Chapter 4

Associations of working memory capacity, inhibition and linguistic closure with speech recognition and the pupil dilation.


Trends in Hearing (2018), Under Revision
Abstract

Context and Objective: The aim of the current study was to examine whether speech recognition performance and the corresponding peak pupil dilation (PPD) are associated with the listeners cognitive abilities and their hearing status.

Methods: These questions were tested for a fixed speech reception threshold (SRT) in a single-talker masker (Experiment 1), and for a range of fixed SNRs for a single-talker and a stationary noise masker (Experiment 2). Verbal working memory capacity, inhibition of interfering information in working memory, and linguistic closure were examined as covariates. Normal-hearing and age-matched hearing-impaired listeners participated in both experiments. Univariate Spearman correlation, association regression models and linear mixed model (LMM) ANOVAs were separately performed for speech recognition performance and PPDs.

Results: The results from Experiment 1 indicated that higher working memory capacity together with better abilities to inhibit interfering information and better linguistic closure abilities are associated to larger pupil responses when listeners were hearing-impaired. The findings from Experiment 2 suggested that the influence of cognitive measures (Reading Span Test (RST), Size Comparison Span (SICspan) and Text Reception Threshold (TRT)) on speech recognition and the pupil response varied across SNR, however the interactions did not exhibit obvious systematic trends. The impact of the listeners hearing ability on the link between cognitive measures, speech recognition and the pupil response did not show a systematic trend either.

Conclusion: The results caution against the appealing assumption that the listeners hearing ability can generally predict the interplay between individual differences in working memory, speech recognition performance and the allocation of cognitive resources.
4.1 Introduction

When people with hearing impairment are engaged in a conversation, they are thought to depend on their cognitive abilities, such as working memory capacity, the inhibition of interfering information and linguistic closure, to compensate for reduced audibility (Pichora-Fuller, 2003; Peelle, 2017). Especially, successful speech recognition in background noise requires a variety of cognitive abilities, granting simultaneous storing and processing of information, determining the target speech from a complex acoustic signal or integrating fragmentary linguistic information. The association between cognitive abilities and speech perception performance has repeatedly been demonstrated in recent research (Lunner and Sundewall-Thorén, 2007; Gatehouse et al., 2003; Ohlenforst et al., 2016; Rudner et al., 2011; Akeroyd, 2008; Kramer et al., 2009; Peelle, 2017; Rönnberg et al., 2008; Lunner, 2003). Higher working memory capacity, better inhibition skills (Koelewijn et al., 2014) and better linguistic closure (Zekveld et al., 2011) skills were related to better speech reception performance (Koelewijn et al., 2014; Koelewijn et al., 2012a; Zekveld et al., 2011). The results from twenty experimental studies, that have measured the relationship between speech recognition in noise and different aspects of cognition showed that working memory together with the individuals hearing ability are the most efficient predictors for speech recognition performance in background noise (Akeroyd, 2008; Humes, 1994).

The association between cognitive abilities and speech recognition performance can be moderated by the level of intelligibility or the interaction between intelligibility and cognitive abilities. The taxing impact of reduced audibility on cognitive demands and memory has for example been demonstrated when memory skills for short stories were tested for normal-hearing and hearing-impaired listeners (Piquado et al., 2012). Normal-hearing listeners were able to deal with less processing time to comprehend and memorize the context of short stories. For hearing-impaired listeners was increased processing time related to improved performance in memorizing the short stories (Piquado et al., 2012). It was consequently suggested that hearing-impaired listeners might perceive more cognitive challenge during speech recognition in background noise than normal-hearing listeners do (Wingfield, 2016). Individual differences in working memory capacity have been suggested to influence speech recognition performance especially in challenging listening conditions (Lunner and Sundewall-Thorén, 2007; Gatehouse et al., 2003; Pichora-Fuller and Singh, 2006; Ohlenforst et al., 2016). That is, when larger working memory capacity can be allocated for task performance, better performance may be achieved. Thus, better cognitive abilities might be related to higher cognitive load, especially when listening is difficult. Even though contradictory findings concerning the association between listening effort and cognitive capacity have been found, evidence about the idea that successful speech recognition performance is related to the engagement of cognitive resources exists (Peelle, 2017).

Another important factor that plays a role in the relationship between cognition, speech recognition and hearing impairment is listening effort. The intense application of cognitive resources for speech understanding is effortful, especially for listeners with impaired hearing (Pichora-Fuller et al., 2016; Ohlenforst et al., 2017b; Ohlenforst et al., 2017a; Peelle, 2017). Listening effort has been defined as the deliberate allocation of mental resources to
overcome obstacles to goal pursuit when carrying out a listening task (Pichora-Fuller et al., 2016). Interestingly, contradictory results were obtained when the relationship between cognitive capacity, listening effort and hearing impairment was examined during speech recognition. On the one hand, high self-reported listening effort was associated with smaller working memory capacity and slower processing speed during a dual task paradigm (Desjardins and Doherty, 2013; Picou et al., 2013; Picou et al., 2011). This is in agreement with the Ease-of-Language Understanding (ELU) model (Rönnberg, 2003; Rönnberg et al., 2008; Stenfelt and Rönnberg, 2009), which suggests that listeners with better cognitive abilities and higher working memory capacity can better deal with challenging listening conditions and therefore perceive lower cognitive load and listening effort (Rönnberg, 2003). On the other hand, using different measures, other studies found larger working memory capacity being significantly associated with a larger pupil response, indicating more cognitive processing load (Koelewijn et al., 2014; Koelewijn et al., 2012b).

The studies described above suggest that the listeners cognitive abilities are associated with listening effort during speech recognition. However, it is not clear yet whether listening effort can as well be associated with the listeners cognitive abilities. Physiological measures, such as the task-evoked pupil dilation, permit the assessment of one aspect of listening effort, namely the allocation of mental resources during speech recognition. Recent research, assessing the interplay between cognitive abilities and speech recognition performance or the pupil response demonstrated that pupillometry can successfully be used to study a variety of factors that affect the allocation of cognitive resources during speech recognition. Those factors include lexical manipulation (Kuchinsky et al., 2013), linguistic complexity (Wendt et al., 2016), cognitive functions (Zekveld et al., 2011), hearing impairment (Zekveld et al., 2011; Kramer et al., 1997), sentence intelligibility (Zekveld et al., 2011) and different masker types (Koelewijn et al., 2012a). Generally speaking, an increase in cognitive demands is reflected in a larger pupil dilation (Beatty, 1982; Engelhardt et al., 2010). The relationship between task demands and changes in the pupil dilation is however not linear as the pupil dilates until task demands exceed the available cognitive resources (Granholm, 1996).

