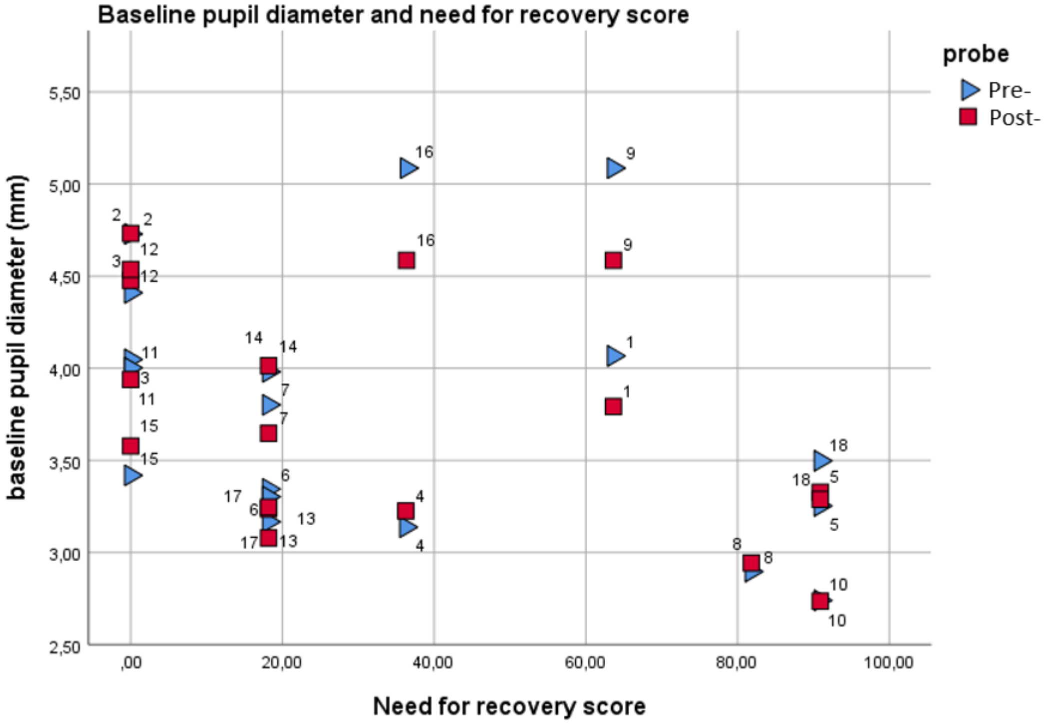


Figure 4.6. Baseline pupil diameter (BPD) in the pre- (blue) and post- (red) load sequence conditions and Need for Recovery scores for each participant blocks that are averaged across the HA description conditions. Observation from the same participant have the same number.

4.3.4. Peak pupil dilation

Figure 4.7. shows the PPD in the pre- and post-load sequence blocks when listening with HAs that are described as conventional and as new technology.

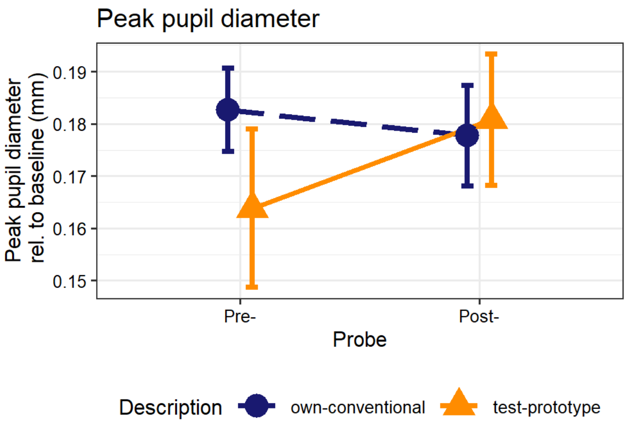


Figure 4.7. Peak pupil diameter in pre- and post- load sequence conditions when listening with hearing aids (HAs) for the own-conventional and test-

prototype conditions. Error bars indicate the within-subject standard error of the mean (Cousineau, 2005; Morey, 2008).

A linear mixed model [notation: “PPD ~ probe + description + NRS + NCS + probe : description + probe : NRS + description : NRS + description : NCS + probe : description : NRS + probe NRS : NCS + (1 | subject)] was constructed to explain the variance in the peak pupil diameter with probe time, HA description, NCS and NRS (*AIC* = -181.5). Overall, except from the HA description x NRS, none of the other variables had a significant contribution to the model.

Table 4.2.			
	β	*t* (54)	*Pr(>|t|)*
Probe time	-0.001	-0.158	0.875
HA description	< 0.0001	0.233	0.817
NRS	0.0001	1.12	0.275
NCS	0.001	0.902	0.378
probe time x HA description	< 0.001	-0.865	0.391
probe time x NRS	0.583	1.07	0.289
HA description x NRS	< -0.001	-2.044	0.045
HA description x NCS	< 0.0005	0.410	0.684
Probe x HA description x NRS	< 0.0001	2.12	0.0312
Probe time 1 x NRS x NCS	<-0.0001	-0.786	0.441
probe time 2 x NRS x NCS	<-0.0001	-1.676	0.109

Figure 4.8. shows the PPD and the NRS of the participants. The Type III Analysis of Variance Table with Satterthwaite's method showed that the probe x NRS interaction [*F* (1, 54) = 7.481, *p* < 0.01] and the probe x description x NRS interaction [*F* (1, 54) = 7.481, *p* < 0.05] contributed to the variance in PPD. The PPD and NRS for the pre and post-load sequence conditions was plot to understand the probe x NRS interaction. Visual inspection of the figure shows larger pre- to post- change in participants with

larger NRS. Due to the small sample size we opt not to further investigate these interactions.

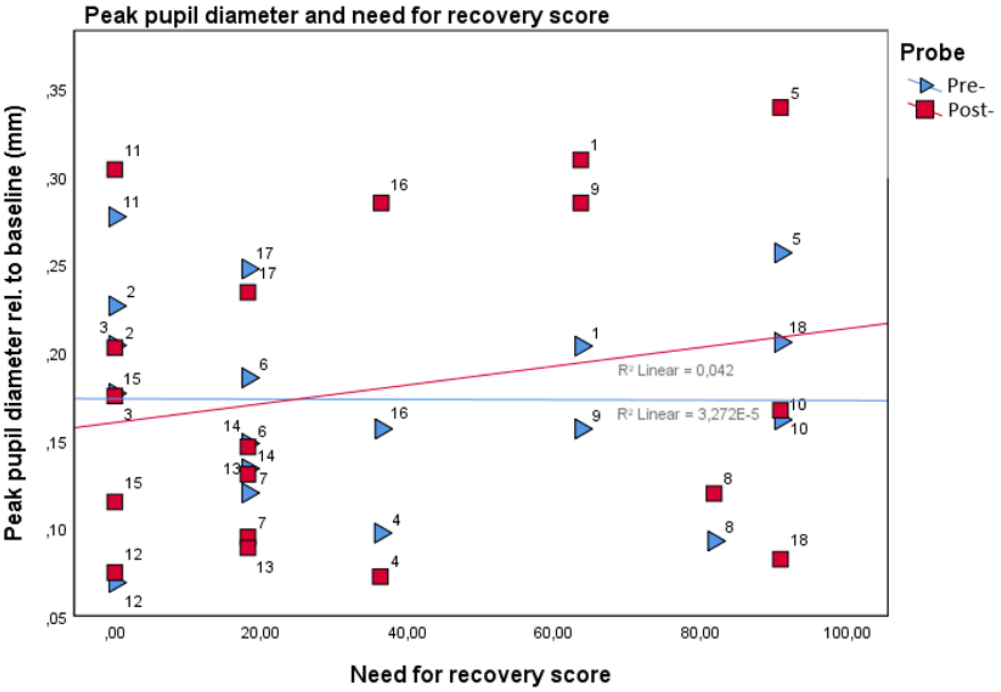


Figure 4.8. Peak pupil diameter (PPD) in the pre- (blue) and post- (red) load sequence conditions and Need for Recovery scores for each participant blocks that are averaged across the HA description conditions. Observations from the same participant have the same number.

4.3.5. Subjective ratings of effort, performance, and quitting

Table 4.1. shows mean self-rated effort, self-rated performance, and self-rated tendency to quit listening across for each probe block and for each HA description condition. The first repeated measures ANOVA with self-reported effort as the dependent variable, and probe time, HA description, and probe time x HA description interaction as the independent variables showed no main effect of probe time [$F(1,17) = 2.145$, $p = .161$, $\eta^2 = .112$] and no main effect of HA description [$F\ (1,17) = 0.602$, $p = .449$, $\eta^2 = .034$]. Nor was the probe time x HA description interaction effect significant [$F(1,17) = 1.699$, $p = .210$, $\eta^2 = .091$]. The second repeated measures ANOVA with self-reported

performance as the dependent variable and probe time, HA description, and probe time x HA description interaction as the independent variables revealed no effect of probe time [F (1,17) = 1.011, $p = .329$, $\eta^2 = .056$], of HA description [F(1,17) = 0.244, $p = .628$, $\eta^2 = .014$], or of probe time x HA description [F(1,17) = 1.151, $p = .703$, $\eta^2 = .009$]. Lastly, the repeated measures ANOVA, with self-reported tendency to quit as the dependent variable and probe time, HA description, and probe time x HA description interaction as the independent variables showed no effect of probe time [F (1,17) = 0.326, $p = .575$, $\eta^2 = .019$], HA description [F (1,17) = 1.454, $p = .244$, $\eta^2 = .079$], or probe time x HA description [F (1,17) = 3.136, $p = .095$, $\eta^2 = .156$], on self-reported effort.

Table 4.1.
Mean (s.d.) self-rated effort, performance, and tendency to quit listening

	Probe time			
	Pre-load sequence		Post-load sequence	
Hearing aid description	own-conventional	test-prototype	own-conventional	test-prototype
Self-rated effort (0-100)	79.4 (19.5)	74.2 (19.8)	79.4 (21.1)	80.3 (16.5)
Self-rated performance (0-100)	59.0 (18.7)	59.3 (15.9)	55.0 (17.3)	57.8 (14.7)
Self-rated tendency to quit listening (0-100)	24.2 (10.7)	27.6 (16.1)	33.1 (15.5)	26.7 (9.7)

4.3.6. Relations between peak pupil diameter and self-reported effort

Figure 4.9. shows a scatterplot of PPD and self-rated effort, where each dot represents an observation from an individual that is the mean PPD per probe block. Due to the larger influence of the incentive x load-sequence interaction on the PPD than that on MPD, PPD was favored for investigating the relationship between pupil and self-report measures. In Figure 4.9. per individual there are 4 observations that are plotted in the same color. The repeated-measures correlation (“rmcorr” Bakdash & Marusich, 2017) between PPD and self-reported effort showed a significant positive association [r (53) = 0.316, p = 0.018, 95 % CI: 0.050 – 0.540].

