Pupil Response as an Indication of Effortful Listening: The Influence of Sentence Intelligibility

Adriana A. Zekveld, Sophia E. Kramer, and Joost M. Festen

Objectives: The aim of this study was to evaluate the influence of sentence intelligibility on the pupil dilation response during listening. Task-induced pupil-dilation reflects explicit effortful processing load. Therefore, pupillometry can be used to examine the listening effort during speech perception in difficult listening conditions. We expected to find increasing pupil dilation as a function of decreasing speech intelligibility.

Design: Thirty-eight young participants (mean age = 23 yrs, SD = 3.2 yrs) with normal hearing were included. They performed three speech reception threshold (SRT) tests in which they listened to sentences in stationary noise. A one-up-one-down, two-up-one-down, or four-up-one-down adaptive procedure was applied, resulting in the correct rehearsal of 50, 71, or 84% of the sentences ($SRT_{50\%}$, $SRT_{71\%}$, and $SRT_{84\%}$, respectively). We examined the peak dilation amplitude, the latency of the peak dilation amplitude, and the mean pupil dilation during the processing of the speech in each of these conditions. The peak dilation amplitude and mean pupil dilation were calculated relative to the baseline pupil diameter during listening to noise alone. For each SRT condition, participants rated the experienced listening effort and estimated their performance level.

Results: The signal to noise ratios (SNRs) in the $SRT_{50\%}$, $SRT_{71\%}$, and $SRT_{84\%}$ conditions increased as a function of the speech intelligibility level. The subjective effort ratings decreased, and the estimated performance increased with increasing speech intelligibility level. Repeated measures analyses of variance indicated that peak dilation amplitude and mean pupil dilation were higher in the $SRT_{50\%}$ condition as compared with the $SRT_{71\%}$ and $SRT_{84\%}$ conditions. The peak dilation amplitude, mean pupil dilation, and peak latency increased with decreasing SNR of the speech in noise, but no effect of noise level by itself on the baseline pupil diameter was observed. Irrespective of SNR, the pupil response was higher for incorrectly repeated sentences than for correctly repeated sentences. The analyses also indicated condition-order effects on the peak dilation amplitude and mean pupil dilatation: the pupil response was higher in the first SRT test than in the second and third tests. Within the first and third test, the baseline pupil diameter and the mean pupil dilation decreased as a function of the sentence number within the test. Spearman correlation coefficients showed no relations among the SNRs at the SRTs, subjective ratings, and the pupil response.

Conclusions: The peak dilation amplitude, peak latency, and mean pupil dilation systematically increase with decreasing speech intelligibility. These results support that listening effort, as indicated by the pupil response, increases with decreasing speech intelligibility. This study indicates that pupillometry can be used to examine how listeners reach a certain performance level. Application of this technique to study listening effort can yield valuable insight into the processing resources required across listening conditions and into the factors related to interindividual differences in speech perception in noise.

(Ear & Hearing 2010;31;480–490)

Department of ENT/Audiology and the EMGO+ Institute for Health and Care Research, VU University Medical Center, Amsterdam, The Netherlands.

INTRODUCTION

When speech understanding is impaired by background noise or hearing loss, speech comprehension becomes more reliant on cognitive (explicit) working memory processes to fill in any incomprehensible or missing information (Pichora-Fuller 2003, 2006; Zekveld et al. 2007; Rönnberg et al. 2008; Kramer et al. 2009). Thus, the use of cognitive processes facilitates listening by allowing the listener to complete parts of the missing information. As such, it helps individuals to compensate for speech perception difficulties. However, the use of "top-down" mechanisms such as working memory requires effort, concentration, and attention (Rabbitt 1968, 1991; Pichora-Fuller et al. 1995; Wingfield et al. 2005). Hearing-impaired people often stress that the listening effort they spend while trying to concentrate and compensate for their hearing loss and the resulting fatigue are extremely demanding (Edwards 2007). A person's attempt to compensate for the hearing loss and to communicate optimally requires vigilance: a constant effort to hear and pay attention. Among those who are occupationally active, mental distress resulting from hearing loss has shown to be associated with a higher need for recovery after work and higher sick leave incidence (Kramer et al. 2006; Nachtegaal et al. 2009). A person with hearing impairment is more tired after an hour of conversing in a noisy situation than someone with normal hearing (Edwards 2007). In other words, the cognitive processes required during speech comprehension are associated with more effortful listening.

To verify this statement further, combining behavioral measures such as speech intelligibility performance with a measure indexing the cognitive effort required when listening to speech would provide valuable information (Stanners et al. 1972). By examining the degree to which a listener relies on cognitive processes, one can measure at what cost a certain performance level has been reached (Wingfield et al. 2005; Edwards 2007; Pichora-Fuller 2007). Just measuring speech intelligibility and working memory capacity is not sufficient to evaluate listening effort in background noise or with hearing loss. For example, when comparing the performances on two conditions of a speech perception task, differences in listening difficulty between the conditions will not become apparent if the participants successfully compensate for increased task demands by increasing the amount of effort. For instance, Verney et al. (2001) combined behavioral and psychophysiologic measures to examine information-processing models and observed that the easiest conditions, as indicated by the behavioral data, however, imposed the greatest demands on the processing resources. This indicates that resource demands can be highest when performance levels are also highest (see also Fish & Granholm 2008). Furthermore, measuring listening effort can be useful in studies examining factors related to interindividual differences in speech perception.

0196/0202/10/3104-0480/0 • Ear & Hearing • Copyright © 2010 by Lippincott Williams & Wilkins • Printed in the U.S.A.

Recently, several studies have shown that hearing-impaired persons rate difficult listening situations more effortful as compared with normal-hearing persons (Hällgren et al. 2005; Larsby et al. 2005). Although a subjective effort-measure gives insight into the experienced listening effort, it is uncertain whether the subjective indication of effort truly reflects the availability or demand on processing resources (Wickens 1992). It may be difficult for persons to discriminate between task dimensions when rating mental effort and interindividual differences in these decision criteria make the interpretation of subjective effort ratings complex (Yeh & Wickens 1988; Recarte et al. 2008).

Several studies have used dual-task paradigms to examine the cognitive resources allocated during speech comprehension in difficult listening conditions. The rationale of this technique is that the performance on the secondary task is assumed to be inversely related to the primary task demands (Rabbitt 1991; Tun et al. 1992; Wickens 1992). Rakerd et al. (1996) used a dual-task paradigm to examine listening effort. In a "speech dual-task trial," participants first memorized a set of nine or 11 digits, then heard a speech fragment of 1-min length, and subsequently had to repeat the memorized digits. After the recall phase, subjects had to answer several questions on the content of the speech fragment. The control condition consisted of "noise dual-task trials" in which the memorization interval was followed by a noise-alone stimulus of 1-min length. Three groups participated: young listeners with normal hearing, young listeners with either congenital or early-onset severe hearing loss, and older listeners with moderate presbyacusic hearing loss. The participants recalled more digits during the noise trials than during the speech trials, indicating that the presentation of speech interrupts the rehearsal of the memorized digits more than the presentation of noise. Importantly, this effect was larger for the two groups of participants with hearing impairment compared with the participants with normal hearing, with the young participants with congenital or early-onset hearing loss performing worst. Although it was not possible to distinguish the effects of ageing and hearing impairment, the authors stated that both factors have likely played a role. This was confirmed by an overall lower memory performance for the older participants, both in the speech and noise dual-task trials. These results suggest that the cognitive costs of speech perception are greater for listeners with hearing loss than for listeners with normal hearing.

McCoy et al. (2005) also applied a dual-task paradigm to study the effect of hearing loss on the listening effort during word comprehension. Individuals with and without hearing loss listened to a word list. Stimulus presentation was interrupted randomly and participants were asked to repeat the last three words presented to them (i.e., a "running memory task"). Data were only included in the analysis when the participant correctly reproduced the last of these three words, as the authors assumed that in that case, each of the words was correctly perceived. The linguistic context was varied: each word in the list was unrelated to the preceding word, associated with the preceding word, or associated with the two preceding words. The recall performance of the participants with hearing loss was lower than that of the participants without hearing loss, except when the words were associated with the two preceding words. The authors concluded that the increased effort allocated to listening by the participants with hearing impairment had reduced the available resources for downstream operations such as encoding the words in memory. Linguistic context reduced the effort required to perceive the speech.

