Impact of Noise and Noise Reduction on Processing Effort: A Pupillometry Study

Dorothea Wendt,[1,2] Renskje K. Hietkamp,[1] and Thomas Lunner[1,2,3,4]

Objectives: Speech perception in adverse listening situations can be exhausting. Hearing loss particularly affects processing demands, as it requires increased effort for successful speech perception in background noise. Signal processing in hearing aids and noise reduction (NR) schemes aim to counteract the effect of noise and reduce the effort required for speech recognition in adverse listening situations. The present study examined the benefit of NR schemes, applying a combination of a digital NR and directional microphones, for reducing the processing effort during speech recognition.

Design: The effect of noise (intelligibility level) and different NR schemes on effort were evaluated by measuring the pupil dilation of listeners. In 2 different experiments, performance accuracy and peak pupil dilation (PPD) were measured in 24 listeners with hearing impairment while they performed a speech recognition task. The listeners were tested at 2 different signal to noise ratios corresponding to either the individual 50% correct (L50) or the 95% correct (L95) performance level in a 4-talker babble condition with and without the use of a NR scheme.

Results: In experiment 1, the PPD differed in response to both changes in the speech intelligibility level (L50 versus L95) and NR scheme. The PPD increased with decreasing intelligibility, indicating higher processing effort under the L50 condition compared with the L95 condition. Moreover, the PPD decreased when the NR scheme was applied, suggesting that the processing effort was reduced. In experiment 2, 2 hearing aids using different NR schemes (fast-acting and slow-acting) were compared. Processing effort changed as indicated by the PPD depending on the hearing aids and therefore on the NR scheme. Larger PPDs were measured for the slow-acting NR scheme.

Conclusions: The benefit of applying an NR scheme was demonstrated for both L50 and L95, that is, a situation at which the performance level was at a ceiling. This opens the opportunity for new means of evaluating hearing aids in situations in which traditional speech reception measures are shown not to be sensitive.

Key words: Hearing aids, Hearing impairment, Noise reduction, Processing effort, Pupillometry

INTRODUCTION

Understanding speech is probably the most important human communication ability in everyday life. People with hearing impairment have particular difficulties in processing and understanding speech under acoustically challenging conditions,

which may cause reduced speech recognition, increased cognitive demands for speech comprehension, or a slowing down of speech processing (Duquesnoy 1983; Plomp 1986; Mattys et al. 2012; Wendt et al. 2015).

Digital hearing aid (HA) technology utilizes several signal processing algorithms, such as wide dynamic range compression and noise reduction (NR), with the goal of facilitating and improving the intelligibility of speech in noise. Specifically, NR algorithms have been developed to reduce the level of the interfering noise and thus improve the effective signal to noise ratio (SNR). For instance, some research examined aggressive NR in the form of ideal binary masks (Wang et al. 2008; Kjems et al. 2009) and showed large intelligibility gains. However, the ideal binary mask requires a priori knowledge about the target and the interfering factor and thus cannot be used for practical applications. Other research combined directional microphones and binary mask reduction to create (nonideal) binary masking schemes that can be used in HAs (Boldt et al. 2008; Ng et al. 2013, 2015). To investigate speech perception in listeners with hearing impairment or to evaluate the benefit of HA signal processing, behavioral measures such as speech reception thresholds (SRTs) are commonly used (Plomp & Mimpen 1979; Nilsson et al. 1994; Hagerman & Kinnefors 1995; Akeroyd 2008). The SRT is typically estimated by applying an adaptive procedure to reach the SNR at which 50% of words are correctly identified (Hagerman & Kinnefors 1995; Brand & Kollmeier 2002). Using traditional speech-in-noise tests, the SRT lies within the range of −10 and 0 dB for listeners with mild to moderate hearing impairment, depending on the speech material and the type of background noise. However, it has been shown that some HA algorithms, such as NR schemes, are most efficient for positive SNRs (Fredelake et al. 2012; Smeds et al. 2015), where SRT measures show ceiling effects. Moreover, the literature indicates that everyday communication situations take place at positive SNRs characterizing situations with high speech intelligibility (Lunner et al. 2016; Haverkamp, Reference Note 1). For instance, Smeds et al. (2015) measured the SNRs of acoustic scenarios for HA users in a realistic environment. Only a few situations were reported in which the SNR was negative or approximately 0 dB, but in fact, the SNR was on average approximately 5 dB or higher. These studies noted that most everyday communication situations take place at positive SNRs, which differ from traditional SRT measures. Moreover, performance is at a ceiling in those situations, and SRT methods are insensitive under those circumstances. Thus, to examine speech perception in hearing-impaired listeners and to test the benefit of HA processing in a more realistic communication situation (i.e., at ecologic SNRs), alternative methods and measures are required.

Even when speech intelligibility is high, people with hearing impairment experience considerable difficulties after conversations in everyday life situations. One reason is that

[1]Eriksholm Research Centre, Snekkersten, Denmark; [2]Hearing Systems, Hearing Systems Group, Department of Electrical Engineering, Technical University of Denmark, Kongens Lyngby, Denmark; [3]Division of Technical Audiology, Department of Clinical and Experimental Medicine, Linköping University, Linköping, Sweden; and [4]Department of Behavioral Sciences and Learning, Linköping University, Linköping, Sweden.

