

1 Task-Induced Mental Fatigue and Motivation Influence Listening Effort as Measured by the 2 Pupil Dilation in a Speech-in-Noise Task

3 Defne Alfandari Menase¹, Michael Richter², Dorothea Wendt^{3,4}, Lorenz Fiedler³, and Graham
4 Naylor¹

5 ¹ University of Nottingham, School of Medicine, Mental Health and Clinical Neurosciences,
6 Hearing Sciences - Scottish Section, Glasgow, UK.

7 ²Liverpool John Moores University, School of Psychology, Liverpool, UK.

8 ³Eriksholm Research Centre, Oticon A/S, Snekkersten, Denmark.

9 ⁴Hearing Systems, Department of Health Technology, Technical University of Denmark,
10 Denmark

11

12 **Objectives:** Listening effort and fatigue are common complaints among individuals with hearing
13 impairment (HI); however, the underlying mechanisms, and relationships between listening effort
14 and fatigue are not well understood. Recent quantitative research suggests that the peak pupil
15 dilation (PPD), which is commonly measured concurrent to the performance of a speech-in-noise
16 task as an index of listening effort, may be informative of daily-life fatigue, but it remains unknown
17 whether the same is true for task-induced fatigue. As fatigue effects are known to manifest
18 differently depending on motivation, the main aim of the present study was to experimentally
19 investigate the interactive effects of task-induced fatigue and motivation on the PPD.

20 **Design:** In a pre-/post- fatigue within-subject design, 18 participants with normal hearing (NH)
21 engaged in a 98-trial-long speech-in-noise task (the 'load sequence', approximately 40 min. long),
22 which either excluded or included additional memory demands (light vs. heavy load sequence).
23 Before and after the load sequence, baseline pupil diameter (BPD) and PPD were measured
24 during shorter 'probe' blocks of speech-in-noise tasks. In these probe blocks, if participants
25 correctly repeated more than 60% of the keywords, they could win vouchers of either 20 or 160
26 Danish kroner worth (low incentive vs. high incentive). After each probe block, participants
27 reported their invested effort, tendency for quitting, and perceived performance.

28 **Results:** The BPD in anticipation of listening declined from pre- to post-load sequence, suggesting
29 an overall decrease in arousal, but the decline did not scale with the magnitude of the load
30 sequence, nor with the amount of monetary incentive. Overall, there was larger pre- to post-load
31 sequence decline in PPD when the load sequence was heavy and when the monetary incentives
32 were low. Post-hoc analyses showed that the decline in PPD was only significant in the heavy-load
33 sequence-low reward condition. The speech-in-noise task performance, self-reported effort, and
34 self-reported tendency to quit listening did not change with the experimental conditions.

35

36 **Conclusions** This is the first study to investigate the influence of task-induced fatigue on BPD and
37 PPD. Whereas BPD was not sensitive to the magnitude of previous load sequence and monetary
38 incentives, the decline in PPD from pre- to post- load sequence was significant after the heavy load
39 sequence when the offered monetary incentives were low. This result supports the understanding
40 that fatigue and motivation interactively influence listening effort.

41

1 **Financial disclosures/conflicts of interest:** This study has received funding from the European
2 Union's Horizon 2020 research and innovation programme under the Marie-Sklodowska-Curie
3 grant agreement No 765329; the Medical Research Council [grant number MR/S003576/1]; and
4 the Chief Scientist Office of the Scottish Government There are no conflicts of interest, financial, or
5 otherwise.

6 **All correspondence should be addressed to:**

7 Graham Naylor, School of Medicine, University of Nottingham, Hearing Sciences – Scottish
8 Section, New Lister Building Level 3, Glasgow Royal Infirmary, 10-16 Alexandra Parade, Glasgow
9 G31 2ER. (m) +44 (0) 77 3648 0077. E-mail: graham.naylor@nottingham.ac.uk

10 **Acknowledgements**

11 We are grateful to Patrycja Książek, Renskje K. Hietkamp, Antoine Favre-Félix, Sara Al-ward, Eva
12 Mathiesen, Julia Yevchenko, and Franziska Behnen for discussions, assistance in data-collection, .
13 and technical experimental set-up.

14 This study has received funding from the European Union's Horizon 2020 research and innovation
15 programme under the Marie-Sklodowska-Curie grant agreement No 765329; the Medical Research
16 Council [grant number MR/S003576/1]; and the Chief Scientist Office of the Scottish Government.
17 There are no conflicts of interest, financial, or otherwise.



19

1 Perceiving speech in noisy environments is known to be more effortful for individuals with hearing
2 impairment (HI) than those with normal hearing (NH; e.g. Dwyer, Firszt, & Reeder, 2014). Such
3 daily listening effort, which taxes cognitive and physiological resources (Pichora-Fuller et al.,
4 2016), is assumed to give rise to listening-related fatigue (Holman, Drummond, Hughes, & Naylor,
5 2019). School children with HI are known to report more fatigue as compared to their peers (Bess
6 & Hornsby, W., Y., 2014). In adults with HI, daily effort is related to increased need for recovery
7 and sick leave from work (Kramer, Kapteyn, & Houtgast, 2006; Nachtegaal et al., 2009). Despite
8 frequent reports of effort and fatigue among individuals with HI, the relationships between and
9 mechanisms underlying listening effort and fatigue are not well understood (Hornsby, Naylor, &
10 Bess, 2016; Hornsby & Kipp, 2016; Pichora-Fuller et al., 2016; Schneider, Bernarding, Francis,
11 Hornsby, & Strauss, 2019).

12 Listening effort has widely been assessed using pupillometry (Kramer, Kapteyn, Festen, & Kuik,
13 1997; Naylor, Koelewijn, Zekveld, & Kramer, 2018; Zekveld, Koelewijn, & Kramer, 2018). The
14 recording of the pupil size is commonly combined with a speech-in-noise (SIN) test, where
15 participants listen to a sentence in noise and repeat the presented sentence or keywords (Nielsen
16 & Dau, 2011; Portnuff & Bell, 2014; Winn, Wendt, Koelewijn, & Kuchinsky, 2018). In this paradigm
17 the intensity difference between the target sentence and background noise can be varied to
18 estimate the signal-to-noise ratio (SNR) that is required for an individual to correctly perceive a
19 certain percentage of sentences (i.e., the speech reception threshold; SRT). The average pupil
20 diameter during a defined period (e.g., 1 s) before the presentation of the target sentence is
21 referred to as the baseline pupil diameter (BPD), and the maximum pupil size relative to baseline
22 elicited during a trial is referred to as the peak pupil dilation (PPD). Whereas the SIN test is
23 informative of the ability of an individual to perceive speech in noise, pupillometry has been argued
24 to capture the processing load (effort) that it takes an individual to perform the task.

25 Most research to quantify listening effort with pupillometry has focused on the relationship between
26 listening difficulty and PPD (see Zekveld et al., 2018 for a review). During SIN tests with

1 background babble noise, the PPD has been shown to be largest where speech recognition is
2 around 50 % (SRT50), and lower for both higher and lower SNRs (Ohlenforst et al., 2018; Wendt,
3 Koelewijn, Książek, Kramer, & Lunner, 2018). The increase in PPD with increasing difficulty at the
4 left-hand side of the SRT – PPD inverted U-shaped curve is in line with the well-known relationship
5 between task demands and effort, where effort increases with increasing task demands, as long as
6 the task is within capacity (Brehm & Self, 1989; Richter, Gendolla, & Wright, 2016). The decrease
7 in PPD at the right-hand side of the curve has previously been interpreted as “giving up” by the
8 participants due to the task not being perceived as achievable (Ohlenforst et al., 2018)

9 In a recent study, the PPD has been shown to be informative of daily-life fatigue (Wang et al.,
10 2018). In that study, a large sample of adults with NH and HI performed a SIN test targeting a 50 %
11 correct response criterion while their pupil diameter was being recorded. A significant negative
12 correlation was found between PPD and self-reported daily-life fatigue. This correlation was
13 independent of hearing acuity. This result suggests that individual differences in daily-life fatigue
14 may be related to individual differences in PPD, but the mechanism driving this association
15 remains unclear. While highly suggestive, the study of Wang et al. (2018) says nothing about how
16 listening effort may interact with task-induced (i.e. transitory) rather than persistent fatigue.
17 Meanwhile, it highlights the potential for confounds arising from daily variations in fatigue, in
18 experiments aiming to study task-induced fatigue.