Recently, researchers recorded pupil dilations and subjective effort ratings to assess the impact of background noise and sentence complexity on cognitive processing demands during sentence comprehension for a group of normal-hearing listeners (Wendt et al., 2016). In line with Koelewijn and colleagues (Koelewijn et al., 2014; Koelewijn et al., 2012b) higher working memory capacity was correlated with larger pupil dilation responses during speech comprehension in one of the conditions. This suggests that listeners with higher working memory capacity may allocate and engage more cognitive resources compared to listeners with smaller working memory capacity. A body of research highlights the importance of assessing listening effort as an extension to commonly applied speech reception performance measures, as measurable changes in intelligibility are not always sensitive enough to reflect changes in listening effort (Pichora-Fuller et al., 2016; Pichora-Fuller and Singh, 2006; Ohlenforst et al., 2017b; Wendt et al., 2017). It has for example been shown that participants have to invest more cognitive resources to maintain comparable intelligibility measures in the presence of a single-talker masker compared to a fluctuating noise masker (Koelewijn et al., 2012b). It has furthermore been shown that the allocation of cognitive resources is related to the signal-to-noise ratio (SNR) level at which speech recognition performance is examined (Wu et al., 2016). Speech recognition performance
decreased as SNR level decreased. Interestingly, purported measures of listening effort did not indicate continuously increasing effort but rather reduced listening effort at the lowest and the highest SNR levels. Thus, when listening becomes very difficult, listeners may decide not to invest intense effort for unrewarding test conditions. Also, intense effort may not be required to maintain high performance in very easy listening conditions. Even when hearing-impaired listeners are able to maintain speech recognition performance comparable to normal-hearing listeners, the expended amount of effort may differ (Ohlenforst et al., 2017b). Measuring listening effort, in addition to commonly used speech reception threshold measures, may uncover changes in cognitive processing load and the allocation of working memory capacity when intelligibility measures are insensitive and at floor or ceiling.

In summary, previous research in the domain of speech perception and listening effort suggests that the pupil response, examined during speech recognition, sensitively reflects the influence of factors assumed to affect cognitive processing load (Kramer et al., 1997; Piquado et al., 2010). For hearing-impaired listeners better linguistic closure abilities were associated with larger PPDs, indicating increased cognitive processing load (Zekveld et al., 2011). This association was not confirmed for young normal-hearing listeners when speech recognition was tested at 50%, 71% and 84% correct performance. A recent review study suggests that individual variations in working memory capacity only explain an insignificantly small amount of variance in speech recognition performance for young normal-hearing listeners (Füllgrabe and Rosen, 2016b). Thus, even though empirical evidence seems to suggest an association between the listener’s working memory capacity and their speech recognition performance in background noise (Larsby et al., 2005; Ohlenforst et al., 2016; Füllgrabe et al., 2015; Füllgrabe and Rosen, 2016a), any such association may differ for normal-hearing and hearing-impaired listeners. In another recent study, speech recognition in the presence of a single-talker masker and the corresponding pupil response were examined, for a range of conditions from 0% to 99% correct performance (Zekveld and Kramer, 2014). The results revealed that participants that had very good linguistic closure abilities showed relatively larger PPDs at low intelligibility levels. The researchers concluded that processing load and presumably cognitive processing overload was reflected by their findings. However, those results hold only for normal-hearing listeners across a range of very difficult conditions. With the knowledge provided by the above mentioned studies, the relationship between cognitive skills, speech recognition and listening effort can only partly be explained for a small range of speech intelligibility conditions and normal-hearing listeners. It is yet unclear how working memory capacity, inhibition and linguistic closure abilities for hearing-impaired listeners are associated with speech recognition and the pupil dilation across a broad range of listening conditions. However, recent research highlights the importance of examining a broad range of listening conditions, including conditions related to low and high speech recognition performance for listeners with normal and impaired hearing (Ohlenforst et al., 2017b).

In the current study, we aimed to clarify how cognitive abilities, including working memory capacity, inhibition and linguistic closure, are associated to speech recognition performance and the corresponding peak pupil dilation (PPD). We intended to investigate the impact of cognitive abilities across a broad range of listening conditions and possible differences in the allocation of cognitive resources between normal-hearing and hearing-impaired listeners.
We tested normal-hearing and hearing-impaired listeners separately, as we expected different results for the two listener groups. A broad variety of listening conditions was distributed across two experiments, included a fixed intelligibility level (Experiment 1) and a range of fixed SNR levels (Experiment 2). For Experiment 1, we hypothesized that larger pupil responses and lower SRTs at 50% correct speech recognition performance are associated with larger working memory capacity and better linguistic closure abilities and that hearing-impaired listeners would depend more on their cognitive resources than normal-hearing listeners (Koelewijn et al., 2014; Koelewijn et al., 2012b; Zekveld et al., 2011; Füllgrabe and Rosen, 2016b; Larsby et al., 2005). Listeners with larger working memory capacity (Van Der Meer et al., 2010; Grady, 2012; Koelewijn et al., 2014; Koelewijn et al., 2012b) and better linguistic closure skills (Zekveld et al., 2011) are presumably able to allocate more cognitive resources and therefore show larger task-evoked PPDs compared to listeners with smaller cognitive abilities (Koelewijn et al., 2014; Koelewijn et al., 2012b). We hypothesized for the outcomes from Experiment 2, that the effect of cognition or linguistic abilities is modulated by the SNR condition, such that better cognitive and linguistic abilities would be related to relatively large pupil responses at low SNRs, when speech intelligibility is low (Zekveld and Kramer, 2014). We assumed that particularly hearing-impaired listeners would be more dependent upon their cognitive abilities than normal-hearing listeners.

4.2 Materials and Methods

Experiments
The current study included two separate experiments (Experiment 1 and 2). The experiments took place during two separate test sessions with the same experimental setup. We measured speech recognition performance and the pupil response during both experiments. In Experiment 1, an adaptive procedure was applied to arrive at 50% correct speech recognition (Wang et al., 2017). In Experiment 2, speech recognition performance was measured for fixed SNRs (Ohlenforst et al., 2017b). The overall recruitment, the experimental setup, the pupillometry measures, the applied pupil data preprocessing and part of the participants were identical for both experiments, and are described first. The speech reception measures and the analysis differed between the two experiments. Those sections are separately described at the end of the following method section.

General methods

Participant recruitment and inclusion criteria
The recruitment of the participants took place at the VU University Medical Center, the VU University, local hearing aid dispensers and community centers in Amsterdam. We recruited the hearing-impaired listeners first, followed by age-matched normal-hearing listeners. The maximum age difference between normal-hearing and hearing-impaired listeners was 5 years. The audiometric inclusion criterion for the hearing-impaired participants was sensorineural hearing impairment (air-bone gap <10 dB between 500Hz and 4kHz) with a pure-tone average (PTA) between 35 and 65 dB HL across 500 Hz and 4k Hz. The hearing impairment had to be roughly symmetrical (difference < 20dB at one frequency, or <15 dB
HL at two frequencies or < 10 dB HL at three frequencies, across 250 Hz to 4000 Hz). The age-matched group of normal-hearing listeners had PTAs below 20 dB HL across 250 Hz to 4kHz. All participants were native speaker of Dutch language and none of the participants reported health problems. We made sure that medical treatment was not sedative and did not have any known or reported effect on the pupil function. None of the participants had cerebral injuries, such as traumatic brain injury or stroke. Further exclusion criteria were a history of psychiatric or neurological diseases or eye diseases, to avoid abnormalities in the pupil response. The experiment was approved by the VU University Medical Center Ethical Committee and all participants provided written informed consent for participation. Thirty-five normal-hearing and 35 hearing-impaired listeners were recruited but only data from 31 hearing-impaired and 34 normal-hearing listeners were included. The data for four hearing-impaired listeners were not further used for the analysis, due to unexpected changes in hearing thresholds with respect to an earlier audiogram (n=1), unexpected cognitive problems (n=1) or other health problems (n=2). One normal-hearing participant did not complete Experiment 1, due to unexpected changes in hearing ability.