hearing-impaired listeners expend extra processing effort to perceive and process speech (McCoy et al. 2005). Processing effort is a measure of the amount of cognitive resources deployed when processing speech. Processing effort depends on the interplay of two factors. On the one hand, it is affected by the processing demands imposed by the listening situation and the task. Processing demands are strongly dependent on stimulus-related factors, such as degraded speech or background noise. The type of background noise further affects processing demands. On the other hand, processing effort is dependent on factors related to the individual listener, such as hearing loss or cognitive abilities (Mattys et al. 2012) and the amount of cognitive resources the listener employs in a (speech recognition) task to compensate for those demands (Rabbitt 1990; Hick & Tharpe 2002; Johnsrude & Rodd 2015). A person's efforts to recognize speech in background noise have been measured with various methods and techniques (McGarrigle et al. 2014 and Ohlenforst et al. 2017 for a review). Self-reported effort has been studied using self-assessment scales and/or questionnaires (Humes 1999; Nachtegaal et al. 2009). Those measures give insight into how a listener perceives his or her effort in a specific listening situation. It has been shown, for instance, that perceived effort due to hearing loss can have various effects on the individual, such as increased susceptibility to fatigue (Hornsby 2013) or increased days of sick leave (Kramer et al. 2006). However, subjective measures are limited since people may differ in their interpretation of effort or have difficulties rating their perceived effort. Furthermore, scales and questionnaires are filled out "after" a task is performed, which makes it hard to monitor the perceived effort "while" performing the task. In contrast, physiologic measures have been used to investigate changes in the activity of the central and autonomic nervous system during speech processing. For instance, changes in pupil dilation have been suggested as an index of locus coeruleus function (Aston-Jones & Cohen 2005). The pupil dilates with increasing demands until processing resources are exceeded (Kahnemann & Beatty 1966; Zekveld et al. 2010, 2011). It is assumed that task-related pupil response reliably reflects changes in cognitive resources allocated by the listener. Thus, if processing efforts increase in speech recognition due to an acoustically challenging situation, this should be reflected by increased pupil dilation (Janisse 1977; Beatty 1982). Several studies have examined the processing effort involved in perceiving speech in background noise (Kramer et al. 1997; Piquado et al. 2010; Zekveld et al. 2010). More recent literature studied the processing effort involved in speech recognition in cases of hearing impairment (Anderson Gosselin & Gagné 2010, 2011; Picou et al. 2013; Koelewijn et al. 2014). Zekveld et al. (2011) investigated the effect of hearing loss, age, and speech intelligibility on effort, as indicated by the pupil dilation. They found less release from effort with increasing speech intelligibility for hearing-impaired people compared with people with normal hearing. Wendt et al. (2015) tested the effect of hearing loss on the duration of sentence processing during an audiovisual task paradigm. To ensure that each participant had roughly the same spectral information available, the spectrum of the noisy speech was adjusted according to the individual hearing loss. By analyzing the participant's eye movements and calculating the speech processing durations, a significant increase in duration due to hearing impairment was reported, even in situations with high speech intelligibility. Interestingly, hearing-impaired

participants who were experienced HA users showed smaller speech processing durations than hearing-impaired participants without HA experience. Furthermore, their speech processing durations were similar to those of the normal-hearing group (Wendt et al. 2015). These findings indicate that experienced HA users benefit from a frequency-specific gain rule, which is commonly used in HAs.

Within recent years, a growing body of research has examined the benefits of HAs and signal processing algorithms on cognitive aspects of speech perception, particularly memory processing and processing effort (Gatehouse & Gordon 1990; Sarampalis et al. 2009; Brons et al. 2013; Picou et al. 2013). Some studies indicated that although HA processing did not result in a significant improvement in speech intelligibility, HA users may still express a preference for certain algorithms or show reduced effort and improved memory performance (Picou et al. 2013; Brons et al. 2013; Ng et al. 2013, 2015; Neher 2014). Brons et al. (2013, 2015) studied the effect of different NR schemes on perceived effort in listeners with normal hearing and those with hearing impairment. They compared the participants' ratings of their effort while listening to speech in babble noise that was processed by one of four HAs. Small but significant differences in perceived effort were reported depending on the NR scheme. Interestingly, no differences in perceived effort were noted when the NR was on versus off. In general, there is growing interest in the concept of listening effort and its relationship with hearing impairment and HA signal processing. However, there are uncertainties and ongoing discussions regarding the benefit of HA signal processing for reducing effort (see Ohlenforst et al. 2017 for a review).

Recent literature has demonstrated that not only hearing impairment but other listener-related abilities, such as working memory, may affect individual speech reception performances and processing effort (Lunner 2003; Akeroyd 2008; Rönnberg et al. 2013; Wendt et al. 2016). Ng et al. (2013) indicated that good cognitive abilities are associated with greater benefit from signal processing. They examined the effects of NR on memory performance of hearing-impaired listeners and reported significantly better memory performance when an NR algorithm was applied. However, this effect was restricted to people with good working memory capacity. In a later study, Ng et al. (2015) again reported that NR had beneficial effects on memory performance; however, this time, the benefit was not associated with the individual's working memory capacity.

The objective of the present study was to evaluate the effects of noise and NR schemes on processing effort in people with hearing impairment and correlate these effects with the individual's working memory capacity. Processing effort was investigated by measuring changes in pupil dilation in a speech recognition task. The NR scheme included directional microphones and a binary mask reduction to create (nonideal) binary masking schemes (Boldt et al. 2008; Ng et al. 2013, 2015). Two different experiments were conducted. In experiment 1, the pupil dilation of each participant was measured at 2 different intelligibility levels corresponding to either the individual's 50% speech recognition (L50) or 95% speech recognition (L95) threshold. The L95 condition was introduced to assess a ceiling for speech recognition performance at which differences in effort as a result of NR processing can still be expected. The effect of the NR system was tested for both intelligibility levels (L50 and L95). The effect of individual differences in cognitive

ability on processing effort and HA processing was further examined. It was hypothesized that:

- Speech intelligibility has an effect on processing effort such that effort is increased at L50 compared with L95. Increased effort is indicated by a significant increase in pupil dilation (according to Zekveld & Kramer 2014).
- By applying an NR scheme (including directional microphone use and NR), effort can be significantly reduced for people with hearing impairment, as indicated by a significant decrease in pupil dilation.
- A benefit of the NR scheme on effort can be measured at ecologic SNRs, when speech recognition performance is at its ceiling.
- A greater benefit of the NR scheme for people with better cognitive abilities is expected. Hearing-impaired participants with good working memory capacity will benefit most from NR in terms of the effort involved in speech recognition (Lunner 2003; Ng et al. 2013).