Besides subjective measures and the application of dual-task paradigms, physiologic measures have also been reported in the literature and shown to be valuable means to study cognitive effort (Wickens 1992).

The results of a study performed by Kramer et al. (1997) demonstrated that task-related pupil dilation can be used to examine the cognitive load required for speech comprehension. The pupil diameter enlarges with increasing mental effort and systematically reflects the task-processing load and resource allocation in multiple domains of cognitive functioning (Beatty 1982; Wickens 1992; Steinhauer & Hakerem 1992; Hoeks 1995; Verney et al. 2004). Beatty (1982) showed that pupillary responses are able to reflect the variations in the cognitive demands evoked by different tasks and interindividual variations in the cognitive load required to perform a task. For example, pupil dilation depends on processing strategies that influence the performance level and increased processing demands on a wide range of tasks (e.g., memory, language, and perception tasks) result in pupil dilation. The pupil response reflects the sum of all processing demands associated with the task in relation to the available resources (Beatty 1982; Just & Carpenter 1993; Granholm et al. 1996; Just et al. 2003; Fish & Granholm 2008).

The peak dilation amplitude to target stimuli commonly occurs within the time window from 0.7 to 1.5 sec. (Verney et al. 2004) and is regulated by the combined activity of the sympathetic and parasympathetic branches of the autonomic nervous system (Lowenstein & Loewenfeld 1962; Steinhauer & Hakerem 1992). Two iris muscles control the size of the pupil; the pupillary sphincter controls pupil contraction and ties directly into the parasympathetic nervous system, whereas the pupillary dilator results in pupil dilation and ties directly into the sympathetic nervous system (Janisse 1977). Through extensive corticoreticular connections, task-evoked pupillary dilation reflects the extent of cortical activation during cognitive processing. Activity in a variety of brain areas can result in pupil dilation; for example, the task-evoked pupil response has been associated with activity in the dorsolateral prefrontal cortex during the performance of tests of working memory and executive functioning (Siegle et al. 2003, 2004). Results of studies examining the neural correlates of the pupil response depend on the applied task and stimuli characteristics; the response can be viewed as an indirect measure of the brain activation evoked by the cognitive demands of a task (Janisse 1977; Stanners et al. 1979; Beatty 1982; Steinhauer & Hakerem 1992). However, this indirectness of the measure does not preclude the exploitation of pupillometry in cognitive neuroscience. As convincingly argued by Beatty and Lucero-Wagoner (2000), the reliable correlation between the task-evoked pupil response and central processing load makes it a valuable research tool, despite the absence of a direct causal link between cognitive intensity and the low face validity of the method. Beatty (1982) provided a thorough review of pupillometric studies demonstrating that the pupil response sensitively indicates between-task and between- and within-individual differences in cognitive processing load. With regard to language processing, Just et al. (2003) provided an overview of studies in which the pupil response has been shown to be sensitive to subtle differences in language complexity. Considering speech perception, more complex listening conditions will require more top-down processing, which is associated with both increased activity in brain regions associated with these top-down functions and pupil dilation (Kramer et al. 1997; Davis & Johnsrude 2003; Scott & Johnsrude 2003; Zekveld et al. 2006; Obleser et al. 2007).

As argued by Kahneman and Wright (1971) and Stanners et al. (1972), an advantage of the application of pupillometry during speech perception is that it does not require adaptation of the stimuli and tasks or negatively affects speech perception. Hence, standard speech perception tests can be applied without interference from the measurement of the pupil dilation. Compared with other physiological measures reflecting changes in cognitive effort, such as heart rate variability, pupillometry is relatively reliable (Kahneman et al. 1969).

Several studies have applied pupillometry to examine the effort required during linguistic processing. Stanners et al. (1972) examined the pupil response during the perception of relatively simple, short sentences. Participants were asked to memorize the sentence for rehearsal or to prepare for paraphrasing the sentences (i.e., repeating the sentence meaning in their own words). The memorize condition was similar to most standard speech reception tests, except that for only part of the stimuli participants were actually asked to repeat the sentence. The pupil size started to increase with the presentation of the third or fourth word in the sentence and continued to increase until shortly after the utterance of the last word in the sentence. The relatively long latency of the pupil response was likely caused by the fact that the first two words of a sentence were always identical, which required no effortful processing. This study supports the use of pupillometry to examine the cognitive effort during listening to short sentences presented in quiet. In addition, Wright and Kahneman (1971) showed that pupil dilation during the processing of auditorily presented sentences increases with increase in the interval between stimulus presentation and the response. Sentence repetition was associated with larger pupil dilation than answering a question about the sentence.

Schluroff (1983) reviewed several studies concerning the pupillary response during listening to sentences. The results suggested that the pupil response reflects both the storage of the information (for rehearsal) and the demand on explicit processes, and hence is related to working memory processes. The amount and complexity of the information and the availability of linguistic context influenced the pupil response. Pupil dilation was highest for the rehearsal of sentences that lacked both syntactic and semantic structure (i.e., random word strings), followed by sentences lacking only semantic structure, and the smallest pupil response was observed for normal sentences. Thus, the use of available context during listening (which uses higher cognitive processes) actually facilitated listening.

Kramer et al. (1997) applied pupillometry during the speech reception threshold (SRT) test (Plomp & Mimpen 1979). Compared with normal-hearing participants, hearing-impaired individuals had a smaller reduction in the allocated effort during easier listening conditions. The poorer the individuals' ability to comprehend speech in noise, the less benefit was obtained from easier listening conditions. This study aimed to extend their results by carefully equalizing the intelligibility level of speech among participants, thereby controlling for interindividual performance differences when comparing the listening effort required for several intelligibility levels. Because little is known about the characteristics of the pupil response for this specific application of pupillometry, the general aim of this study was to gain more insight into several factors influencing the pupil response during listening to speech presented in noise. This insight may aid the application of pupillometry in future studies to the cognitive load during listening in difficult situations. The task performance (SRT), subjective ratings, and the pupil response are three different supplementary types of taskload indices (Wickens 1992). Because these indices may differentially reflect both differences in task difficulty and differences between individuals, it is interesting to obtain and directly compare these measures, which was done in this study.

The primary aim was to examine the influence of speech intelligibility on the pupil response. Therefore, we created three conditions with fixed performance levels (50, 71, or 84%) during which we aimed to assess the pupil response. The secondary aim was to examine the relation between the pupil response at these three intelligibility levels and (1) subjective effort and performance ratings and (2) the signal to noise ratio (SNR) required for these three speech performance levels. We also aimed to investigate the effects of the noise level and the order of conditions on the pupil measures and to examine differences in the pupil response between incorrectly and correctly repeated sentences. Young listeners with normal hearing were included in this study.

Based on the results described by Kramer et al. (1997), we expected that the listening effort reflected by the pupil response would be higher for the more difficult SRT conditions (i.e., at lower SNRs).

PATIENTS AND METHODS

Participants

Thirty-eight normal-hearing students (28 women and 10 men) of the Vrije Universiteit (VU University) in Amsterdam participated in this study. Their ages ranged from 19 to 31 yrs, with a mean age of 23 yrs (SD = 3.2 yrs). All participants were native Dutch speakers who reported normal or corrected to normal vision, no dyslexia, and no history of neurological disease. None of the participants wore eye make-up during the tests, and seven participants used contact lenses during the tests. All participants provided written informed consent in accordance with the Ethical Committee of the VU University Medical Center.

Stimuli and Tests

Pure-Tone Audiometry • Pure-tone hearing thresholds of the participants were measured at the start of the test session to ensure that the thresholds of both ears were ≤20 dB HL at the frequencies 250, 500, 1000, 2000, and 4000 Hz.

