



# The effect of monetary reward on listening effort and sentence recognition



Thomas Koelewijn<sup>a,b,c,\*</sup>, Adriana A. Zekveld<sup>a</sup>, Thomas Lunner<sup>d,e,f,g</sup>, Sophia E. Kramer<sup>a</sup>

<sup>a</sup> Department of Otolaryngology – Head and Neck surgery, Amsterdam UMC, Vrije Universiteit Amsterdam, Ear & Hearing, Amsterdam Public Health research institute, De Boelelaan 1117, Amsterdam, the Netherlands

<sup>b</sup> Department of Otorhinolaryngology/Head and Neck Surgery, University Medical Center Groningen, University of Groningen, 9700 RB Hanzeplein 1, Groningen 9713GZ, the Netherlands

<sup>c</sup> Research School of Behavioral and Cognitive Neuroscience, Graduate School of Medical Sciences, University of Groningen, Groningen, the Netherlands

<sup>d</sup> Eriksholm Research Centre, Sneekkersten, Denmark

<sup>e</sup> Department of Electrical Engineering, Hearing Systems, Hearing Systems group, Technical University of Denmark, Kgs., Lyngby, Denmark

<sup>f</sup> Division of Technical Audiology, Department of Clinical and Experimental Medicine, Linköping University, Linköping, Sweden

<sup>g</sup> Department of Behavioural Sciences and Learning, Linköping University, Linköping, Sweden

## ARTICLE INFO

### Article history:

Received 22 January 2020

Revised 6 April 2021

Accepted 13 April 2021

Available online 24 April 2021

### Keywords:

Monetary reward

Listening effort

Pupil response

Speech reception

Sentence recognition

## ABSTRACT

Recently we showed that higher reward results in increased pupil dilation during listening (listening effort). Remarkably, this effect was not accompanied with improved speech reception. Still, increased listening effort may reflect more in-depth processing, potentially resulting in a better memory representation of speech. Here, we investigated this hypothesis by also testing the effect of monetary reward on recognition memory performance. Twenty-four young adults performed speech reception threshold (SRT) tests, either hard or easy, in which they repeated sentences uttered by a female talker masked by a male talker. We recorded the pupil dilation response during listening. Participants could earn a high or low reward and the four conditions were presented in a blocked fashion. After each SRT block, participants performed a visual sentence recognition task. In this task, the sentences that were presented in the preceding SRT task were visually presented in random order and intermixed with unfamiliar sentences. Participants had to indicate whether they had previously heard the sentence or not. The SRT and sentence recognition were affected by task difficulty but not by reward. Contrary to our previous results, peak pupil dilation did not reflect effects of reward. However, post-hoc time course analysis (GAMMs) revealed that in the hard SRT task, the pupil response was larger for high than low reward. We did not observe an effect of reward on visual sentence recognition. Hence, the current results provide no conclusive evidence that the effect of monetary reward on the pupil response relates to the memory encoding of speech.

© 2021 The Author(s). Published by Elsevier B.V.

This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>)

## 1. Introduction

The Framework for Understanding Effortful Listening (FUEL, Pichora-Fuller et al., 2016), postulates that the amount of effort expended during listening depends on our motivation to achieve goals and attain rewards. In addition, the motivational intensity theory (Brehm and Self, 1989) states that motivational arousal occurs when a task is sufficiently difficult, within one's capacity, and is justified by the magnitude of reward. Consistent with FUEL, the motivational intensity theory, and previous re-

search (Bijleveld et al., 2009; Knapen et al., 2016; Richter, 2016), we recently observed an effect of monetary reward on listening effort when the speech reception test was sufficiently difficult (Koelewijn et al., 2018). The results showed a larger peak pupil dilation (PPD), indicating more effort, while processing spoken sentences in a high compared to a low reward condition. These results are consistent with Richter (2016) who showed an effect of reward on pre-ejection period (PEP) reactivity, an indicator of sympathetic activity and associated with effort, in a difficult task condition. Also, a study by Bijleveld et al. (2009) showed an effect of reward on the PPD when participants had to listen to, memorize, and report back 2 or 5 digits. Together these studies provide compelling evidence showing that monetary reward can affect the amount of effort spent during a listening task.

\* Corresponding author at: Department of Otorhinolaryngology/Head and Neck Surgery, University Medical Center Groningen, University of Groningen, 9700 RB Hanzeplein 1, Groningen 9713GZ, the Netherlands.

E-mail address: [t.koelewijn@rug.nl](mailto:t.koelewijn@rug.nl) (T. Koelewijn).

Interestingly, although participants in our recent study (Koelewijn et al., 2018) had a larger PPD in the high reward condition as compared to the low reward condition, their speech reception threshold (SRT) did not improve. This was surprising as one may expect that the additional effort invested in listening associated with the high reward would result in better performance. On the other hand, Richter (2016) did not show an effect of reward on performance either, despite the effect of reward on PEP. Bijleveld et al. (2009) observed a 96.9% mean task accuracy over all conditions, suggesting that the performance they measured was at ceiling. We suggested that the absent effect of reward on performance may indicate that speech perception is largely automatic and highly efficient, so trying even harder will not result in improved performance (Koelewijn et al., 2018). However, the additional effort may be associated with the ability to stay engaged in a conversation (Brehm and Self, 1989), which may justify the investment of additional effort in the listening task. Also, in the context of an experiment, reward itself (e.g., monetary reward) could be the primary motivator for participants to spend more effort irrespective of their actual task performance. However, it is also possible that the additional effort invested does affect processes involved in sentence recognition or recall that do not necessarily lead to a change in the SRT. Candidate processes are cognitive processes that are affected by the allocation of resources such as memory processes (Johansson et al., 2018; Pichora-Fuller et al., 2016).