The objective of experiment 2 was to examine the effect of NR schemes on effort using 2 commercially available HAs. For the one HA (HA1), the NR scheme relied on a multi-microphone noise estimate, an adaptive minimum-variance distortionless response (MVDR) beamformer combined with a postfilter that produces fast-acting NR (Kjems & Jensen 2012; Jensen & Pedersen 2015). For the other HA (HA2), the NR scheme relied on a single-channel noise estimate, a first-order directionality effect and slow-acting NR. While directionality effects, such as those used in HA1 and HA2, are known to improve speech understanding, slow-acting NR, such as those used in HA2, does not provide such benefits and is often considered a comfort feature of modern HAs (Bentler et al. 2008). The NR scheme employed in HA1 used a more efficient directionality effect that aims to minimize the noise variance and postfilter-based NR that better approximates the effect of a NR based on an ideal binary masker. Ideal binary masker NR systems require a priori knowledge of the noise and are therefore unrealistic for use in HAs, but they have been shown to reduce the negative effect of noise on memory processing for people with normal hearing (Sarampalis et al. 2009) and those with hearing loss (Ng et al. 2013). It was therefore hypothesized that the NR strategies employed in HA1 provide benefits not only in terms of speech understanding but also in terms of cognitive processing and processing effort.

EXPERIMENT 1

Materials and Methods

In experiment 1, the effect of an NR scheme (inspired by Wang et al. 2008; Boldt et al. 2008; Kjems & Jensen 2012; Jensen & Pedersen 2015) on processing effort was tested using pupillometry during a speech recognition task. The participants were asked to listen to and repeat back Danish sentences played in 4-talker babble. The effect of NR on effort was investigated at 2 different SNRs corresponding to 2 different individual performance levels.

Participants • Twenty-four hearing-impaired listeners with an average age of 59 years (ranging from 35 to 80 years) were included in the experiment. The participants were native speakers of Danish and had a symmetrical sensorineural hearing loss (Fig. 1). Their pure-tone average from 500 to 4000 Hz ranged
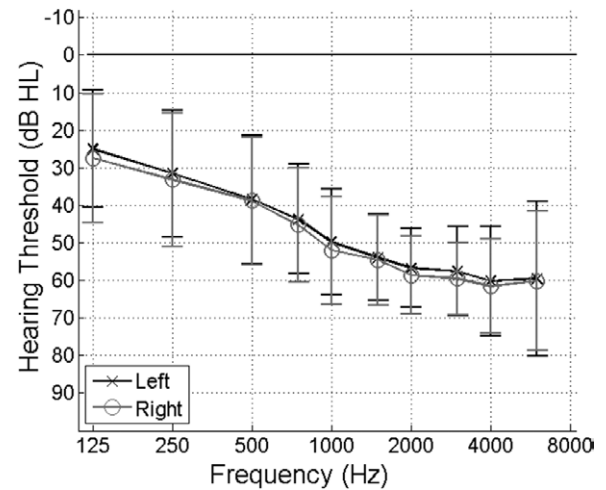

Fig 1. Mean audiogram of the 24 listeners (left and right ear), including error bars (±1 STD).

from 34 to 70 dB HL with an average of 47 dB HL; the averaged maximum difference between the left and right ear from 125 to 6000 Hz was 15 dB. The participants had no history of eye diseases or eye operations. They were all habitual binaural HA users with at least 1 year of experience (ranging from 1.1 to 13.7 years). The experiment was carried out without the use of glasses or contact lenses. Ethical approval for the study was obtained from the Research Ethics Committees of the Capital Region of Denmark.

Speech Material and Noise Conditions • In a spatial setup of 5 loudspeakers, Danish sentences from the Hearing In Noise Test (HINT) (Nielsen & Dau 2011) were presented in 4-talker babble created by 4 overlapping talkers. To construct the 4-talker masker of continuous speech, 4 single audio files were created (2 male and 2 female nonprofessional speakers reading text from a newspaper). All the audio files had the same long-term average frequency spectrum as the Danish HINT sentences. Speech pauses longer than 0.05 seconds were removed.

For each trial, a random mixture of the 4 speech audio files was created. A single trial was defined as the duration of the presentation of the 4-talker babble that started 3 seconds before the onset of the HINT sentences and ended 3 seconds after sentence offset. The HINT sentences were presented from a loudspeaker positioned in front of the listener (at 0°). The 4-talker masker was presented from the side/back of the participants. This was realized by presenting each competing talker spatially via one of the four loudspeakers with a distance of 1.2 m to the listener's side or back (at ±90° and ±150°, Fig. 2). The position of the 4 competing talkers was randomized across conditions. One male speaker and 1 female speaker were always positioned at the ±90 azimuth position. Thus, the effect of a competing speaker with the same gender position was balanced across all conditions.

NR Scheme • The participants were tested while wearing HAs under 2 different conditions. In the first condition, no NR scheme was applied, and only the amplification using Voice Aligned Compression (VAC) (Le Goff 2015) was used. In this condition, called the NoNR condition, the HAs provided quasi-linear amplification according to each participant's hearing thresholds based on the VAC rationale to assure audibility. The VAC approach falls within the family of curvilinear wide
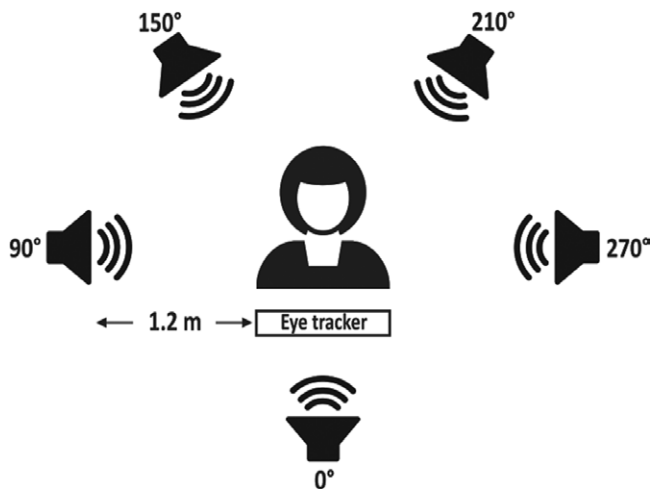
Fig. 2. Spatial setup of the loudspeakers in experiments 1 and 2. Four loudspeakers were positioned 1.20 m from the listeners to their sides and at their back, that is, at ±90° and ±150°, to present the competing talkers. The target loudspeaker and the eye tracker (with a distance of 0.6 m) were positioned in the front at 0°.

dynamic range compression. Compared with many other amplification strategies, the VAC rationale provides less compression at high input levels and more compression at low input levels through lower compression kneepoints (varying between 30 and 40 dB SPL depending on the frequency region and amount of hearing loss). This compression model is based partly on loudness data presented by Buus and Florentine (2001) and is intended to ensure improved sound quality without the loss of speech intelligibility rather than loudness compensation per se.