SRT in Quiet • In the SRT tests in noise (see below), the speech level was fixed at 55 dB SPL, and the SNR was varied adaptively by changing the noise level. We first measured the SRT in quiet. This score is required to check whether the speech level (i.e., 55 dB SPL) was about 20 dB higher than the SRT in quiet to ensure that noise and speech details were sufficiently audible for each listener (Plomp 1986). During the SRT test in quiet, we adaptively estimated the sound level required for reproducing 50% of the sentences without error. One list of 13 short, everyday Dutch sentences was presented (Plomp & Mimpen 1979) through a single loudspeaker located in front of the participant. Although it is possible to wear both the pupillometric equipment and headphones, we decided to use a loudspeaker for auditory stimulus presentation to ensure the physical comfort of the participants and avoid measurement error due to discomfort. The first sentence was presented at a sound level below the expected SRT and was presented repeatedly; increasing the sound level in increments of 4 dB until the participant repeated the sentence correctly. Each of the other sentences was presented once; if the sentence was reproduced correctly, the sound level of the next sentence was decreased by 2 dB, and if it was recognized incorrectly, the sound level was increased by 2 dB. Participants were asked to repeat each sentence and encouraged to make their best guess for sentences they had not understood entirely. A sentence was scored as correct if the participant was able to repeat each word of the sentence without error. No feedback was given during the tests. The SRT in quiet was the mean sound level of sentences 5 to 13. For all participants, the speech level applied in the speech in noise tests (i.e., 55 dB SPL) was at least 30 dB higher than the SRT in quiet.

SRT in Noise • In each SRT in noise test, 45 short, everyday Dutch sentences as developed by Versfeld et al. (2000) were presented. These sentences are similar to those developed by Plomp and Mimpen (1979). The mean duration of the selected sentences was 1.8 secs (range, 1.4 to 2.7 secs). The sentences contained eight to nine syllables, words in the sentences did not contain more than three syllables, and the grammatical and syntactical structure of the sentences was relatively simple and similar across sentences. Sentences from this set are pronounced clearly, in a natural tempo. An example sentence (translated into English) is "the shop is within walking distance." See the article by Versfeld et al. for more details. The use of this standardized sentence material minimized differences between sentences in, for example, syntactic complexity, which can affect both intelligibility (Stewart & Wingfield 2009) and the pupil response (Just & Carpenter 1993).

Stationary noise with a long-term average spectrum of the speech served as masker. For each sentence, the presentation of the noise started 3 secs before the presentation of the sentence in the noise. An answer prompt (i.e., a 1-sec, 1000 Hz tone at 55 dB SPL) was presented 4.8 secs after the start of the speech (i.e., on average, 3 secs after the end of the speech, range 3.5 to 2.1 secs) after which the participant repeated the sentence aloud. The experimenter scored the response (correct or incorrect) after each sentence. Then, the next sentence was given (Fig. 1).

Three SRT conditions were applied. In each condition, the speech level was fixed at 55 dB SPL and the SNR was varied adaptively by changing the noise level. In one condition, we applied the one-up-one-down adaptive procedure as was applied in the SRT test in quiet (Plomp & Mimpen 1979), thereby estimating the SNR required for 50% sentence intelligibility. In the other two conditions, the adaptive procedure was modified (Levitt 1971) to target two additional intelligibility levels: 71% (two-up-one-down procedure) and 84% (four-up-one-down procedure). In the two-up-one-down procedure, participants

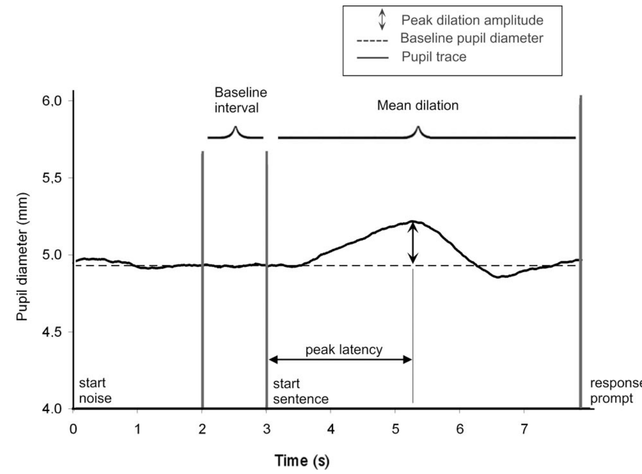

Fig. 1. Illustration of the pupil outcome measures. The mean pupil diameter in the 1-sec interval before the start of the speech (during listening to noise alone) is the baseline pupil diameter. The maximum pupil diameter between the start of the speech and the response prompt, relative to the baseline pupil diameter, is the peak dilation amplitude. The peak latency is the interval between the start of the sentence and the peak dilation amplitude. The mean pupil dilation is the average pupil diameter between the start of the speech and the response prompt, relative to the baseline pupil diameter.

had to repeat two subsequent sentences correctly before adapting the SNR of the following sentence with −2 dB. In the four-up-one-down procedure, participants had to correctly rehearse four subsequent sentences for a 2-dB SNR reduction of the following sentence. In each condition, the SNR was increased by 2 dB after an incorrect response. The rationale for applying these intelligibility levels was that they will cover a substantial range in listening effort, without resulting in effortless speech perception (i.e., ~100% correct perception) or too difficult perception leading to overload and pupil constriction (Poock 1973; Granholm et al. 1996).

The SRT was the mean SNR of sentences 5 to 45. We will refer the one-up-one-down, two-up-one-down, and four-up-one-down conditions as the $SRT_{50\%}$, $SRT_{71\%}$, and $SRT_{84\%}$ conditions, respectively.

Subjective Listening Effort and Performance Rating • Immediately after completing an SRT test, participants were asked to rate the listening effort and their performance during the preceding test. For the effort rating, we used a continuous rating scale similar to the one described by Larsby et al. (2005). The following are the English translations of the Dutch labels used (the corresponding numbers on the scale are given between parentheses): no effort (1), low effort (3), moderate effort (5), high effort (7), and very high effort (9). For the performance rating, a similar scale was presented with the following labels: None of the sentences was intelligible (1), some of the sentences were intelligible (3), about half of the sentences were intelligible (5), a lot of the sentences were intelligible (7), and all sentences were intelligible (9).

Pupillometry

The pupillometric data of the left eye were measured with a SensoMotoric Instruments (Berlin, Germany) recording system (2D Video-Oculography, version 4). The system uses infrared video-based tracking technology to measure the pupil diameter. The spatial resolution of the pupillometer was 0.03 mm. The location and size of the pupil were automatically recorded at 50 Hz and a PC connected to the pupillometer stored the data together with time stamps, indicating the start of the trials and the stimuli, the prompt signal, and the response of the experimenter. During testing, the experimenter inspected the raw pupil data to check for a valid recording of the pupil diameter; if necessary, corrective action was taken. The pupillometric equipment was calibrated using printed pupils of known size positioned on a head model.

Procedure

Test administration took place in a sound-treated room. Each participant sat in a comfortable chair. During the SRT tests, they were instructed to maintain focus at a fixation dot to reduce the possibility that the light reflex affected the task-evoked pupil response. The fixation dot was positioned at a white wall at 4 m distance. All auditory stimuli were presented through a loudspeaker placed in front of the participant, positioned below the fixation dot.

First, participants performed the SRT test in quiet. Then, the pupillometric equipment was positioned on the head of the participants and the camera focus was adapted. Participants were not allowed to wear eyeglasses during the tests, but contact lenses were allowed, as they did not affect the recording of the pupil by reflecting the infrared illumination. Similar to the study by Chapman et al. (1999), the room illumination was adjusted to prevent ceiling and floor effects in the pupil response. Therefore, first the pupil diameter was measured in maximum illumination (230 lx) and subsequently in darkness by placing the front cover plate of the pupillometric equipment in front of the eyes. The room illumination was adapted such that the pupil diameter was around the middle of its dynamic range. This makes the task-evoked pupil response independent of the baseline pupil diameter (Janisse 1977; Hyönä et al. 1995; Chapman et al. 1999; Beatty & Lucero-Wagoner 2000). The mean room illumination across participants was 50 lx. Participants were asked to inhibit eye blinks. The positioning of the pupillometric equipment and the light adjustments took 5 mins per subject.