Auditory stimuli and spectral shaping
The same target stimulus was used for the speech recognition task in both experiments. Everyday Dutch sentences (Versfeld et al., 2000) were spoken by a female talker and diotically presented via headphones. An example sentence is: “de winkel is op loopafstand” (translation: “the shop is within walking distance”). Each sentence consisted of eight or nine syllables and each word contained one or two syllables. Sentence duration ranged between 1.4 and 2 seconds. The same presentation procedure of the stimuli was applied for both experiments and the different masker types. The masker started for each trial three seconds before the onset of the target sentence and ended four seconds after the sentence offset. The participants were asked to repeat the sentence aloud, after the offset of the post-sentence noise, which was indicated by an answer prompt tone. The answer prompt tone was presented for 1 second at 55 dB SPL at a frequency of 1000 Hz. In Experiment 2, speech recognition performance was measured in the presence of a stationary noise or a single-talker masker. The single-talker masker consisted of concatenated sentences (Versfeld et al., 2000) spoken by a male talker. The long-term average spectrum of both masker types was identical to the long-term average speech spectrum of the target sentences (Versfeld et al., 2000), and the masker was always presented at 65 dB SPL. In Experiment 1, only the single-talker masker was used. As the masker was always presented prior to the target sentence, we wanted to avoid that the participants could learn to estimate task difficulty from changing noise levels. Keeping the masker levels constant ensured also that the masker would not be played too loud at very low SNRs. The participants did not wear hearing aids during the experimental measures. However, we wanted to make sure that equal audibility was provided amongst all participants. For each hearing-impaired participant, the sound files for the speech recognition tests were individually amplified according to the hearing thresholds at each ear. The National Acoustic Laboratories’ linear fitting procedure, revised version (NAL-R) (Byrne and Dillon, 1986) was applied in 1/3 octave steps within the frequency range of 315 Hz to 6300 Hz. The stimuli were bilaterally presented via headphones. Headphone saturation was avoided by setting the maximum sound pressure level within each frequency band to 95 dB SPL. We verified that all participants would reach 100 % correct sentence
recognition performance when the target sentences were presented at 65 dB SPL in quiet. The auditory stimuli and the applied spectral shaping described in the current study were similar to one of our previous pupillometry studies (Ohlenforst et al., 2017b).

Outcome measures
Speech recognition performance
Participants’ responses were manually scored by the experimenter, and each sentence was accounted as correct if the entire sentence was repeated without mistakes. In both experiments, the same sentence material (female talker) was presented as target stimulus (Versfeld et al., 2000). In Experiment 1 the SRT (Festen and Plomp, 1990) for 50% correct performance was applied to measure sentence recognition in the presence of a single-talker masker background. In Experiment 2, sentence recognition performance was measured across a range of eight SNRs for a stationary noise masker and across nine SNRs for the same single-talker masker used in Experiment 1. The SNR levels in Experiment 2 corresponding to a range of intelligibility conditions from 0% to 100% correct performance.

Pupillometry
The pupil diameter of both participants eyes (left and right) were recorded during the speech recognition test session within each experiment, using an eye tracking system by SensoMotoric Instruments (Berlin, Germany, 2D Video-Oculography, version 4). The eye-tracking system had a spatial resolution of 0.03 mm and a sampling frequency of 60 Hz. The pupil location, the pupil size between the masker onset and the masker offset and the speech recognition scores were recorded and stored at a connected computer. The experimenter observed the real-time pupil data during the experiment. Some participants started to blink more often, some lowered their eye lids or moved their head more during the experiment. The experimenter applied corrective actions, such as the adjustment of the distance to the screen or reminded the participants to remain focused during the test.

Covariate factors
Reading Span Test (RST)
The RST serves as assessment of verbal working memory capacity in the visual domain (Daneman and Carpenter, 1980). The test originates from an English version, which was translated to Swedish. The sentence material tested in the current study consisted of 5-word Dutch sentences (Besser et al., 2013), which were equivalent to the Swedish version (Andersson et al., 2001; Rönberg et al., 1989). The test consisted in total of 54 sentences, which were presented on a computer screen with an increasing number of sentences in sets of 3 to 6 sentences. Half of the sentences were semantically correct, the other half were semantically incorrect. First, the participants had to read each sentence out loud and verbally judge its semantic correctness. The participants were told to recall either the first or the last noun of all sentences in each set. The task was strategically difficult, as the participants did not know beforehand which noun (first or last) of all sentences they would be asked to recall. After the presentation of each set of sentences, the participants had to repeat all first or all last nouns of all sentences in a set in correct presentation order. The
The overall RST score is independent of the recall order and corresponds to the total number of correctly recalled nouns, which results in a maximum score of 54. Better performance is indicated by a higher overall RST score.

**Text reception threshold (TRT) test**
The ability to recognize masked speech information is associated with the individuals’ hearing abilities and the cognitive processing abilities. Commonly, the SRT test is used to measure the listeners ability to recognize masked speech in the auditory domain. The listeners working memory capacity in speech recognition research is frequently assessed by means of the RST test. However, the ability to recognize masked speech information can also be assessed in the visual domain by means of linguistic closure, which describes the ability to read masked text. The TRT test is an assessment of linguistic closure ability (Zekveld et al., 2007) and is also known as the visual analogue to the SRT test. The TRT test consists of 13 sentences, which are masked by a bar pattern and presented on a computer screen. The text is of red color and masked by black bars on white background. Each sentence is presented cumulatively word by word in a timing comparable to the word onset of the corresponding recorded SRT sentences. Once the whole sentence is completed, all words remain visible for 500 ms on the computer screen. The participant repeated aloud as much of the sentence as they could. The experimenter scored performance as correct if the whole sentence was repeated correctly. The bar pattern covered 42% of the first sentence. If a sentence was not entirely correctly repeated by the participant, the bar pattern became more sparse (i.e. the percentage of unmasked text was increased) for the following sentence. A 1-up-1-down procedure with a step size of 6% was applied to converge on the percentage of unmasked text required to read around 50% of the sentences entirely correctly. The overall TRT score corresponded to the average proportion of unmasked text for sentences 5 to 14, with lower TRT scores corresponding to better performance. One practice and three regular TRT tests were performed by each participant. We used the mean TRT score of the actual test session for each participant as a covariate or correlation factor in the analyses.