In the second condition, the NR condition, a NR scheme was applied in 2 different processing blocks. In the first block, the 2 microphone signals were combined via 3 fixed beamformers to create enhanced omnidirectional and rear cardioid signals. In the second block, a 2-channel MVDR beamformer was applied (Kjems & Jensen 2012) to use spatial filtering to attenuate interfering signals that did not come from in front of the listener, where the target was located. Afterwards, the signal was postprocessed using a single-channel postfilter (Jensen & Pedersen 2015) to further remove interfering noise.

Estimation of L50 and L95 • To ensure comparable speech intelligibility levels, the SNRs for 50% speech recognition (L50) and 95% speech recognition (L95) were measured for each participant. The individual L50s and L95s were estimated using correct-word scoring for words presented in 4-talker babble. The participants were tested using HAs without NR (i.e., in the NoNR condition). To obtain the L50, an adaptive procedure was applied (Brand & Kollmeier 2002); after a correct response (5 words), the SNR was decreased by 2 dB, and after an incorrect response (0 words), the SNR was increased by 2 dB. The step size for 1 to 4 correct words was relative to the maximum step size, for example, 2 correct words at L50 resulted in a 0.8 dB decrease in SNR. However, for the first 5 sentences, the step size was doubled. To estimate the L95, the SNR at 80% correct (henceforth referred to as SRT80) was measured first with an adaptive procedure (Levitt 1971), with a 3.2 dB increase in SNR after an incorrect response and a 0.8 dB decrease in SNR after a correct response. Again, the step size for 1 to 4 correct words was relative to the maximum

step size, for example, 2 correct words resulted in a 2.4 dB increase in SNR. For L95, the step size was also doubled for the first 5 sentences. From the SRT80, the L95 was estimated by fitting a psychometric function to the data. The masking onset was 3 seconds before the onset of each sentence and continued for 3 seconds after the sentence offset. Therefore, the length of each trial varied depending on the length of the presented HINT sentence, which had a mean duration of 1.5 seconds. After noise offset, participants were asked to repeat back the sentence. At the beginning of the session, each participant performed 3 training lists consisting of 20 sentences each. The first list was presented to familiarize the participant with the procedure. Afterwards, the participants performed 2 more test lists for the estimation of L50 and L95. The average L50 was 1.3 dB SNR (±2.3), and the average L95 was 7.1 dB SNR (±2.3) for all participants. After training, the participants completed 4 test lists: 2 without NR scheme (NoNR at L50 and L95) and 2 with active NR scheme (NR at L50 and L95). Each test list contained 25 sentences. The order of list presentation was randomized for each participant using a Latin square design. The participants were wearing HAs throughout the test procedure (during both training and testing). One participant was unable to complete all 4 conditions and was excluded from further data analysis. While the participants were performing the speech recognition task, an eye-tracking camera recorded their pupil dilation.

Reading Span (RS) Test • The RS test (originally developed by Daneman & Carpenter 1980) measures working memory capacity. A modified version of the working memory test that taxes memory storage and processing simultaneously was applied in this study (developed by Rönnberg et al. 1989). The participants' task was to listen to and comprehend a sequence of sentences. Half of the sentences were semantically incorrect (e.g., "The train sang a song"), whereas the other half were semantically correct (e.g., "The girl brushed her teeth"). The participants were asked to indicate verbally whether the sentence was meaningful after each sentence (within 1.75 seconds after sentence offset). After a sequence of sentences, the participants were asked to recall either the first or the final word of each sentence, as indicated by the word "First" or "Final" presented on the monitor. The first or the final word was requested in a randomized order. Sets of 3, 4, 5, and 6 sentences were presented in ascending order and repeated 3 times. The maximum possible score was 54 correctly recalled words. The RS scores were calculated for each participant as the percentage of the maximum number of recalled words.

Apparatus and Spatial Setup

An eye-tracker system (iView X RED System; Senso-Motoric Instruments, Teltow, Germany) was used to record the participants' pupil dilation. The sampling rate was 120 Hz throughout the experiment. An infrared eye camera with an automatic eye and head tracker was placed in front of the listener to measure both eyes remotely, that is, without contact. The presentation of the stimuli was controlled by a PC using MATLAB-based programming (MathWorks, Natick, MA). Signals were routed through a sound card (RME Hammerfall DSB multiface II; Audio AG, Haimhausen, Germany). Auditory signals were then played back via speakers (Genelec 8030B; Genelec Oy, Iisalmi, Finland). The experiment was

conducted in a double-walled, sound-treated IAC Acoustics booth. The participants were seated 60 cm from the eye tracker. During each trial, pupil size and pupil *x* and *y* traces of both eyes were recorded to detect horizontal and vertical eye movements, respectively. Only the pupil size of the left eye was used for further analysis (see description about the pupil data analysis below).

Pupil Data Analysis

Pupil data from the first 5 trials at the beginning of each list were excluded from further analysis. For all the remaining sentences, the averaged pupil diameter for each participant and each condition was calculated as follows: first, diameter values more than 3 SDs below the mean pupil diameter were coded as eye blinks or movements. Trials for which more than 20% of the data consisted of blinks and movements were excluded from further analysis. Following the application of this criterion, not more than 3% of all trials (across all participants) were removed, which was on average less than 1 trial per condition. For the remaining trials, blinks were removed using a linear interpolation that started 5 samples before and ended 8 samples after the blinks. A 5-point moving average smoothing filter was passed over the deblinked trials to remove any high-frequency artifacts. For 1 participant, more than 50% of the trials required interpolation; therefore, this participant was excluded from further data analysis (Siegle et al. 2003). All remaining traces were baseline corrected by subtracting the baseline value. This value was estimated using the mean pupil size within the 1 second before the onset of the sentence where the participant listened to the noise alone (Fig. 3). The pupil responses were averaged across all remaining trials for each condition. The peak pupil dilation (PPD) was calculated for each participant and each condition (NoNR L50; NR L50; NoNR L95; NR L95). The PPD was defined as the maximum pupil dilation within the time interval between the sentence onset and the noise offset (Fig. 3).