Subsequently, participants performed a practice SRT in noise test (two-up-one-down procedure) in which 13 sentences were presented. After this test, they indicated the subjective listening effort and estimated their performance level. Data from the practice tests were not included in the analyses. Then, they performed the three SRT conditions. The order of the SRT tests and the assignment of sentence lists to conditions were balanced separately across participants. Sentences were presented only once to each participant. The duration of the test session was 1.5 hrs with a 10-min break halfway.

Data Selection, Cleaning, and Reduction

Data Selection • Similar to the SNRs of sentences one to four of each list (see Section SRT in Noise), the pupil traces of these sentences were omitted from the analyses. The mean and SD of the pupil diameter between the start of the speech and the answer prompt was calculated per trace; pupil diameter values below 3SDs of the mean were coded as a blink. Trials in which the data contained >15% of blinks between the start of the baseline (see below) and the prompt signal were excluded from the data analysis. Furthermore, the pupil data for each sentence (i.e., each pupil trace) was visually inspected for artifacts because of eye movements by examining the recorded diameter, and the x- and y-coordinates of the pupil center. For each participant, at least 68% of the traces per condition were included in the data analysis (on average 96%). Linear interpolation beginning four samples before and ending eight samples after a blink replaced blinks throughout the data set to reduce artifacts resulting from partial eyelid closures (Siegle et al. 2008).

Calculation of Pupil Dilation Indices • A five-point moving average smoothing filter was passed over the data of the selected and deblinked traces after which the data were averaged separately for each participant per condition. To calculate the pupil dilation, the average pupil diameter in the 1.0 sec preceding the onset of the speech (during the noise-alone presentation) was used as a baseline. We determined three pupil outcome measures for the average trace of each condition: (1) the peak dilation amplitude relative to the baseline pupil diameter, (2) the latency of the peak dilation amplitude, and (3) the mean pupil dilation between the end of the baseline interval (i.e., the start of the speech) and the response prompt, relative to the baseline pupil diameter. These outcome measures are illustrated in Figure 1. The peak dilation amplitude was examined to enable comparing the current results with those of earlier studies (e.g., Kramer et al. 1997). The mean pupil dilation was additionally examined as this measure provides greater reliability, stability, and a better index of cognitive resource allocation than peak measures (Ahern & Beatty 1979; Verney et al. 2001). The pupil data during the response interval (after the response prompt) were omitted from the analysis, because responding itself leads to a pupil response with peak dilation amplitude occurring around 0.5 secs after participants start responding (Richer & Beatty 1985).

Statistical Analyses

Within-subjects effects of SRT condition, condition order, and noise level on the behavioral and pupillometric data were all tested with repeated measures analyses of variance (ANOVA). If statistically significant effects were observed, pairwise comparisons were performed in which we applied Bonferroni-corrected p values. A p value <0.05 was considered significant.

RESULTS

Behavioral Data

The mean pure-tone hearing threshold at 500, 1000, 2000, and 4000 kHz, averaged over both ears, was 4.0 dB SPL (SD = 3.0 dB) and the mean SRT in quiet was 18.2 dB SPL (SD = 2.0 dB). Table 1 shows the descriptive statistics of the SRTs in noise, the actual performance level in the SRT tests (percentage correctly repeated sentences), and the subjective ratings. The performance level, SRTs, and subjective performance ratings increased with increasing speech intelligibility, and the effort ratings decreased with increasing intelligibility. The repeated measures ANOVA indicated significant effects of condition on the SRTs and on the subjective effort and performance ratings. Subsequent pairwise comparisons indicated statistically significant differences in SRTs between the $SRT_{50\%}$ and $SRT_{71\%}$

TABLE 1. Descriptive statistics of the SRTs and the subjective ratings

	SRT Condition					
	$SRT_{50\%}$		$SRT_{71\%}$		$SRT_{84\%}$	
	M	SD	M	SD	M	SD
SRT, dB SNR	−4.4	0.6	−2.6	1.0	−0.8	1.0
Performance (% correct)	50.8	1.5	70.4	2.3	85.4	1.8
Subjective effort (0…10)	7.0	1.5	5.9	1.5	5.0	1.8
Subjective performance (0…10)	4.8	1.1	6.4	1.2	7.4	0.9

SNR, signal-to-noise ratio; SRT, speech reception threshold.

TABLE 2. Descriptive statistics of the pupil measures

	SRT Condition					
	$SRT_{50\%}$		$SRT_{71\%}$		$SRT_{84\%}$	
	M	SD	M	SD	M	SD
Baseline pupil diameter, mm	5.23	0.87	5.07	0.73	5.11	0.86
Peak dilation, mm	0.17	0.11	0.12	0.10	0.11	0.10
Latency of peak dilation, sec	2.48	0.65	2.55	0.78	2.32	0.66
Mean dilation, mm	0.05	0.06	0.02	0.06	0.00	0.06

Peak and mean dilation are calculated relative to the individual mean baseline pupil diameter. The peak latency is relative to the onset of the speech.
SRT, speech reception threshold.

conditions and between the $SRT_{71\%}$ and $SRT_{84\%}$ conditions ($p < 0.001$). Additionally, the subjective effort ratings differed between the $SRT_{50\%}$ and $SRT_{71\%}$ conditions and between the $SRT_{71\%}$ and $SRT_{84\%}$ conditions ($p < 0.001$ and $p < 0.01$, respectively), as well as the performance ratings ($p < 0.001$). The balanced test order enabled us to examine the effects of test order on the behavioral data. The repeated measures ANOVA indicated no statistically significant effects of the order position of the test on the SRT and the subjective effort and performance ratings.

Pupil Data: Order Effects • No statistically significant effect of condition order on the baseline diameter and the peak latency was observed. The order effect on the peak dilation amplitude and mean pupil dilation was statistically significant ($F_{[2,74]}$ =0 8.2, $p < 0.001$ and $F_{[2,74]} = 9.1$, $p < 0.001$, respectively). Pairwise comparisons indicated that both outcome measures were significantly larger in the first test compared with the second and third tests ($p = 0.016$). No statistically significant differences in the pupil response between the second and third tests were observed.

To examine the effect of sentence number on the pupil measures, we averaged the pupil traces separately for each SRT condition (i.e., first, second, and third) and sentence number (5 to 45), over participants. After visual inspection of the data, we calculated the Pearson correlation coefficients between the sentence number and the pupil measures. The baseline pupil diameter and the mean dilation decreased statistically significantly during the course of the first SRT test ($r = -0.83$, $p <$ 0.001 and $r = -0.50$, $p < 0.001$, respectively). Additionally, the baseline pupil diameter decreased statistically significant during the third test ($r = -0.61$, $p < 0.001$).

Relationship Between Individual SRTs, Subjective Ratings, and the Pupil Response • Spearman correlation coefficients between the individual SRTs, the subjective effort and performance ratings, and the pupil measures (peak dilation amplitude, peak latency, and mean pupil dilation) were calculated separately for each performance condition. No statistically significant relationships were observed. However, note that the differences in intelligibility performance (i.e., 50, 71, or 84%) were systematically reflected by the averaged subjective effort and performance ratings (Table 1).

Effect of SRT Condition on the Pupil Response • Table 2 shows the descriptive results of the pupil measures for each of the three SRT conditions, and Figure 2 shows the mean pupil response in each condition. For the creation of Figure 2, first the mean pupil trace for each participant and condition was calculated. Then, these traces were shifted to the average peak latency (over participants) for the corresponding condition (Just & Carpenter 1993). The rationale for this shift was that we aimed to illustrate the differences in peak dilation amplitude between the conditions in Figure 2. Because of interindividual differences in the peak latency, without shifting, the peak dilation amplitude of the average trace over participants would be smaller than the data used in the analyses (i.e., as provided in Table 2). After shifting the traces, the pupil data were averaged over participants and displayed in Figure 2.