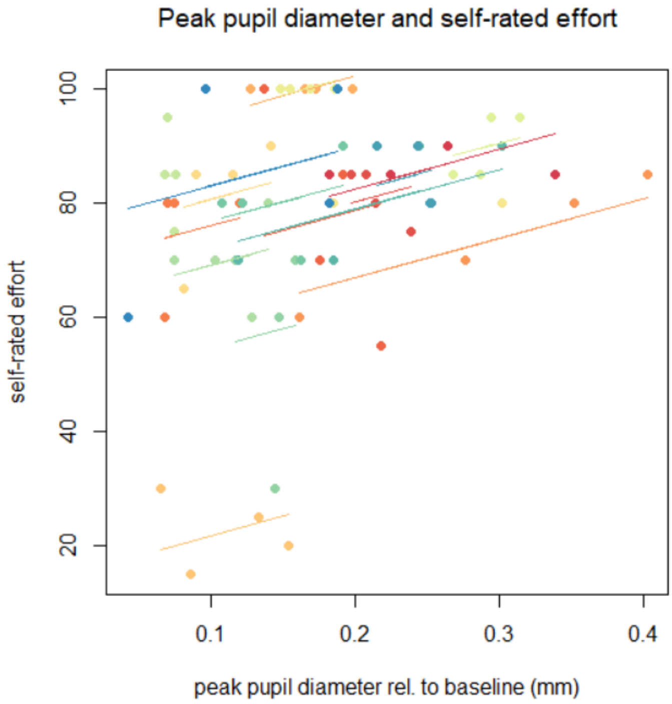


Figure 4.9. Scatterplot of peak pupil diameter and self-rated effort. Each dot represents the averaged value per HA description and probe time condition per participant. Observations from the same participant are given the same color and the corresponding lines show the rmcorr fit for each participant.

Figure 4.10 shows a scatterplot of peak pupil diameter and self-reported tendency to quit listening. The repeated measures correlation showed no significant association between the measures [r(53) =0.151, *p* = 0.268, 95% CI: -0.123- 0.405].

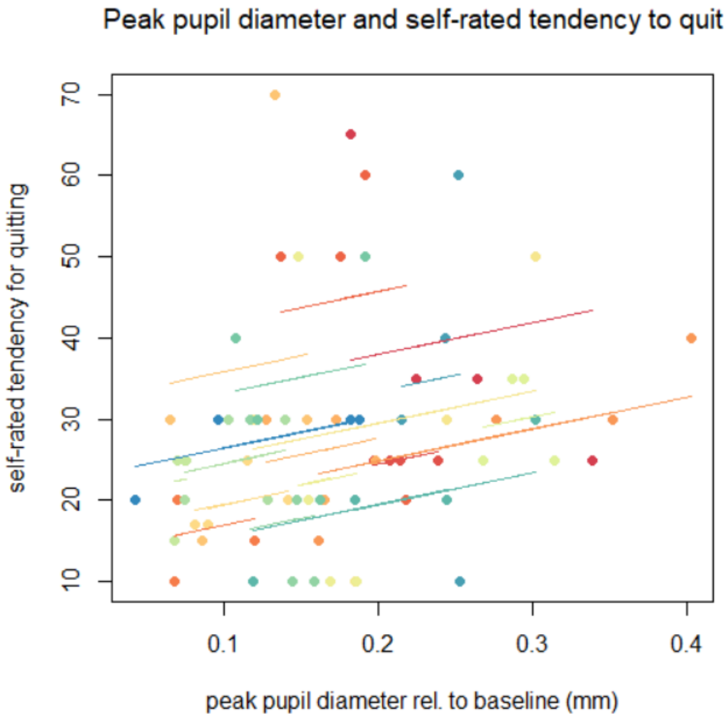


Figure 4.10. Scatterplot of peak pupil diameter and self-rated effort. Each dot represents the averaged value per HA description and probe time condition per participant. Observations from the same participant are given the same color and the corresponding lines show.

4.3.7. Lexical decision for semantically fatigue-related words

Table 4.2 shows the average RT per probe time and semantic category. The repeated measures ANOVA showed no significant main effect of probe time [*F* (1,17) = 0.052, *p* = .822, η^2 = .003]. There was a significant main effect of semantic category [*F* (1,17) = 124.586, *p* < .001, η^2 = .880]. On average RTs for fatigue-related words were longer than RTs for other words. Lastly, there was no significant probe time x semantic category interaction effect on RT [*F* (1,17) = 0.629, *p* = .439, η^2 = .036].

Table 4.2.

Reaction times for correct lexical decisions (SD in parenthesis)

Probe time Semantic category	Pre-load sequence	Post-load sequence
Fatigue-related word	797.41 (105.11)	790.97 (103.13)
Other word	706.07 (79.64)	706.21 (89.33)

4.4. Discussion

The aim of the current study was to investigate through pupillometry the influence of fatigue and motivation on listening effort in adults with HI. In a pre- /post- fatigue experiment, a speech-in-noise task with concurrent memory load (i.e. load sequence) was used to induce fatigue. To manipulate motivation for listening effort, HAs were (vs. were not) described as novel technology that would enhance the sounds that users focus on. Following previous research on the effects of time-on-task on BPD, we hypothesized a decline in BPD from pre- to post- load sequence (hypothesis 1). Following motivational theories of fatigue, we hypothesized that any decline in PPD should be larger when motivation for listening is low (hypothesis 2). Given that existing daily-life fatigue may decrease PPD and may moderate the influence of load sequence on PPD, we expected a main effect of NRS (hypothesis 3) and a probe time x NRS interaction effect on PPD (hypothesis 4). Similarly, we expected that larger NCS may increase PPD (hypothesis 5), boost the effect of HA descriptions on PPD (hypothesis 6) and moderate the NRS x load sequence interaction effect on PPD (hypothesis 7). As previous studies report inconsistent results on the relationship between pupillometric and self-report measures of effort, whether there would be a correlation between the pupillometric and self-report measures of effort was an open question. Lastly, the implicit subjective state of fatigue was investigated with a lexical decision task.

4.4.1. Relation between baseline pupil dilation, fatigue, and motivation

The analyses did not show a relationship between probe time and BPD. Furthermore, overall, the lexical decision task did not show reduced RT to fatigue-related words. These results suggest that overall the load sequence may not have been successful in inducing a decline in arousal or a subjective experience of fatigue. However, the ANOVA showed a significant interaction effect of probe time and NRS on BPD. The NRS and BPD plot (Figure 4.6) suggests that individuals who experienced moderate levels of daily-life fatigue showed the largest decline in BDP from pre- to post- load sequence. Individuals with smallest and largest levels of daily-life fatigue showed little change in BDP from pre- to post-load sequence. Furthermore, whereas the pre-load sequence BPD of the individuals with moderate daily life fatigue was the largest, that of the individuals with high daily life fatigue was smallest. Thus, to the extent that BPD is informative on pre-exiting fatigue, it could be speculated that the large diminish in BPD for individuals with moderate NRS may reflect a large impact of load sequence on the anticipatory resources mobilized for listening. Speculatively, the relatively small decline of BPD for the individuals with the largest daily-life fatigue may reflect floor effects. That is, the load sequence may have had little impact on anticipatory recruitment of resources as these participants already showed little preparatory capacity before performing the load sequence. Due to the small sample size no post-hoc test were run to investigate the source of the probe time x NRS interaction and the current results do not allow further speculation; however future studies could investigate the nature of the relationship between anticipatory pupil size and existing daily life fatigue.

To the best of our knowledge this is the first study to investigate the relationship between task-induced fatigue and pupil diameter metrics during a speech-in-noise task in a sample of HI adults. Wang et al. (2018) investigated relationships between daily life fatigue and pupil diameter metrics during a speech-in-noise task adaptively targeting 50% correct sentence recognition. They report no relationship between BPD and daily life fatigue. Wang et al. report results from a sample of mixed HI and NH adults, with on average larger BPD (4.95 mm) compared to that in the current sample (3.75 mm). The NRS scores are comparable on average (*M* = 39.35 and *M* = 39.84, respectively), although it is to note that the current sample had a slightly larger spread (*SD* = 28.7 and *SD* = 34.92, respectively). Thus, it could be speculated that existing moderate daily life fatigue may be most evident in the change of BPD after effortful exertion, rather than the in the absolute value. However, this conclusion should be interpreted with caution as the association between NRS and the change in BPD is of correlational nature.

4.4.2. Relation between peak pupil dilation, fatigue, and motivation

The analyses did not show a relationship between probe time and PPD, or an interaction between probe time and HA descriptions and PPD. However, the ANOVA showed a significant interaction effect of probe time and NRS on PPD. Visual inspection of the PPD against NRS (Figure 4.8) suggests a more positive relationship between NRS and PPD in the post-load sequence block than that in the pre-load sequence block. That is, for participants with larger NRS, on average visual inspection suggests a larger increase in PPD from pre- to post-load sequence. Considering the load sequence in the framework of understanding effortful listening and in the context of fatigue would predict a decrease in effort from pre- to post-load sequence that should be stronger for participants with stronger need for recovery. The observed increase in PPD from pre- to post- load sequence with increasing need for recovery, is thus, the opposite of what we expected. One interpretation of this trend could be increased effort in the post-load sequence to compensate for the fatigue induced by the load sequence. It could be speculated that this need to compensate could be stronger for participants who experience daily life fatigue. Inconsistent with this fatigue interpretation are, however, the lack of change in the tendency to quit listening and the lack of change in the RT to fatigue-related words. Together, these subjective and implicit measures suggest a lack of subjective fatigue experience. Another interpretation could be a habituation into the task that may have been stronger for the participants with larger daily life fatigue. Speculatively, for individuals with larger NRS, being asked to perform long load sequence may be uncomfortable and arousing in the beginning of the session. The dissipation of such feelings post-load sequence may have led to smaller BPD that allows larger PPD. The fatigue, learning, and habituation interpretations are of course not mutually exclusive. The change in PPD likely reflects a combination of these, which is common in investigations of fatigue (e.g. Faber, Maurits, & Lorist, 2012). Visual inspection of the PPD and NRS plot (Figure 4.8) suggests that the relationship between the change in PPD from pre-to post-load sequence may not be completely linear. Here due to the small sample size we opt not to fit non-linear functions. Future meta-analyses could further explore the nature of the interactive influence of task-induced and daily-life fatigue on PPD.

One explanation for the lack of decline in PPD from pre- to post- load sequence may be related to how the HA descriptions were interpreted by the participants. Before performing the load sequence, the variance in PPD when listening with HAs described as test-prototype was considerably larger than when listening with HAs that were described as the same as their own (Figure 4.7). After performing the load sequence, the variance in the test-prototype condition was diminished and comparable with that of the own-conventional condition (Figure 4.7). This suggests the influence of HA descriptions was larger in the pre- load sequence condition than post-load sequence. Although the intended effect of the test-prototype description was to motivate participants to listen to the target sentence, the description has a certain amount of ambiguity on whether the device makes it easier or more difficult to listen to the target sentence. One way to interpret the test-prototype hearing aid by the participants could be to deduct that listening with the test-prototype hearing aid should demand less effort as compared to that with their own. Speculatively, this interpretation may have potentially contributed to the overall large variance and, although not statistically significant, relatively low pre-load sequence PPD. The convergence of the test-prototype and own-conventional condition PPD means post-load sequence suggest an implicit learning that listening in both conditions requires roughly the same amount of mental effort. The lack of a difference between self-rated effort and self-rated tendency to quit listening supports the argument that the HA descriptions may not have the intended motivation manipulation. The lack of difference in performance between the HA descriptions could be used to argue for the non-success of the motivation manipulation. The lack of success in the motivation manipulation may have led to a lack of dispersion in PPD between a compensation strategy and suppression of effort after fatigue. Future studies may explore different ways of motivating participants to mobilize listening effort and may employ manipulation checks to test whether their manipulation worked.