Results Experiment 1

To analyze the effect of intelligibility level and NR scheme, 2 separate repeated-measures analyses of variance (ANOVA) were performed. One ANOVA was conducted for the speech recognition performances; the other used the PPD data. To examine whether cognitive abilities were related to individual processing effort, nonparametric Spearman correlation coefficients were calculated for the RS performance and the PPDs. The coefficients were calculated separately for each of the 4 conditions (2 intelligibility levels × NR on versus off).

Speech Recognition Performance • Figure 4 shows the mean response accuracy across participants for the speech recognition task. In general, the participants' speech recognition performance was very high; therefore, recognition rates were transformed to rationalized arcsine transformed [rationalized arcsine transform units (rau)] scores (Studebaker 1985). The highest accuracy was measured for the L95 conditions (between 104.5 and 117.3 rau). For L50, the recognition performance was between 65.7 rau (NoNR) and 101.0 rau (NR). Interestingly, the recognition performance under the NoNR L50 condition was quite high. The performance on the speech
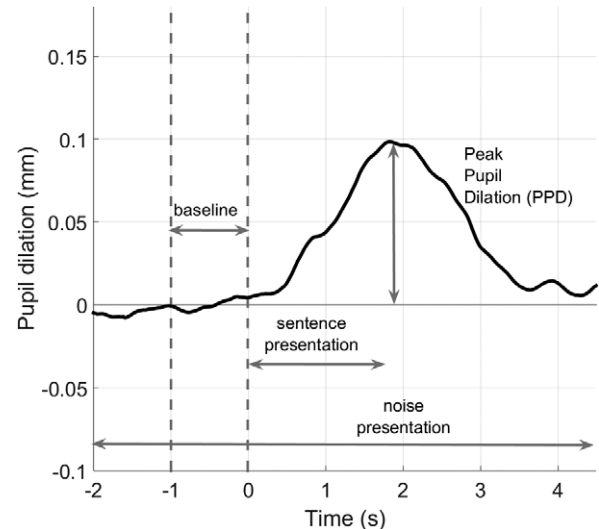

Fig. 3. Example of a normalized pupil curve averaged across all listeners. Pupil size was normalized according to the baseline. The baseline value was estimated using the time-averaged pupil size 1 second before the sentence presentation.

recognition task was analyzed using an ANOVA with intelligibility level (L50, L95) and NR scheme (NoNR, NR) as within-subject factors. The ANOVA revealed a main effect of intelligibility level [$F_{(1,22)} = 147.2$, $p < 0.001$, $\omega = 0.87$] indicating significant improvement in speech recognition at L95. In addition, an NR effect was measured [$F_{(1,22)} = 94.1$, $p < 0.001$, $\omega = 0.81$], indicating significantly higher performances under the NR conditions. Moreover, an interaction between intelligibility level and NR scheme was found [$F_{(1,22)} = 48.7$, $p < 0.001$, $\omega = 0.69$]. Post hoc analysis revealed differences in recognition rates between NoNR and NR in the L50 condition ($p < 0.001$). However, no difference in performances was found between NoNR and NR in the L95 condition ($p = 0.07$).

Peak Pupil Dilation • The PPD was calculated over the remaining trials for each condition. The PPDs are plotted in Figure 5 for all 4 test conditions. The effect of intelligibility level and NR on
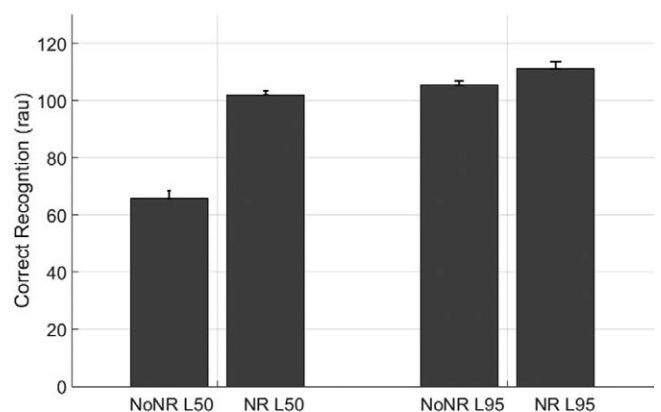

Fig. 4. Correct recognized words (in rationalized arcsine transform units) on the speech recognition task for all 4 conditions, that is, NoNR L50, NR L50, NoNR L95, and NR L95. NoNR L50 indicates without noise reduction at L50; NoNR L95, without noise reduction at L95; NR L50, with noise reduction at L50; NR L95, with noise reduction at L95. Error bars show standard errors.
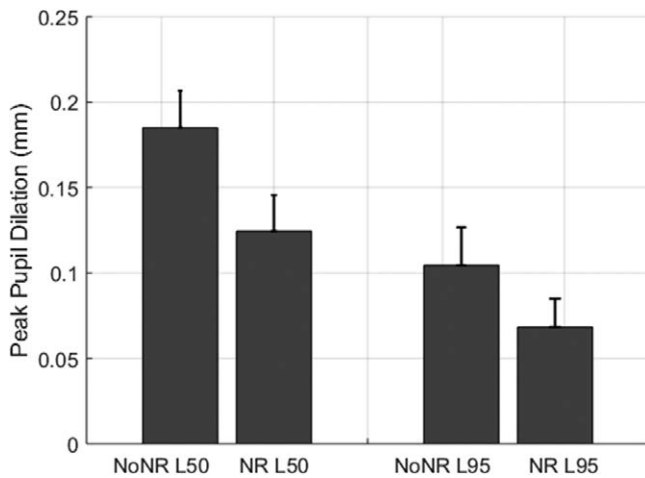
Fig. 5. Peak pupil dilation (mm) averaged across all participants. PPD is shown for 4 different conditions. Error bars show standard errors.