The mean baseline pupil diameter differed significantly between conditions ($F_{[2,74]} = 3.9$, $p = 0.025$). However, a pairwise comparison indicated no statistically significant differences in the baseline diameter between the SRT conditions. We observed statistically significant effects of SRT condition on the peak dilation amplitude and mean pupil dilation; $F_{[2,74]} =$ 11.0, $p < 0.001$ and $F_{[2,74]} = 10.2$, $p < 0.001$, respectively. Pairwise comparisons indicated that for both outcome measures, the pupil response in the $SRT_{50\%}$ condition was significantly larger than the pupil response in the $SRT_{71\%}$ and $SRT_{84\%}$ conditions ($p < 0.05$). The pupil response in the $SRT_{71\%}$ condition did not differ from that in the $SRT_{84\%}$ condition.

Effect of Noise Level on the Pupil Response • The SNR was adaptively varied between the sentences, and the baseline pupil diameter was determined during listening to noise alone. We examined the effects of noise level on the mean baseline pupil diameter, the peak dilation amplitude, the peak latency, and the mean pupil dilation. For this, for each participant, the pupil traces were averaged separately for each of the presented noise levels, over conditions. Then, for each noise level, the baseline diameter and pupil response measures were determined and averaged over participants. For 37 participants, data were available for each of the following noise levels: 55, 57, 59, 61, and 63 dB SPL (Table 3 and Fig. 3).

No statistically significant effect of the noise level on the baseline pupil size was observed, but the peak dilation amplitude, the peak latency, and the mean pupil dilation significantly increased with increasing noise levels ($F_{[4,144]} = 24.02$, $p <$ 0.001; $F_{[4,144]} = 3.49$, $p < 0.01$; $F_{[4,144]} = 15.66$, $p < 0.001$, respectively). A pairwise comparison showed that the peak dilation amplitude at each of the five noise levels differed significantly from the peak dilation amplitude at the other noise levels ($p < 0.05$). The pairwise comparison did not indicate statistically significant differences in the peak latency between

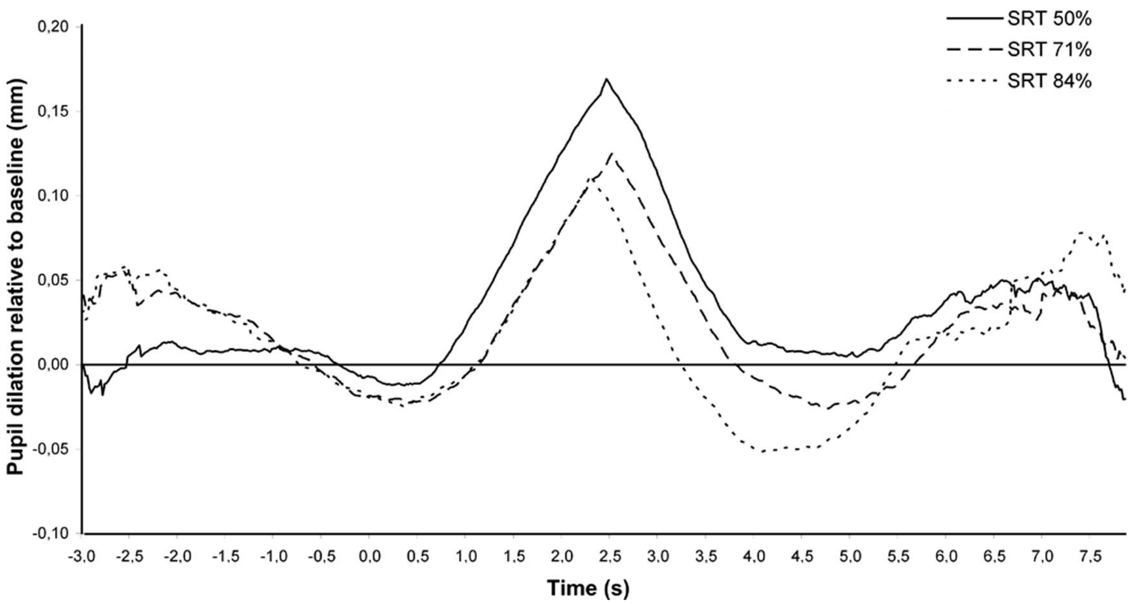

Fig. 2. Mean pupil response for the three speech reception threshold (SRT) conditions resulting in 50, 71, or 84% sentence recognition. First, the pupil response for each participant and condition was calculated. The resulting traces were shifted to the average peak latency (over participants) for the corresponding condition. The pupil data were then averaged over participants, separately for each condition. The value 0 on the time scale indicates the start of the sentence.

the noise levels, but the mean pupil dilation at the noise levels 55 and 57 dB SPL was significantly smaller than the mean pupil dilation at 59, 61, and 63 dB SPL ($p < 0.05$). The mean pupil dilation did not statistically differ between these three highest noise levels.

Pupil Response: Incorrectly Versus Correctly Repeated Sentences • For each individual, we selected the SNR at which the largest number of correct and incorrect responses was observed over conditions (i.e., close to the $SRT_{50\%}$), and averaged the pupil data separately for the incorrect and correct responses. We then determined the peak dilation amplitude, the peak latency, and the average pupil response for both types of trials.

Table 4 shows the descriptive statistics of the pupil response separately for the correctly and incorrectly repeated sentences. Pairwise *T* tests indicated that the peak dilation amplitude and the mean pupil dilation were higher and the peak latency was longer for the incorrectly repeated sentences ($t_{[37]} = -5.0, p < 0.001$; $t_{[37]} = -6.2$, $p < 0.001$; $t_{[37]} = -3.9$, $p < 0.001$, respectively) compared with the correctly repeated sentences.

DISCUSSION

The most important result of this study is that the pupil response during listening to sentences systematically varied as a function of the speech intelligibility. Consistent with numerous studies that measured mental effort by means of pupillometry, task manipulations assumed to increase processing load were associated with a larger pupil response. Similar to Kramer et al. (1997), pupil dilation was largest in the $SRT_{50\%}$ condition, and decreased as a function of increasing speech intelligibility. The pupil response in the $SRT_{71\%}$ condition did not differ from the response in the $SRT_{84\%}$ condition. This result can be based on the larger difference in intelligibility between the $SRT_{50\%}$ and $SRT_{71\%}$ conditions (i.e., ≈20%) than the intelligibility difference between the $SRT_{71\%}$ and the $SRT_{84\%}$ conditions (i.e., ≈15%). It indicates that increasing the intelligibility from 50 to 71% reduces the listening effort more than further increasing the intelligibility level from 71 to 84%. Interindividual differences in the listening effort as reflected by the pupil response were not related to differences in speech

TABLE 3. Mean pupil baseline diameter, peak dilation, peak latency, and mean dilation in pupil response interval (N = 37)

	Pupil Measure							
	Baseline Diameter, mm		Peak Dilation, mm		Peak Latency, sec		Mean Dilation, mm	
Noise Level, dB SPL	M	SD	M	SD	M	SD	M	SD
55	5.14	0.81	0.08	0.08	2.16	0.92	0.03	0.06
57	5.14	0.79	0.12	0.10	2.35	0.99	0.01	0.05
59	5.19	0.77	0.17	0.11	2.58	0.54	0.05	0.06
61	5.22	0.82	0.21	0.14	2.65	0.76	0.08	0.09
63	5.21	0.89	0.34	0.27	2.70	1.01	0.10	0.17

The noise level was varied in the SRT test and the speech level was fixed at 55 dB SPL. During the baseline interval, only noise was presented. The peak and mean dilation are based on the pupil response to speech in noise. Baseline pupil diameters are raw values (mm); peak and mean dilation are calculated relative to the baseline.