**Size-Comparison Span test (SICspan)**
The participants’ working memory capacity and their ability to suppress irrelevant linguistic information was measured in terms of the size-comparison span test (Sörqvist et al., 2010; Sörqvist and Ronnberg, 2012). The participants had to respond to a set of questions that was presented on a computer screen, which required relative size judgments between two items (e.g. “is a BUSH larger than a TREE?”). The participants responded that the judgment was correct by pressing “J” or “N” respectively for correct or incorrect statements, using a QWERTY keyboard. A single semantically related word, that the participants had to remember, was presented after each question on the computer screen (e.g. “FLOWER”). The test covered ten sets in total and each set consisted of 2 to 6 size comparison questions. Between sets, different nouns were presented. Within a set, the presented nouns and those that had to be remembered were always from the same semantic category (e.g. plants). After the presentation of a whole set of questions was finished, the participants were asked to orally repeat all the single nouns to remember. In total, 40 items could be remembered correctly. The order in which the single nouns were recalled did not matter. The total number of correctly recalled nouns resulted in the overall SICspan score, with lower scores
corresponding to lower performance. Correctness of size comparisons was not scored.

**General procedure**

Each test session started by seating the participant on a fixed chair inside a sound proof measurement booth. Participants were instructed to sit in a comfortable position that they could stay in for at least 10 minutes. A computer screen and the eye tracking camera were placed in front of the participant. The distance between the center of the computer screen and the midpoint of the participants’ eyes was approximately 55 cm. After the participant was seated properly, the light condition in the measurement booth was calibrated. The pupil size was measured during a condition of maximum illumination (approximately 230 lux) followed by a condition of complete darkness. To avoid floor or ceiling effects in the pupil response, individual adaptation of the illumination was applied for each participant until a medium pupil size was measured (Hyönä and Ekholm, 2016). The illumination during the test sessions was on average 13.3 lux (SD=3.2 lux). Each test session started with a practice session to ensure that the participants were confident with the experimental procedure, such as the inhibition of head movements and blinking during the sentence recognition task. The participants’ task was to focus on a white fixation dot on the computer screen in front of them, and to blink as little as possible during the sentence presentation and the response interval. The actual experiment started with a practice session during which one single sentence for each SNR and masker type condition was presented in random order. After the practice session, the first experimental session started and for each presented sentence the pupil diameter was recorded. For Experiment 1, 25 sentences were presented in one block, and during Experiment 2, 17 conditions (8 SNRs for stationary masker, 9 SNRs for single-talker masker) were recorded, consisting of 10 sentences per condition.

**Pupil data selection, cleaning and data reduction**

We examined the task-evoked pupil diameter when the speech signal was present with respect to the baseline pupil diameter when only the masker signal was present. The baseline was defined as the average pupil diameter recorded during the final second of the three second presentation of the masker, before target speech onset. In the current study, we focused on the peak pupil dilation (PPD), which is the maximum pupil diameter between the onset of the sentence and the response prompt, relative to the baseline pupil diameter. Pupil diameter values that were 3 standard deviations below the mean pupil diameter (between sentence onset and prompt tone relative to the baseline) were categorized as blinks. Blinks in the pupil traces were replaced by linear interpolation (5 samples before and 7 samples after the blink). We excluded pupil traces with more than 15% of blinks between the start of the baseline (last second of pre-noise before sentence onset) and the prompt tone from the analysis. The pupil response within each selected and de-blinked trace was smoothed by a five-point moving average filter. For each participant, all the included de-blinked and smoothed traces for each condition were time-aligned and averaged. The PPD of this averaged pupil trace provided the data for the statistical analysis.
Experiment 1

Materials and method
Participants
In experiment 1, 19 hearing-impaired (n=13 females, n=6 males) and 27 normal-hearing (n=17 females, n=10 males) native Dutch speaking participants were included. The hearing-impaired listeners were on average 47.2 years old (SD=10.9) and the normal-hearing participants’ mean age was 46.3 years (SD=12.4). The age-matched group of normal-hearing listeners had PTAs below 20 dB HL across 250 Hz to 4 kHz. The average PTA for the normal-hearing participants was 8.8 dB HL (4.6 dB HL). We scaled the participants educational level into seven levels, with 1 corresponding to a low and 7 to a high (e.g. university degree) education. The mean educational level for the hearing-impaired listeners was 5.81 (SD=1.33). The mean educational level for the normal-hearing listener groups was 5.56 (SD=1.56). The accepted age range for participation was between 18 and 60 years. Hearing-impaired listeners were on average 47.2 years old (SD=10.8) and normal-hearing listeners were on average almost equally old but had a slightly larger standard deviation (mean=46.3 years, SD=12.4).

Speech Reception Threshold Test (SRT)
The SRT (Plomp and Mimpen, 1979) was measured by presenting speech in a single-talker masker background (Festen and Plomp, 1990). The target speech consisted of one block of 25 sentences (female-talker) presented in a stream of concatenated single-talker masker sentences (male-talker) (Versfeld et al., 2000). The participants did not receive feedback on their performance and the whole sentence had to be repeated without errors to be scored as correct. Both stimuli were played via headphones and the SNR was initially set to -10 dB. The single-talker masker was presented at 65 dB SPL. Only the first target sentence was adjusted in steps of 4 dB and the level of the remaining sentences was adaptively adjusted by 2-dB in a one-up-one-down staircase procedure, targeting 50% intelligibility (Plomp and Mimpen, 1979). First, we estimated the starting SNR level for the adaptive procedure. This was done by repeatedly presenting the first sentence with increasing SNR levels in steps of 4 dB, until the participant was able to correctly repeat the whole sentence. The resulting SNR level was the individual starting level for the remaining sentences in the adaptive procedure. The dependent variable henceforth labelled as ‘SRT’ was the mean SNR of the sentences 5 to 25. Pupil traces were recorded for each SRT test track. The smoothed pupil traces were time-aligned to sentence onset and averaged. We calculated the PPD from the averaged pupil trace, as the maximum pupil dilation between sentence onset and noise offset, relative to the baseline pupil diameter. The resulting SRT and PPD values were the dependent variables in the statistical analysis described in the following.

Statistical analyses
To start with, a linear Spearman correlation analysis between the three cognitive factors (RST, TRT and SICspan), the outcome variables (SRT and PPD) and three possible confounders (PTA, age and educational level) were separately applied for each listener group. Afterwards, we built 36 separate association models to investigate the relationship between cognition, the PPD and the SRT. Eighteen models were built for each listener
group (normal-hearing and hearing-impaired), including three cognitive measures (RST, TRT and SICspan), three possible confounders (PTA, age and educational level) and two outcome measures (SRT and PPD). The dependent variables in the models were the SRTs or PPDs. The independent variables in the model were the cognitive measures obtained during the RST, the TRT or the SICspan tests. For each model, we tested separately if PTA, age or educational level were possible confounders. We aimed to investigate whether the relationship between the cognitive factors and SRT and PPD can be partly explained by the participants PTA, age or educational level. A critical α level of p < .05 was applied for the association models. To be a confounding factor, each variable (PTA, age or educational level) had to fulfill two criteria. First, the possible confounder had to be correlated (p < .20) with the corresponding independent (RST, TRT or SICspan scores) and dependent (SRT or PPD) variables in the model. If the correlation criteria was fulfilled, the regression model was tested with and without including the possible confounder variable. The regression coefficient of the independent variable (RST, TRT or SICspan) had to change by at least 10% when the confounder variable was added to the model compared to when the confounder variable was not included in the model (Grayson, 1987). If and only if both criteria were fulfilled, the confounder variable was kept in the regression model. If two or more of the possible confounders met both criteria, the strongest confounder was added to the model. If adding an additional confounder resulted in a change of at least 10% of the regression coefficient of the independent variable, the additional confounder was also included in the model.