4.4.3. Relation between PPD, self-reported effort, and self-reported tendency to quit

We observed a positive within-subject relationship between PPD and self-reported effort, and no relationship between PPD and self-reported tendency to quit. Previous research on the relationship between the subjective and objective measurements of effort mostly focused on between-individual differences and report no correlation (e.g. Koelewijn, Zekveld, Festen, & Kramer, 2012; Strand, Brown, Merchant, Brown, & Smith, 2018). Recently within-individual differences in the task-evoked pupil response in a speech-in-noise task have been shown to be associated with within-individual differences in subjectively reported tiredness (McGarrigle et al., 2021), but not effort. Our results support the argument that the analysis of within-individual differences may be more effective in revealing the relationship between subjective experience and objective observations during listening (McGarrigle et al., 2021). More research is needed to decipher the conditions under which such relationship holds.

4.4.4. Study limitations and future research

The study has a number of limitations which should be mentioned. Firstly, the experimenters (i.e. research clinicians) who informed participants of the HA conditions were not blind to the fact that both HAs were in reality programmed as the same. This may have affected the persuasiveness of the HA descriptions. Secondly, participants in the study knew how much time the experiment would take and may have thus regulated their arousal accordingly. That is, towards the end of the experiment, the knowledge that the experiment is about to finish may have led to increased motivation for listening in the post-load sequence trial. Increased motivation may have occluded the hypothesized reduction in effort. Lastly, the small sample size may have reduced statistical power to observe some of the hypothesized effects and did not allow us to run post-hoc tests; however existing results suggest an interactive influence of daily-life and task-induced fatigue on BPD and PPD. Future studies could further explore the nature of these relationships. In the current experiment ethical considerations limited the amount of deception that was used in the motivation manipulation and limited the length of the load sequence that was used as fatigue manipulation. Therefore, the motivation and fatigue manipulations were mild only, and may not be enough for participants to consciously experience or report.

4.4.5. Conclusions

To conclude, this study shows using pupillometry that need for recovery and previous load sequence interactively influence the BPD and PPD in a speech-in-noise task. We believe that this result is best interpreted within the framework of understanding effortful listening (Pichora-Fuller et al., 2016). That is, existing daily life fatigue and task-induced fatigue interactively influence listening effort. The influence of HA descriptions on the pupil metrics was not significant. Furthermore, we observed a positive relationship between PPD and self-reported effort, suggesting that objective and subjective measures of effort may capture shared variance. More research with more effective motivation manipulations is needed to investigate the nature of the interactive relationships between task-induced fatigue, daily-life fatigue, motivation, and listening effort.

Chapter 5. General discussion

5.1. Overview

The main aim of this work was to investigate fatigue and motivation effects in pupillometric measures of listening effort. These effects were examined over 3 experimental studies with speech-in-noise paradigms. For all experiments, a long, uninterrupted speech-in-noise task block (“load sequence”, 30- 40 minutes duration) was used to induce fatigue. Listening effort was examined during the pre- and post- load sequence speech-in-noise task blocks. The baseline pupil diameter (BPD) was defined as the average pupil diameter 1 sec before the onset of the target sentences. The mean pupil dilation (MPD; Experiment 1 and 2) and the peak pupil dilation (PPD; Experiment 2 and 3) were investigated as indices of listening effort. In Experiment 1, pre- and post-load sequence BPD and MPD were investigated in easy and difficult SNRs, and in high and low monetary incentive conditions in adults with normal hearing. In Experiment 2, the magnitude of the load sequence was varied in 2 levels (“heavy” and “light” load sequence) in sessions with adults with normal hearing. Experiment 3 investigated pre-and post-load sequence (heavy load sequence only) PPD in adults with hearing impairment. To manipulate motivation, the hearing aids (HAs) worn during the pre- and post-load sequence blocks were described to the participants as prototypes under evaluation. This condition was contrasted with the condition where the HAs were described as having been programmed in the same manner to the participants’ own ones (“test prototype” vs. “own conventional”). In addition to the pupil metrics, self-reported effort, performance estimation, and tendency to quit listening were collected throughout all experiments using self-report scales. Furthermore, need for recovery (daily-life fatigue) levels were collected from all participants. Correlations between pupillometric and self-report outcomes were computed.

5.2. Key findings relating to listening demands and motivation

5.2.1. The influence of listening demands on the BPD and MPD

In Experiment 1, MPD was larger with larger listening demands. Although the influence of listening demands on MPD was not the main interest of this work, this replication of the influence of listening demands on MPD (Borghini & Hazan, 2018; Koelewijn et al., 2015; Ohlenforst, Zekveld, Lunner, et al., 2017; Winn et al., 2016) shows that the experiment was set up successfully. The finding is in line with the predictions of the Framework for Understanding Effortful Listening (Pichora-Fuller et al., 2016). From the perspective of the FUEL, assuming sufficient willingness to perform, the capacity allocated to listening is predicted to be proportional to the demands of the listening task (Brehm & Self, 1989; Pichora-Fuller et al., 2016; Richter et al., 2016). In case of low SNR, larger mismatch between the sensory encoding of the target words and long-term memory representations, plausibly leads to larger explicit, top-down processing (Rönnberg et al., 2019), which places greater demand on working memory (Wingfield, 2016), and results in larger concomitant pupil dilation (Hess & Polt, 1964; Zenon, 2019) during listening.

In addition to MPD, the BPD was also larger in anticipation of listening in the difficult SNR condition. This is not uncommon in pupillometric studies of listening effort (Zekveld et al., 2010). One explanation for larger BPD could be the lack of time for the pupil diameter to return to lower levels. Another explanation could be preparation for the upcoming trial that was announced by the experimenter to the participant as difficult (Ayasse & Wingfield, 2020; Kahneman & Beatty, 1966). In Experiment 1, participants were asked to respond 500 ms after the presentation of the target sentence. This relatively short (Ohlenforst, Zekveld, Lunner, et al., 2017; Wendt et al., 2018; Winn et al., 2018) time frame to respond may have led participants to prepare for listening before the start of the target sentence, which may have led to larger BPD in difficult trials. In sum, in line with the predictions of the FUEL, in the difficult SNR condition, both the available capacity (here, BPD) and the allocation of capacity for listening (here, MPD) were larger (Pichora-Fuller et al., 2016).

5.2.2. The influence of listening demands in interaction with monetary incentives on the BPD and MPD

In Experiment 1, there was a listening demand x monetary incentive interaction effect on the BPD. Post-hoc analyses showed significant evidence for larger BPD with larger listening demand, only for the high monetary incentive condition. This suggests anticipatory effort in trials of high value. In the context of listening effort, the function of preparation could be considered in two ways. Firstly, preparation may be argued to facilitate the investment of maximum effort during listening. Individuals may regulate their arousal in anticipation of a trial to be at mid-level (Aston-Jones & Cohen, 2005) to facilitate working memory encoding and maximize listening performance (Nobre & Van Ede, 2018; Shalev & Nobre, 2021). However, if the pupil dilation during listening reflects effort, than one would expect better preparation to be accompanied by a larger task evoked dilation. There was no evidence for such a relationship in Experiment 1. Another interpretation of anticipatory BPD could be to consider it as a way to reduce the effort that individuals invest during listening. Pro-active use of control, for example by activating the properties of the target sentences in advance of their presentation (Braver, 2012; Miller & Cohen, 2001), which is known to be used particularly when willingness to perform well is high (Jimura et al., 2010), may be argued to reduce the effort required to attend to the target sentences during their presentation (Bonnefond & Jensen, 2012; Fukuda & Vogel, 2009; Jahfari et al., 2012). In experiment 1, there was no evidence for a monetary incentive effect or for a listening demand x monetary incentive interaction effect on the MPD. This observation may support the interpretation that anticipatory arousal may function to reduce the need to mobilize effort during listening, nullifying the effects of monetary incentives on the MPD. In sum, the results of Experiment 1 suggest an influence of motivation on capacity during difficult listening conditions.

5.3. Key findings relating to fatigue and motivation

5.3.1. The influence of probe time (pre- vs post- load sequence) on BPD, MPD, and PPD

Both in Experiment 1 and 2, there was significant evidence for a decline in BPD from pre-to-post load sequence. In Experiment 3, the mean BPD at post-load sequence was lower than that during pre-load sequence, but this was not statistically significant. From the perspective of FUEL, the decline in BPD could be considered as a decline in available capacity (Pichora-Fuller et al., 2016). The decline in BPD is consistent with previous reports of decline in BPD within experimental sessions or blocks of trials (“time-on-task” effects; see Zekveld et al., 2018 for a review). It can be interpreted as declining arousal that is plausibly due to habituating to the speech-in-noise task and to the environment (Nunnally et al., 1967; Peavler, W, 1974). Previously, declines in arousal during sustained speech-in-noise tasks have been interpreted as tiredness, drowsiness, and fatigue (McGarrigle et al., 2017b). It has been suggested that acute mental fatigue could result from the accumulation of inflammation in the same brain networks that implement top-down control, leading to reductions in cognitive control and gain (Aston-Jones & Cohen, 2005; Salamone et al., 2016). Although fatigue is a plausible interpretation for the declining BPD in the current work, the lack of subjective reports (and implicit indications) of fatigue, and lack of declines in performance from pre- to post- load sequence suggest lack of reportable or severe fatigue. Thus, this interpretation should be taken cautiously. In sum, Experiment 1 and 2 showed a decline in BPD from pre- to post- load sequence, suggesting a general decline in arousal with the fatigue manipulation.

Averaged across the monetary incentive and load sequence magnitude conditions, there wasn’t any sufficient evidence for an overall decline in MPD or PPD in any of the experiments, suggesting overall no evidence for a decline in listening effort. The lack of evidence for any decline in the self-reported effort and tendency to quit is in line with the pupillometric results.

5.3.2. The influence of probe time (pre- vs. post- load sequence) in interaction with load sequence magnitude on BPD, MPD, and PPD

In Experiment 2 the hypothesis that larger fatigue should lead to larger pre- to post-load sequence decline in anticipatory arousal (BPD) and listening effort (MPD and PPD) was investigated. To manipulate larger fatigue, the load sequence included a working memory task in addition to the speech-in-noise task (cf. Hornsby, 2013). There was no evidence for any interaction effect of probe time (i.e. pre- to post-load sequence decline) and load sequence magnitude on the BPD. If the decline in BPD reflected mental fatigue, then, assuming that heavier load sequence leads to larger fatigue, the decline in BPD should have been larger in the heavy load sequence condition. The current result thus suggests that the decline in BPD (arousal) is heavily governed by habituation (Aston-Jones & Cohen, 2005; Harve-Minvielle & Susan, 1995). Although habituation is not explicitly referred to in the quantitative model of listening related fatigue, the decline in BPD could be considered in a similar manner to the hypothesized decline in the “base motivation” parameter (Schneider et al., 2019). Base motivation is hypothesized to be large in the beginning of sessions due to interest in the task. It is hypothesized to decline throughout sessions independently from the accumulation of mental fatigue (Schneider et al., 2019).

Experiment 2 showed evidence for larger pre- to post-load sequence decline PPD in the heavy load sequence condition. Although the mean decline in MPD from pre- to post- load sequence was larger after the heavy load sequence, this effect was not statistically significant. Larger decline in PPD from pre- to post- load sequence suggests larger decline in listening effort after engaging in the heavy load sequence. This is in line with the theorizing that engaging in a task with larger demands leads to greater effort, which in turn reduces the willingness for further effort (Müller & Apps, 2019). From the perspective of the quantitative model of listening-related fatigue (Schneider et al., 2019), this finding supports the feedback loop between effort and fatigue. That is, our results support the prediction that after a period of effortful listening, the increase in listening-related fatigue leads to decreased motivation, which in turn decreases the mobilization of further effort for listening (Schneider et al., 2019). By experimentally demonstrating the link between mental fatigue and PPD, the current results extend the previous reports of correlations between daily-life fatigue and PPD (Wang et al., 2018). From the perspective of the FUEL, together, these results support the argument that mental fatigue should be taken into account in the audiological assessment of listening effort (Pichora-Fuller et al., 2016).