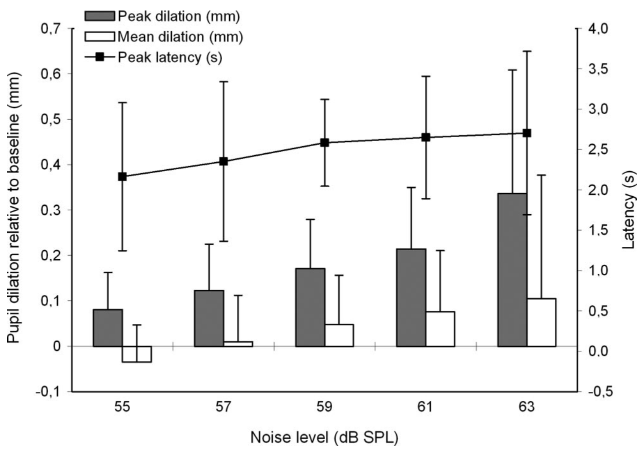

Fig. 3. Peak dilation amplitude, peak latency, and mean pupil dilation for five noise levels (SNRs). The speech level was 55 dB SPL. The peak dilation amplitude and mean pupil dilation were calculated relative to baseline. Data are based on 37 participants. Error bars equal 1 SD.

intelligibility across participants (i.e., the SRTs) or to the interindividual differences in the subjective effort and performance ratings. The absence of interindividual relations between the pupil response, the test performance, and the subjective ratings indicates that measuring the pupil response supplements these behavioral measures. Note that, on average, the differences in intelligibility performance (i.e., 50, 71, or 84%) were systematically reflected by the subjective effort and performance ratings (see Table 1).

In contrast to this study, Recarte et al. (2008) did observe significant correlations between the reported task load and pupil size during listening to a 2-min audio message, but the strength of this association decreased substantially when listeners had to perform an additional visual task. The authors suggested that the reported workload for complex (dual) tasks might reflect the total amount of allocated effort, whereas the effort reflected by the pupil response reflects the magnitude of the maximum activation in the brain areas involved in the task rather than the sum of the activated areas. Speech perception in noise involves complex processing (i.e., both processing and storage of the incoming auditory information; Pichora-Fuller et al. 1995). Although we did not vary the complexity of the task in other ways than by manipulating speech intelligibility, the current results suggest that subjective and objective measures reflect different aspects of the processing load. This is consistent with the results described by Schluroff (1983), Yeh and Wickens (1988), and Zekveld et al. (2009a,b). The subjective effort ratings and the pupil response can be influenced by several individual and task characteristics. Among these there are common factors such as age (Tomporowsi 2003; van Gerven et al. 2004), cognitive abilities and IQ (Ahern & Beatty 1979; Zekveld et al. 2009b), and fatigue (Janisse 1977). Other mechanisms, such as response bias, have a more specific influence. Furthermore, task characteristics are important (Recarte et al. 2008). The interpretation of the current results is furthermore complicated by the fact that these factors can interact or can have a different impact on subjective effort ratings and on physiologic measures of the allocated mental resources. The results indicate that persons who rate the SRT test as effortful do not tend to have a larger pupil response than persons indicating that the task was easy or vice versa. Only few studies have combined pupillometry and subjective ratings, and these studies generally did not focus on language perception. Therefore, future research is required to provide more insight into the relation between physiologic and subjective measures of mental processing load, and speech perception.

TABLE 4. Descriptive statistics of the pupil response for the correctly and incorrectly repeated sentences

	Pupil Measure					
	Peak Dilation, mm		Peak Latency, sec		Mean Dilation, mm	
	M	SD	M	SD	M	SD
Correct	0.15	0.12	2.35	0.57	0.02	0.12
Incorrect	0.24	0.08	2.97	0.80	0.09	0.07

The peak dilation amplitude, peak latency, and the mean pupil dilation systematically increased with increasing noise level (i.e., decreasing SNR of the speech in noise), in contrast to the baseline pupil diameter, which did not differ significantly between the noise levels. These results confirm the value of using pupillometry to study listening effort during speech perception in noise. As expected, listening was more effortful when the SNR of the speech in noise was lower (less intelligible speech). Moreover, with decreasing SNRs, the latency of the peak dilation amplitude increased, indicating that the time required for effortful (cognitive) processes increases with lower speech intelligibility levels. This relation between cognitive load and the peak latency has been observed in other studies as well (Beatty & Wagoner 1978; Ahern & Beatty 1979; Just & Carpenter 1993; Krüger 2000).

In contrast to the absent effect of noise level on the baseline pupil diameter as observed in this study, Antikainen and Niemi (1983) observed that higher noise levels resulted in a larger pupil response. They examined the pupil dilation as a function of the intensity of broadband noise stimuli (10-sec duration). Participants did not have to perform a task, and hence, the focus of that study was at arousal effects rather than at the influence of cognitive load on the pupillary response. Noise levels were 80, 90, and 100 dBA, covering a wider range than the noise levels presented in this study. Consistent with the results of Antikainen and Niemi (1983), Nunnally et al. (1967) observed that the pupil response to pure tones (10 secs duration) increased with increasing sound levels. The response to a tone presented at 94 dB SPL was significantly larger than the response to tones at lower levels (in steps of 10 dB), except for the response to the tone with the lowest intensity level (64 dB) that was always presented in the first condition.

In general, it is assumed that the general arousal level of the subject is reflected by the "tonic" pupil size and that phasic changes of the pupil size in response to stimuli index cognitive or emotional processes (Janisse 1977; Antikainen & Niemi 1983). An alternative interpretation of the current results is that the larger pupil response at lower intelligibility levels reflected an emotional response (e.g., increased anxiety or stress) to the higher task difficulty. Although we cannot rule out emotional processes as explanation for the current results, several factors make it more likely that the pupil response reflected cognitive processing load. First, the emotional demands of the task were low and neutral stimuli were used. Second, an emotional response to increased task difficulty would also likely influence the baseline pupil diameter. Although we did not observe an effect of noise level on the baseline diameter, we did observe a condition effect. By calculating the pupil dilation relative to the baseline pupil diameter, we controlled for such emotional or anticipatory effects (Kahneman & Beatty 1966; Stanners et al. 1979). Furthermore, Stanners et al. (1979) argued that the pupil response during language interpretation more likely reflects cognitive instead of emotional processes.

The pupil response was higher and the peak latency was longer during listening to sentences that were incorrectly repeated compared with the sentences that were correctly repeated. Janisse (1977) reviewed several studies that examined differences in pupil response associated with correct versus incorrect responses. In contrast to the current findings, in most of these studies, correct responses were associated with a larger pupil response. However, Boersma et al. (1970) and Chritchley et al. (2005) observed larger pupil dilation during incorrect trials. Janisse suggests that during correct trials, subjects exert more mental effort and thus achieve success. However, our data suggest that specifically during listening to (partly) unintelligible sentences, participants try to increase speech perception by effortful cognitive processes. In other words, the current data indicate that listening to sentences that were completely repeated correctly was relatively easy.

The balanced test order applied in this study enabled us to examine the effects of order position (both across and within tests) on the baseline pupil diameter and the pupil response. The peak dilation amplitude and the mean pupil dilation were higher in the first test than those observed in the second and third tests. This indicates that the first test was more effortful than the subsequent tests. No effects of order position on the SRTs, effort, and performance ratings were observed, indicating that the higher amount of effort allocated in the first test was not associated with better speech intelligibility. Similar results have been observed in other studies (Stanners et al. 1972; Steinhauer & Hakerem 1992; Hyönä et al. 1995). In general, the reduction in pupil response after the first tests is assumed to reflect the fact that participants become more familiar with the stimuli, procedure, pupillometric equipment, and other test characteristics, rather than reduced pupil variability.

Additionally, we observed a significant relation between the sentence number and the baseline pupil diameter in the first and third tests. Furthermore, the pupil dilation decreased during the course of the first SRT test. Before each test, participants had a short break in which they completed the effort and performance ratings for the previous test. This may have resulted in a larger baseline diameter and larger mean pupil dilation compared with the end of each test. The task may become relatively monotonous after listening to 30 to 40 sentences.