Processes that are likely affected by reward are related to the encoding and storage of speech information as measured by recognition or (free) recall of this information (e.g., Koeritzer et al., 2018; Ng et al., 2013). Although speech recall and recognition can be considered functionally different (e.g., Balota and Neely, 1980), they are considered to rely on similar cognitive processes and reflect information being memorized (Anderson and Bower, 1974; Yonelinas, 2002). Research shows that stimuli preceded by a higher reward are more likely to be remembered (e.g., Adcock et al., 2006; Gruber and Otten, 2010; Madan and Spetch, 2012). Madan and Spetch (2012) showed that a high relative to a low reward was associated with better free recall of visually presented words. Interestingly, a recent study by Zhang et al. (2019) showed an effect of reward on both the pupil dilation response as well as on performance. In this study, the auditory sentences/questions were presented at five different speech rates and the design included five levels of reward. Furthermore, participants answered auditorily presented questions in the form of relatively long speech stimuli (e.g., 'The stone is on the left of the ball, the kite is on the right of the ball, where is the stone in relation to the kite?'). The length of these stimuli might result in more memory related processing making them more susceptible for reward, than would be the case for shorter every day sentences (Versfeld et al., 2000). Although previous work didn't show an effect of reward on SRT, effects of reward may emerge in tasks in which participants are explicitly asked to encode information.

Research has shown that providing a high reward can increase effort allocation (e.g., Koelewijn et al., 2018; Richter, 2016) and memory performance (e.g., Adcock et al., 2006; Gruber and Otten, 2010; Madan and Spetch, 2012). Nevertheless, this does not imply that increased effort allocation during listening causes or predicts higher memory performance. Listening effort is defined as the deliberate allocation of mental resources to overcome obstacles in goal pursuit when listening (FUEL, Pichora-Fuller et al., 2016). This means that listening effort reflects the total processing load of a multitude of hearing and language related processes that may only partially relate to memorizing what is heard. Research shows that listening effort related to increasing age and hearing loss, coincides with a decline in recall memory, which is in line with the effortfulness hypothesis. This hypothesis states that extra

resources needed for the perceptual processing of speech in listeners with hearing loss, are no longer available for memory encoding (McCoy et al., 2018; Tun et al., 2009; 2012). Thus, effort can reflect resource use at various speech processing stages and can be related to specific challenges of specific listener groups. This means that we should be careful when generalizing outcomes over these groups.

Speech recall is known to be adversely affected by the level of background noise (Kjellberg et al., 2008; Rabbitt, 1968; Surprenant, 2007) and interfering speech or babble (Heinrich et al., 2018; Ng et al., 2013). Ng et al. (2013) used a sentence final word identification and recall (SWIR) test to assess memory processing. The SWIR test consisted of two consecutive tasks. In the 'identification task' participants were asked to report the final word of each sentence immediately after listening to it. After eighth sentences, participants performed a free recall task in which they were asked to recall, in any order, all the sentence-final words that they had previously reported. The outcomes showed that interfering speech (compared to speech shaped noise) especially disrupts recall performance. Note that interfering speech is also known to strongly affect listening effort (Koelewijn et al., 2012; Ohlenforst et al., 2017; Wendt et al., 2018). Koeritzer et al. (2018) applied a speech perception test followed by a visual sentence recognition task. In this task, participants had to report whether a visually presented sentence had appeared previously in an auditory SRT task. In the recognition task, SRT sentences were mixed with novel sentences that shared several words with the target sentences. Participants responded with yes or no, by pressing a keypad, to indicate whether that exact sentence was presented in the preceding SRT task. Both hits and false alarms were scored from which a  $d'$  prime ( $d'$ ) was calculated as outcome measure. Speech recognition was adversely affected by increases in semantic ambiguity, by decreases in signal to noise ratio (SNR) when masked with eight-talker babble, and for masked sentences compared to sentences presented in quiet. In summary, in addition to hearing loss, auditory task difficulty also seems to affect memory processing (Koeritzer et al., 2018).

In the current study we explored the effect of reward on visual sentence recognition. We applied the same SRT task as in our previous study (Koelewijn et al., 2018) but in addition implemented a visual sentence recognition task, which was administered directly after the SRT task (Koeritzer et al., 2018). The main purpose of this study was to investigate the effect of monetary reward on recognition memory. Additionally, we aimed to replicate our previous findings (Koelewijn et al., 2018) indicating an increase in PPD for high as compared to low monetary reward. Hence, in the current study participants performed the same SRT task as previously (Koelewijn et al., 2018). The SRT task was either difficult (hard) or easy and participants received either a high or low monetary reward when they repeated at least 70% of the sentences correctly. During the recognition memory task, the list of 25 sentences presented just before in the SRT task were mixed with a list of 25 unfamiliar (foil) sentences. Sentences were visually presented one at a time on a computer screen. Participants had to respond whether or not they had heard the sentence during the SRT task. During the SRT task we recorded the pupil dilation response and calculated the SRT, and for the recognition memory task a  $d'$  was calculated. In line with our previous study, we hypothesized a lower SRT and larger PPD (more effort) in the hard compared to the easy SRT task. Based on previous research (e.g., Rabbitt, 1968) we also expected better recognition of sentences presented in the easy compared to the difficult SRT tasks. While we did not expect an effect of reward on the SRT based on our previous results (Koelewijn et al., 2018), we expected larger pupil responses (i.e. increased listening effort) for high compared to low reward, reflecting more in-depth processing of speech information, potentially resulting in a better

memory representation of the presented stimuli. Hence, we hypothesized better visual sentence recognition in the high reward than in the low reward conditions.

## 2. Methods

In this study, we used the same setup and mostly the same design as in our previous study (Koelewijn et al., 2018). Differences compared to the previous design were the exclusion of a speech in quiet SRT task and the inclusion of the visual sentence recognition task.

### 2.1. Participants

Twenty-four normal hearing adults (3 males, 21 females), recruited at the VU University and Amsterdam UMC, participated in the study. The sample size for this study was based on a moderate effect size of (intentional) attention related processes on the PPD (at a significance of .05 and a statistical power of .80), as observed in a previous study (Koelewijn et al., 2015). Ages of the participants ranged from 18 to 26 years with a median age of 21 years. Normal hearing was defined as pure-tone thresholds less than or equal to 20 dB HL at the octave frequencies 0.25–4 kHz. Participants' pure-tone hearing thresholds averaged over both ears and over the octave frequencies 1–4 kHz (three-frequency pure-tone average), ranged from -3.3 to 10.8 dB hearing level (HL) (mean = 1.9 dB HL, standard deviation (SD) = 3.4 dB). Participants had no history of neurological diseases and reported normal or corrected-to-normal vision. They were native Dutch speakers and provided written informed consent in accordance with the Ethics Committee of the VU University Medical Center, Amsterdam.