PPD was analyzed by conducting an ANOVA with intelligibility level (L50, L95) and NR scheme (NoNR, NR) as within-subject factors. The ANOVA revealed a main effect of intelligibility level [$F_{(1,21)} = 26.1$, $p < 0.001$, $\omega = 0.58$] indicating greater PPD at L50. An effect of the NR scheme on pupil dilation was found [$F_{(1,21)} = 16.6$, $p = 0.001$, $\omega = 0.48$], indicating significantly reduced PPD for the NR condition. Moreover, a small but significant interaction effect was measured [$F_{(1,21)} = 4.9$, $p = 0.04$, $\omega = 0.2$]. A paired t test revealed differences between NoNR and NR at L50 ($t = 5.7$, $p < 0.001$) and L95 ($t = 2.2$, $p < 0.036$). No significant differences in the baseline value were found among all 4 conditions.

RS Data • The RS test was performed to measure the participants' working memory capacity. The average test result was 42% (STD = 8.8%). This is in line with Lunner (2003) and Petersen et al. (2016). Petersen et al. reported a median RS value of 42.6% for a group of 283 participants 27 to 87

years of age. According to Ng et al. (2013, 2015), NR can reduce the adverse effect of noise on memory performance for people with good working memory performances. Thus, the beneficial effect of NR scheme on processing effort, as indicated by smaller PPD, was expected to be particularly strong for people with good RS performances. The Spearman rank correlation coefficients between the RS scores and the PPD in each of the 4 conditions showed small but significant negative correlations in the NR L50 condition ($r = -0.37$, $p = 0.043$) and the NR L95 condition ($r = -0.4$, $p = 0.027$). That is, higher (better) RS scores were associated with lower PPD. No statistically significant associations were observed for the conditions without the NR scheme, that is, NoNR L50 ($r = -0.02$) and NoNRL 95 ($r = -0.007$). These data may suggest that the PPD was reduced for the participants with good working memory capacity when the NR scheme was applied. However, the correlation coefficients were rather small (between $r = -0.3$ and -0.4, see Fig. 6).

EXPERIMENT 2

Materials and Methods

In experiment 1, the contrast between NR on versus off was tested with signal processing in a research setting. In other words, this setting is not automatically prescribed to patients—although clinicians can prescribe it if necessary—but it was used for research purposes. The objective of experiment 2 was to compare 2 different NR schemes (including directional microphone use and NR) that are used in commercially available HAs. For that purpose, 2 HAs were tested that used different NR schemes with different automatic control (see the following sections describing the NR scheme in detail). The first HA (HA1) had properties similar to the NR condition in experiment 1. Experiment 2 was conducted with the same participants and followed the same procedure as experiment 1.
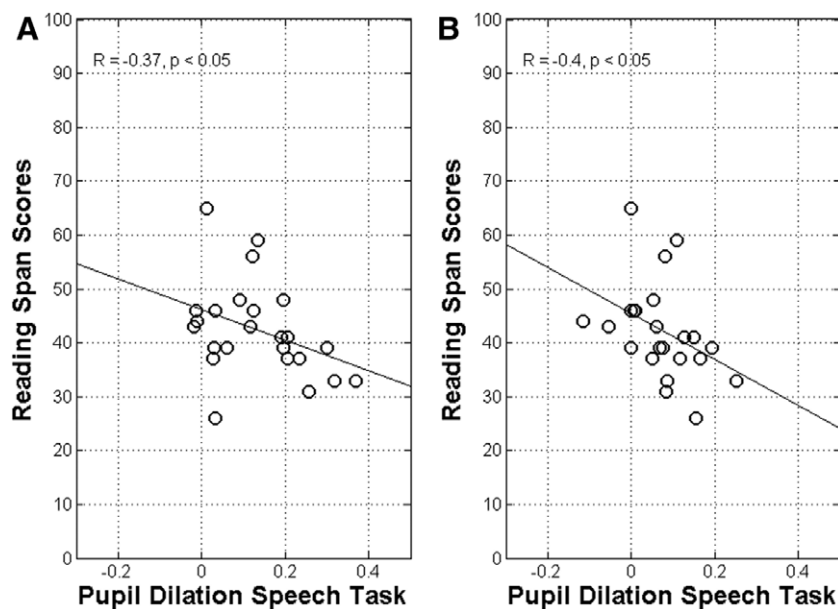

Fig. 6. Reading span scores as a function of the peak pupil dilation for NR L50 condition (A) and the NR L95 condition (B). NR L50 indicates with noise reduction at L50; NR L95, which noise reduction at L95.

Noise Conditions • Danish HINT sentences were presented in 4-talker babble (same talkers as in experiment 1) over a spatial loudspeaker setup. The 4-talker babble was presented via 4 different loudspeakers positioned at ±90° and ±150° (Fig. 2). Unmodulated speech-shaped noise (SSN) was added to the 4-talker babble to simulate a diffuse noise environment and to trigger the automatic control of the NR algorithms. The SSN was added to the 2 competing talkers presented from the back at ±150° with an SNR of −1.8 dB. The overall SNR of the 4-talker babble and the SSN was 4 dB.

NR Schemes • The NR schemes used in HA1 and HA2 differed considerably. HA2 was an Oticon Alta 2 Pro instrument. Its NR scheme uses a single microphone noise estimate and consists of an adaptive first-order directionality system and a slow-acting NR system. HA1 was an Oticon Opn instrument. Its NR system uses a multi-microphone noise estimate and consists of an adaptive MVDR beamformer (Kjems & Jensen 2012) combined with a postfilter that provides fast-acting NR (Jensen & Pedersen 2015). Here, the multi-microphone noise estimator uses an adaptive beamformer to create a back-facing cardioid response that serves as a noise estimator for both the MVDR beamforming action and the postfilter.

Procedure • The pupillometry paradigm was administered at the SNRs corresponding to the participant's 95% correct speech recognition (L95), as in experiment 1. In the paradigm, the noise masker started 6 seconds before the onset of each sentence and continued for 3 seconds after speech offset. Therefore, the length of each trial varied depending on the length of the presented sentence, which had a mean duration of 1.5 seconds. The 6 seconds of noise before the sentence onset was applied to allow the automatic control of the NR algorithm to stabilize. After the noise offset, the participants were asked to repeat the sentence. Two different HAs were tested (HA1 and HA2). The participants completed 2 test lists of 25 sentences each, one using HA1 and the other using HA2.