5.3.3. The influence of probe time (pre- post- load sequence) in interaction with monetary incentives on BPD, MPD, and PPD

Both in Experiment 1 and in 2 there was no evidence for the interactive effect of probe time x monetary incentives on BPD. This suggests little influence of monetary incentives on the pre- to post- load sequence change in anticipatory arousal. Similarly, in Experiment 3 there was no sufficient evidence for the probe time x HA description interaction effect on the BPD. Similarly, previous findings report no influence of monetary incentive on the BPD in a speech-in-noise task (Koelewijn et al., 2018). Together, these results suggest little influence of motivation on the time-on-task effects that are observed in BPD. To the extent that the BPD could be argued to be indicative of the base motivation value in the Quantitative Model of Listening-Related Fatigue, these results are in line with the decrease in base motivation that is independent of any external rewards and independent of any mental fatigue (Schneider et al., 2019). They suggest that the decline in BPD is somewhat governed by habituation (Aston-Jones & Cohen, 2005; Harve-Minvielle & Susan, 1995; Pichora-Fuller et al., 2016; Schneider et al., 2019).

In Experiment 1 there was no sufficient evidence for a probe time x monetary incentive interaction effect on MPD. In Experiment 2, the pre- to post-load sequence decline in MPD was larger with low monetary incentives. Similarly, in Experiment 2, the PPD declined from pre- to post- load sequence only in the low monetary incentive condition. In Experiment 3, there was overall no decline in PPD, and no sufficient evidence for a probe time x HA description interaction effect on PPD. The inconsistency regarding the probe time x motivation interaction effect across the experiments could be explained by the effectiveness of the monetary incentives. The offered monetary incentives were much larger in Experiment 2 than in Experiment 1. In addition, there is no previous empirical evidence for the motivating effect of HA descriptions on listening effort (Carolan et al., 2021b). Thus, it could be argued that the motivation manipulation was the most effective in this Experiment 2.

The probe time x monetary incentive interaction effect on MPD and PPD in Experiment 2 suggests lower listening effort after the load sequence, particularly when the willingness to perform well is low. This result supports the understanding that the mobilization of effort in a fatigued state depends on the motivation to perform (Hockey, 2010; Müller & Apps, 2019) and that this is also the case for listening effort (Schneider et al., 2019). It supports the argument that motivation and fatigue should be taken into account in the audiological measurement of listening effort (Pichora-Fuller et al., 2016).

5.3.4. The influence of probe time (pre- post- load sequence) in interaction with need for recovery on BPD and PPD

Another key finding in this body of work is the evidence for the probe time x need for recovery interaction effect on the BPD and PPD in Experiment 3. These effects suggest that the effectiveness of the load sequence manipulation may have depended on the need for recovery of the participants. Although the small sample size did not permit post-hoc analyses, visual inspection of the need for recovery x PPD plot does suggest a meaningful pattern that is worth reporting. For participants who report low need for recovery, the pre- to post- load sequence change in PPD was small, suggesting little influence of the fatigue manipulation on listening effort. For participants who reported larger need for recovery, the pre- to post- change in PPD was considerably larger, suggesting that fatigue manipulations may be more effective for participants who do usually experience larger need for recovery. This pattern illustrates the value of taking need for recovery into account when examining fatigue in samples with HI (Pichora-Fuller et al., 2016; Wang et al., 2018)

5.4. Limitations and future work

One limitation of the current work is the lack of evidence as to whether the fatigue and motivation manipulations worked. Although the current results show influences of load sequence and monetary incentive manipulations on the pupil metrics that are in the expected directions, the self-report measures did not show any indication of motivation or fatigue effects. One way to consider the discrepancy between the pupil and self-report findings could be to think that the pupil metrics reveal information that is in a finer scale and in greater precision than what is consciously accessible and reportable (Brouwer et al., 2015; Grassini & Laumann, 2020; Richter & Slade, 2017). Due to ethical considerations, the amount of the monetary incentives, the deception in the HAs descriptions, and the duration of the load sequence were not large. Thus, one could argue that the observed results in the pupil metrics do reveal information on fatigue and motivation that is not captured by self-report (Brouwer et al., 2015; Grassini & Laumann, 2020; Richter & Slade, 2017). Nevertheless, to help rule out non-fatigue or motivation-related explanations regarding the observed changes in the pupil metrics, future research could aim to develop or include independent verifications of the fatigue and motivation manipulations (for example, by using cognitive (executive) control tasks (van der Linden et al., 2003).

In this work extrinsic (money) and intrinsic (hearing aid description) rewards were used to manipulate motivation (Di Domenico & Ryan, 2017). One remaining challenge within hearing research is to manipulate social motivation in speech-in-noise tasks. It could be argued that social rewards are important motivators for conversation (Pichora-Fuller, 2016; Pichora-Fuller et al., 2016; Pielage et al., 2021). Future research with speech-in-noise tasks could incorporate paradigms from social psychology that tap into the fundamental and powerful need to belong to demonstrate the influence of social rewards on the pupil metrics (Roy F. Baumeister & Leary, 1995).

This work investigated the influence of fatigue and motivation only on the pupil metrics. One remaining question for investigation is the underlying mechanisms that lead to the increase or decrease of the BPD or PPD with fatigue and motivation. Previous studies show increased engagement of attentional and cognitive control networks with increasing difficulty in a speech-in-noise task (Zekveld et al., 2014). One could hypothesize that fatigue and motivation should influence the recruitment of those networks. To note is that although PPD changed with motivation and fatigue in Experiment 2, there was no evidence of any change in performance. One explanation for the changes in the pupil metrics without any changes in performance is that participants may be using different strategies to perform the task. These may be varying degrees of reliance on semantics, visual imagination, or phonology when keeping the words in mind for later repetition. Future research may explore whether motivation and fatigue affect the use of different strategies and the recruitment associated brain networks.

Lastly, a construct that is closely related to listening-related fatigue, but not covered by the current body of work is stress-related effort and fatigue. Qualitative research suggests that in addition to effort-related fatigue, fatigue that is experienced as a consequence of the frustration and stress, and the associated coping has a substantial influence on the daily-life fatigue of individuals with hearing impairment (Holman et al., 2019). Physiological research shows increased stress-related hormones in speech-in-noise tasks with HI adults after they have received stressing verbal nudges about their performance level (Zekveld et al., 2019). Recently, social observation during the performance of a speech-in-noise task has been shown to lead to larger PPD (Pielage et al., 2021). Thus, future research could investigate whether a load sequence that is heavy due to the requirement to manage social stress leads to similar declines in BPD and PPD.

5.5. Summary

Overall, this piece of work demonstrates the influence of motivation and fatigue on the pupil metrics (BPD, MPD, PPD) of listening effort. We show that particularly in difficult listening conditions, the anticipatory arousal and listening effort may be influenced by listening-related fatigue and motivation. This is in line with frameworks and models of listening effort and listening-related fatigue that consider motivation and fatigue as separate parameters in predicting effort.

References

Ackerman, P. L. (2010). 100 Years Without Resting. In P. L. Ackerman (Ed.), *Decade of Behavior/Science Conference. Cognitive fatigue: Multidisciplinary perspectives on current research and future applicationsCognitive fatigue: Multidisciplinary perspectives on current research and future applications.* (pp. 11–43). American Psychological Association. https://doi.org/10.1037/12343-001

Alavash, M., Lim, S. J., Thiel, C., Sehm, B., Deserno, L., & Obleser, J. (2018). Dopaminergic modulation of hemodynamic signal variability and the functional connectome during cognitive performance. *NeuroImage*, *172*(September 2017), 341–356. https://doi.org/10.1016/j.neuroimage.2018.01.048

Alhanbali, S., Dawes, P., Lloyd, S., & Munro, K. J. (2017). Self-Reported Listening-Related Effort and Fatigue in Hearing-Impaired Adults. *Ear and Hearing*, *38*(1), e39–e48. https://doi.org/10.1097/AUD.0000000000000361

Alhanbali, S., 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*, *0*(0), 1–11. https://doi.org/10.1080/14992027.2020.1853262

Ambler, B. A., Fisicaro, S. A., & Proctor, R. W. (1976). Temporal characteristics of primary-secondary message interference in a dichotic listening task. *Memory & Cognition*, *4*(6), 709–716. https://doi.org/10.3758/BF03213238

Aston-Jones, G., & Cohen, J. D. (2005). Adaptive gain and the role of the locus coeruleus-norepinephrine system in optimal performance. *Journal of Comparative Neurology*, *493*(1), 99–110. https://doi.org/10.1002/cne.20723

Ayasse, N. D., & Wingfield, A. (2020). Anticipatory Baseline Pupil Diameter Is

Sensitive to Differences in Hearing Thresholds. *Frontiers in Psychology*, *10*(January), 1–7. https://doi.org/10.3389/fpsyg.2019.02947

Bakdash, J. Z., & Marusich, L. R. (2017). Repeated measures correlation. *Frontiers in Psychology*, *8*(MAR), 1–13. https://doi.org/10.3389/fpsyg.2017.00456

Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Beatty, J. (1982). Phasic Not Tonic Pupillary Responses Vary With Auditory Vigilance Performance. *Psychophysiology*, *19*(2), 167–172. https://doi.org/10.1111/j.1469-8986.1982.tb02540.x

Bench, J., Kowal, A., & Bamford, J. (1979). The BKB (Bamford-Kowal-Bench) Sentence Lists for Partially-Hearing Children. *British Journal of Audiology*, *13*(3), 108–112. https://doi.org/10.3109/03005367909078884

Bentler, R. A., Niebuhr, D. P., Johnson, T. A., & Flamme, G. A. (2003). Impact of digital labeling on outcome measures. *Ear and Hearing*, *24*(3), 215–224. https://doi.org/10.1097/01.AUD.0000069228.46916.92

Bernarding, C., Strauss, D. J., Hannemann, R., Seidler, H., & Corona-Strauss, F. I. (2013). Neural correlates of listening effort related factors: Influence of age and hearing impairment. *Brain Research Bulletin*, *91*, 21–30. https://doi.org/10.1016/j.brainresbull.2012.11.005

Bess, H. F., & Hornsby, W., Y., B. (2014). Commentary: Listening Can Be Exhausting—Fatigue in Children and Adults With Hearing Loss Fred. *Ear and Hearing*, *35*(6), 592–599. https://doi.org/10.1097/AUD.0000000000000099. Commentary

Bijleveld, E., Custers, R., & Aarts, H. (2009). The Unconscious Eye Opener. *Psychological Science*, *20*(11), 1313–1315. https://doi.org/10.1111/j.1467-9280.2009.02443.x

Bland, J. M., & Altman, D. G. (1995). Calculating correlation coefficients with repeated observations. Part1- correlation within subjects. *Statistics Notes*, *310*. https://doi.org/: https://doi.org/10.1136/bmj.310.6977.446

Bo Nielsen, J., Dau, T., & Neher, T. (2014). A Danish open-set speech corpus for competing-speech studies. *The Journal of the Acoustical Society of America*, *135*(1), 407–420. https://doi.org/10.1121/1.4835935

Boksem, M. A. S., Meijman, T. F., & Lorist, M. M. (2006). Mental fatigue, motivation and action monitoring. *Biological Psychology*, *72*(2), 123–132. https://doi.org/10.1016/j.biopsycho.2005.08.007

Boksem, M. A. S., & Tops, M. (2008). Mental fatigue: Costs and benefits. *Brain Research Reviews*, *59*(1), 125–139. https://doi.org/10.1016/j.brainresrev.2008.07.001