### 2.2. SRT task and Reward

During the SRT task, participants had to listen to and repeat everyday Dutch sentences from the VU98 corpus (Versfeld et al., 2000), uttered by a female talker. An example of one of these sentences is 'Het is drie uur in de middag', which directly translates to 'It is three 'clock in the afternoon'. The sentences were presented simultaneously with a single-talker masker. This study used a within-subject design with four blocks: difficulty (2 levels) x reward (2 values). The 'easy' condition was tracking 85% correct speech reception in noise and the 'hard' condition was tracking 50% correct speech reception. At the start of the session, participants practiced the easy and hard listening conditions by listening and responding to 10 practice sentences each. The conditions were presented in a blocked fashion and while participants were unaware of the actual intelligibility levels, they were informed about the difficulty level (i.e. hard or easy) at the start of each block. Additionally, they were told that they could earn a high (5 euros) or low (0.20 euro) reward when repeating at least 70% of the sentences correctly, independent of their performance on the subsequent visual sentence recognition task. The participants did not receive any feedback, neither after individual trials nor at the end of the block. The order of the blocks (conditions) was counterbalanced (Latin square) over participants. Each block contained 25 trials (sentences) and during each trial, the pupil diameter of each participant was recorded.

The single-talker masker used in the SRT task contained concatenated sentences from another set of the VU98 corpus uttered by a male talker. The masker was spectrally shaped in order to obtain the same long-term average frequency spectrum as the female speech. Note that this shaping did not affect the fundamental or formant frequencies of the male voice, which made it still easy to differentiate the male and female voice. The SRT (dB SNR) was adaptively estimated separately for each reward value, targeting a

sentence intelligibility level of either 50% or 85% correct using a weighted up-down method (Kaernback, 1991). The sentence was only scored as correct if each word was correctly repeated in the right order. Each trial started with the onset of the masker, which was 3 s prior to the onset of the target sentence and continued for 3 s after the offset of the target sentence. The length of each trial co-varied with the length of the presented sentence, which had a mean duration of 1.90 s (range 1.38 to 2.70 s). At the end of the masker a 1 kHz prompt tone was presented for 0.5 s after which participants were instructed to repeat the sentence.

After each block, participants were asked to rate their effort, performance, and tendency to quit performing the task during the last block (Koelewijn et al., 2012; Zekveld et al., 2010). Effort was rated on a visual analogue scale from 0 ('no effort') to 10 ('very effortful'). For the performance rating, participants estimated the proportion of sentences they had perceived correctly, which was scaled from 0 ('none of the sentences were intelligible') to 10 ('all sentences were intelligible'). For quitting, participants indicated how often they had abandoned the task because it was too difficult. This was rated from 0 ('this happened for none of the sentences') to 10 ('this happened for all of the sentences').

### 2.3. Visual sentence recognition task

After listening to each block and after filling out the rating scales, participants performed a visual sentence recognition task. In line with Koeritzer et al. (2018), in this task, the 25 sentences presented in the previous SRT task were mixed with 25 foil sentences that were unfamiliar to the participants. These foil sentences were selected from the same set (Versfeld et al., 2000) such that they had one overlapping key word with one of the target SRT sentences from the preceding block. Key words were nouns, adjectives, or additives. The overlap in key words between the SRT task sentences and foil sentences, was done to avoid ceiling effects in the sentence recognition performance for the easy conditions (Koeritzer et al., 2018). During the test, a small number of key words appeared more than twice. The 50 sentences that were presented in each sentence recognition block, were individually presented in random order on a computer screen. The sentence was displayed until participants indicated whether or not they had heard the sentence by pressing the 'J' (yes) or 'N'(no) key on a qwerty keyboard. The task was not speeded, although participants were told not to take too much time when in doubt. A  $d'$  score was calculated for each condition to serve as an outcome measure. This was done using the loglinear approach (Hautus, 1995) that deals with hit and false alarm rates of zero or one by calculating them for each data cell as  $\frac{\Sigma x - 0.5/n + 1}{n}$  where  $\Sigma x$  equals the sum of hits or false alarms respectively, and  $n$  equals the number of trials. The visual sentence recognition task was practiced by the participants using a block of 20 sentences following each of the two SRT practice blocks. Before the start of the experiment participants were explicitly told that the reward related to the SRT task performance only.

### 2.4. Apparatus and procedure

The test setup was identical to the one used in our previous study (Koelewijn et al., 2018), and all testing was performed in the same sound treated room. After recording the participant's audiogram and testing for near-vision acuity using Bailey and Lovie (1980) word charts, they performed the SRT and visual sentence recognition tasks. During these tasks, participants were seated at 65 cm viewing distance in front of a computer screen (Dell, 17 inch). During the SRT tasks, the pupil diameter of both eyes was measured at a 60 Hz sampling rate using an infrared eye tracker (SMI RED250mobile System). The light intensity of

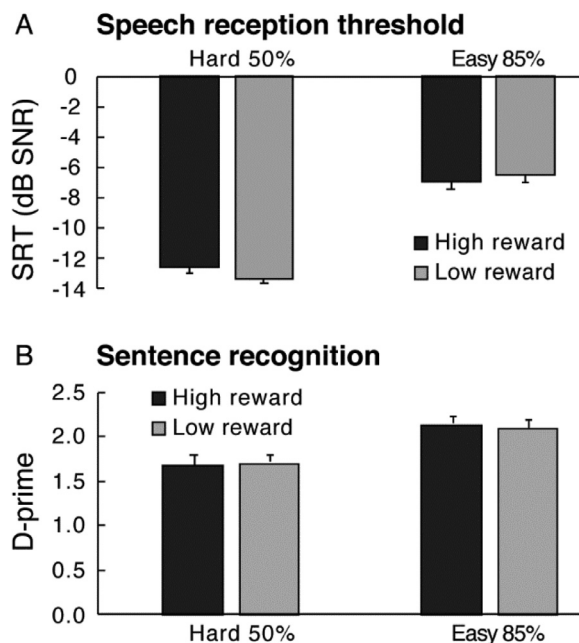
the LEDs attached to the ceiling of the room was adjusted such that, for each participant, the pupil diameter was around the middle of its dynamic range. For the SRT task, audio in the form of wave files (44.1 Hz, 16 bit) was presented diotically through headphones (Sennheiser, HD 280, 64  $\Omega$ ) by an external soundcard (asus Xonar Essence One). All tests were presented by a lap-top computer running Windows 10 (HP ZBook, SMI RED250mobile System). The whole procedure, including measurement of pure-tone hearing thresholds, near vision acuity, calibrating the eye-tracker, practicing and performing the SRT and visual sentence recognition tasks, and a 10 min break halfway through the SRT task took 2 h. At the end of the session participants were informed about their performance and were debriefed about the task. They all received 5.20 euros reward, for their performance in the two easy conditions, in addition to the 7.50 euros hourly rate.