19 Mental fatigue is well known to arise after a period of sustained effortful exertion, expressing as a
20 subjective struggle to maintain the exertion required (Boksem, Meijman, & Lorist, 2006; see Müller
21 & Apps, 2019 for a review). Fatigue has previously been associated with decreased arousal and
22 hence decreased BPD in cognitively challenging tasks (Hopstaken, van der Linden, Bakker, &
23 Kompier, 2015a). In previous SIN tasks, BPD has been observed to decline with time-on-task
24 (Alhanbali, Munro, Dawes, Carolan, & Millman, 2020; Ayasse & Wingfield, 2020; Zekveld, Kramer,
25 & Festen, 2010). Fluctuations in BPD have commonly been attributed to the locus coeruleus –
26 noradrenergic functional dynamics (Murphy, Robertson, Balsters, & O’connell, 2011). Although

1 changes in BPD have been argued to reflect a variety of states other than fatigue such as learning,
2 habituation, and task (dis)engagement (Ayasse & Wingfield, 2020; Gilzenrat, Nieuwenhuis, Jepma,
3 & Cohen, 2010; McGarrigle, Dawes, Stewart, Kuchinsky, & Munro, 2017) reduced arousal is
4 considered a hallmark of fatigue (McGarrigle et al., 2017; van der Linden, 2011).

5 In a fatigued state, whether individuals continue to invest effort is known to depend on their
6 willingness to perform the task at hand (Boksem et al., 2006; Boksem & Tops, 2008; Hockey,
7 1997, 2010). Indeed, a study using pupillometry during a mentally challenging task showed
8 diminishing PPD over time-on-task in a 2-hour long task, but restoration to initial levels after
9 participants were offered rewards for correct performance (Hopstaken, van der Linden, Bakker, &
10 Kompier, 2015b). This result suggests that fatigue and the willingness to perform well (motivation),
11 as manipulated by rewards, together influence the PPD in mentally challenging tasks.

12 Previous research with pupillometry has shown a positive relationship between rewards and PPD
13 during the SIN paradigm (Koelewijn, Zekveld, Lunner, & Kramer, 2018, although see 2021). PPD
14 was measured in NH adults during a SIN task in quiet and targeting SRT80 and SRT50,
15 corresponding to low, medium and high listening task demands respectively. In blocks of 25
16 sentences, participants were promised either 5.0 or 0.2 Euro in exchange for correctly repeating 17
17 (68%) or more of the sentences. When listening demand was low, the size of the monetary
18 incentive had no effect on the PPD. With background noise present, greater incentive led to larger
19 PPD in both task demand conditions. This suggests that particularly when listening demands are
20 high, motivation may have an influence on listening effort.

21 The aim of the current study was to experimentally investigate the joint influence of fatigue and
22 motivation on the BPD and PPD during a SIN task in a pre-/ post-fatigue paradigm. Pre- and post-
23 fatigue listening effort was measured during SIN tasks in two motivation conditions. Following
24 previous findings of reduced arousal over time on task in SIN tasks (Alhanbali et al., 2020; Ayasse
25 & Wingfield, 2020; Zekveld et al., 2018, 2010), we predicated that BPD would decline from pre- to
26 post- fatigue (Hypothesis 1). As greater fatigue is thought to lead to greater suppression in effort

1 (Müller & Apps, 2019; Schneider et al., 2019), we expected a larger decline in PPD after larger
2 fatigue (Hypothesis 2). Because in a fatigued state the investment of effort is thought to depend on
3 the willingness to exert effort (Müller & Apps, 2019), we expected that any decrease in PPD from
4 pre- to post-fatigue should be larger when motivation is low (Hypothesis 3). Following the thinking
5 that fatigue lowers the willingness for further effort, we expected that the difference in the decline in
6 PPD between the smaller and greater fatigue should be larger when motivation is low (Hypothesis
7 4).

8 Methods

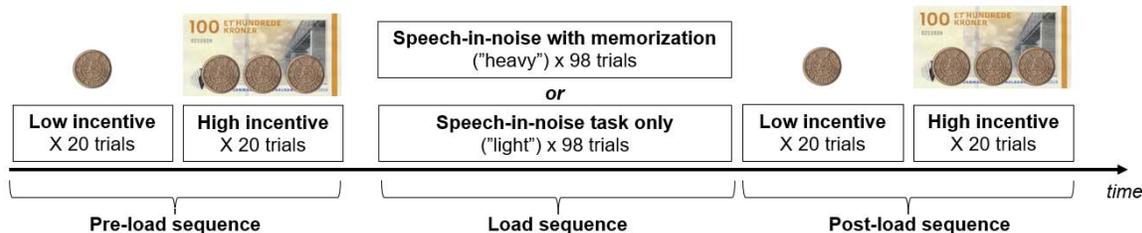
9 Participants

10 Thirty-two adults with normal hearing (i.e. hearing thresholds were less than or equal to 25 dB HL
11 over the frequencies from 0.25 to 4 kHz for both ears) participated. The data of 18 adults (Mean
12 age = 42; $SD = 16$; 7 female) was included in the analyses. Due to excess missing samples (see
13 trial inclusion criteria below), the data of 12 participants were excluded from the analyses. All
14 participants had normal hearing. Participants reported normal or corrected-to normal vision, Danish
15 as native language, and no neurological or psychiatric disorders. Participants were naïve to the
16 fatigue-related aim of the experiment. This was to avoid interoceptive awareness interfering with
17 the natural motivation to perform during the experiment. Later they were debriefed about the
18 purpose of the experiment. The study was approved by the local ethics committee and all
19 participants provided written informed consent.

20 Experimental design

21 Figure 1 illustrates the experimental design. To investigate the interactive influence of fatigue and
22 motivation on listening effort, a within-subject design with two levels of previous mental load and
23 two levels of monetary incentives was used. Listening effort was measured using pupillometry and
24 self-report during the pre- and post-load sequence probes. In this study the load sequence was
25 only used to induce fatigue. The assumptions that the load sequence induces fatigue and that the

- 1 load sequence with memorization is more effortful were made. No investigations were made on
- 2 any pupil behavior during the load sequence.



3

Figure 1 Experimental design. To induce fatigue, a long speech-in-noise test of 98 trials (i.e. “load sequence”) was used. Shorter blocks of 20 trials were administered pre- and post- load sequence to probe the effect of previous mental load on listening effort.

4 To investigate the influence of previous mental load magnitude on subsequent listening effort, a
5 SIN task with 98 sentences was used (i.e. “load sequence”). The load sequence was
6 approximately 40 minutes in duration. Each participant completed the experiment in 2 sessions. In
7 these sessions, the speech-in-noise task during the load sequence either excluded or included a
8 concurrent memory task (i.e. “light load sequence” and “heavy load sequence”, conditions
9 respectively). In the light load sequence condition, participants only repeated the final words of the
10 sentences that were presented. In the heavy load sequence condition, in addition to repeating the
11 final words of the sentences, participants were also asked to memorize the final words in blocks of
12 7 sentences. At the end of each 7th sentence, participants were asked to recall the 7 words that
13 were to be memorized.