Although this study indicates that the application of pupillometry within the field of audiology may provide valuable insights into the effect of speech intelligibility on the cognitive processing load during listening, it is currently unknown to what extent and in which conditions the method could be clinically applicable. More research is required to the application of pupillometry within the field of audiology. In a future study, we intend to study the effects of the between-subject variables age and hearing acuity on the listening effort as reflected by pupil dilation during listening to speech in noise. Increasing age and hearing loss both increase the reliance on effortful cognitive processing and interact with the effects of stimulus quality and background noise (Pichora-Fuller 2007; Shinn-Cunningham & Best 2008). In addition, age has been found to be associated with the task-evoked pupil response (van Gerven et al. 2004, but also see Granholm et al. 2000). Furthermore, as described in the *Introduction* section, Kramer et al. (1997) observed that hearing-impaired individuals had a smaller reduction in the allocated effort during easier listening conditions. However, their results may have been confounded by age differences between the groups. Therefore, it is important to examine whether listening is more effortful for participants with hearing impairment than listeners with normal hearing when the actual performance level is matched between the groups, as was observed by McCoy et al. (2005) and Rakerd et al. (1996), using dual-task paradigms.

In conclusion, this study demonstrates the value of pupillometry as an objective measure of listening effort during speech perception in noise. We systematically varied speech intelligibility between conditions, and observed that the peak dilation amplitude, mean pupil dilation during sentence processing, and the peak latency all systematically increased with decreasing speech intelligibility. These results support that in difficult listening conditions, speech perception is more reliant on explicit, effortful cognitive processes such as utilization of working memory capacity. This study supports the predictions made by current language comprehension models (e.g., the ELU model; Rönnberg 2003; Rönnberg et al. 2008) and capacity theories of language comprehension (Just & Carpenter 1992). In addition, the relation between speech intelligibility and the pupil response is consistent with the results of studies that examined the influence of speech intelligibility on reaction times (Zekveld et al. 2006). Application of pupillometry to study-listening effort can yield valuable insight into the processing resources required across listening conditions (e.g., in different types of background noise or when using hearing aids). Additionally, it may provide more insight into the factors related to interindividual differences in speech perception (e.g., age, hearing loss, or working memory capacity). By examining how listeners reach a certain performance level, pupillometry provides an additional dimension to the assessment of speech perception ability besides behavioral measures. The pupillometric equipment is relatively easy to use compared with, for example, neuroimaging data, and the fact that it does not interfere with speech comprehension makes the application of this technique within the field of audiology appealing.

ACKNOWLEDGMENTS

The authors thank J. H. M. van Beek for his technical assistance in the development of the test and data analysis software and D. van der Raaij for her assistance in the data collection.

This work was supported by grants from Stichting Het Heinsius-Houbolt Fonds.

Address for correspondence: Adriana A. Zekveld, PhD, Department of ENT/Audiology and the EMGO+ Institute for Health and Care Research, VU University Medical Center, P.O. Box 7057, 1007 MB Amsterdam, The Netherlands. E-mail: aa.zekveld@vumc.nl.

Received August 10, 2009; accepted January 2, 2010.

REFERENCES

Ahern, S., & Beatty, J. (1979). Pupillary responses during information processing vary with scholastic aptitude test scores. *Science*, *205*, 1289–1292.

Antikainen, J., & Niemi, P. (1983). Neuroticism and the pupillary response to a brief exposure to noise. *Biol Psychol*, *17*, 131–135.

Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychol Bull*, *91*, 276–292.

Beatty, J., & Lucero-Wagoner (2000). The pupillary system. In J. T. Cacioppo, L. G., Tassinary, G. G., Berntson (Eds). *Handbook of Psychophysiology* (2nd ed.). New York: Cambridge University Press.

Beatty, J., & Wagoner, B. L. (1978). Pupillometric signs of brain activation vary with level of cognitive processing. *Science*, *199*, 1216–1218.

Boersma, F., Wilton, K., Barham, R., et al. (1970). Effects of arithmetic problem difficulty on pupillary dilation in normals and educable retardates. *J Exp Child Psychol*, *9*, 142–155.

Chapman, C. R., Oka, S., Bradshaw, B. H., et al. (1999). Phasic pupil dilation response to noxious stimulation in normal volunteers: Relationship to brain evoked potentials and pain report. *Psychophysiology*, *36*, 44–52.

Chritchley, H. D., Tang, J., Glaser, D., et al. (2005). Anterior cingulate activity during error and autonomic response. *Neuroimage*, *27*, 885–895.

Davis, M. H., & Johnsrude, I. S. (2003). Hierarchical processing in spoken language comprehension. *J Neurosci*, *23*, 3423–3431.

Edwards, B. (2007). The future of hearing aid technology. *Trends Amplif*, *11*, 31–46.

Fish, S. C., & Granholm, E. (2008). Easier tasks can have higher processing loads: Task difficulty and cognitive resource limitations in schizophrenia. *J Abnorm Psychol*, *117*, 355–363.

Granholm, E., Asarnow, R. F., Sarkin, A. J., et al. (1996). Pupillary responses index cognitive resource limitations. *Psychophysiology*, *33*, 457–461.

Granholm, E., Morris, S., Asarnow, R. F., et al. (2000). Accelerated age-related decline in processing resources in schizophrenia: Evidence from pupillary responses recorded during the span of apprehension task. *J Int Neuropsychol Soc*, *6*, 30–43.

Hällgren, M., Larsby, B., Lyxell, B., et al. (2005). Speech understanding in quiet and noise, with and without hearing aids. *Int J Audiol*, *44*, 574–583.

Hoeks B. (1995). *The Pupillary Response as a Measure of Processing Load.* Thesis, Katholieke Universiteit Nijmegen, The Netherlands.

Hyönä, J., Tommola, J., Alaja, A. -M. (1995). Pupil dilation as a measure of processing load in simultaneous interpretation and other language tasks. *Q J Exp Psychol*, *48A*, 598–612.

Janisse, M. P. (1977). *Pupillometry. The Psychology of the Pupillary Response.* Washington, D. C.: Hemisphere Publishing Corporation.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychol Rev*, *99*, 122–149.

Just, M. A., & Carpenter, P. A. (1993). The intensity dimension of thought: Pupillometric indices of sentence comprehension. *Can J Exp Psychol*, *47*, 310–339.

Just, M. A., Carpenter, P. A., Miyake, A. (2003). Neuroindices of cognitive workload: Neuroimaging, pupillometric and event-related potential studies of brain work. *Theor Iss Ergon Sci*, *4*, 56–88.

Kahneman, D., & Beatty, J. (1966). Pupil diameter and load on memory. *Science*, *154*, 1583–1585.

Kahneman, D., Tursky, B., Shapiro, D., et al. (1969). Pupillary, heart rate and skin resistance changes during a mental task. *J Exp Psych*, *79*, 164–167.

Kahneman, D., & Wright, P. (1971). Changes of pupil size and rehearsal strategies in a short-term memory task. *Q J Exp Psychol*, *23*, 187–196.

Kramer, S. E., Kapteyn, T. S., Festen, J. M., et al. (1997). Assessing aspect of hearing handicap by means of pupil dilation. *Audiology*, *36*, 155–164.

Kramer, S. E., Kapteyn, T., Houtgast, T. (2006). Occupational performance: Comparing normally-hearing and hearing-impaired employees using the Amsterdam Checklist for Hearing and Work. *Int J Audiol*, *45*, 503–512.

Kramer, S. E., Zekveld, A. A., Houtgast, T. (2009). Measuring cognitive factors in speech comprehension: The value of using the text reception threshold test as a visual equivalent of the SRT test. *Scand J Psychol*, *50*, 507–515.

Krüger, F. (2000). *Coding of Temporal Relations in Semantic Memory: Cognitive Load and Task-evoked Pupillary Response*. Berlin: Waxmann.