Pupil Data Analysis

The pupil data analysis method was similar to that used in experiment 1. The first 5 trials were removed from further analysis. For all remaining sentences, the averaged pupil diameter was calculated. The pupil data were normalized by subtracting a baseline value, defined as the mean pupil size during the 1 second before the sentence onset. The PPD was calculated during the interval between the sentence onset and the noise offset (Fig. 3).

Results Experiment 2

To analyze the differences between the 2 NR schemes in terms of performance level and pupil size, 2 separate paired t tests were performed, one for the recognition rates and the other for the PPD data.

Speech Recognition Performance • Figure 7 shows the mean response accuracy across participants for both NR schemes (HA1: 117,3 rau; HA2: 111,9 rau). The t test revealed small but significant differences in performance between the 2 conditions ($t = 2.4$, $p = 0.03$, $\omega = 0.2$), indicating higher response accuracy with HA1 (Fig. 7).

Peak Pupil Dilation • A t test was conducted to compare the PPD with HA1 and HA2. The t test revealed significant
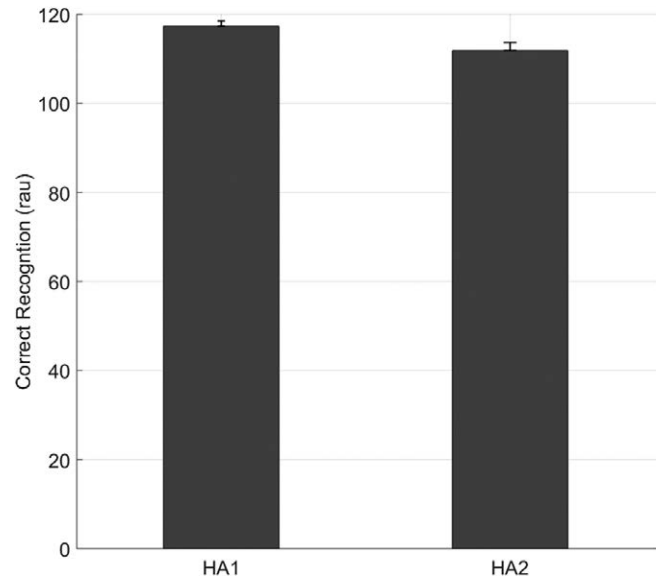

Fig. 7. Correct recognized words (in rationalized arcsine transform units) in the speech recognition task under both conditions, that is, with HA1 and HA2. Error bars show standard errors. HA indicates hearing aid.

differences between the PPDs ($t = 2.2$, $p = 0.04$, $\omega = 0.2$), indicating significant larger PPDs with HA2 (PPD = 0.093 mm) compared with HA1 (PPD = 0.069 mm). In general, these results indicate that the PPD and, thus, the processing effort were significantly reduced with the use of HA1 (Fig. 8).

DISCUSSION

This study investigated the effect of intelligibility level and NR schemes on processing effort as indicated by the PPD in a group of people with hearing impairment. Our results from experiment 1 indicated that processing effort and recognition performance were affected by both intelligibility level (L50 versus L95) and NR scheme (NoNR versus NR). Increased PPD was found for the L50 compared with the L95 condition, suggesting increased processing effort in the L50 condition. When applying an NR scheme, processing effort was reduced as indicated by significant smaller PPDs. To the best of the authors' knowledge, this is the first study to demonstrate that NR processing has a beneficial effect on effort, as indicated by pupil dilation, in hearing-impaired listeners. Furthermore, a beneficial effect of NR processing on speech recognition performance was demonstrated in situations with high and positive SNRs (average SNR in the L95 condition was +7 dB ± 2 dB), which is in line with the ecologic SNRs reported by Smeds et al. (2015). Experiment 2 showed that in those situations reflecting realistic SNRs, effort can also change as a result of a particular NR scheme of the HA.

An effect of speech intelligibility level on processing effort has been shown previously. For instance, Zekveld et al. (2010, 2011) investigated the influence of SNR and speech intelligibility on effort. The authors reported that effort increased with deceasing intelligibility level. This is in line with the results of the present study. Significant reductions in PPD were measured when the speech recognition performance increased. These results support the idea that when the quality of auditory input is reduced either by hearing impairment or an adverse
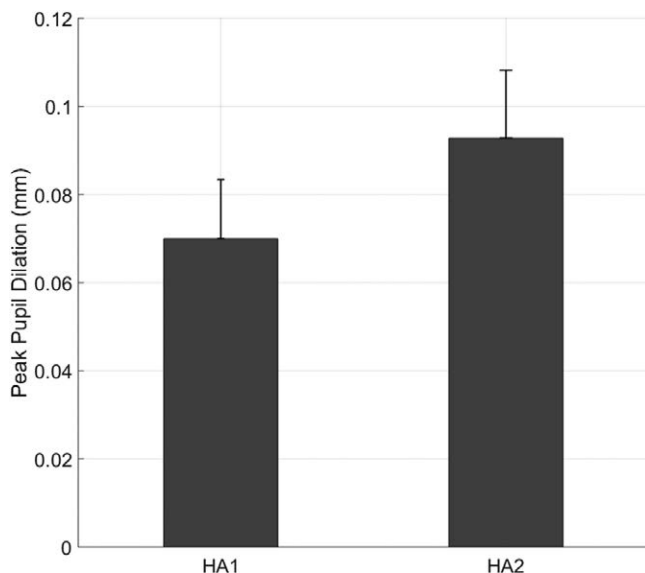
Fig. 8. Peak pupil dilation averaged across all participants. Error bars show standard errors.

acoustic environment, listeners may allocate more cognitive resources to process speech. The utilization of greater cognitive resources will then lead to higher effort requirements for processing suboptimal and degraded speech signals. This is predicted by theories regarding the ease of language comprehension, such as the Ease of Language Understanding (ELU) model (Rönnberg 2003; Rönnberg et al. 2013), or by capacity theories of language comprehension (Just & Carpenter 1992). In a consensus article, Pichora-Fuller et al. (2016) present a Framework for Understanding Effortful Listening (FUEL) for understanding the interplay of cognitive demands, motivation, and processing effort. The FUEL is an adaptation of the classic model by Kahneman (1973), and it suggests that processing effort is modulated primarily by 2 factors: the cognitive demands imposed by the task and the motivation of the individual. In the present study, participant's motivation is assumed to be constant; however, task demands were varied across conditions. When task demands were decreased, due to increased speech intelligibility, reduced processing effort was found.