### 2.5. Pupil data analysis

Consistent with the scoring of the SRT, the pupil data of the first five trials of each block were excluded from the pupil data analyses. For the remaining 20 trials per condition, zero values in the pupil diameter traces (pupil diameter data over time for each trial) within the time window of 1 s before and 4.3 s after sentence onset were coded as eye-blinks. Traces with 20% or more eye-blinks were excluded from further analysis (1.1% of all trials). In the remaining pupil traces, as well as in the  $x$ - and  $y$ -coordinate traces of the pupil center (reflecting eye-gaze direction), blinks were removed by linear interpolation between the fifth sample before and the eighth sample after the blink.  $X$  and  $y$ -coordinate traces, within 1 s before and 4.3 s after sentence onset, that contained change in eye-gaze coordinates more than 10 degrees from fixation (reflecting moderate to large eye-movements) were removed from analysis (5.5% of the remaining trials). After de-blinking and the exclusion of trials containing eye-movements, a five-point moving average smoothing filter was passed over the pupil data to remove any high-frequency artifacts. For each remaining trace a baseline value was calculated. The baseline was the mean pupil size within the 1 s period before sentence onset, during which participants were listening to the speech masker. The baseline period is shown by the left and middle dotted vertical lines in both plots in Fig. 2. All remaining traces were baseline corrected by subtracting the trial's baseline value from the value for each time point within that trace. Traces were averaged over the trials in each condition, separately for each participant. Within the average trace, the MPD (mm) was defined as the average pupil dilation relative to baseline in the interval between target speech onset and the start of the response prompt, shown by the middle and right dotted vertical lines in both plots in Fig. 2. Within this same time window, the PPD (mm) was defined as the largest value relative to the baseline. PPD latency (ms) was defined relative to the sentence onset. Lastly, for each participant and each condition, the average pupil diameter at baseline was calculated. For seven participants the data of the right eye were analyzed, as the data for the left eye was of insufficient quality.

### 2.6. Statistics

For all dependent behavioral (SRT), pupil (MPD, PPD, PPD latency, baseline), self-rating scales (effort, performance, and quitting), and visual sentence recognition ( $d'$ ) measures, we performed a  $2 \times 2$  analyses of variance (ANOVA) with condition (easy and hard) and reward (high and low) as the within-subject factors (Table 3). Additionally, we applied a non-linear regression analysis, in the form of Generalized Additive Mixed Models (GAMMs) to analyze the effect of reward on pupil size over time within



**Fig. 1.** (A) SRTs (dB SNR) at two difficulty levels for each reward, averaged over participants. The error bars show standard errors. (B) Average outcomes ( $d'$ ) for the visual sentence recognition task as function of the reward and difficulty conditions in the preceding SRT tasks, averaged over participants. The error bars show standard errors.

the easy and hard listening conditions. GAMMs assessed the effect from speech onset until response prompt.

GAMMs can model non-linear regressions by using smooth functions that consist of non-linear patterns consist of a combination of basis functions of which the number is pre-specified (Wieling, 2018). GAMMs describe the data over time rather than looking at a significant difference between conditions at the PPD or in the average pupil size over a predefined time window. An advantage of GAMMs over other time course analyses, like Growth Curve Analysis, is that it can include an autoregressive error model to deal with autocorrelation in the residuals, which is shown to occur strongly in pupil data (van Rij et al., 2017).

The analysis was performed in R using the packages *mgcv* (Wood, 2018) and *itsadug* (van Rij et al., 2017). In line with van Rij et al., (2019) and Boswijk et al. (2020) who performed GAMMs on changes in pupil size over time as well, we used the 'bam' function to fit GAMMs. Within this function we defined 'Pupil' (pupil diameter in mm over time) as the dependent variable, and included the four conditions (easy – low reward, easy – high reward, hard – low reward, hard – high reward). The GAMMs used by van Rij et al., (2019) and Boswijk et al., (2020), included X and Y eye-position data, to account for changes in pupil size caused by gaze position. Eye-position data was not included in the GAMMs of the current study, because participants were instructed to fixate their eyes during listening. Instead, conditions containing large eye-movement artifacts were excluded from analysis.

## 3. Results

Table 1 shows the average SRTs (Fig. 1A), pupil response parameters, and subjective ratings as a function of reward for both the hard and easy difficulty conditions. Average outcomes for the visual sentence recognition task as function of the reward and difficulty conditions in the preceding SRT tasks are presented in Table 2 and Fig. 1B. The average pupil traces over participants for each condition are plotted in Fig. 2.

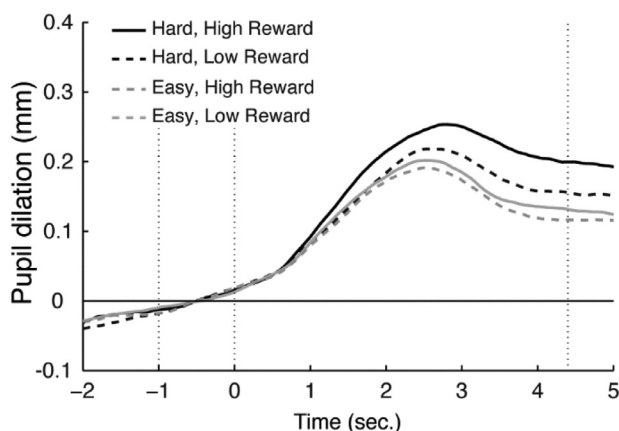
**Table 1.**  
Average speech reception threshold (SRT), Pupil data, self-rated effort, performance, and quitting scores.