Results
The data distribution was checked for all variables and Spearman correlation analysis was applied. The cognitive measures (RST, TRT and SICspan), the dependent variables SRT and PPD, and the PTA values were all normally distributed for both listener groups. The participants’ mean educational level and the mean age was above the normal distribution and a Spearman’s correlation coefficient was applied.
To assess whether the SRT and the PPD were associated with working memory capacity (RST), working memory capacity and the ability to inhibit irrelevant linguistic information (SICspan) and linguistic closure (TRT), linear regression analyses were separately applied for the SRTs and the PPDs.

Spearman correlation analysis for normal-hearing listeners
For the normal-hearing listeners no significant (p<0.05) correlation was found between any of the three cognitive measures (RST, TRT and SICspan) and the SRT or the PPD.

Spearman correlation analysis for hearing-impaired listeners
For the hearing-impaired listeners, we found that higher (better) SICspan performance was correlated with larger PPD (r=0.54, p=0.02). A significant positive Spearman correlation between SRT and PTA suggested that higher (worse) hearing thresholds are associated with higher (worse) SRT scores (rho=0.5, p=0.03).
Association analysis for normal-hearing listeners

None of the linear regression models showed a significant association between the cognitive measures (RST, TRT and SiCspan) and the SRT or the PPD for the normal-hearing listeners.

Association analysis for hearing-impaired listeners

The association models (see Table 1) between RST and SRT and RST and the PPD were not significant, and no confounders were identified (PTA, age, educational level). The association model between SiCspan and PPD (β=0.012; R²=0.34; F(1,17)=8.67; p<0.01) suggests that HI individuals with a higher level of working memory capacity may allocate more cognitive resources when listening at 50% performance. A significant association between the TRT and the PPD was also found in the linear regression analysis (β=0.007; R²=0.30; F(1,17)=7.3; p<0.05). Better linguistic closure ability (lower TRT score) was associated with larger PPD, suggesting the allocation of more cognitive resources.

Table1: Significant linear regression models with PPD as dependent variable, and SiCspan and TRT as independent variables. Shown are the unstandardized regression coefficients including the standard error β (SE), the standardized regression coefficients βstand, the t-values, the p-values of significant associations (p<0.05) and the variance R².

<table>
<thead>
<tr>
<th>Dependent variable</th>
<th>Independent variable</th>
<th>β (SE)</th>
<th>βstand</th>
<th>t</th>
<th>P</th>
<th>R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>PPD</td>
<td>SiCspan</td>
<td>0.012 (0.004)</td>
<td>0.58</td>
<td>2.94</td>
<td>0.009</td>
<td>0.34</td>
</tr>
<tr>
<td>PPD</td>
<td>TRT</td>
<td>0.007 (0.003)</td>
<td>0.55</td>
<td>2.70</td>
<td>0.015</td>
<td>0.30</td>
</tr>
</tbody>
</table>

In summary, a significant association between working memory capacity and inhibition skills (SiCspan) and linguistic closure (TRT) with the PPD at 50% correct speech performance was shown for hearing-impaired listeners but not for normal-hearing listeners. This finding is in line with our first hypothesis and recent research, suggesting that better cognitive skills enable hearing-impaired listeners to allocate more cognitive resources for task performance, but that working memory capacity may be a less important factor in speech recognition by normal-hearing listeners (Van Der Meer et al., 2010; Grady, 2012; Koelewijn et al., 2014; Koelewijn et al., 2012b; Zekveld et al., 2011; Füllgrabe and Rosen, 2016b). The results showed that PTA, age and education level were not significant confounders on the relationship between the cognitive abilities tested within this study and the SRT or PPD. The results from Experiment 1 do not suggest a significant association between the SRT and the listeners’ cognitive abilities, as has been suggested by previous research (Koelewijn et al., 2014; Zekveld et al., 2011).
Experiment 2

Materials and method

Participants
In experiment 2, 31 hearing-impaired (n= 24 females, n= 7 males) and 34 normal-hearing (n= 22 females, n= 13 males) native Dutch speaking participants were included. The hearing-impaired listeners were on average 47.2 years old (SD=10.9) and the normal-hearing participants’ mean age was 46.3 years (SD=12.4). The mean PTA for the hearing-impaired group was 42.1 dB HL (SD=9.3 dB HL). The average PTA for the normal-hearing participants was 8.8 dB HL (4.6 dB HL).

Speech recognition at fixed SNRs
Based on previous research, a large range of SNRs was chosen to estimate speech recognition performance across the whole psychometric function separately for each masker type (Zekveld and Kramer, 2014; Festen and Plomp, 1990; Smeds et al., 2015; Wu et al., 2016). We aimed to cover an SNR range corresponding to a majority of daily life sound environments and communication situations for HI listeners and therefore included a large range of positive SNRs for both masker types (Wu et al., 2016; Smeds et al., 2015). The target sentences were presented at eight SNRs in steps of 4 dB, between -12 dB and +16 dB, when the stationary noise masker was present. In the presence of the single-talker masker the target sentences were presented at nine SNRs, in steps of 5 dB, between -25 dB and +15 dB. At each SNR, a block of 10 sentences was presented. The order of the SNRs for each masker type was randomized.

Statistical analyses
Pupil data from 34 normal-hearing and 31 hearing-impaired listeners underwent pupil data cleaning and data reduction as described earlier. Per participant, 170 pupil traces were presented and on average 15.8 (SD=16.2) pupil traces included more than 15% blinks and were excluded. For six hearing-impaired and two normal-hearing participants pupil data for one (or more) SNRs was excluded as less than 5 valid traces were measured per condition. We further analyzed pupil data and sentence recognition performance for 25 hearing-impaired (mean age 47.9 years) and 32 normal-hearing (mean age 47.8 years) participants. We analyzed the pupil and speech recognition data by means of linear mixed models (LMM) ANOVAs, as this type of analysis can deal with missing data across conditions. The package lme4 (Bates et al., 2014) and the function lmer in the statistics software R-studio were used for the analysis. The averaged PPD or percentage correct sentence recognition scores for each SNR level were the dependent factors in the model. The fixed factors in the model were SNR and the three cognitive measures (RST, TRT or SICspan) were separately tested as covariates. We tested the effect of SNR and the three different covariates on the PPD and sentence recognition for each participant group separately, with participants as random factor. In total, 24 LMM models were tested including 2 outcome measures (sentence recognition, PPD), 2 masker types (stationary noise, single-talker), 2 listener groups (normal-hearing, hearing-impaired) and 3 possible covariates (RST, TRT, SICspan).
Results
Linear mixed-model (LMM) ANOVAs for the normal-hearing listeners

The results from the LMM ANOVAs on speech recognition performance and the PPD including SNR and each of the three cognitive measures as covariates (RST, TRT and SICspan) for the normal-hearing listeners are shown in Table 2.