14 To probe the influence of previous mental load on subsequent listening effort, both before and after
15 the load sequence, participants performed two SIN tasks (i.e. pre- and post-load sequence
16 ‘probes’). The pre- and post-load sequence probe SIN tasks consisted of 2 blocks of 20 trials each.
17 To investigate the moderating effect of monetary incentives on the influence of previous mental
18 load on subsequent listening effort, in the pre- and post-load sequence blocks, participants were
19 offered either 20 or 160 Danish Kroner per block (“high monetary incentive” and “low monetary

1 incentive” conditions, respectively) on the condition that they correctly repeated at least 12 of the
2 20 sentences (i.e. 60% correct). Per participant, the order of the monetary incentive conditions was
3 the same in the pre- and post-load sequence blocks, and the same across both experimental
4 sessions.

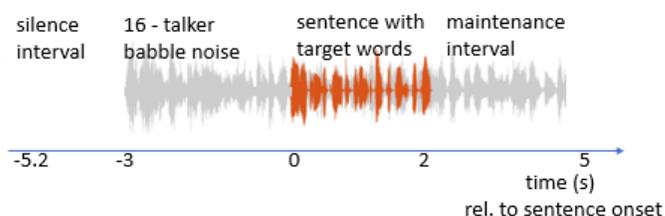
5 During the load sequence, the SNR within 7-sentence blocks was constant at either +1 or -4 dB,
6 and changed across 7-sentence blocks within sessions in a pre-determined sequence. These
7 SNRs were chosen to span a range of expected recognition rates around 65-100 % (Wendt et al.,
8 2018). To prevent participants from predicting the upcoming SNR during the load sequence and to
9 make the load sequence relatively homogenously taxing across time, the SNR sequence of +1, +1,
10 -4, -4, +1, -4, +1, +1, -4, -4, -4, +1, -4, and +1 dB was created. The SNR was cycled through the
11 sequence with a different starting point for each participant to prevent the possible influence of any
12 sequence-specific effects on subsequent listening effort. To allow comparison of the sessions for
13 the effect of memory load, per participant the SNR order was the same for the light load sequence
14 and heavy load sequence conditions (i.e. in both sessions of the experiment),.

15 The SNR in the pre- and post-load sequence blocks was individually pre-determined at SRT50 (as
16 described below). This is because, fatigue effects are known to be most evident while performing
17 difficult but achievable tasks (Lagory, Dearen, Tebo, & Wright, 2011). Previous research has
18 shown that the SNR at SRT50 elicited the largest pupil dilation as compared to the other SRT
19 levels, suggesting maximal effort at SRT50 in a rested state (e.g. Wendt, Koelewijn, Książek,
20 Kramer, & Lunner, 2018; Ohlenforst et al., 2018). Therefore, the SNR that corresponds to SRT50
21 was chosen for the pre- and post- load sequence probe blocks.

22 Stimuli and trial sequence

23 Target stimuli in the pre- and post-load sequence blocks consisted of sentences from the Danish
24 Dagmar-Asta-Tine corpus(DAT; Nielsen, Dau, & Neher, 2014) , whereas target stimuli in the load
25 sequence consisted of everyday sentences from the Danish Hearing in Noise Test (HINT; Nielsen
26 & Dau, 2011). The 16-talker babble masker noise consisted of speech segments from 2 female

1 and 2 male talkers. The long-term average frequency spectrum of the masker noise was identical
2 to that of the target speech signal of the corresponding speech corpus, separately for the DAT and
3 HINT material. In this study we only investigated listening effort during the pre- and post-load
4 sequence block. Therefore, only the sequence of trials in the probe blocks are described here.
5 Figure 2. illustrates the trial sequence in the pre- and post-load sequence blocks. 2.2 seconds after
6 the experimenter pressed a button to start a trial, the babble noise started. The target sentence
7 started 3 seconds after the start of the babble noise. All sentences were 2 seconds long and had
8 the same carrier words “Dagmar tænkte på (~550 ms) ... og en (~166 ms) ... I går (~333 ms)”
9 (English: Dagmar thought of ... and ... yesterday”); and 2 keywords (~450 ms and ~500 ms,
10 respectively). The babble noise continued for 3 seconds after the ending of the target sentence.
11 Thereafter participants were asked to repeat the keywords. Answers were counted as correct only
12 if both keywords had been repeated correctly. The experimenter scored the answers by a button-
13 press (“correct or incorrect”), which also triggered the start of the next trial.



14
15 Figure 2. Example trial sequence during the pre- and post-load sequence blocks. All trials started with 2.2
16 seconds silence, followed by a 16-talker babble noise. Then participants were presented with a sentence
17 (e.g. Dagmar tænkte på en aften o gen festdag I går). The noise continued for 3 seconds after the end of the
18 sentence. Participants were asked to repeat the keywords as accurately and as quickly as possible (e.g. “en
19 aften, festdag”).

20 Subjective report scales

21 After each of the pre- and post- load sequence blocks, participants were asked to report how much
22 effort they mobilized for listening, how well they think they performed, and how much they felt an
23 inclination to quit listening (cf. Koelewijn et al., 2018). Three 100-point visual analog scales were

1 printed in 18 font size on an A4 paper. The questions were (1) “*Hvor meget anstrengte du dig for at*
2 *høre sætningerne? –(How much effort did you put into hearing the sentences?);* 0 = Ingen, 25 =
3 Lav, 50 = Moderat, 75 = Høj, 100 = Meget høj – EN: 0 = None, 25 = Low, 50 = Moderate, 75 =
4 High, 100 = Very High), (2) “*Hvor mange af ordene tror du, at du forstod korrekt?*” (*How many of*
5 *the words do you think you understood correctly?*) ; 0 = Ingen, 25 = Mindre end halvdelen, 50 =
6 Halvdelen, 75 = Mere end halvdelen, 100 = Alle – EN: 0 = None, 25 = Less than half, 50 = Half, 75
7 = More than half, 100 = All), and (3) “*Hvor ofte måtte du opgive at forstå sætningen? (How often*
8 *did you have to give up understanding the phrase? (0 = Aldrig, 25 = Mindre end halvdelen af tiden,*
9 *50 = Halvdelen af tiden, 75 = Mere end halvdelen af tiden, 100 = Altid – EN: 0 = Never, 25 = Less*
10 *than half the time, 50 = Half the time, 75 = More than half the time, 100 = all the time).* Participants
11 marked their answers on paper.

12 Procedure and apparatus

13 Prior to the experiment participants were informed that the study was about listening effort and
14 motivation. All participants visited the laboratory twice: Once for the session with the light load
15 sequence condition, and once to complete the heavy load sequence condition. The visits were
16 maximum 3 weeks apart. Sessions began at either 9 am or 1 pm, and per participant, both visits
17 were scheduled at the same time of the day to avoid measurements during circadian spikes and to
18 minimize variability of diurnal changes in cortisol (Kirschbaum & Hellhammer, 1989; Lovallo &
19 Buchanan, 2016). Participants were advised to get a good night’s sleep the night before, not take
20 any alcohol 24 hours prior to, and not drink any caffeinated beverages 2 hours prior to the
21 experiment. These precautions were presumed also to minimise any confounding effects from
22 daily-life fatigue as observed by Wang et al. (2018).