Reward	Difficulty			
	Hard		Easy	
	High	Low	High	Low
Behavioral	SRT (SD), dB SNR			
	-12.85 (1.65)	-13.29 (1.75)	-6.99 (2.23)	-6.55 (2.07)
Pupil	MPD (SD), mm			
	0.17 (0.07)	0.14 (0.07)	0.13 (0.09)	0.12 (0.07)
	PPD (SD), mm			
	0.27 (0.11)	0.24 (0.10)	0.22 (0.11)	0.20 (0.09)
Subjective	PPD latency (SD), s			
	2.92 (0.64)	2.95 (0.64)	2.71 (0.61)	2.57 (0.43)
	Baseline (SD), mm			
	5.30 (0.59)	5.26 (0.53)	5.21 (0.50)	5.22 (0.57)
	Self-rated Effort (SD) (low = 0, high = 10)			
	7.64 (1.12)	7.39 (1.08)	5.55 (1.55)	5.19 (1.29)
Self-rated Performance (SD) (low = 0, high = 10)				
	5.20 (1.06)	5.04 (1.27)	6.95 (0.98)	6.87 (1.13)
Self-rated Quitting (SD) (low = 0, high = 10)				
	2.25 (1.29)	2.36 (1.29)	1.24 (0.95)	1.46 (1.51)

**Table 2.**  
Average visual sentence recognition task data in the form of a  $d'$  and response time (TR), for each reward level for the easy and hard conditions.

Reward	Difficulty			
	Hard		Easy	
	High	Low	High	Low
Recognition	$d'$ (SD)			
	1.67 (0.52)	1.70 (0.44)	2.13 (0.42)	2.08 (0.42)
	Hits % (SD)			
	64.83 (9.87)	64.17 (11.10)	77.00 (8.61)	76.33 (9.13)
	Misses % (SD)			
	35.17 (9.87)	35.83 (11.10)	23.00 (8.61)	23.50 (9.08)
	Correct Rejections % (SD)			
	90.17 (6.46)	90.67 (8.40)	92.00 (7.64)	92.00 (5.00)
False Alarms % (SD)				
	9.83 (6.46)	9.33 (8.40)	8.00 (7.64)	8.00 (5.00)
Average Response Time over Hits and Correct Rejections (SD, s)				
	1.60 (0.26)	1.65 (0.29)	1.66 (0.40)	1.56 (0.28)

**Average pupil responses**



**Fig. 2.** Average pupil responses for the hard and easy conditions for low and high rewards. The onset of the sentences was at 0 s. The baseline was the average pupil diameter over one second preceding the start of the sentence. The area between the second and third dotted lines indicates the time window used for calculating the mean pupil dilation.

**3.1. Behavioral outcomes**

In line with the results of our previous study, the analysis of the SRTs (see Table 3) revealed a significant main effect of task difficulty ( $F_{[1,23]} = 669.82, p < .001$ ), showing lower SRTs for the hard compared to the easy conditions. No effect of reward ( $F_{[1,23]} = 0.64, p = .637$ ) or interaction between task difficulty and reward ( $F_{[1,23]} = 2.32, p = .141$ ) on the SRT was found. The recognition  $d'$  showed a significant main effect of task difficulty ( $F_{[1,23]} = 24.30, p < .001$ ), revealing a higher  $d'$  score for the easy compared to the hard conditions. There was no effect of reward ( $F_{[1,23]} = 0.02, p = .886$ ) or interaction between task difficulty and reward ( $F_{[1,23]} = 0.27, p = .609$ ) on the  $d'$  score for visual sentence recognition.

**3.2. Pupil measures analysis**

Analysis of the PPDs (see Table 3) revealed a significant main effect of task difficulty ( $F_{[1,23]} = 13.03, p = .001$ ) with larger PPDs in the hard than in the easy conditions. No effect of reward ( $F_{[1,23]} = 2.92, p = .101$ ) or an interaction between task difficulty and reward ( $F_{[1,23]} = 0.52, p = .478$ ) was observed. For the MPDs similar outcomes were observed. Analysis of the PPD latencies showed a significant effect of task difficulty ( $F_{[1,23]} = 7.11, p = .014$ ) with larger PPD latencies in the hard than in the easy conditions. Again, no effect or reward ( $F_{[1,23]} = 0.69, p = .415$ ) or

**Table 3.** Outcomes of SRT, pupil measures, subjective ratings, and Recognition performance, using  $2 \times 2$  ANOVAs with repeated measures 'task difficulty' (hard, easy) and 'reward' (high, low).

	Difficulty		Reward		Difficulty * Reward	
	F <sub>[1,23]</sub>	p	F <sub>[1,23]</sub>	p	F <sub>[1,23]</sub>	p
SRT	669.82	< .001	.229	.637	2.32	.141
MPD	7.88	.010	2.85	.105	.77	.391
PPD	13.03	.001	2.92	.101	.52	.478
PPD latency	7.11	.014	.69	.415	.69	.416
Baseline	2.08	.163	.24	.627	.20	.658
Self-rated Effort	70.08	< .001	2.779	.109	.10	.753
Self-rated Performance	66.23	< .001	.63	.434	.06	.817
Self-rated Quitting	19.53	< .001	1.23	.279	.11	.748
Recognition d'	24.30	< .001	.02	.886	.27	.609

**Table 4.** Summary of the outcomes of the GAMM model for the effects of Condition and Time.

Parametric coefficients	Estimates	Std. error	t-value	Pr(>  t )	
Intercept	0.122786	0.007342	16.724	< 2e-16	***
Condition85%high	0.009991	0.010348	0.966	0.3343	
Condition50%low	0.019661	0.010303	1.908	0.0564	
Condition50%high	0.045387	0.010320	4.398	1.9e-05	***
Smooth terms	Edf	Ref.df	F	p-value	
s(Time): Condition85%low	8.847	8.9995	430.5	<2e-16	***
s(Time): Condition85%high	8.860	8.996	445.0	<2e-16	***
s(Time): Condition50%low	8.853	8.995	436.2	<2e-16	***
s(Time): Condition50%high	8.853	8.995	533.7	<2e-16	***
s(Event)	1547.325	1794.000	280.3	<2e-16	***
s(Time, Event)	1700.062	1792.000	332.1	<2e-16	***

Deviance explained = 81.9%  
 \* =  $p < 0.05$ , \*\* =  $p < 0.01$ , \*\*\* =  $p < 0.001$

interaction ( $F_{[1,23]} = 0.69, p = .416$ ) was observed. There were no significant main effects or interactions on the pupil baseline.

### 3.3. GAMMs

Although the main effect of reward was not significant, visual inspection of the pupil dilation responses (Fig. 2) suggested a slightly higher PPD for the high as compared to the low reward conditions in the second half of the pupil dilation response function (after the maximum dilation). Winn and Moore (2018) referred to the time period after the PPD until response as the 'retention interval', or the time period where listeners prepare their verbal response (Piquado et al., 2010). Hence, our post-hoc hypothesis of the effect of reward was that it may specifically affect this retention interval and therefore, that it is more time specific than the effect of task difficulty.