When RST and SNR were included in the LMM ANOVAs, we found for the stationary (F(7,200)=115.3; p<0.001) and the single-talker masker (F(8,200)=64.0; p<0.001), that speech recognition performance was significantly affected by the SNR. We found a significant main effect of the RST on speech recognition (better RST associated with better recognition) for the stationary noise (RST: F(200,7)=115.3; p=0.01;) and the single-talker (RST: F(1,25)=13.1; p=0.001). A significant interactive effect between RST and SNR on speech recognition performance was obtained for both masker types (stationary: RST x SNR: F(200,1)=6.3; p=0.01; single-talker: SNR x RST: F(8,200)=4.3; p<0.001). To illustrate the interactive effect between SNR and RST on speech recognition and the PPD, we split the normal-hearing listeners into two groups depending on their RST performance. RST performance equal to or below the median RST performance (18 (SD 5.89)) was included in the low RST performance group, while RST scores above the median performance corresponded to the high RST performance group (n=13 in low and high RST group). In Figure 1, the averaged sentence recognition performance in stationary noise (subfigure A) and the single-talker masker (subfigure C) and the averaged PPD for the single-talker masker (subfigure D) are shown, grouped for high/low RST performance. The interactions between SNR and RST on speech recognition performance (subfigure A and C) can broadly be described as RST having most effect at SNRs where the recognition scores were not at ceiling or floor values. In graphical terms, the psychometric function shifted slightly towards lower SNRs for the high RST group. The significant interaction between SNR and RST on PPD for the single-talker masker condition ([F(35,194)=2.88; p=0.004], subfigure D) was of a different character, indicating larger PPDs in the high-RST group across the negative SNR range between -25 dB and -5 dB SNR , but lower or equal PPDs for the positive SNRs.

When linguistic closure performance (TRT scores) and SNR were included in the LMM ANOVAs, a significant main effect of the TRT on speech recognition performance was revealed only for the single-talker masker (F(8.29)=3.7; p=<0.01). There was no effect of TRT on speech recognition performance for the stationary noise masker and the PPD for either masker type.

When the listeners inhibition skills (SICspan scores) and the SNR were included in the LMM ANOVAs, a significant main effect of SICspan on sentence recognition (F(8,208)=15.2; p<0.001) and a significant interaction effect between SICspan and SNR on sentence recognition (F(8,208)=3.0; p<0.005) resulted for the single-talker masker. The interaction between SICspan and SNR on sentence recognition performance in the presence of a single-talker masker showed a pattern very similar to that shown by the interaction of RST and SNR on speech recognition. The effects of SICspan and TRT on the SRTs were not accompanied by any effects on the PPD.
Figure 1: Subfigure A (top left) and B (top right) show the averaged sentence recognition performance in stationary noise (A) and the single-talker masker (B), grouped for high/low RST and SICspan performance. The subfigures C and D (bottom left and right) show the averaged sentence recognition performance (C) and the averaged PPD (D) for the single-talker masker background, grouped for high/low RST performance.

Overall, five of twelve possible effects involving cognitive covariates on sentence recognition performance or PPD indicated the influence of individual cognitive variables on the allocation of cognitive resources in the speech recognition task in the normal-hearing group.
Table 2: Results for the normal-hearing listeners. LMM ANOVAs on the sentence recognition performance and the PPD as outcome measures, the SNR levels for both masker types (stationary and single-talker masker) and cognitive performance scores (RST, TRT and SICspan) as fixed effects and participants as random factor.

### Normal-hearing listeners

<table>
<thead>
<tr>
<th>Possible covariates</th>
<th>Stationary noise masker</th>
<th>Single-talker masker</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sentence recognition</strong></td>
<td>SNR: F(7,200)=115.3, p&lt;0.001</td>
<td>SNR: F(8,200)=64.0, p&lt;0.001</td>
</tr>
<tr>
<td></td>
<td>RST: F(1,200)=6.3; p=0.01</td>
<td>RST: F(1,25)=13.1; p=0.001</td>
</tr>
<tr>
<td></td>
<td>SNR x RST: F(7,200)=3.6; p=0.001</td>
<td>SNR x RST: F(8,200)=4.3; p&lt;0.001</td>
</tr>
<tr>
<td><strong>PPD</strong></td>
<td>SNR: n. s.</td>
<td>SNR: n. s.</td>
</tr>
<tr>
<td></td>
<td>RST: n. s.</td>
<td>RST: n. s.</td>
</tr>
<tr>
<td></td>
<td>SNR x RST: n. s.</td>
<td>SNR x RST: n. s.</td>
</tr>
<tr>
<td><strong>SNR</strong></td>
<td>SNR: F(7,204)=18.0, p&lt;0.001</td>
<td></td>
</tr>
<tr>
<td></td>
<td>SNR: F(8,232)=6.6, p&lt;0.001</td>
<td></td>
</tr>
<tr>
<td></td>
<td>SNR: F(8,219)=2.15, p&lt;0.05</td>
<td></td>
</tr>
<tr>
<td></td>
<td>SNR x RST: n. s.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>SNR x TRT: n. s.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>TRT: n. s.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>TRT: F(8,29)=3.7, p&lt;0.01</td>
<td></td>
</tr>
<tr>
<td></td>
<td>TRT: n. s.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>SNR x TRT: n. s.</td>
<td></td>
</tr>
<tr>
<td><strong>TRT</strong></td>
<td>SNR: n. s.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>SNR x TRT: n. s.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>TRT: n. s.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>SNR x TRT: n. s.</td>
<td></td>
</tr>
<tr>
<td><strong>SICspan</strong></td>
<td>SNR: F(7,112)=10.9, p&lt;0.001</td>
<td></td>
</tr>
<tr>
<td></td>
<td>SNR: F(8,208)=15.2, p&lt;0.001</td>
<td></td>
</tr>
<tr>
<td></td>
<td>SNR x SICspan: n. s.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>SICspan: n. s.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>SICspan: F(1,26)=14.2, p&lt;0.01</td>
<td></td>
</tr>
<tr>
<td></td>
<td>SNR x SICspan: n. s.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>SNR x SICspan: n. s.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>SNR x SICspan: n. s.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>SNR x SICspan: n. s.</td>
<td></td>
</tr>
</tbody>
</table>

Linear mixed-model (LMM) ANOVAs for the hearing-impaired listeners

When RST and SNR were included in the LMM ANOVAs, the results for the hearing-impaired listeners showed (unsurprisingly) that speech recognition performance for both the stationary noise and the single-talker masker was significantly affected by the SNR (stationary noise: F(147,7)=34.3, p<0.001; single-talker masker F(176,8)=15.6, p<0.001) (see
Table 3). We did not find a significant effect of RST on speech recognition performance for any masker type and we did not find a significant effect of SNR or RST on the PPD for either masker type.