Bonnefond, M., & Jensen, O. (2012). Alpha oscillations serve to protect working memory maintenance against anticipated distracters. *Current Biology*, *22*(20), 1969–1974. https://doi.org/10.1016/j.cub.2012.08.029

Borghini, G., & Hazan, V. (2018). Listening effort during sentence processing is increased for non-native listeners: A pupillometry study. *Frontiers in Neuroscience*, *12*(MAR), 1–13. https://doi.org/10.3389/fnins.2018.00152

Braver, T. S. (2012). The variable nature of cognitive control: A dual-mechanisms framework Shifting the emphasis to variability in cognitive control. *Trends in Cognitive Sciences*, *16*(2), 106–113. https://doi.org/10.1016/j.tics.2011.12.010.The

Brehm, J. W., & Self, E. A. (1989). The intensity of motivation. *Annual Review of Psychology*, *40*(1973), 109–131. https://doi.org/10.1146/annurev.ps.40.020189.000545

Brouwer, A. M., Zander, T. O., van Erp, J. B. F., Korteling, J. E., & Bronkhorst, A. W. (2015). Using neurophysiological signals that reflect cognitive or affective state: Six recommendations to avoid common pitfalls. *Frontiers in Neuroscience*, *9*(APR), 1–11. https://doi.org/10.3389/fnins.2015.00136

Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology*, *42*(1), 116–131. https://doi.org/10.1037/0022-3514.42.1.116

Cacioppo, J. T., Petty, R. E., Feinstein, J. A., & Jarvis, W. B. G. (1996). Dispositional differences in cognitive motivation: The life and times of individuals varying in need for cognition. *Psychological Bulletin*, *119*(2), 197–253. https://doi.org/10.1037/0033-2909.119.2.197

Carolan, P. J., Heinrich, A., Munro, K. J., & Millman, R. E. (2021a). Financial reward has differential effects on behavioural and self-report measures of listening effort. *International Journal of Audiology*, *0*(0), 1–11. https://doi.org/10.1080/14992027.2021.1884907

Carolan, P. J., Heinrich, A., Munro, K. J., & Millman, R. E. (2021b). Quantifying the effects of motivation on listening effort: a systematic review and meta-analysis. *PsyArXic Preprints*, 1–60.

Carver, C. S., & White, T. L. (1994). Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: The BIS/BAS scales. *Journal of Personality and Social Psychology*, *67*(2), 319–333.

Cousineau, D. (2005). Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson's method. *Tutorials in Quantitative Methods for Psychology*, *1*(1), 42–45. https://doi.org/10.20982/tqmp.01.1.p042

Davis, H., Schlundt, D., Bonnet, K., Camarata, S., Bess, F. H., & Hornsby, B. (2020). Understanding Listening-Related Fatigue: Perspectives of Adults with Hearing Loss. *International Journal of Audiology*, *0*(0), 1–11. https://doi.org/10.1080/14992027.2020.1834631

Dawes, P., Hopkins, R., & Munro, K. J. (2013). Placebo effects in hearing-aid trials are reliable. *International Journal of Audiology*, *52*(7), 472–477. https://doi.org/10.3109/14992027.2013.783718

Dawes, P., Powell, S., & Munro, K. J. (2011). The placebo effect and the influence of participant expectation on hearing aid trials. *Ear and Hearing*, *32*(6), 767–774. https://doi.org/10.1097/AUD.0b013e3182251a0e

De Vries, J., Michielsen, H. J., & Van Heck, G. L. (2003). Assessment of fatigue among working people: A comparison of six questionnaires. *Occupational and Environmental Medicine*, *60*(SUPPL. 1), 10–15. https://doi.org/10.1136/oem.60.suppl_1.i10

Dhingra, I., Zhang, S., Zhornitsky, S., Le, T. M., Wang, W., Chao, H. H., Levy, I., & Li, C. S. R. (2020). The effects of age on reward magnitude processing in the monetary incentive delay task. *NeuroImage*, *207*(November 2019), 116368. https://doi.org/10.1016/j.neuroimage.2019.116368

Di Domenico, S. I., & Ryan, R. M. (2017). The emerging neuroscience of intrinsic motivation: A new frontier in self-determination research. *Frontiers in Human Neuroscience*, *11*(March), 1–14. https://doi.org/10.3389/fnhum.2017.00145

Drew, D. S., Muhammed, K., Baig, F., Kelly, M., Saleh, Y., Sarangmat, N., Okai, D., Hu, M., Manohar, S., & Husain, M. (2020). Dopamine and reward hypersensitivity in Parkinson’s disease with impulse control disorder. *Brain*, *143*(8), 2502–2518. https://doi.org/10.1093/brain/awaa198

Earle, F., Hockey, B., Earle, K., & Clough, P. (2015). Separating the effects of task load and task motivation on the effort–fatigue relationship. *Motivation and Emotion*, *39*(4), 467–476. https://doi.org/10.1007/s11031-015-9481-2

Eckert, M. A., Teubner-Rhodes, S., & Vaden, K. I. (2016). Is listening in noise worth it? the neurobiology of speech recognition in challenging listening conditions. *Ear and Hearing*, *37*, 101S-110S. https://doi.org/10.1097/AUD.0000000000000300

Faber, L. G., Maurits, N. M., & Lorist, M. M. (2012). Mental Fatigue Affects

Visual Selective Attention. *PLoS ONE*, *7*(10), 1–10. https://doi.org/10.1371/journal.pone.0048073

Francis, A. L., & Love, J. (2020). Listening effort: Are we measuring cognition or affect, or both? *Wiley Interdisciplinary Reviews: Cognitive Science*, *11*(1), 1–27. https://doi.org/10.1002/wcs.1514

Fukuda, K., & Vogel, E. K. (2009). Human variation in overriding attentional capture. *Journal of Neuroscience*, *29*(27), 8726–8733. https://doi.org/10.1523/JNEUROSCI.2145-09.2009

Gergelyfi, M., Jacob, B., Olivier, E., & Zénon, A. (2015). Dissociation between mental fatigue and motivational state during prolonged mental activity. *Frontiers in Behavioral Neuroscience*, *9*(JULY), 1–15. https://doi.org/10.3389/fnbeh.2015.00176

Gosselin, P. A., & Gagné, J.-P. (2010). Use of a Dual-Task Paradigm to Measure Listening Effort Utilisation d’un paradigme de double tâche pour mesurer l’attention auditive. *Revue Canadienne d’orthophonie et d’audiologie*, *34*, 43–51.

Graham, B., Cottey, L., Smith, J. E., Mills, M., & Latour, J. M. (2020). Measuring Need for Recovery as an indicator of staff well-being in the emergency department: A survey study. *Emergency Medicine Journal*, *37*(9), 555–561. https://doi.org/10.1136/emermed-2019-208797

Grassini, S., & Laumann, K. (2020). Questionnaire Measures and Physiological Correlates of Presence: A Systematic Review. *Frontiers in Psychology*, *11*(March). https://doi.org/10.3389/fpsyg.2020.00349

Grimm, P. (2010). Social Desirability Bias. In *Marketing Research*. John Wiley & Sons Ltd. https://doi.org/https://doi.org/10.1002/9781444316568.wiem02057

Gustafson, S. J., Key, A. P., Hornsby, B. W. Y., & Bess, F. H. (2018). Fatigue related to speech processing in children with hearing loss: Behavioral, subjective, and electrophysiological measures. *Journal of Speech,* *Language, and Hearing Research*, *61*(4), 1000–1011. https://doi.org/10.1044/2018_JSLHR-H-17-0314

Harve-Minvielle, A., & Susan, J, S. (1995). Rapid habituation of auditory responses of locus coeruleus cells in anaesthetized and awake rats. *NeuroReport*, *6*(10), 1235–1240.

Herlambang, M. B., Cnossen, F., & Taatgen, N. A. (2021). The effects of intrinsic motivation on mental fatigue. *PLoS ONE*, *16*(1 January). https://doi.org/10.1371/journal.pone.0243754

Herlambang, M. B., Taatgen, N. A., & Cnossen, F. (2019). The Role of Motivation as a Factor in Mental Fatigue. *Human Factors*, *61*(7), 1171–1185. https://doi.org/10.1177/0018720819828569

Hess, E. H., & Polt, J. M. (1964). Pupil size in relation to mental activity during simple problem-solving. *Science*, *143*(3611), 1190–1192. https://doi.org/10.1126/science.143.3611.1190

Hétu, R., Riverin, L., Lalande, N., Getty, L., & St-Cyr, C. (1988). Qualitative analysis of the handicap associated with occupational hearing loss. *British Journal of Audiology*, *22*(4), 251–264. 10.3109/03005368809076462

Hicks, C. B., & Tharpe, A. M. (2002). Listening effort and fatigue in school-age children with and without hearing loss. *Journal of Speech, Language, and Hearing Research*, *45*(3), 573–584. https://doi.org/10.1044/1092-4388(2002/046)

Hockey, G. R. J. (1997a). Compensatory control in the regulation of human performance under stress and high workload: A cognitive-energetical framework. *Biological Psychology*, *45*(1–3), 73–93. https://doi.org/10.1016/S0301-0511(96)05223-4

Hockey, G. R. J. (1997b). Compensatory control in the regulation of human performance under stress and high workload: A cognitive-energetical framework. *Biological Psychology*, *45*(1–3), 73–93. https://doi.org/10.1016/S0301-0511(96)05223-4

Hockey, G. R. J. (2010). A motivational control theory of cognitive fatigue. *Cognitive Fatigue: Multidisciplinary Perspectives on Current Research and Future Applications.*, *January 2011*, 167–187. https://doi.org/10.1037/12343-008

Hofmann, W., Gawronski, B., Gschwendner, T., Le, H., & Schmitt, M. (2005). A meta-analysis on the correlation between the Implicit Association Test and explicit self-report measures. *Personality and Social Psychology Bulletin*, *31*(10), 1369–1385. https://doi.org/10.1177/0146167205275613

Holman, J. A., Drummond, A., Hughes, S. E., & Naylor, G. (2019). Hearing impairment and daily-life fatigue: a qualitative study. *International Journal of Audiology*, *58*(7), 408–416. https://doi.org/10.1080/14992027.2019.1597284

Hopstaken, J. F., van der Linden, D., Bakker, A. B., & Kompier, M. A. J. (2015a). A multifaceted investigation of the link between mental fatigue and task disengagement. *Psychophysiology*, *52*(3), 305–315. https://doi.org/10.1111/psyp.12339

Hopstaken, J. F., van der Linden, D., Bakker, A. B., & Kompier, M. A. J. (2015b). The window of my eyes: Task disengagement and mental fatigue covary with pupil dynamics. *Biological Psychology*, *110*, 100–106. https://doi.org/10.1016/j.biopsycho.2015.06.013

Hornsby, B. W. Y. (2013). The effects of hearing aid use on listening effort and mental fatigue associated with sustained speech processing demands. *Ear and Hearing*, *34*(5), 523–534. https://doi.org/10.1097/AUD.0b013e31828003d8

Hornsby, B. W. Y., & Kipp, A. M. (2016). Subjective ratings of fatigue and vigor in adults with hearing loss are driven by perceived hearing difficulties not degree of hearing loss. *Ear and Hearing*, *37*(1), e1–e10. https://doi.org/10.1097/AUD.0000000000000203

Houben, R., Van Doorn-Bierman, M., & Dreschler, W. A. (2013). Using response time to speech as a measure for listening effort. *International Journal of Audiology*, *52*(11), 753–761. https://doi.org/10.3109/14992027.2013.832415

Jahfari, S., Verbruggen, F., Frank, M. J., Waldorp, L. J., Colzato, L., Richard Ridderinkhof, K., & Forstmann, B. U. (2012). How preparation changes the need for top-down control of the basal ganglia when inhibiting premature actions. *Journal of Neuroscience*, *32*(32), 10870–10878. https://doi.org/10.1523/JNEUROSCI.0902-12.2012

Jimura, K., Locke, H. S., & Braver, T. S. (2010). Prefrontal cortex mediation of cognitive enhancement in rewarding motivational contexts. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(19), 8871–8876. https://doi.org/10.1073/pnas.1002007107

Kahneman, D. (1973). Attention and Effort. In *The American Journal of Psychology* (Vol. 88, Issue 2). PRENTICE-HALL INC, C., Englewood Cliffs, New Jersey. https://doi.org/10.2307/1421603

Kahneman, D., & Beatty, J. (1966). Pupil Diameter and Load on Memory. *Science*, *154*(3756), 1583–1585.