To test this post-hoc hypothesis, we used Generalized Additive Mixed Models (GAMMs) to assess the effect of reward over time, from speech onset until response prompt, for both the easy and hard listening conditions. For the GAMMs, the pre-processing of the pupil data was the same as for the pupil measures analysis, except that the data was down-sampled from 60 Hz to 30 Hz and the five-point moving average smoothing filter was excluded to avoid introducing extra autocorrelation between the successive samples. The following model was used: Pupil ~ Condition + s(Time, by = Condition) + s(Event, bs = "re") + s(Time, Event, bs = "re"). This model specifies that pupil size depends on condition and for each condition a nonlinear regression line was estimated (see Fig. 3A). A random intercept and a random slope were included for each unique trial-participant combination (Event). To control for autocorrelation in the residuals, the *acf\_resid* function of the *itsadug* package was used. Finally, the parameters *rho* (output from the *acf\_resid* function) and *AR.start* were added to the revised *bam* model. For more details on how to perform GAMMs on pupillometry data see van Rij et al. (2019).

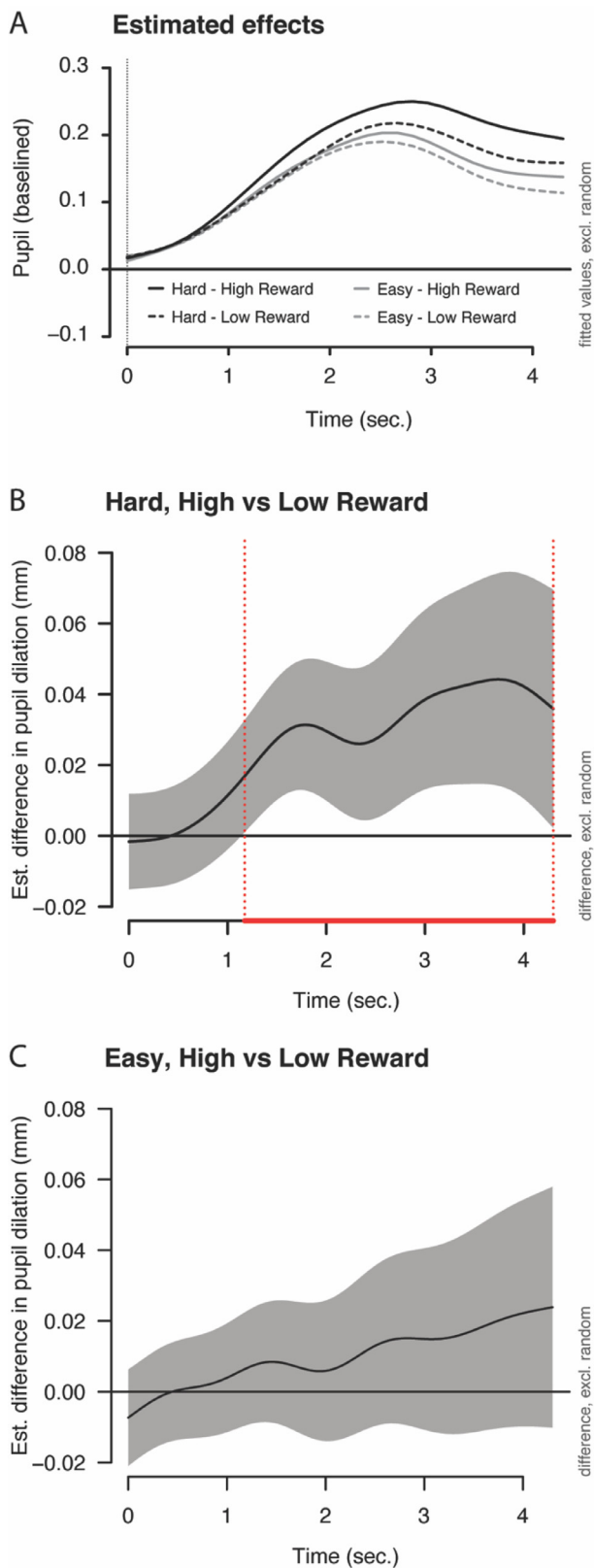
The final model accounted for 81.9% of the deviance explained. The parametric coefficients and the approximate significance of smooth terms are reported in Table 4. Fig. 3 shows the estimated effects and the estimated difference in pupil size over time for the high minus low reward conditions, for the hard (Fig. 3B) and easy (Fig. 3C) listening conditions separately. Within these difference plots, significant differences are determined based on visual inspection of where the difference plots deviate from the 0 line. This is indicated by the red line on the x-axis of the differences plot in Fig. 3B. It shows significant estimated differences, over time, between high and low reward for the hard listening condition.

### 3.4. Self-rated scores

In line with our previous study, self-rated effort, performance, and quitting all showed a significant main effect of task difficulty (see Table 3), with a hard SRT task resulting in higher effort and quitting ratings and a lower performance rating relative to the easy SRT task. No significant main effect of reward or an interaction effect was found.

## 4. Discussion

This study replicated known effects of difficulty level on the pupil dilation response. However, in contrast to Koelewijn et al. (2018), no main effect of reward on the PPD was observed. Post-hoc time course analysis (GAMMs) did however show an effect of reward on the pupil size in the hard listening condition. The results indicated larger pupil dilation for the high reward relative to the low reward condition for the hard SRT task. This suggested that reward has an effect on the amount of effort exerted when the task is sufficiently difficult which is in line with the motivation intensity theory (Brehm and Self, 1989). Note that the time period in which reward had a significant effect



**Fig. 3.** (A) The estimated effects for pupil dilation in the hard and easy conditions for low and high rewards. (B) Estimated difference between high and low reward in the hard condition with pointwise 95% confidence intervals. (C) Estimated difference between high and low reward in the easy condition with pointwise 95% confidence intervals. The onset of the sentences was at 0 s. Significant differences (deviations from the 0 line) are marked in red.

on the pupil size continues until after the PPD, which possibly includes processes related to response preparation.