When the linguistic closure performance (TRT scores) and the SNR were included in the LMM ANOVA, a significant main effect of SNR on sentence recognition performance was still obtained for both masker types (stationary noise: F(147,7)=8.6, p<0.001; single-talker masker F(176,8)=3.7; p<0.001). There was no main effect of TRT on speech recognition performance in either noise type. However, a significant interaction effect between SNR and TRT on the PPD for the single-talker masker (F(155,8)=2.07; p<0.05) indicated that linguistic closure may have an impact on the allocation of cognitive resources. However, no clear pattern was evident in this interaction (see Figure 2).

When the listeners inhibition skills (SICspan scores) and the SNR were included in the LMM ANOVA, no significant effects involving SICspan were obtained for any masker type or outcome measure.

Overall, only one of twelve possible effects involving cognitive covariates and PPD (a significant interaction between TRT and SNR on the PPD in the single-talker masker) indicated any influence of individual cognitive variables on the allocation of cognitive resources in the speech recognition task when listeners were hearing-impaired.

**Peak pupil dilation (PPD) in single-talker masker for hearing-impaired listeners with high/low TRT**

![Graph showing peak pupil dilation (PPD) across SNRs for the single-talker masker condition, grouped for high and low TRT performance for hearing-impaired listeners.]

**Figure 2:** Averaged peak pupil dilation (PPD) across SNRs for the single-talker masker condition, grouped for high and low TRT performance for hearing-impaired listeners.
Table 3: Results for the LMM ANOVAs on the sentence recognition performance and the PPD as outcome measures, the SNR levels for both masker types (stationary and single-talker masker) and cognitive performance scores (RST, TRT and SICspan) as fixed effects and participants as random factor for the hearing-impaired listeners.

<table>
<thead>
<tr>
<th>Possible covariates</th>
<th>Sentence recognition</th>
<th>PPD</th>
<th>Sentence recognition</th>
<th>PPD</th>
</tr>
</thead>
<tbody>
<tr>
<td>RST</td>
<td>SNR: F(7,147)=34.3, p&lt;0.001</td>
<td>SNR: n. s.</td>
<td>SNR: F(8,176)=15.6, p&lt;0.001</td>
<td>SNR: n. s.</td>
</tr>
<tr>
<td></td>
<td>RST: n. s.</td>
<td>RST: n. s.</td>
<td>RST: n. s.</td>
<td>RST: n. s.</td>
</tr>
<tr>
<td></td>
<td>SNR x RST: n. s.</td>
<td>SNR x RST: n. s.</td>
<td>SNR x RST: n. s.</td>
<td>SNR x RST: n. s.</td>
</tr>
<tr>
<td>TRT</td>
<td>SNR: F(7,147)=8.6, p&lt;0.001</td>
<td>SNR: n. s.</td>
<td>SNR: F(8,176)=3.7, p&lt;0.001</td>
<td>SNR: n. s.</td>
</tr>
<tr>
<td></td>
<td>TRT: n. s.</td>
<td>TRT: n. s.</td>
<td>TRT: n. s.</td>
<td>TRT: n. s.</td>
</tr>
<tr>
<td></td>
<td>SNR x TRT: n. s.</td>
<td>SNR x TRT: n. s.</td>
<td>SNR x TRT: n. s.</td>
<td>SNR x TRT: F(8,155)=2.07, p&lt;0.05</td>
</tr>
<tr>
<td>SICspan</td>
<td>SNR: F(7,112)=10.9, p&lt;0.001</td>
<td>SNR: n. s.</td>
<td>SNR: F(8,128)=8.1, p&lt;0.001</td>
<td>SNR: n. s.</td>
</tr>
<tr>
<td></td>
<td>SICspan: n. s.</td>
<td>SICspan: n. s.</td>
<td>SICspan: n. s.</td>
<td>SICspan: n. s.</td>
</tr>
<tr>
<td></td>
<td>SNR x SICspan: n. s.</td>
<td>SNR x SICspan: n. s.</td>
<td>SNR x SICspan: n. s.</td>
<td>SNR x SICspan: n. s.</td>
</tr>
</tbody>
</table>
4.3 General discussion

This study aimed to clarify the association between cognitive abilities and speech recognition performance and the corresponding pupil response, as a measure of allocation of cognitive resources. We aimed to investigate the impact of cognition across a broad variety of listening conditions and possible differences in the allocation of cognitive resources between normal-hearing and hearing-impaired listeners.

Our first hypothesis was that larger pupil responses at 50% correct speech recognition performance are associated with larger working memory capacity and better linguistic closure abilities and that hearing impaired listeners would depend more on their cognitive resources than normal-hearing listeners (Koelewijn et al., 2014; Koelewijn et al., 2012b; Zekveld et al., 2011; Füllgrabe and Rosen, 2016b; Larsby et al., 2005). The Spearman correlation analysis for the normal-hearing listeners revealed no significant association between measures of working memory capacity or linguistic closure abilities with speech recognition or the corresponding PPD. For the hearing-impaired listeners, the Spearman correlation analysis revealed that higher working memory capacity and better inhibition abilities (higher SICspan score) were correlated with larger PPD at 50% correct speech recognition performance. The results for the hearing-impaired listeners, suggest that higher working memory capacity together with good abilities to inhibit interfering information (SICspan) are related to an increased allocation of cognitive resources. The results from the Spearman correlation analysis showed that larger pupil responses at 50% correct speech recognition performance are associated with larger working memory capacity and better inhibition of interfering information, when listeners were hearing-impaired, but not when they were normal-hearing.