Kirschbaum, C., & Hellhammer, D. H. (1989). Salivary Cortisol in Psychobiological Research: An Overview. *Biological Psychology/Pharmacopsychology*, *22*, 150–169. https://doi.org/https://doi.org/10.1159/000118611

Knapen, T., De Gee, J. W., Brascamp, J., Nuiten, S., Hoppenbrouwers, S., & Theeuwes, J. (2016). Cognitive and ocular factors jointly determine pupil responses under equiluminance. *PLoS ONE*, *11*(5). https://doi.org/10.1371/journal.pone.0155574

Koch, X., & Janse, E. (2016). Speech rate effects on the processing of conversational speech across the adult life span. *J. Acoust. Soc. Am*, *139*(3), 16–18. https://doi.org/http://dx.doi.org/10.1121/1.4944032

Koelewijn, T., de Kluiver, H., Shinn-Cunningham, B. G., Zekveld, A. A., & Kramer, S. E. (2015). The pupil response reveals increased listening effort when it is difficult to focus attention. *Hearing Research*, *323*, 81–90. https://doi.org/10.1016/j.heares.2015.02.004

Koelewijn, T., Zekveld, A. A., Festen, J. M., & Kramer, S. E. (2012). Pupil dilation uncovers extra listening effort in the presence of a single-talker masker. *Ear and Hearing*, *33*(2), 291–300. https://doi.org/10.1097/AUD.0b013e3182310019

Koelewijn, T., Zekveld, A. A., Lunner, T., & Kramer, S. E. (2018). The effect of reward on listening effort as reflected by the pupil dilation response. *Hearing Research*, *367*, 106–112. https://doi.org/10.1016/j.heares.2018.07.011

Koelewijn, T., Zekveld, A. A., Lunner, T., & Kramer, S. E. (2021). The effect of monetary reward on listening effort and sentence recognition. *Hearing Research*, *406*, 108255. https://doi.org/10.1016/j.heares.2021.108255

Kramer, S. E., Kapteyn, T. S., Festen, J. M., & Kuik, D. J. (1997). Assessing aspects of auditory handicap by means of pupil dilatation. *International Journal of Audiology*, *36*(3), 155–164. https://doi.org/10.3109/00206099709071969

Kramer, S. E., Kapteyn, T. S., & Houtgast, T. (2006). Occupational performance: Comparing normally-hearing and hearing-impaired employees using the Amsterdam Checklist for Hearing and Work. *International Journal of Audiology*, *45*(9), 503–512. https://doi.org/10.1080/14992020600754583

Krueger, M., Schulte, M., Zokoll, M. A., Wagener, K. C., Meis, M., Brand, T., & Holubec, I. (2017). Relation Between Listening Effort and Speech Intelligibility in Noise. *American Journal of Audiology*, *26*(1), 378–392.

Kuchinsky, S. E., Ahlstrom, J. B., Vaden, J. K. I., Cute, S. L., Humes, L. E., Dubno, J. R., & Eckert, M. A. (2013). Pupil size varies with word listening and response selection difficulty in older adults with hearing loss.

Psychophysiology, *50*(1), 23–24. https://doi.org/10.1111/j.1469-8986.2012.01477.x.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models . *Journal of Statistical Software*, *82*(13). https://doi.org/10.18637/jss.v082.i13

Larsen, R. S., & Waters, J. (2018a). Neuromodulatory correlates of pupil dilation. *Frontiers in Neural Circuits*, *12*(March), 1–9. https://doi.org/10.3389/fncir.2018.00021

Larsen, R. S., & Waters, J. (2018b). Neuromodulatory correlates of pupil dilation. *Frontiers in Neural Circuits*, *12*(March), 1–9. https://doi.org/10.3389/fncir.2018.00021

Lawrence, R. J., Wiggins, I. M., Anderson, C. A., Davies-Thompson, J., & Hartley, D. E. H. (2018). Cortical correlates of speech intelligibility measured using functional near-infrared spectroscopy (fNIRS). *Hearing Research*, *370*, 53–64. https://doi.org/10.1016/j.heares.2018.09.005

Lorist, M. M. (2008). Impact of top-down control during mental fatigue. *Brain Research*, *1232*, 113–123. https://doi.org/10.1016/j.brainres.2008.07.053

Lorist, M. M., Klein, M., Nieuwenhuis, S., De Jong, R., Mulder, G., & Meijman, T. F. (2000). Mental fatigue and task control: Planning and preparation. *Psychophysiology*, *37*(5), 614–625. https://doi.org/10.1017/S004857720099005X

Lovallo, W. R., & Buchanan, T. W. (2016). Stress hormones in psychophysiological research: Emotional, behavioral, and cognitive implications. In *Handbook of Psychophysiology, Fourth Edition* (Issue February, pp. 465–494). https://doi.org/10.1017/9781107415782.021

Luke, S. G. (2017). Evaluating significance in linear mixed-effects models in R. *Behavior Research Methods*, *49*(4), 1494–1502. https://doi.org/10.3758/s13428-016-0809-y

Luts, H., Eneman, K., Wouters, J., Schulte, M., Vormann, M., Buechler, M., Dillier, N., Houben, R., Dreschler, W. A., Froehlich, M., Puder, H., Grimm, G., Hohmann, V., Leijon, A., Lombard, A., Mauler, D., & Spriet, A. (2010). Multicenter evaluation of signal enhancement algorithms for hearing aids. *The Journal of the Acoustical Society of America*, *127*(3), 1491–1505. https://doi.org/10.1121/1.3299168

MacLeod, A., & Summerfield, Q. (1990). A procedure for measuring auditory and audio-visual speech-reception thresholds for sentences in noise: rationale, evaluation, and recommendations for use. *Journal of British Audiology*, *24*(1), 29–43. https://doi.org/10.3109/03005369009077840

Massar, S. A. A., Csathó, Á., & Van der Linden, D. (2018). Quantifying the motivational effects of cognitive fatigue through effort-based decision making. *Frontiers in Psychology*, *9*(MAY), 1–5. https://doi.org/10.3389/fpsyg.2018.00843

Matthen, M. (2016). Effort and displeasure in people who are hard of hearing. *Ear and Hearing*, *37*, 28S-34S. https://doi.org/10.1097/AUD.0000000000000292

Mattys, S. L., Davis, M. H., Bradlow, A. R., & Scott, S. K. (2012). Speech recognition in adverse conditions: A review. *Language and Cognitive Processes*, *27*(7–8), 953–978. https://doi.org/10.1080/01690965.2012.705006

McGarrigle, R., Dawes, P., Stewart, A. J., Kuchinsky, S. E., & Munro, K. J. (2017a). Measuring listening-related effort and fatigue in school-aged children using pupillometry. *Journal of Experimental Child Psychology*, *161*, 95–112. https://doi.org/10.1016/j.jecp.2017.04.006

McGarrigle, R., Dawes, P., Stewart, A. J., Kuchinsky, S. E., & Munro, K. J. (2017b). Pupillometry reveals changes in physiological arousal during a sustained listening task. *Psychophysiology*, *54*(2), 193–203. https://doi.org/10.1111/psyp.12772

McGarrigle, R., Munro, K. J., Dawes, P., Stewart, A. J., Moore, D. R., Barry, J. G., & Amitay, S. (2014). Listening effort and fatigue: What exactly are we measuring? A British Society of Audiology Cognition in Hearing Special Interest Group “white paper.” *International Journal of Audiology*, *53*(7), 433–445. https://doi.org/10.3109/14992027.2014.890296

McGarrigle, R., Rakusen, L., & Mattys, S. (2021). Effortful listening under the microscope: Examining relations between pupillometric and subjective markers of effort and tiredness from listening. *Psychophysiology*, *58*(1), 1–22. https://doi.org/10.1111/psyp.13703

McLaughlin, D. J., Braver, T., & Peelle, J. (2020). Measuring the subjective cost of listening effort using a discounting task. *The Journal of the Acoustical Society of America*, *148*(4), 2623–2623. https://doi.org/10.1121/1.5147295

Metalis, S. A., & Hess, E. H. (1981). Psychophysiological index attenuation to pictures. *Physiological Psychology*, *9*(2), 235–243. https://doi.org/10.3758/BF03332930

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience*, *24*(February 2001), 167–202. https://doi.org/10.1146/annurev.neuro.24.1.167

Moreno, M. A., & van Orden, G. C. (2001). Word Recognition, Cognitive Psychology of. In *International Encyclopedia of the Social & Behavioral Sciences* (pp. 16556–16561). https://doi.org/10.1016/b0-08-043076-7/01553-9

Morey, R. D. (2008). Confidence Intervals from Normalized Data: A correction to Cousineau (2005). *Tutorials in Quantitative Methods for Psychology*, *4*(2), 61–64. https://doi.org/10.20982/tqmp.04.2.p061

Muhammed, K., Manohar, S., Ben Yehuda, M., Chong, T. T. J., Tofaris, G., Lennox, G., Bogdanovic, M., Hu, M., & Husain, M. (2016). Reward sensitivity deficits modulated by dopamine are associated with apathy in

Parkinson’s disease. *Brain*, *139*(10), 2706–2721. https://doi.org/10.1093/brain/aww188

Müller, T., & Apps, M. A. J. (2019). Motivational fatigue: A neurocognitive framework for the impact of effortful exertion on subsequent motivation. *Neuropsychologia*, *123*(October 2017), 141–151. https://doi.org/10.1016/j.neuropsychologia.2018.04.030

Nachtegaal, J., Kuik, D. J., Anema, J. R., Goverts, S. T., Festen, J. M., & Kramer, S. E. (2009). Hearing status, need for recovery after work, and psychosocial work characteristics: Results from an internet-based national survey on hearing. *International Journal of Audiology*, *48*(10), 684–691. https://doi.org/10.1080/14992020902962421

Naylor, G., Koelewijn, T., Zekveld, A. A., & Kramer, S. E. (2018). The Application of Pupillometry in Hearing Science to Assess Listening Effort. *Trends in Hearing*, *22*, 1–3. https://doi.org/10.1177/2331216518799437

Naylor, G., Oberg, M., Wanstrom, G., & Lunner, T. (2015). Exploring the Effects of the Narrative Embodied in the Hearing Aid Fitting Process on Treatment Outcomes. *Ear and Hearing*, *36*(5), 517–526. http://linker.worldcat.org/?rft.institution_id=129664&spage=517&pkgName=lwwhic2016&issn=0196-0202&linkclass=to_article&jKey=00003446-000000000-00000&issue=5&provider=Ovid&date=2015-09&aulast=Naylor+G.;+Oberg+M.;+Wanstrom+G.;+Lunner+T.&atitle=Explori