23 All sessions took place in a sound-proof booth. The luminance in the booth was dim (approx. 100
24 lux) for all participants during all experimental sessions to avoid floor or ceiling effects in the pupil
25 diameter measures. Pupil size was recorded through the Tobii Pro Spectrum eye tracker at a
26 sampling frequency of 1200 Hz. All auditory stimuli were presented via five loudspeakers. The

1 loudspeakers (Genelec 8040A; Genelec Oy, Iisalmi, Finland) were distributed around the
2 participant along an imaginary circle with 1 meter radius. One loudspeaker was positioned in front
3 of the participant (0°) and the four other loudspeakers were positioned at the side and back of the
4 participant ($\pm 90^\circ$ and $\pm 270^\circ$). The target stimuli were presented from the loudspeaker in front
5 of the participants, while the 16-talker background noise was presented through the remaining
6 loudspeakers. Pupil size video recording and estimation was performed by a table mounted Tobii
7 Spectrum at 1200 Hz. All presentation of stimuli and collection of data was controlled using
8 software running on MATLAB 2018b (MathWorks, Natick, MA).

9 First, the speech reception threshold corresponding to 50% correct sentence recognition (SRT50)
10 was determined. This included a practice round and a test round with 20 DAT (Nielsen et al., 2014)
11 sentences each. The background masker was always presented at 70 dB SPL (as averaged
12 across 30 seconds). During the estimation of SRT50 the first sentence was presented at 70 dB and
13 (if the keywords were not repeated correctly) repeated with its level increased by 4dB until the
14 participants repeated all keywords correctly. Sentences 2-4 were presented at levels decreased or
15 increased by 4 dB, depending on whether the participants repeated all the keywords in the
16 previous sentence correctly or not, respectively. The level of the fifth sentence was calculated by
17 taking the average of the levels of the first four sentences and the level at which the fifth sentence
18 would have been presented with a step size of ± 4 dB. For the rest of the sentences, the level of
19 the target sentence was decreased or increased by 2 dB depending on correct or incorrect
20 (incomplete) repetition of the keywords, respectively. The estimated SRT50 was calculated as the
21 average SNR of the last five sentences and what a twenty-first sentence would have been
22 presented at.

23 Prior to the start of the experiment, to provide participants with an understanding of the listening
24 demands (i.e., SIN at the pre-determined and fixed SRT50), participants performed an additional
25 practice round with 20 sentences in blocks of 10. Participants received feedback from the
26 experimenter per block on how many sentences were repeated correctly.

1 At the start of each pupillometric data collection session, a 5-point eye-tracker calibration was
2 performed. Thereafter, participants were told that they could earn additional money by correctly
3 repeating minimum 12 of the sentences out of blocks of 20 sentences. Participants were instructed
4 to fixate on a cross on a black screen that was approx. 60 cm away from them and as far as
5 possible to schedule their blinks at the end of their response period. Before each of the 20-trial
6 blocks, participants were informed about the amount of monetary incentives through a verbal
7 message from the experimenter (e.g. “Now we are proceeding with a block of 20 sentences. If you
8 repeat more than 12 correctly, you will receive 160 Danish kroner”). In addition, visual reminders of
9 the amount were placed at approximately 20 cm away from the fixation cross on the computer
10 screen. During this session, no feedback was provided to the participants about their performance.
11 Participants were informed beforehand that this session would take approximately 2 hours, but
12 were not informed about how many blocks of trials the experiment had in total. After completing the
13 pre-load sequence probe blocks, participants took a 5-minute break. At the beginning of the load
14 sequence, participants received written and verbal information on the load sequence. To prevent
15 recovery from fatigue, immediately after the load sequence, participants continued with the post-
16 load sequence probe blocks without a break.

17 Pre-processing and calculation of the pupil dilation indices

18 Pupil data were processed using MATLAB R2018b (MathWorks, Natick, MA), similarly to the
19 procedure reported in Wendt et al. (2018). Pupil traces starting 1 s preceding sentence onset until
20 the end of the babble noise were selected for analyses. Traces with more than 30 % of missing
21 data were considered invalid. Per participant, the data from the eye with most valid traces were
22 selected for analysis. The data of a given participant were included in the analyses if 10 or more of
23 the trials per block of 20 trials were valid for all experimental conditions. Pupil diameter values
24 more than 2 standard deviations away from the trial median were coded as blinks. Blinks were later
25 interpolated in a linear fashion. The interpolation started for the samples within 35 ms before and
26 ended for the samples within 100 ms after the blink to account for blink-related changes in pupil

1 size. Later a moving average filter with a window of samples that would fall within 150 ms length
2 was used to smooth the de-blinked trials and to remove any high-frequency artefacts. The trials
3 were averaged per condition, per participant: probe time (pre- vs. post- load sequence) load
4 sequence magnitude (light vs. heavy) x monetary incentive (low vs. high). That is, there were 8
5 averaged traces per participant. For each averaged trace, the mean value of the trace
6 corresponding to the 1-second interval before the onset of the target sentence was taken as the
7 baseline pupil diameter (BPD). The maximum pupil size during the interval from the onset of the
8 sentence until the end of the presentation of the background babble, minus the BPD, was saved as
9 peak pupil dilation (PPD).

10 **Statistical analyses**

11 To the best of our knowledge this was the first study to investigate the influence of varying levels of
12 fatigue on the pupil dilation response in varying levels of motivation in a SIN task. Therefore, it was
13 not possible *a priori* to estimate the magnitudes of the main effects of fatigue and its interactions
14 with motivation on the PPD. To capture the contributions of all possible effects to the variance in
15 the PPD, and taking the small sample size into account, we took the approach of testing the
16 mentioned hypotheses with Type III ANOVA's instead of singling out specific contrasts.

17 To assess whether there was an overall change from pre- to post- load sequence, t-tests were
18 conducted, where the dependent measures of interest were speech recognition performance, BPD,
19 PPD, self-rated effort, self-rated performance, and self-rated tendency to quit, and the independent
20 variable was probe time (i.e. pre- vs. post-load sequence).

21 The effects of load sequence magnitude (light and heavy) and monetary incentive amount (low and
22 high) on the change in listening effort from pre- to post- load sequence were investigated by
23 calculating the change in the pupil dilation metrics and self-report scores from the pre- to the post-
24 load sequence blocks (i.e. $\Delta_{\text{post-pre}}$).

25 To control for any effect of the participant's initial state of fatigue (which may have been different
26 between the two sessions), we calculated the difference scores between pre- and post- load

1 sequence for the pupil and self-report measures. Using these difference scores as dependent
2 variables, we conducted 2-by-2 repeated-measures ANOVAs with the factors load sequence
3 magnitude (light and heavy) and monetary incentive amount (high and low).

4 Results

5 Speech reception thresholds used in the experiment

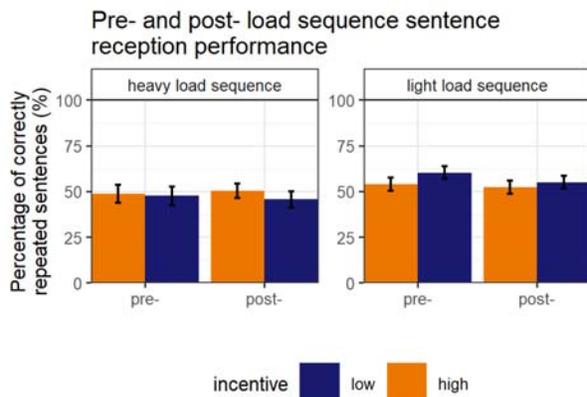
6 The average speech level corresponding to the SRT50 of the participants in a background noise of
7 70 dB was 65.2 dB (range: 62.6 - 68.4; $SD = 1.5$ dB), i.e. SRT50 = -4.8 dB SNR.

8 Performance in the load sequence

9 For the light load sequence condition, average performance for repeating the last words of the
10 sentences was 83.72 % (range: 67 – 94, $SD = 7.95$), whereas for the heavy sequence, average
11 performance for repeating the last words of the sentences was 79.78 % (range: 66 – 95, $SD =$
12 9.09). There was no evidence for a difference in repetition performance between the conditions
13 [$t_{paired}(17) = 1.967, p = .066$]. These results are reported in more detail in Książek et al. (under
14 review).