Larsby, B., Hällgren, M., Lyxell, B., et al. (2005). Cognitive performance and perceived effort in speech processing tasks: Effects of different noise backgrounds in normal-hearing and hearing-impaired subjects. *Int J Audiol*, *44*, 131–143.

Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *J Acoust Soc Am*, *49*, 467–477.

Lowenstein, O., & Loewenfeld, I. E. (1962). The pupil. In H. Davson (Ed.), *The Eye* (Vol. 3). New York: Academic Press.

McCoy, S. L., Tun, P. A., Cox, L. C., et al. (2005). Hearing loss and perceptual effort: Downstream effects on older adults' memory for speech. *Q J Exp Psychol*, *58*, 22–33.

Nachtegaal, J., Kuik, D. J., Anema, J. R., et al. (2009). Hearing status, need for recovery after work, and psychosocial work characteristics: Results from an Internet based National Survey on Hearing. *Int J Audiol*, *48*, 684–691.

Nunnally, J. C., Knott, P. D., Duchnowski, A. (1967). Pupillary response as a general measure of activation. *Percept Psychophys*, *2*, 149–155.

Obleser, J., Wise, R. J. S., Dresner, M. A., et al. (2007). Functional integration across brain regions improves speech perception under adverse listening conditions. *J Neurosci*, *27*, 2283–2289.

Pichora-Fuller, M. K. (2003). Processing speed and timing in aging adults: Psychoacoustics, speech perception, and comprehension. *Int J Audiol*, *42*, S59–S67.

Pichora-Fuller, M. K. (2006). Perceptual effort and apparent cognitive decline: Implications for audiological rehabilitation. *Semin Hear*, *27*, 284–293.

Pichora-Fuller, M. K. (2007). Audition and cognition: What audiologists need to know about listening. In C. Palmer, & R. Seewald (Eds). *Hearing Care for Adults* (pp, 71–85). Stäfa, Switzerland: Phonak.

Pichora-Fuller, M. K., Schneider, B. A., Daneman, M. (1995). How young and old adults listen to and remember speech in noise. *J Acoust Soc Am*, *97*, 593–608.

Plomp, R. (1986). A signal-to-noise ratio model for the speech-reception threshold of the hearing impaired. *J Speech Hear Res*, *29*, 146–154.

Plomp, R., & Mimpen, A. M. (1979). Improving the reliability of testing the speech reception threshold for sentences. *Audiology*, *18*, 43–52.

Poock, G. K. (1973). Information processing versus pupil diameter. *Percept Mot Skills*, *37*, 1000–1002.

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

Rabbitt, P. M. A. (1991). Mild hearing loss can cause apparent memory failures which increase with age and reduce with IQ. *Acta Otolaryngol Suppl*, *176*, 167–176.

Rakerd, B., Seitz, P., Whearty, M. (1996). Assessing the cognitive demands of speech listening for people with hearing loss. *Ear Hear*, *17*, 97–106.

Recarte, M. Á., Pérez, E., Conchillo, Á., et al. (2008). Mental workload and visual impairment: Differences between pupil, blink, and subjective rating. *Span J Psychol*, *11*, 374–385.

Richer, F., & Beatty, J. (1985). Pupillary dilations in movement preparation and execution. *Psychophysiology*, *22*, 204–207.

Rönnberg, J. (2003). Cognition in the hearing impaired and deaf as a bridge between signal and dialogue: A framework and a model. *Int J Audiol*, *42*, S68–S76.

Rönnberg, J., Rudner, M., Foo, C., et al. (2008). Cognition counts: A working memory system for ease of language understanding. *Int J Audiol*, *47*, S99–S105.

Schluroff, M. (1983). In the eye of the beholder: Cognitive effort during sentence processing. In R. Bäuerle, Ch. Schwarze, A. von Stechow (Eds). *Meaning, Use, and Interpretation of Language* (pp, 302–323). New York: de Gruyter.

Scott, S. K., & Johnsrude, I. S. (2003). The neuroanatomical and functional organization of speech perception. *Trends Neurosci*, *26*, 100–107.

Shinn-Cunningham, B. G., & Best, V. (2008). Selective attention in normal and impaired hearing. *Trends Amplif*, *12*, 283–299.

Siegle, G., Ichikawa, N., Steinhauer, S. (2008). Blink before and after you think: Blinks occur prior to and following cognitive load indexed by pupillary responses. *Psychophysiology*, *45*, 679–687.

Siegle, G. J., Steinhauer, S. R., Stenger, V. A., et al. (2003). Use of concurrent pupil dilation assessment to inform interpretation and analysis of fMRI data. *Neuroimaging*, *20*, 114–124.

Siegle, G. J., Steinhauer, S. R., Thase, M. E. (2004). Pupillary assessment and computational modeling of the Stroop task in depression. *Int J Psychophysiol*, *52*, 63–76.

Stanners, R. F., Coulter, M., Sweet, A. W., et al. (1979). The pupillary response as an indicator of arousal and cognition. *Motiv Emot*, *3*, 319–340.

Stanners, R. F., Headley, D. B., Clark, W. R. (1972). The pupillary response to sentences: Influences of listening set and deep structure. *J Verb Learning Verb Behav*, *11*, 257–263.

Steinhauer, S. R., Hakerem, G. (1992). The pupillary response in cognitive psychophysiology and schizophrenia. In D. Friedman & G. Bruder (Eds), *Psychophysiology and Experimental Psychopathology: A Tribute to Samuel Sutton* (vol. 658, pp, 182–204). New York: NY Academy of Sciences.

Stewart, R., & Wingfield, A. (2009). Hearing loss and cognitive effort in older adults' report accuracy for verbal materials. *J Am Acad Audiol*, *20*, 147–154.

Tomporowski, P. D. (2003). Performance and perceptions of workload among young and older adults: Effects of practice during cognitively demanding tasks. *Educ Gerontol*, *29*, 447–466.

Tun, P. A., Wingfield, A., Stine, E. A. L., et al. (1992). Rapid speech processing and divided attention: Processing rate versus processing resources as an explanation of age effects. *Psychol Aging*, *7*, 546–550.

van Gerven, P. W. M., Paas, F., Merriënboer, J. J. G., et al. (2004). Memory load and the cognitive pupil response in aging. *Psychophysiology*, *41*, 167–174.

Verney, S. P., Granholm, E., Dionisio, D. P. (2001). Pupillary responses and processing resources on the visual backward masking task. *Psychophysiology*, *38*, 76–83.

Verney, S. P., Granholm, E., Marshall, S. P. (2004). Pupillary responses on the visual backward masking task reflect general cognitive ability. *Int J Psychophysiol*, *52*, 23–26.

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

Wickens, C. D. (1992). *Engineering Psychology and Human Performance.* NY: HarperCollins Publishers Inc.

Wingfield, A., Tun, P. A., McCoy, S. L. (2005). Hearing loss in older adulthood. What it is and how it interacts with cognitive performance. *Curr Dir Psychol Sci*, *14*, 144–148.

Wright, P., & Kahneman, D. (1971). Evidence for alternative strategies of sentence retention. *Q J Exp Psychol*, *23*, 197–213.

Yeh, Y. -Y., & Wickens, C. D. (1988). Dissociation of performance and subjective measures of workload. *Hum Factors*, *30*, 111–120.

Zekveld, A. A., George, E. L. J., Kramer, S. E., et al. (2007). The development of the text reception threshold test: A visual analogue of the speech reception threshold test. *J Speech Lang Hear Res*, *50*, 576–584.

Zekveld, A. A., Heslenfeld, D. J., Festen, J. M., et al. (2006). Top-down and bottom-up processes in speech comprehension. *Neuroimage*, *32*, 1826–1836.

Zekveld, A. A., Kramer, S. E., Kessens, J. M., et al. (2009a). User evaluation of a communication system that automatically generates captions to improve telephone communication. *Trends Amplif*, *13*, 44–68.

Zekveld, A. A., Kramer, S. E., Kessens, J. M., et al. (2009b). The influence of age, hearing, and working memory on the speech comprehension benefit derived from an automatic speech recognizer. *Ear Hear*, *30*, 262–272.