Ng, E. H. N., Rudner, M., Lunner, T., & Rönnberg, J. (2013). Relationships between self-report and cognitive measures of hearing aid outcome. *Speech, Language and Hearing*, *16*(4), 197–207. https://doi.org/10.1179/205057113X13782848890774

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

Nobre, A. C., & Van Ede, F. (2018). Anticipated moments: Temporal structure in attention. *Nature Reviews Neuroscience*, *19*(1), 34–48. https://doi.org/10.1038/nrn.2017.141

Nunnally, J. C., Knott, P. D., Duchnowski, A., & Parker, R. (1967). Pupillary response as a general measure of activation. *Perception & Psychophysics*, *2*(4), 149–155. https://doi.org/10.3758/BF03210310

Obleser, J., Wöstmann, M., Hellbernd, N., Wilsch, A., & Maess, B. (2012). Adverse listening conditions and memory load drive a common alpha oscillatory network. *Journal of Neuroscience*, *32*(36), 12376–12383. https://doi.org/10.1523/JNEUROSCI.4908-11.2012

Ohlenforst, B., Wendt, D., Kramer, S. E., Naylor, G., Zekveld, A. A., & Lunner, T. (2018). Impact of SNR, masker type and noise reduction processing on sentence recognition performance and listening effort as indicated by the pupil dilation response. *Hearing Research*, *365*, 90–99. https://doi.org/10.1016/j.heares.2018.05.003

Ohlenforst, B., Zekveld, A. A., Jansma, E. P., Wang, Y., Naylor, G., Lorens, A., Lunner, T., & Kramer, S. E. (2017). Effects of hearing impairment and hearing aid amplification on listening effort: A systematic review. *Ear and Hearing*, *38*(3), 267–281. https://doi.org/10.1097/AUD.0000000000000396

Ohlenforst, B., Zekveld, A. A., Lunner, T., Wendt, D., Naylor, G., Wang, Y., Versfeld, N. J., & Kramer, S. E. (2017). Impact of stimulus-related factors and hearing impairment on listening effort as indicated by pupil dilation. *Hearing Research*, *38*(3), 267–281. https://doi.org/10.1097/AUD.0000000000000396

Olafson, K. M., & Ferraro, F. R. (2001). Effects of emotional state on lexical decision performance. *Brain and Cognition*, *45*(1), 15–20. https://doi.org/10.1006/brcg.2000.1248

Pals, C., Sarampalis, A., Dijk, M. Van, & Ba, D. (2019). Effects of Additional

Low-Pass–Filtered Speech on Listening Effort for Noise-Band–Vocoded Speech in Quiet and in Noise. *Ear & Hearing*, *40*(1), 3–17. https://doi.org/10.1097/AUD.0000000000000587

Peavler, W, S. (University of D. (1974). Pupil size, information overload, and perfrmance differences. *Psychophysiology*, *11*(5), 559–566.

Peelle, J. E. (2018). Listening effort: How the cognitive consequences of acoustic challenge are reflected in brain and behavior. *Ear and Hearing*, *39*(2), 204–214. https://doi.org/10.1097/AUD.0000000000000494

Pichora-Fuller, M. K. (2016). How social psychological factors may modulate auditory and cognitive functioning during listening. *Ear and Hearing*, *37*, 92S-100S. https://doi.org/10.1097/AUD.0000000000000323

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

Picou, E. M., & Ricketts, T. A. (2014). Increasing motivation changes subjective reports of listening effort and choice of coping strategy. *International Journal of Audiology*, *53*(6), 418–426. https://doi.org/10.3109/14992027.2014.880814

Pielage, H., Zekveld, A. A., Saunders, G. H., Versfeld, N. J., Lunner, T., & Kramer, S. E. (2021). The Presence of Another Individual Influences Listening Effort, But Not Performance. *Ear & Hearing*, *Publish Ah*(April). https://doi.org/10.1097/aud.0000000000001046

Plain, B., Richter, M., Zekveld, A. A., Lunner, T., Bhuiyan, T., & Kramer, S. E. (2020). Investigating the Influences of Task Demand and Reward on Cardiac Pre-Ejection Period Reactivity During a Speech-in-Noise Task. *Ear*

& Hearing, *Publish Ah*, 1–14. https://doi.org/10.1097/aud.0000000000000971

Richter, M., Gendolla, G. H. E., & Wright, R. A. (2016). *Three Decades of Research on Motivational Intensity Theory*. *December*, 149–186. https://doi.org/10.1016/bs.adms.2016.02.001

Richter, M., & Slade, K. (2017). Interpretation of physiological indicators of motivation: Caveats and recommendations. *International Journal of Psychophysiology*, *119*(April), 4–10. https://doi.org/10.1016/j.ijpsycho.2017.04.007

Rönnberg, J., Holmer, E., & Rudner, M. (2019). Cognitive hearing science and ease of language understanding. *International Journal of Audiology*, *58*(5), 247–261. https://doi.org/10.1080/14992027.2018.1551631

Rovetti, J., Goy, H., Pichora-Fuller, M. K., & Russo, F. A. (2019). Functional Near-Infrared Spectroscopy as a Measure of Listening Effort in Older Adults Who Use Hearing Aids. *Trends in Hearing*, *23*. https://doi.org/10.1177/2331216519886722

Roy F. Baumeister, & Leary, M. R. (1995). The Need to Belong: Desire for Interpersonal Attachments as a Fundamental Human Motivation. *Psychological Bulletin*, *117*(3), 497–529. https://doi.org/10.1109/ITME.2008.4743876

Rudner, M., Lunner, T., Behrens, T., Thorén, E. S., & Rönnberg, J. (2012). Working memory capacity may influence perceived effort during aided speech recognition in noise. *Journal of the American Academy of Audiology*, *23*(8), 577–589. https://doi.org/10.3766/jaaa.23.7.7

Rund, N. (2018). *Are we as tired as we think we are? : A study over the relation of the explicit and implicit within-day variability of fatigue*. http://essay.utwente.nl/75245/

Sadowsky, C. J. (1993). An Examination of the Short Need for Cognition Scale. *The Journal of Psychology*, *127*(4), 451–454.

https://doi.org/https://doi.org/10.1080/00223980.1993.9915581

Salamone, J. D., Yohn, S. E., López-Cruz, L., San Miguel, N., & Correa, M. (2016). Activational and effort-related aspects of motivation: Neural mechanisms and implications for psychopathology. *Brain*, *139*(5), 1325–1347. https://doi.org/10.1093/brain/aww050

Sarampalis, A., Kalluri, S., Edwards, B., & Hafter, E. (2009). Objective measures of listening effort: Effects of background noise and noise reduction. *Journal of Speech, Language, and Hearing Research*, *52*(5), 1230–1240. https://doi.org/10.1044/1092-4388(2009/08-0111)

Schneider, E. N., Bernarding, C., Francis, A. L., Hornsby, B. W. Y., & Strauss, D. J. (2019). A Quantitative Model of Listening Related Fatigue. *International IEEE/EMBS Conference on Neural Engineering, NER*, *2019-March*, 619–622. https://doi.org/10.1109/NER.2019.8717046

Schneider, M., Elbau, I. G., Nantawisarakul, T., Pöhlchen, D., Brückl, T., Czisch, M., Saemann, P. G., Lee, M. D., Binder, E. B., & Spoormaker, V. I. (2020). Pupil dilation during reward anticipation is correlated to depressive symptom load in patients with major depressive disorder. *Brain Sciences*, *10*(12), 1–15. https://doi.org/10.3390/brainsci10120906

Shalev, N., & Nobre, A. C. (2021). *Eyes wide open: regulation of arousal by temporal expectations*. https://doi.org/https:doi.org/10.1101/2020.07.24.219428

Silvestrini, N., & Gendolla, G. H. E. (2019). Affect and cognitive control: Insights from research on effort mobilization. *International Journal of Psychophysiology*, *143*, 116–125. https://doi.org/10.1016/j.ijpsycho.2019.07.003

Smets, et. a. (1995). The multidimensional fatigue inventory (MFI) Psychometric qualities of an instrument to assess fatigue. *Journal of Psychosomatic Research*, *39*(5), 315–325. https://www.academia.edu/5238606/The_multidimensional_Fatigue_In ventory_MFI_psychometric_qualities_of_an_instrument_to_assess_fatigue

Stanners, R. F., Coulter, M., Sweet, A. W., & Murphy, P. (1979). The pupillary response as an indicator of arousal and cognition. *Motivation and Emotion*, *3*(4), 319–340. https://doi.org/10.1007/BF00994048

Strand, J. F., Brown, V. A., Merchant, M. B., Brown, H. E., & Smith, J. (2018). Measuring listening effort: Convergent validity, sensitivity, and links with cognitive and personality measures. *Journal of Speech, Language, and Hearing Research*, *61*(6), 1463–1486. https://doi.org/10.1044/2018_JSLHR-H-17-0257

van den Brink, R. L., Pfeffer, T., & Donner, T. H. (2019). Brainstem Modulation of Large-Scale Intrinsic Cortical Activity Correlations. *Frontiers in Human Neuroscience*, *13*(October), 1–18. https://doi.org/10.3389/fnhum.2019.00340

van der Linden, D., Frese, M., & Meijman, T. F. (2003). Mental fatigue and the control of cognitive processes: Effects on perseveration and planning. *Acta Psychologica*, *113*(1), 45–65. https://doi.org/10.1016/S0001-6918(02)00150-6

Van Veldhoven, M., & Broersen, S. (2003a). Measurement quality and validity of the “need for recovery scale.” *Occupational and Environmental Medicine*, *60*(SUPPL. 1), 3–9. https://doi.org/10.1136/oem.60.suppl_1.i3

Van Veldhoven, M., & Broersen, S. (2003b). Measurement quality and validity of the “need for recovery scale.” *Occupational and Environmental Medicine*, *60*(SUPPL. 1), 2–9. https://doi.org/10.1136/oem.60.suppl_1.i3

Verbaten, M. N., Roelofs, J. W., Sjouw, W., & Verbaten, M. N. (1986). Habituation of early and late ERP components and the orienting reaction: the effect of stimulus formation. *International Journal of Psychophysiology*, *3*, 287–298.

Vercoulen, J. H. M. M., Swanink, C. M. A., Fennis, J. F. M., Galama, J. M. D., van der Meer, J. W. M., & Bleijenberg, G. (1994). Dimensional assessment of chronic fatigue syndrome. *Journal of Psychosomatic Research*, *38*(5), 383–392. https://doi.org/10.1016/0022-3999(94)90099-X

Wang, Y., Naylor, G., Kramer, S. E., Zekveld, A. A., Wendt, D., Ohlenforst, B., & Lunner, T. (2018). Relations Between Self-Reported Daily-Life Fatigue, Hearing Status, and Pupil Dilation During a Speech Perception in Noise Task. *Ear and Hearing*, *39*(3), 573–582. https://doi.org/10.1097/AUD.0000000000000512

Wegman, J., van Loon, I., Smeets, P. A. M., Cools, R., & Aarts, E. (2018). Top-down expectation effects of food labels on motivation. *NeuroImage*, *173*(February), 13–24. https://doi.org/10.1016/j.neuroimage.2018.02.011

Wendt, D., Koelewijn, T., Książek, P., Kramer, S. E., & Lunner, T. (2018). Toward a more comprehensive understanding of the impact of masker type and signal-to-noise ratio on the pupillary response while performing a speech-in-noise test. *Hearing Research*, *369*, 67–78. https://doi.org/10.1016/j.heares.2018.05.006

Wild, C. J., Yusuf, A., Wilson, D. E., Peelle, J. E., Davis, M. H., & Johnsrude, I. S. (2012). Effortful listening: The processing of degraded speech depends critically on attention. *Journal of Neuroscience*, *32*(40), 14010–14021. https://doi.org/10.1523/JNEUROSCI.1528-12.2012

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., Edwards, J. R., & Litovsky, R. Y. (2016). The Impact of Auditory Spectral Resolution on Listening Effort Revealed by Pupil Dilation. *Ear and Hearing*, *36*(4), 1–31. https://doi.org/10.1097/AUD.0000000000000145.