With exception of the pupil baseline, all pupil measures (MPD, PPD, and PPD latency) were larger for the hard than for the easy SRT condition (e.g., Zekveld et al., 2010), suggesting that the average and maximum processing load, as well as the time at which the maximum processing load was reached, increased with task difficulty. Analysis of the behavioral outcomes showed no effect of reward on the SRT scores, which is in line with previous results (Koelewijn et al., 2018). Surprisingly, and against our hypothesis, reward did not affect visual sentence recognition ( $d'$ ). In previous studies, hard listening conditions resulted in lower SRT scores and lower sentence recognition (Koeritzer et al., 2018) compared to easy listening conditions. In all, the current results did not replicate our previous outcomes (Koelewijn et al., 2018) showing an effect of reward on the PPD. Instead, the (exploratory) post-hoc GAMM analysis suggested that participants did exploit more effort in processing speech in the high reward compared to the low reward conditions in the hard listening conditions (see Fig. 3). However, this occurred in the absence of behavioral benefits as reflected by the SRT and visual sentence recognition scores.

There were some notable differences between the current and our previous study (Koelewijn et al., 2018), which might have resulted in the absence of a main effect of reward on the PPD as observed before. Inclusion of the visual sentence recognition task in the current study made the test session for participants longer and more complex. Although visual sentence recognition was a separate task, participants were aware that it always followed the SRT task. Moreover, participants were explicitly told that the reward could be earned by performing well at the SRT task and was unrelated to recognition performance. Increase in task complexity between the previous (Koelewijn et al., 2018) and current study, and maybe also the increase in duration of the task instruction between studies, could have decreased the significance of the reward instruction for the participant, which might explain the absence of a significant main effect of reward on the PPD. Note that Madan and Spetch (2012), who used a 'surprise' recall task, for which participants receive no prior instruction, did show an effect of reward on the free recall of (visually presented) words. Hence, decreased significance of the reward instruction, could have resulted in a smaller effect of reward on the PPD.

The current study showed no effect of reward on visual sentence recognition, even though previous research did show that stimuli preceded by a higher reward are more likely to be remembered (e.g., Adcock et al., 2006; Gruber and Otten, 2010; Madan and Spetch, 2012). The data furthermore revealed that the visual sentence recognition task was relatively easy. Most sentences that were repeated completely correctly in the SRT task also resulted in hits during the recognition task (86%). Although the foil sentences in the recognition task contained key words overlapping with the target sentences, correct rejections rates in the recognition task were > 90% in all conditions. The visual sentence recognition task might have been too easy and therefore not sensitive to the effect of reward. Even when using a more sensitive task, research indicates that not everyone shows an effect of reward on performance. Callan and Schweighofer (2008) revealed that anticipation of reward affects dopaminergic-modulated brain plasticity in such way that based on the level of reward-induced anxiety, reward either results in a positive or negative effect on word learning behavior. In other words, personal traits like anxiety, not tested for in the current study, could account for variance in the effect of reward on performance.

Interestingly, some previous studies have investigated brain response during the time period at which a monetary reward was presented (e.g., Adcock et al., 2006; Gruber and Otten, 2010). Gruber and Otten (2010), showed electrical brain activity during

the presentation of a monetary reward to be predictive of how well a visually presented word was remembered. An fMRI study by Adcock et al. (2006) suggests reward to positively affect memory formation through dopamine release in the hippocampus before the learning stage. In the current study, the amount of reward (0.20 or 5 euros) was only presented once at the start of each block. Therefore, we did not collect pupillometry data during these events. In contrast to the current study, Zhang et al. (2019) presented the amount of reward on a trial basis by means of an auditory cue. Their results show that the pupil response to the auditory reward-cue increased with the amount of the reward. However, the baseline correction performed on the pupil response to the auditory sentences/questions makes these results difficult to interpret as this baseline was influenced by the auditory cue. Future research using a design that rewards participants on a trial basis and that allows to assess the pupil data during these reward periods (Zhang et al., 2019), might provide more insights in the effect of reward on speech processing performance and memory.

In addition to the previous results (Koelewijn et al., 2018), PPD latency was affected by task difficulty, showing a longer latency for the hard compared to the easy condition. Previous findings showing larger PPDs being related to larger PPD latencies (Koelewijn et al., 2015), which is suggested to reflect increased processing time related to the complexity or amount of information. Finally, self-rated effort, performance, and quitting, showed an effect of difficulty but not of reward, which is in line with previous results (Koelewijn et al., 2018).

To conclude, the increase in listening effort as shown for high reward conditions (Koelewijn et al., 2018), reflects an increase in cognitive processing load also during the retention period. The effect of monetary reward on listening effort most likely relates to changes in the level of task engagement. This is in line with the results of the GAMM analysis showing an effect of reward only for the hard listening conditions where task engagement is most necessary, and predicted by the motivation intensity theory (Brehm and Self, 1989). Although increased task engagement has shown to affect memory (e.g., Madan and Spetch, 2012), the current study using a visual sentence recognition task did not show this effect. Future research using a more optimized design might allow these effects to be shown.

### CRedit authorship contribution statement

**Thomas Koelewijn:** Conceptualization, Methodology, Software, Formal analysis, Investigation, Data curation, Writing - review & editing, Visualization, Project administration, Funding acquisition. **Adriana A. Zekveld:** Conceptualization, Methodology, Formal analysis, Writing - original draft, Writing - review & editing. **Thomas Lunner:** Conceptualization, Methodology, Writing - original draft. **Sophia E. Kramer:** Conceptualization, Methodology, Writing - original draft, Writing - review & editing, Supervision.

### Acknowledgments

This research was funded by Oticon Fonden (Foundation) Grant 16-0463. We thank Marita Everhardt and Vincent Boswijk (University of Groningen, The Netherlands) for providing their expertise on GAMMs.

### References

Adcock, R.A., Thangavel, A., Whitfield-Gabrieli, S., Knutson, B., Gabrieli, J.D.E., 2006. Reward-motivated learning: mesolimbic activation precedes memory formation. *Neuron* 50 (3), 507–517.

Anderson, J.R., Bower, G.H., 1974. A propositional theory of recognition memory. *Mem. Cognit.* 2, 406–412.

Bailey, I.L., Lovie, J.E., 1980. The design and use of a new near-vision chart. *Am. J. Optom. Phys. Opt.* 57, 378–387.

Balota, D.A., Neely, J.H., 1980. Test-expectancy and word-frequency effects in recall and recognition. *J. Exp. Psychol. Hum. Learn. Mem.* 6, 576–587.