15 Speech recognition performance

16 Figure 3 shows pre- and post-load sequence speech recognition performance as a function of load
17 sequence magnitude and monetary incentive amounts. The average performance was 51.7 %
18 ($Median = 52.5, SD = 10.9$) and did not change from pre- to post- load sequence, [$t_{paired}(17) =$
19 0.695, $p = .497$]. The repeated measures ANOVA with Δ post-/pre- performance score as the
20 dependent variable, and load sequence magnitude (light vs. heavy) and monetary incentive
21 amount (high vs. low) as independent variables showed no main effect of load sequence
22 magnitude, [$F(1,17) = 0.716, p = .409, \eta_p^2 = .040$], no main effect of monetary incentive amount, [F
23 (1,17) = 0.250, $p = .624, \eta_p^2 = .014$], and no load sequence magnitude x monetary incentive
24 amount interaction effect, [$F(1,17) < 0.001, p = 1, \eta_p^2 < .001$]. In sum, there was no significant
25 difference in performance depending on any experimental condition.

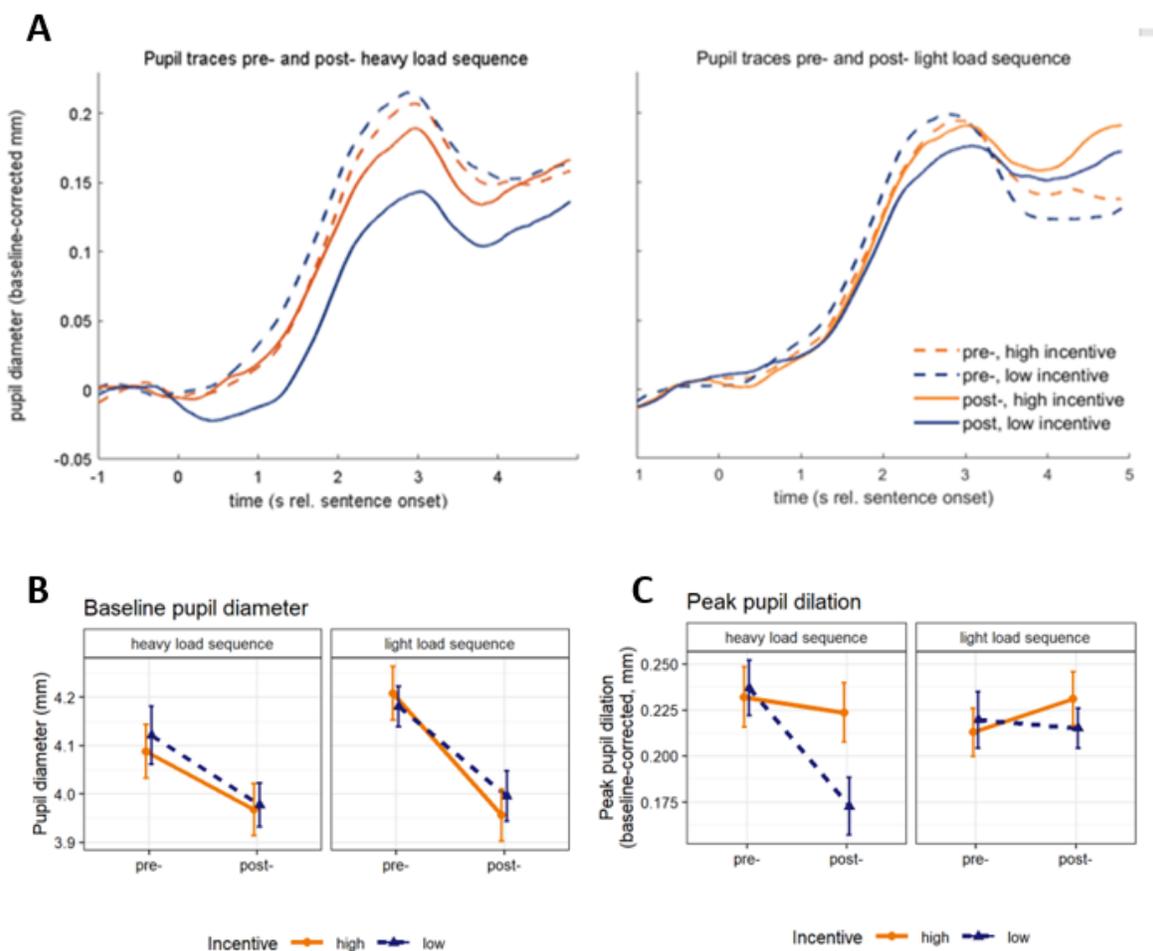


1

2 Figure 3. Sentence recognition performance during the pre- and post-load sequence probe blocks as a
3 function of load sequence magnitude and monetary incentive amounts.

4 Pupil traces, baseline pupil diameter (BPD), and peak pupil dilation (PPD)

5 Figure 4A shows the baseline-corrected pre- and post-load sequence pupil traces starting from the
6 beginning of the baseline interval, until the end of the maintenance interval, averaged across the
7 participants, and grouped together according to load sequence magnitude and monetary incentive
8 amount conditions. By visual inspection alone we observe that all traces show a peak around 1
9 second after the end of the target sentence (i.e., 3 seconds relative to target onset). Furthermore,
10 the peak is larger in the pre- load sequence condition as compared to post-load sequence. The
11 difference between pre- and post- load sequence peaks seems larger for the heavy load sequence
12 condition as compared to that of light, and larger for the low incentive condition as compared to
13 that of high.



1

2 Figure 4. Pre- and post- load sequence (A) baseline-corrected pupil traces (B) the baseline pupil diameter,
 3 and (C) baseline-corrected peak pupil dilation as a function of load sequence magnitude and monetary
 4 incentive amount. The error bars in (B) and (C) show the within-subject standard error of the mean.

5 Figure 4.B. shows the pre- and post- load sequence BPD as a function of load sequence
 6 magnitude and monetary incentive amount. On average pre-load sequence BPD was larger than
 7 post- load sequence BPD, [$t_{paired}(17) = 4.599, p < .001$]. The repeated measures ANOVA with
 8 $\Delta_{post-/pre-}$ of BPD as dependent variable, and load sequence magnitude and monetary incentive
 9 amount as the independent variables showed no effect of monetary incentive, [$F(1,17) = 0.639, p$
 10 $= .435, \eta_p^2 < .036$], load sequence, [$F(1,17) = 2.187, p = .157, \eta_p^2 < .114$], or monetary incentive x
 11 load sequence interaction effect, [$F(1,17) = 1.2, p = .289, \eta_p^2 < .066$].

1 Figure 4.C. shows the pre- and post- load sequence PPD as a function of load sequence
 2 magnitude and monetary incentive amount. Averaged across the monetary incentive and load
 3 sequence conditions, overall, the difference in PPD between the pre- and post- load sequence
 4 conditions was not significant, [$t_{paired}(17) = 1.195, p = .248$]. The repeated measures ANOVA with
 5 Δ post-/pre- of PPD as dependent variable, and monetary incentive and load sequence as the
 6 independent variables showed a main effect of load sequence, [$F(1,17) = 4.899, p = .041, \eta_p^2 <$
 7 $.224$], as the decline in PPD from pre- to post-load sequence probe was larger after the heavy load
 8 sequence as compared to the light load sequence. In addition, there was a main effect of monetary
 9 incentive, [$F(1,17) = 11.886, p = .003, \eta_p^2 < .411$], as the decline from pre- to post- was larger
 10 when monetary incentives were low. Although the mean difference between the monetary incentive
 11 conditions in the decline from pre- to post load sequence was larger after the heavy load sequence
 12 as compared to the light load sequence, the monetary incentive x load sequence interaction effect
 13 on the Δ post-/pre- of PPD was not significant, [$F(1,17) = 1.286, p = .273, \eta_p^2 < .070$]. However,
 14 post-hoc analyses with Tuckey's method showed that the Δ post-/pre- differed from zero only in the
 15 heavy-load-sequence-low-monetary-incentive condition [$p = 0.05$; all other p 's > 0.05].