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

Wood, S. L., & Swait, J. (2002). Psychological indicators of innovation adoption: Cross-classification based on need for cognition and need for change. *Journal of Consumer Psychology*, *12*(1), 1–13. https://doi.org/10.1207/153276602753338108

Wöstmann, M., Herrmann, B., Wilsch, A., & Obleser, J. (2015). Neural alpha dynamics in younger and older listeners reflect acoustic challenges and predictive benefits. *Journal of Neuroscience*, *35*(4), 1458–1467. https://doi.org/10.1523/JNEUROSCI.3250-14.2015

Wu, Y.-H., Stangl, E., Zhang, X., Perkins, J., & Eilers, E. (2017). Psychometric functions of dual-task paradigms for measuring listening effort. *Physiology & Behavior*, *176*(5), 139–148. https://doi.org/10.1016/j.physbeh.2017.03.040

Zekveld, A. A., Heslenfeld, D. J., Johnsrude, I. S., Versfeld, N. J., & Kramer, S. E. (2014). The eye as a window to the listening brain: Neural correlates of pupil size as a measure of cognitive listening load. *NeuroImage*, *101*, 76–86. https://doi.org/10.1016/j.neuroimage.2014.06.069

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. (2014). Cognitive processing load across a wide range of listening conditions: Insights from pupillometry. *Psychophysiology*, *51*(3), 277–284. https://doi.org/10.1111/psyp.12151

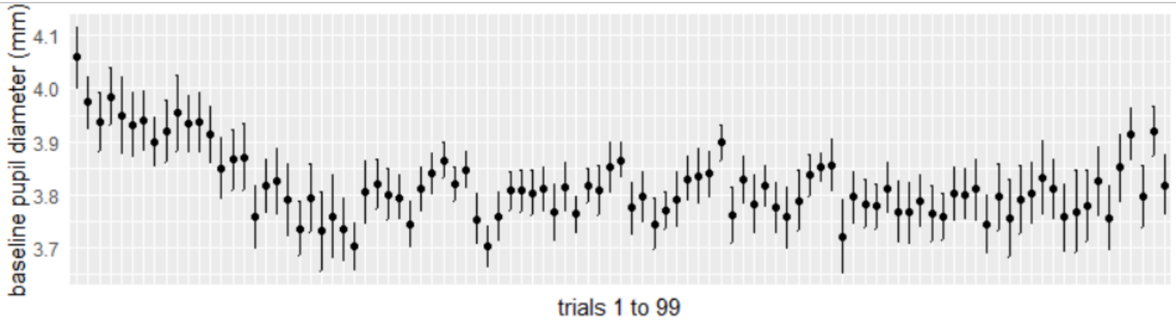
Zekveld, A. A., Kramer, S. E., & Festen, J. M. (2010). Pupil response as an indication of effortful listening: The influence of sentence intelligibility. *Ear and Hearing*, *31*(4), 480–490. https://doi.org/10.1097/AUD.0b013e3181d4f251

Zekveld, A. A., van Scheepen, J. A. M., Versfeld, N. J., Veerman, E. C. I., & Kramer, S. E. (2019). Please try harder! The influence of hearing status and evaluative feedback during listening on the pupil dilation response, saliva-cortisol and saliva alpha-amylase levels. *Hearing Research*, *381*, 107768. https://doi.org/10.1016/j.heares.2019.07.005

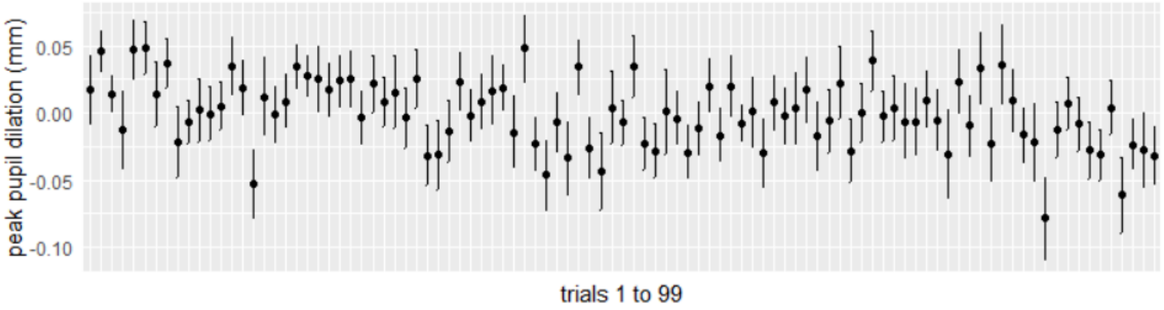
Zenon, A. (2019). Eye pupil signals information gain. *Royal Society Publishing, Neuroscience and Cognition*.

Appendices

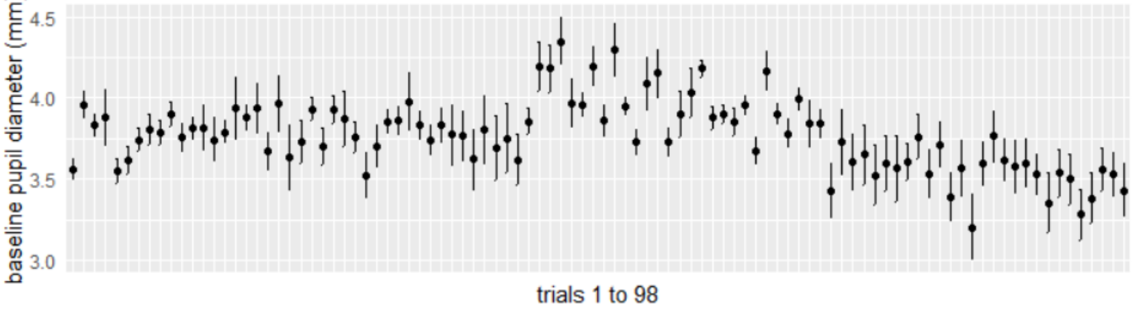
1. Baseline pupil diameter during the load sequence block of Experiment 1



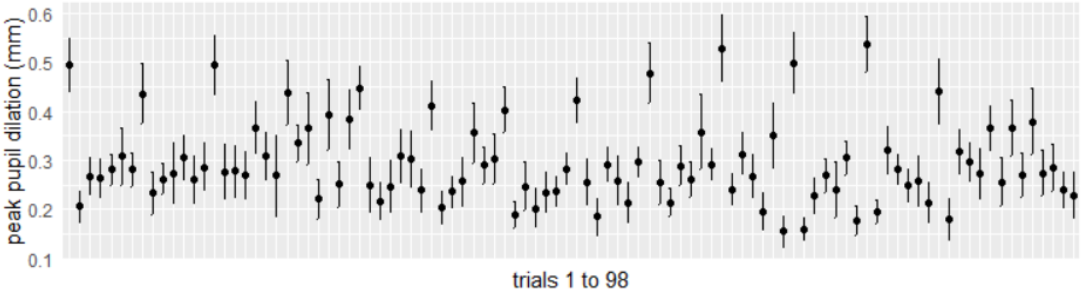
2. Peak pupil dilation during the load sequence block of Experiment 1



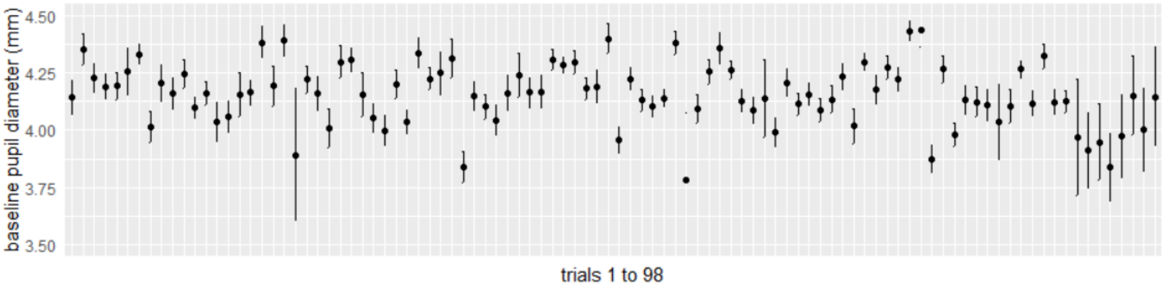
3. Baseline pupil diameter during the light load sequence block of Experiment 2



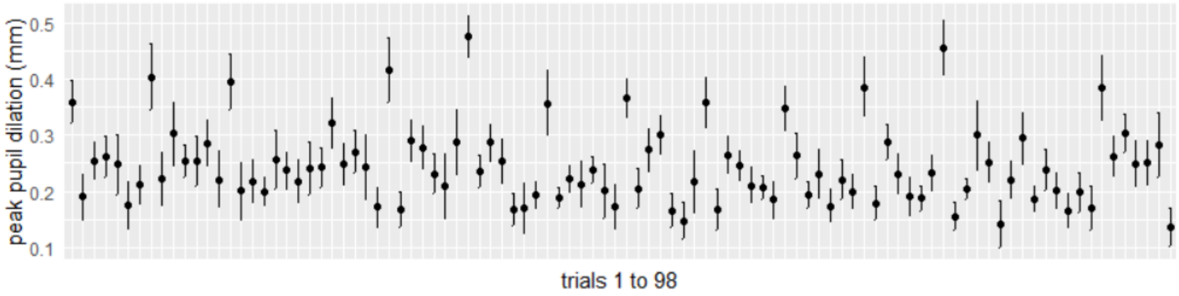
4. Peak pupil dilation during the light load sequence block of Experiment 2



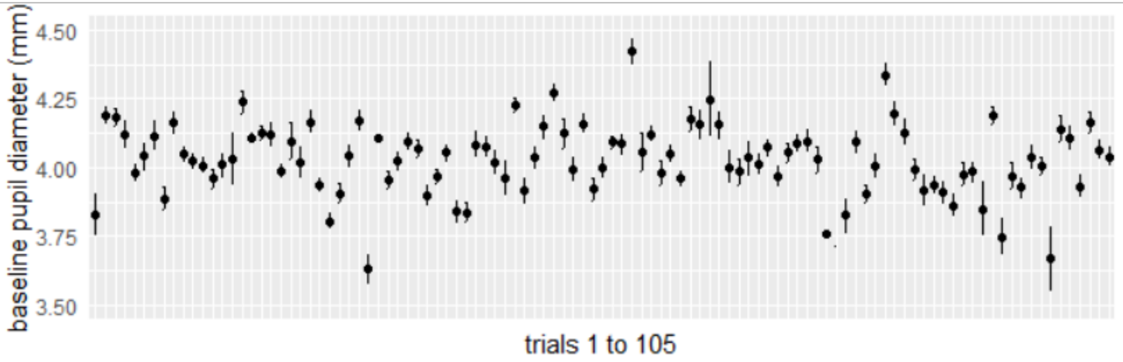
5. Baseline pupil diameter during the heavy load sequence block of Experiment 2



6. Peak pupil dilation during the heavy load sequence block of Experiment 2



7. Baseline pupil diameter during the load sequence block of Experiment 3



8. Peak pupil dilation during the load sequence block of Experiment 3