Bijleveld, E., Custers, R., Aarts, H., 2009. The unconscious eye opener: pupil dilation reveals strategic recruitment of resources upon presentation of subliminal reward cues. *Psychol. Sci.* 20 (11), 1313–1315.

Boswijk, V., Loerts, H., Hilton, N.H., 2020. Salience is in the eye of the beholder: increased pupil size reflects acoustically salient variables. *Ampersand* 7, 1000061.

Brehm, J.W., Self, E.A., 1989. The intensity of motivation. *Annu. Rev. Psychol.* 40, 109–131.

Callan, D.E., Schweighofer, N., 2008. Positive and negative modulation of word learning by reward anticipation. *Hum. Brain Mapp.* 29 (2), 237–249.

Gruber, M.J., Otten, L.J., 2010. Voluntary control over prestimulus activity related to encoding. *J. Neurosci.* 30 (29), 9793–9800.

Hautus, M.J., 1995. Corrections for extreme proportions and their biasing effects on estimated values of  $d'$ . *Behav. Res. Meth. Ins. C.* 27 (1), 46–51.

Heinrich, A., Schneider, B.A., Craik, F.I.M., 2018. Investigating the influence of continuous babble on auditory short-term memory performance. *Q. J. Exp. Psychol.* 61 (5), 735–751.

Johansson, R., Pärnamets, P., Bjernstedt, A., Johansson, M., 2018. Pupil dilation tracks the dynamics of mnemonic interference resolution. *Sci. Rep.* 1–8.

Kaernback, C., 1991. Simple adaptive testing with the weighted up-down method. *Percept. Psychophys.* 49 (3), 227–229.

Kjellberg, A., Ljung, R., Hallman, D., 2008. Recall of words heard in noise. *Appl. Cogn. Psychol.* 22 (8), 1088–1098.

Knapen, T., de Gee, J.W., Brascamp, J., Nuiten, S., Hoppenbrouwers, S., Theeuwes, J., 2016. Cognitive and ocular factors jointly determine pupil responses under equi-luminance. *PLoS ONE* 11, e0155574–e0155613.

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. *Hear. Res.* 323, 81–90.

Koelewijn, T., Zekveld, A.A., Festen, J.M., Kramer, S.E., 2012a. Pupil dilation uncovers extra listening effort in the presence of a single-talker masker. *Ear Hear.* 33, 291.

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. *Hear. Res.* 367, 106–112.

Koeritz, M.A., Rogers, C.S., Van Engen, K.J., Peelle, J.E., 2018. The impact of age, background noise, semantic ambiguity, and hearing loss on recognition memory for spoken sentences. *J. Speech Lang. Hear. Res.* 61 (3) 740–12.

Madan, C.R., Spetch, M.L., 2012. Is the enhancement of memory due to reward driven by value or salience? *Acta Psychol.* 139 (2), 343–349.

McCoy, S.L., Tun, P.A., Cox, L.C., Colangelo, M., Stewart, R.A., Wingfield, A., 2018. Hearing loss and perceptual effort: downstream effects on older adults' memory for speech. *Q. J. Exp. Psychol. A Hum. Exp. Psychol.* 58, 22–33.

Ng, E.H.N., Rudner, M., Lunner, T., Pedersen, M.S., Rönneberg, J., 2013. Effects of noise and working memory capacity on memory processing of speech for hearing-aid users. *Int. J. Audiol.* 52, 433–441.

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. *Hear. Res.* 351, 68–79.

Pichora-Fuller, K.M., 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 Hear.* 37 (Suppl 1), 5S–27S.

Piquado, T., Cousins, K.A.Q., Wingfield, A., Miller, P., 2010. Effects of degraded sensory input on memory for speech: behavioral data and a test of biologically constrained computational models. *Brain Res.* 1365, 48–65.

Rabbitt, P.M.A., 1968. Channel-capacity, intelligibility and immediate memory. *Q. J. Exp. Psychol.* 20 (3), 241–248.

Richter, M., 2016. The moderating effect of success importance on the relationship between listening demand and listening effort. *Ear Hear.* 37 (Suppl 1), 111S–117S.

van Rij, J., Wieling, M., Baayen, H., & van Rijn, H., 2017. Package "Itsadug". <https://cran.r-project.org/web/packages/itsadug/itsadug.pdf>

van Rij, J., Hendrix, P., van Rijn, H., Baayen, R.H., Wood, S.N., 2019. Analyzing the time course of pupillometric data. *Trends Hear.* 23, 1–22.

Surprenant, A.M., 2007. Effects of noise on identification and serial recall of nonsense syllables in older and younger adults. *Aging Neuropsychol.* C 14 (2), 126–143.

Tun, P.A., McCoy, S., Wingfield, A., 2009. Aging, hearing acuity, and the attentional costs of effortful listening. *Psychol. Aging* 24, 761–766.

Tun, P.A., Williams, V.A., Small, B.J., Hafter, E.R., 2012. The effects of aging on auditory processing and cognition. *Am. J. Audiol.* 21, 344–350.

Versfeld, N., Daalder, L., Festen, J., Houtgast, T., 2000. Method for the selection of sentence materials for efficient measurement of the speech reception threshold. *J. Acoust. Soc. Am.* 107, 1671–1684.

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. *Hear. Res.* 1–13.

Wieling, M., 2018. Analyzing dynamic phonetic data using generalized additive mixed modeling: a tutorial focusing on articulatory differences between L1 and L2 speakers of English. *J. Phon.* 70, 86–116.



- Winn, M.B., Moore, A.N., 2018. Pupillometry reveals that context benefit in speech perception can be disrupted by later-occurring sounds, especially in listeners with cochlear implants. *Trends Hear.* 22 (4), 1–22.
- Wood, S. N., 2018. Package "Mgcv". <https://cran.r-project.org/web/packages/mgcv/mgcv.pdf>
- Yonelinas, A.P., 2002. The nature of recollection and familiarity: a review of 30 years of research. *J. Mem. Lang.* 46, 441–517.
- Zekveld, A., Kramer, S., Festen, J., 2010. Pupil response as an indication of effortful listening: the influence of sentence intelligibility. *Ear Hear.* 31, 480–490.
- Zhang, M., Siegle, G.J., McNeil, M.R., Pratt, S.R., Palmer, C., 2019. The role of reward and task demand in value-based strategic allocation of auditory comprehension effort. *Hear. Res.* 381, 1–15.