16 **Subjective ratings of effort, performance, and quitting**

17 Table 1 shows pre- and post-load sequence self-rated effort, performance, and tendency to quit
 18 listening as a function of load sequence magnitude and monetary incentive amount. Shapiro-Wilk
 19 tests of normality confirmed that all the rating scores were normally distributed [all p 's > 0.05].

Table 1.					
<i>Self-rated effort, performance, and tendency to quit listening. Mean (s.d.) across participants</i>					
		Fatigue manipulation			
		heavy load sequence		light load sequence	
Rating scale (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)

1

2 *Self-rated effort*

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

9 *Self-rated performance*

10 On average there was no evidence for a change in self-rated performance from pre- to post-
 11 fatigue, [$t_{\text{paired}}(17) = 1.797$; $p = .090$]. The repeated measures ANOVA with $\Delta\text{post-/pre-}$ of self-rated
 12 performance as dependent variable, and load sequence and monetary incentive as the
 13 independent variables showed no effect of load sequence, [$F(1,17) = 0.773$, $p = .392$, $\eta_p^2 = .073$],
 14 or monetary incentive [$F(1,17) = 1.909$, $p = .185$, $\eta_p^2 = .101$]. Nor was the load sequence x
 15 monetary incentive interaction effect significant, [$F(1,17) = 4.360$, $p = .052$, $\eta_p^2 = .204$].

16 *Self-rated tendency to quit listening*

17 On average there was no evidence for a change in self-reported tendency for quitting from pre- to
 18 post- fatigue, [$t_{\text{paired}}(17) = 0.200$; $p = .844$]. The repeated measures ANOVA with $\Delta\text{post-/pre-}$ of self-
 19 rated tendency for quitting as dependent variable, and monetary incentive and load sequence as
 20 the independent variables showed no effect of load sequence, [$F(1,17) = 0.961$, $p = .341$, $\eta_p^2 =$

1 .053], monetary incentive [$F(1,17) = 1.156, p = .697, \eta_p^2 = .009$], or a load sequence x monetary
2 incentive interaction effect, [$F(1,17) = 0.836, p = .373, \eta_p^2 = .047$].

3 Discussion

4 The main aim of the current study was to investigate fatigue and motivation effects in pupillometric
5 measures of listening effort. This was realized through the assessment of listening effort in a pre-
6 /post- fatigue experiment, where two levels of load sequence (light and heavy) were used to induce
7 mental fatigue, and applying two levels of monetary incentives (high and low). We hypothesized a
8 decline in BPD from pre- to post- load sequence (Hypothesis 1). Furthermore, following the view
9 that the willingness to invest effort when in a fatigued state depends on the perceived value of
10 investing effort, we hypothesized that any decline in PPD should be larger after the heavier load
11 sequence (Hypothesis 2), and larger with smaller incentives (Hypothesis 3). The difference in
12 decline between the heavy and light conditions should be particularly larger when incentives are
13 low (Hypothesis 4).

14 Supporting Hypothesis 1, there was a decline in BPD from pre- to post- load sequence. In the
15 current experiment, the decline in BPD did not depend on whether the load sequence was heavy
16 or light, or whether the incentives were high and low, suggesting little influence of motivation and
17 magnitude of fatigue on the BPD. Previously Hopstaken et al., (2015) report a steeper decline over
18 the course of a visual 3-back task as compared to 2-back and 1-back tasks. This effect was
19 accompanied by performance improvement in the 3-back task and performance declines in the 2-
20 back and 1-back tasks. The steeper decline in BPD in the 3-back task was therefore attributed to
21 learning effects. Similarly, Ayasse & Wingfield (2020) report a decline in BPD over the course of a
22 SIN task that was steeper in individuals with poorer hearing thresholds. Inconsistent with the
23 fatigue interpretation, there was an improvement in performance. The authors concluded that the
24 BPD reflects a combination of different factors. In the current experiment, there was on average no
25 change in performance from pre- to post- load sequence, independently of whether the load
26 sequence was heavy or light. Self-reported tendency to quit listening did not change from the pre-

1 to the post-load sequence trials either. This suggests that there is no evidence of any subjective
2 experience of fatigue. On the other hand, self-report scale of tendency to quit may have been
3 insensitive to capture the experience of fatigue due to a response bias. That is, participants may
4 have been unwilling to admit their tendency to give up as part of social politeness (Grimm, 2010).
5 Although it is difficult to completely rule out or pinpoint the contributions of fatigue, habituation, or
6 learning to the overall decrease in BPD (arousal) that developed over the experiment, a decline in
7 baseline is consistent with the decline of capacity, which is viewed as a hallmark of fatigue
8 (Kahneman, 1973). Future studies could employ different (e.g. implicit) subjective measures to
9 better assess the development of the subjective experience of fatigue.

10 Supporting Hypothesis 2, the average decline in PPD was larger after the heavy load sequence as
11 compared to that after the light load sequence. One interpretation of this finding is that heavier load
12 sequence results in larger mental fatigue, which diminishes the perceived value of investing effort
13 in a greater extent than that in the light load sequence condition. Therefore, individuals
14 subsequently invest less effort in listening. Previously Hornsby (2013) showed greater increase in
15 secondary task RTs across a SIN task with concurrent memory load when participants were not
16 wearing hearing aids as compared to when they were, suggesting that greater mental demands
17 result in greater fatigue over time. Our results extend the findings of Hornsby (2013) with
18 pupillometric measures, and support the understanding that reductions in mental demands of
19 previous load can reduce fatigue effects. Another explanation for the larger decline in PPD after
20 the heavy load sequence may be the learning of a listening strategy that is more economical on
21 effort during the heavy load sequence condition. That is, to manage concurrent listening and
22 memory demands, participants may have adopted an approach in listening that requires minimum
23 effort and have continued with it throughout the post- load sequence blocks.

24 Supporting Hypothesis 3, on average, the decline in PPD from pre- to post- load sequence was
25 larger when the incentives were low. This result suggests that motivational factors play a role in
26 determining how much listeners exert effort after a period of effortful exertion. The moderating

1 effect of monetary incentives can be interpreted in consideration of the FUEL and complementary
2 models of motivational fatigue (Hockey, 2010; Müller & Apps, 2019; Pichora-Fuller et al., 2016).
3 That is, after performing the load sequence, investing effort may have been perceived as
4 worthwhile on trials with high monetary incentives and less worthwhile on trials with lower
5 monetary incentives.

6 Overall, we did not observe any interaction between load sequence magnitude and monetary
7 incentives on the PPD (Hypothesis 4 not supported); however, post-hoc analyses showed that the
8 change in PPD from pre- to post- load sequence was only significant in the heavy-load-low-
9 incentive condition. This result is in line with the understanding that the effect of fatigue is
10 moderated by motivational factors (Hockey, 2010; Müller & Apps, 2019; Pichora-Fuller et al.,
11 2016). Although learning effects on the decline in PPD from pre- to post- load sequence in the
12 heavy load sequence condition should not be ruled out, they are insufficient to explain the effect of
13 monetary incentives on the change in PPD. The current finding suggests that in a SIN task, fatigue
14 effects are only visible when the willingness to perform well is low. In other words, the effect of
15 fatigue on the PPD may be masked by motivational factors.