Linear regression analysis suggests, in line with the Spearman correlation analysis and recent evidence, that normal-hearing listeners may not depend on their working memory capacity during speech recognition in background noise (Füllgrabe and Rosen, 2016b). A meta-analysis on currently available evidence revealed that individual variations in working memory capacity can only explain a small portion of variance (approximately 2%) in speech recognition performance when listeners are young and of normal-hearing. The researchers caution to generally assume that individual differences in working memory capacity determine speech understanding independently of the listeners’ age or hearing abilities (Füllgrabe and Rosen, 2016b). Studies included in this review assessed speech recognition dominantly for 50% correct performance. However, listeners may actually depend to a greater extend on their working memory capacity when speech understanding is more challenging than at 50% performance. The Ease of Language Understanding (ELU) model suggests that speech recognition in less favorable SNR conditions may tax the individuals working memory capacity more, as intense identification processes are required to distinguish between sensory and mental representations (Rönnberg et al., 2013). Overall, the results from linear regression analysis suggest that higher working memory capacity together with good abilities to inhibit interfering information (SICspan) are associated with an increased allocation of cognitive resources (indicated by the pupil response), particularly when listeners were hearing-impaired.
Our second hypothesis was that the effect of cognition or linguistic abilities is modulated by the SNR condition, in the way that better cognitive and linguistic abilities would be altered to relatively large pupil responses at low SNRs. We assumed that particularly hearing-impaired listeners would depend upon their cognitive abilities compared to normal-hearing listeners. The results of Experiment 2, suggest that cognitive measures are significantly associated with speech recognition performance and the pupil response. Four interaction effects between cognitive measures (RST and SICspan) and the SNR on sentence recognition performance or the pupil response resulted for the group of normal-hearing listeners. For the hearing-impaired listeners a significant interaction between the listeners linguistic closure abilities (TRT) and the SNR on the pupil response was found. We expected to find a stronger association between cognitive measures and the SNR on speech recognition and the pupil response for the hearing-impaired listeners. According to the ELU model are explicit, effortful processing mechanisms in working memory required when the perceptual speech information and the phonological representation in long-term memory don’t match (Rönnberg et al., 2013). A mismatch can be caused by internal distortions, such as hearing impairment, or the integrity of linguistic or cognitive processing and by external distortions, such as background noise. The ELU model predicts that a greater mismatch would result in more effortful listening. It is intuitively appealing to assume that hearing impairment causes a greater mismatch between the auditory speech input and the stored information in long-term memory. We assumed based on the predictions by the ELU model (Rönnberg et al., 2013) and recent evidence (Füllgrabe and Rosen, 2016b), that working memory capacity is less important for normal-hearing listeners, while hearing-impaired listeners would particularly depend on their cognitive abilities during speech recognition in background noise. The results from our second experiment do not only suggest a stronger association between cognitive measures and the SNR on sentence recognition performance and the corresponding pupil response, they are also contradictory to the results from our first experiment. They main difference between both experiments is that in Experiment 1, an adaptive procedure was applied to measure 50% correct speech recognition, while in Experiment 2 speech recognition performance was measured for fixed SNRs. Even though the link between cognitive measures and speech understanding may be reliably and strong, differences in speech recognition tasks may create different processing demands that may activate different sub-components in working memory. The currently applied tests to assess the interplay between the working memory system, speech recognition performance and the allocation of cognitive resources may not be sensitive enough to reflect the association between different task and different components in the working memory system. It was for example not possible to show a constant trend of performance across all three cognitive measures (RST, SICspan, TRT), applied within this study. For future research several measures that presumably tax the same components in the working memory system (e.g. RST and dual-task paradigm) should be applied and performance within participants should be compared to draw more reliable conclusions.

Within this current study we intended to investigate the impact of cognitive abilities across a broad range of listening conditions, including different SNRs and masker types, and possible differences in the allocation of cognitive resources between normal-hearing and hearing-impaired listeners. Based on apriori knowledge from previous evidence and our aim to
investigate groups differences, a number of separate statistical analysis was carried out to test our research questions. Separate analysis for each masker type (stationary noise and single-talker masker) were motivated by recent findings, suggesting that the maintenance of comparable speech recognition performance may require more cognitive processing load for a single-talker compared to a fluctuating noise masker condition (Koelewijn et al. 2012b, 2014; Zekveld et al. 2011). It was furthermore suggested that cognitive processing load during speech recognition in the presence of a single-talker masker may evoke larger pupil dilation response compared to fluctuating or stationary noise conditions (Koelewijn et al. 2012b). A number of studies suggested in addition that cognitive processing load or listening effort may be modulated by the listeners’ hearing ability and that the interaction between SNR and hearing ability may affect the pupil dilation response (Koelewijn et al. 2012a, 2014; Ohlenforst et al. 2017b). We analyzed the contribution of the three cognitive measures (RST, SICspan and TRT) separately, based on evidence suggesting that the TRT depends on cognitive factors of speech recognition that differ from cognitive aspects measured by the RST (Besser et al. 2012, 2013). By testing the separate contribution of different cognitive measures, we aimed to promote a better understanding of the specific cognitive abilities that those tests tap into and their predictive values for speech recognition and listening effort in noise. Recent research indicated for example that the TRT test measures cognitive processes relevant for speeded lexical decision making during recognizing partly masked sentences and that these processes require working memory capacity (Zekveld et al. 2018). Better TRT performance was associated with better hearing abilities and better working memory and verbal processing speed and generally with better speech understanding (Humes et al. 2013). Recent hearing research that examined the relationship of TRT and working memory on SRT performance in various maskers was reviewed (Besser et al. 2013). It appeared that TRT and working memory capacity measures such as the RST are related to each other, but differ in their relationships with SRT performance. For normal-hearing listeners TRT better predicted speech recognition in fluctuating nonspeech maskers whereas TRT and RST scores were associated with SRTs in speech maskers (Besser et al. 2013). Previous research suggested furthermore that the association between cognitive measures and speech recognition performance may be affected by the listeners’ hearing ability (George et al. 2007; Besser et al. 2013; Zekveld et al. 2018). It was indicated that cognitive load during speech recognition was larger for listeners with better TRT performances for 50% correct performance when listeners were normal-hearing and the 50% and the 84% condition when listeners were hearing-impaired (Zekveld et al. 2011).

Splitting the normal-hearing listeners into groups of high and low cognitive performance, differences in sentence recognition appeared across a large range of SNRs for the single-talker masker condition. Listeners with high working memory capacity (RST) or high working memory capacity and good inhibition abilities (SICspan) showed better speech recognition performance across a range of SNR conditions, ranging from 0% to 100% intelligibility. For the stationary noise masker, the influence of RST on speech recognition performance varied across a small range of negative SNRs (-4 dB to -8 dB), however the interaction did not exhibit any obvious systematic trend. The benefit of better cognitive functionality (high RST, high SICspan) may be reflected by better speech recognition performance for the single-talker masker across a large range of SNRs, compared to the stationary masker. More cognitive
processing load is perhaps required for the intelligibility of speech when interfering speech information is present compared to a stationary noise masker (Koelewijn et al., 2012a; Larsby et al., 2005).

The results of the current study cation against the intuitively appealing assumption that individual differences in cognitive measures, such as working memory capacity may equally explain speech recognition performance for different groups of listeners. The contradictory findings between Experiment 1 and Experiment 2 in the current study may partly be explained by the variance in the outcome measures. Additionally, the variance across SNRs was three times smaller for the normal-hearing than for the hearing impaired listeners. Overall, our findings and the applied methods within this study can only explain a small part of the interplay between cognitive measures such as working memory capacity, inhibition of interfering information and linguistic closure on speech recognition performance and the pupil response depending on the listeners hearing ability.

4.4 General conclusion

In summary, our findings provide at least partial support for our hypothesis. The results strongly indicate that higher working memory capacity, inhibition of interfering information and linguistic closure abilities can have a positive impact on speech recognition but may also come at the cost of increased allocation of cognitive resources to a given listening task, when listeners are hearing-impaired. The impact of the listeners hearing ability on the link between cognitive measures, speech recognition and the pupil response did not show a systematic trend. The association between cognition, speech recognition and listening effort are not straightforward. The contributions of cognitive ability to speech recognition and listening effort depend on someone’s hearing status and the listening conditions. Associations may differ or even be reversed when comparing easy, difficult, and undoable conditions. The results caution to conclude that the listeners hearing ability can generally predict how individual differences in working memory, speech recognition performance and the allocation of cognitive resources interact.