16 Whereas Wang et al (2018) have shown correlational evidence for the relationship between daily-
17 life fatigue and PPD, the current results are the first to refine this understanding by demonstrating
18 the impact of task-induced fatigue on PPD.

19 Limitations

20 The study has some limitations that should be mentioned. Technical difficulties in the collection of
21 pupil data led to the exclusion of 12 participants from the study, potentially limiting the power of our
22 statistical analyses. The self-report measures that were used in the study may have not been
23 adequate to capture all dimensions and components of motivation and fatigue, such as sensitivity
24 to monetary rewards, or sleepiness and tiredness. Better such measures would have been useful
25 in evaluating the extent to which the offered monetary incentives were effective in manipulating
26 motivation, and to the nature of the mental state that the load sequence task has elicited. In this

1 study the effects of motivation and task-induced fatigue were investigated in a sample of NH
2 adults. The generalizability of the mechanisms explored here to listening-related daily-life fatigue in
3 adults with HI is left unclear. Longitudinal studies utilizing physiological methods would be required
4 if one wished to shed light on how task-induced fatigue relates to the development of daily-life
5 fatigue in adults with HI.

6 **Conclusions**

7 The present study demonstrates that task-induced fatigue and motivation interactively affect the
8 capacity mobilized for listening. The BPD declined from pre- to post- load sequence, suggesting an
9 overall decrease in arousal, but the decline did not scale with the magnitude of the previous mental
10 load. The PPD declined from pre- to post- load sequence significantly only in the heavy load
11 sequence-low monetary incentive condition. This supports the understanding that the investment of
12 listening effort depends on an individual's listening goals (Pichora-Fuller et al., 2016). The current
13 results show that the influence of fatigue on effort is not uniform across different motivational
14 states, suggesting that models of listening effort and listening-related fatigue may need to consider
15 motivation (the willingness to invest effort) and (listening-related) fatigue as different parameters
16 that can have a range of values.

17

- 1 References
- 2 Alhanbali, S., Munro, K. J., Dawes, P., Carolan, P. J., & Millman, R. E. (2020). Dimensions of self-
- 3 reported listening effort and fatigue on a digits-in-noise task, and association with baseline
- 4 pupil size and performance accuracy. *International Journal of Audiology*, *0*(0), 1–11.
- 5 <https://doi.org/10.1080/14992027.2020.1853262>
- 6 Ayasse, N. D., & Wingfield, A. (2020). Anticipatory Baseline Pupil Diameter Is Sensitive to
- 7 Differences in Hearing Thresholds. *Frontiers in Psychology*, *10*(January), 1–7.
- 8 <https://doi.org/10.3389/fpsyg.2019.02947>
- 9 Bess, H. F., & Hornsby, W., Y., B. (2014). Commentary: Listening Can Be Exhausting—Fatigue in
- 10 Children and Adults With Hearing Loss Fred. *Ear and Hearing*, *35*(6), 592–599.
- 11 <https://doi.org/10.1097/AUD.000000000000099>. Commentary
- 12 Bo Nielsen, J., Dau, T., & Neher, T. (2014). A Danish open-set speech corpus for competing-
- 13 speech studies. *The Journal of the Acoustical Society of America*, *135*(1), 407–420.
- 14 <https://doi.org/10.1121/1.4835935>
- 15 Boksem, M. A. S., Meijman, T. F., & Lorist, M. M. (2006). Mental fatigue, motivation and action
- 16 monitoring. *Biological Psychology*, *72*(2), 123–132.
- 17 <https://doi.org/10.1016/j.biopsycho.2005.08.007>
- 18 Boksem, M. A. S., & Tops, M. (2008). Mental fatigue: Costs and benefits. *Brain Research Reviews*,
- 19 *59*(1), 125–139. <https://doi.org/10.1016/j.brainresrev.2008.07.001>
- 20 Brehm, J. W., & Self, E. A. (1989). The intensity of motivation. *Annual Review of Psychology*,
- 21 *40*(1973), 109–131. <https://doi.org/10.1146/annurev.ps.40.020189.000545>
- 22 Dwyer, N. ., Firszt, J. B., & Reeder, R. M. (2014). Effects of unilateral input and mode of hearing in
- 23 the better ear: self-reported performance using the speech, spatial and qualities of hearing
- 24 scale. *Ear and Hearing*, *35*(1), 126–136. <https://doi.org/10.1097/AUD.0b013e3182a3648b>

- 1 Gilzenrat, M. S., Nieuwenhuis, S., Jepma, M., & Cohen, J. D. (2010). Pupil diameter tracks
2 changes in control state predicted by the adaptive gain theory of locus coeruleus function.
3 *Cognitive, Affective and Behavioral Neuroscience*, 10(2), 252–269.
4 <https://doi.org/10.3758/CABN.10.2.252>
- 5 Grimm, P. (2010). Social Desirability Bias. In *Marketing Research*.
6 <https://doi.org/https://doi.org/10.1002/9781444316568.wiem02057>
- 7 Hockey, G. R. J. (1997). Compensatory control in the regulation of human performance under
8 stress and high workload: A cognitive-energetical framework. *Biological Psychology*, 45(1–3),
9 73–93. [https://doi.org/10.1016/S0301-0511\(96\)05223-4](https://doi.org/10.1016/S0301-0511(96)05223-4)
- 10 Hockey, G. R. J. (2010). A motivational control theory of cognitive fatigue. *Cognitive Fatigue:*
11 *Multidisciplinary Perspectives on Current Research and Future Applications.*, (January 2011),
12 167–187. <https://doi.org/10.1037/12343-008>
- 13 Holman, J. A., Drummond, A., Hughes, S. E., & Naylor, G. (2019). Hearing impairment and daily-
14 life fatigue: a qualitative study. *International Journal of Audiology*, 58(7), 408–416.
15 <https://doi.org/10.1080/14992027.2019.1597284>
- 16 Hopstaken, J. F., van der Linden, D., Bakker, A. B., & Kompier, M. A. J. (2015a). A multifaceted
17 investigation of the link between mental fatigue and task disengagement. *Psychophysiology*,
18 52(3), 305–315. <https://doi.org/10.1111/psyp.12339>
- 19 Hopstaken, J. F., van der Linden, D., Bakker, A. B., & Kompier, M. A. J. (2015b). The window of
20 my eyes: Task disengagement and mental fatigue covary with pupil dynamics. *Biological*
21 *Psychology*, 110, 100–106. <https://doi.org/10.1016/j.biopsycho.2015.06.013>
- 22 Hornsby, W., Y., B., Naylor, G., & Bess, H. F. (2016). A Taxonomy of Fatigue Concepts and Their
23 Relation to Hearing Loss. *Ear and Hearing*, 37(1), 136–144.
24 <https://doi.org/doi:10.1097/AUD.0000000000000289>.

- 1 Hornsby, B. W. Y. (2013). The effects of hearing aid use on listening effort and mental fatigue
2 associated with sustained speech processing demands. *Ear and Hearing*, *34*(5), 523–534.
3 <https://doi.org/10.1097/AUD.0b013e31828003d8>
- 4 Hornsby, B. W. Y., & Kipp, A. M. (2016). Subjective ratings of fatigue and vigor in adults with
5 hearing loss are driven by perceived hearing difficulties not degree of hearing loss. *Ear and*
6 *Hearing*, *37*(1), e1–e10. <https://doi.org/10.1097/AUD.0000000000000203>
- 7 Kahneman, D. (1973). Attention and Effort. In *The American Journal of Psychology* (Vol. 88).
8 <https://doi.org/10.2307/1421603>
- 9 Kirschbaum, C., & Hellhammer, D. H. (1989). Salivary Cortisol in Psychobiological Research: An
10 Overview. *Biological Psychology/Pharmacopsychology*, *22*, 150–169.
11 <https://doi.org/https://doi.org/10.1159/000118611>
- 12 Koelewijn, T., Zekveld, A. A., Lunner, T., & Kramer, S. E. (2018). The effect of reward on listening
13 effort as reflected by the pupil dilation response. *Hearing Research*, *367*, 106–112.
14 <https://doi.org/10.1016/j.heares.2018.07.011>
- 15 Koelewijn, T., Zekveld, A. A., Lunner, T., & Kramer, S. E. (2021). The effect of monetary reward on
16 listening effort and sentence recognition. *Hearing Research*, *406*, 108255.
17 <https://doi.org/10.1016/j.heares.2021.108255>
- 18 Kramer, S. E., Kapteyn, T. S., Festen, J. M., & Kuik, D. J. (1997). Assessing aspects of auditory
19 handicap by means of pupil dilatation. *International Journal of Audiology*, *36*(3), 155–164.
20 <https://doi.org/10.3109/00206099709071969>
- 21 Kramer, S. E., Kapteyn, T. S., & Houtgast, T. (2006). Occupational performance: Comparing
22 normally-hearing and hearing-impaired employees using the Amsterdam Checklist for Hearing
23 and Work. *International Journal of Audiology*, *45*(9), 503–512.
24 <https://doi.org/10.1080/14992020600754583>

- 1 Lagory, J., Dearen, B. B., Tebo, K., & Wright, R. A. (2011). Reported fatigue, difficulty, and
2 cardiovascular response to an auditory mental arithmetic challenge. *International Journal of*
3 *Psychophysiology*, 81(2), 91–98. <https://doi.org/10.1016/j.ijpsycho.2011.05.005>
- 4 Lovallo, W. R., & Buchanan, T. W. (2016). Stress hormones in psychophysiological research:
5 Emotional, behavioral, and cognitive implications. In *Handbook of Psychophysiology, Fourth*
6 *Edition* (pp. 465–494). <https://doi.org/10.1017/9781107415782.021>
- 7 McGarrigle, R., Dawes, P., Stewart, A. J., Kuchinsky, S. E., & Munro, K. J. (2017). Pupillometry
8 reveals changes in physiological arousal during a sustained listening task. *Psychophysiology*,
9 54(2), 193–203. <https://doi.org/10.1111/psyp.12772>
- 10 Müller, T., & Apps, M. A. J. (2019). Motivational fatigue: A neurocognitive framework for the impact
11 of effortful exertion on subsequent motivation. *Neuropsychologia*, 123(October 2017), 141–
12 151. <https://doi.org/10.1016/j.neuropsychologia.2018.04.030>
- 13 Murphy, P. R., Robertson, I. H., Balsters, J. H., & O'connell, R. G. (2011). Pupillometry and P3
14 index the locus coeruleus-noradrenergic arousal function in humans. *Psychophysiology*,
15 48(11), 1532–1543. <https://doi.org/10.1111/j.1469-8986.2011.01226.x>
- 16 Nachtegaal, J., Kuik, D. J., Anema, J. R., Goverts, S. T., Festen, J. M., & Kramer, S. E. (2009).
17 Hearing status, need for recovery after work, and psychosocial work characteristics: Results
18 from an internet-based national survey on hearing. *International Journal of Audiology*, 48(10),
19 684–691. <https://doi.org/10.1080/14992020902962421>
- 20 Naylor, G., Koelewijn, T., Zekveld, A. A., & Kramer, S. E. (2018). The Application of Pupillometry in
21 Hearing Science to Assess Listening Effort. *Trends in Hearing*, 22, 1–3.
22 <https://doi.org/10.1177/2331216518799437>
- 23 Nielsen, J. B., & Dau, T. (2011). The Danish hearing in noise test. *International Journal of*
24 *Audiology*, 50(3), 202–208. <https://doi.org/10.3109/14992027.2010.524254>

- 1 Ohlenforst, B., Wendt, D., Kramer, S. E., Naylor, G., Zekveld, A. A., & Lunner, T. (2018). Impact of
2 SNR, masker type and noise reduction processing on sentence recognition performance and
3 listening effort as indicated by the pupil dilation response. *Hearing Research*, 365, 90–99.
4 <https://doi.org/10.1016/j.heares.2018.05.003>
- 5 Pichora-Fuller, M. K., Kramer, S. E., Eckert, M. A., Edwards, B., Hornsby, B. W. Y., Humes, L. E.,
6 ... Wingfield, A. (2016). Hearing impairment and cognitive energy: The framework for
7 understanding effortful listening (FUEL). *Ear and Hearing*, 37, 5S-27S.
8 <https://doi.org/10.1097/AUD.0000000000000312>
- 9 Portnuff, C., & Bell, B. (2014). Effective Use of Speech-in-Noise Testing in the Clinic. *Hearing*
10 *Journal*, 72(5). <https://doi.org/10.1097/01.HJ.0000559502.51932.b1>
- 11 Richter, M., Gendolla, G. H. E., & Wright, R. A. (2016). *Three Decades of Research on*
12 *Motivational Intensity Theory*. (December), 149–186.
13 <https://doi.org/10.1016/bs.adms.2016.02.001>
- 14 Schneider, E. N., Bernarding, C., Francis, A. L., Hornsby, B. W. Y., & Strauss, D. J. (2019). A
15 Quantitative Model of Listening Related Fatigue. *International IEEE/EMBS Conference on*
16 *Neural Engineering, NER, 2019-March*, 619–622. <https://doi.org/10.1109/NER.2019.8717046>
- 17 van der Linden, D. (2011). The urge to stop: The cognitive and biological nature of acute mental
18 fatigue. In P. L. Ackerman (Ed.), *Cognitive fatigue: Multidisciplinary perspectives on current*
19 *research and future applications*. Retrieved from <https://doi.org/10.1037/12343-007>
- 20 Wang, Y., Naylor, G., Kramer, S. E., Zekveld, A. A., Wendt, D., Ohlenforst, B., & Lunner, T. (2018).
21 Relations Between Self-Reported Daily-Life Fatigue, Hearing Status, and Pupil Dilation During
22 a Speech Perception in Noise Task. *Ear and Hearing*, 39(3), 573–582.
23 <https://doi.org/10.1097/AUD.0000000000000512>
- 24 Wendt, D., Koelewijn, T., Książek, P., Kramer, S. E., & Lunner, T. (2018). Toward a more

- 1 comprehensive understanding of the impact of masker type and signal-to-noise ratio on the
2 pupillary response while performing a speech-in-noise test. *Hearing Research*, 369, 67–78.
3 <https://doi.org/10.1016/j.heares.2018.05.006>
- 4 Winn, M. B., Wendt, D., Koelewijn, T., & Kuchinsky, S. E. (2018). Best Practices and Advice for
5 Using Pupillometry to Measure Listening Effort: An Introduction for Those Who Want to Get
6 Started. *Trends in Hearing*, 22, 1–32. <https://doi.org/10.1177/2331216518800869>
- 7 Zekveld, A. A., Koelewijn, T., & Kramer, S. E. (2018). The Pupil Dilation Response to Auditory
8 Stimuli: Current State of Knowledge. *Trends in Hearing*, 22, 1–25.
9 <https://doi.org/10.1177/2331216518777174>
- 10 Zekveld, A. A., Kramer, S. E., & Festen, J. M. (2010). Pupil response as an indication of effortful
11 listening: The influence of sentence intelligibility. *Ear and Hearing*, 31(4), 480–490.
12 <https://doi.org/10.1097/AUD.0b013e3181d4f251>

13

14