Whereas lower speech intelligibility negatively affected processing effort, our results indicate that NR schemes have a beneficial effect on processing effort. Significantly reduced PPDs were measured with NR processing on. Most interesting, the effect of NR processing on PPD was shown in the L95 condition in experiment 1. Even when speech recognition was at almost 100% and no significant differences in the recognition performance occurred, the effort was reduced when the NR was applied. This is in line with literature demonstrating a benefit of HA signal processing. Picou et al. (2013) tested the effect of HA processing and background noise on listening effort in hearing-impaired listeners. Effort was examined using a dual-task paradigm in which participants had to perform a primary task (speech recognition) and a secondary task (visual task) simultaneously. An effect of background noise and HA processing on effort was demonstrated by changes in the reaction time in a secondary task. Picou et al. concluded that background noise increased effort, while HA processing reduced the processing effort in hearing-impaired listeners. Similarly,

Sarampalis et al. (2009) showed that the Ephraim–Malah algorithm (Ephraim & Malah 1984, 1985) can reduce cognitive effort related to speech processing. In a dual-task paradigm, it was demonstrated that reaction times (measured in a secondary task) significantly decreased when recognizing speech with the NR algorithm (primary task), suggesting reduced effort. However, the benefit of the NR algorithm was only demonstrated for participants with normal hearing and at a negative SNR. Despite the findings of a few recent studies (Sarampalis et al. 2009; Picou et al. 2013), the effect of HA processing on effort is still strongly debated in literature. Ohlenforst et al. (2017) undertook a systematic review to find evidence of an effect of hearing impairment and HA amplification on processing effort. Literature was reviewed with regard to studies applying different methodologies, including self-report, behavioral, and physiologic measures, to examine if and how HA amplification impacts processing effort. Although several studies indicated a change in processing effort associated with HA amplification (most of those studies using the self-report or behavioral measures), Ohlenforst et al. drew the conclusion that the existing evidence for reduced effort due to HA amplification was not significant. According to the authors, the absence of an effect might be due to a great diversity of tests within each measurement type (subjective/self-report, behavioral, and physiological).

In the present study, the benefit of the NR was found at more ecologic SNRs (approximately 7 dB). According to Smeds et al. (2015), this SNR range reflects acoustic scenarios in everyday conversation for HA users. Other studies also indicate that signal processing has a beneficial effect on cognitive measures at ecologic SNRs (Ng et al. 2013, 2015; Lunner et al. 2016). For instance, Ng et al. (2013, 2015) introduced a memory test, called Sentence final Word Identification and Recall (SWIR) test, to examine the impact of an NR algorithm on memory performance in ecologically valid listening situations. They demonstrated that the performance in memory can be improved when applying a NR processing. Ng et al. (2013) further reported that participants with good cognitive abilities benefit most from the NR algorithm. Thus, the impact of working memory capacity on the benefit of a NR scheme was examined in the present study. Only a small negative correlation between the PPD under the NR conditions and working memory capacity was found. In other words, the participants with better working memory capacity tended to have smaller PPDs. Although all the participants performed within the expected range of standard RS scores (according to Petersen et al. 2016; Lunner 2003), the participants with higher scores, suggesting higher working memory capacity, tended to have lower PPDs compared with the participants with lower RS scores. Interestingly, significant correlations were only measured for the conditions in which the NR algorithm was applied. These results suggest that greater working memory capacity may help to reduce the effort involved in speech perception. This is in line with the findings by Ng et al. (2015) and the idea that better cognitive abilities, such as working memory capacity, can actually help to reduce cognitive demands involved in processing speech in aided conditions. Furthermore, Souza et al. (2015) suggested that cognitively low-performing hearing-impaired participants may be more susceptible to signal processing artifacts than cognitively high-performing participants. Hence, the correlation found in our study may suggest that the participants

with smaller working memory capacity were more affected by artefacts from the NR scheme than the participants with higher capacity.

In experiment 2, it was demonstrated that the processing effort involved in speech recognition in noise further depended on the type of NR scheme used. Two HAs with different NR schemes were compared. With the HA (HA1) that used a MVDR beamformer combined with fast-acting NR, recognition performances and PPD were significantly reduced. This effect is assumed to stem not only from the higher gain in SNR from the output of HA1 but further from the fact that the postfilter gain adjustment that provided NR was faster and more accurate compared with the NR used in HA2. Thus, the level of the noise between speech pauses could be reduced with HA1 (Le Goff et al. 2016). This degree of accuracy was not achievable with the slow-acting NR scheme of HA2, which had a reaction time in the order of several seconds. Although the participants performed at high recognition levels, the PPD was significantly reduced (approximately 0.024 mm) for HA1 and its fast-acting NR compared with the slow-acting NR of HA2.

To our knowledge, this study showed for the first time a benefit of a NR scheme on processing effort by pupillometry. This opens a new perspective on using pupillometry to evaluate and develop high-performance HAs and test the benefits of HA signal processing in situations where traditional speech reception measures fail because of ceiling effects. Furthermore, the presented results underline the importance of using alternative outcome measures, such as processing effort, in HA research.

Several cognitive processes have been related to changes in the pupil size such as emotional response or arousal (Einhäuser 2017). The present study paradigm has been carefully developed and extensively used in several studies before to disentangle different phenomena and processes affecting the pupil size during speech processing (Zekveld et al. 2010, 2011, 2014; Koelewijn et al. 2014). Although an impact of other cognitive processes on the pupil size cannot be excluded, some factors were minimized within this study design. For instance, the emotional demands of the task were balanced across the experiment and are expected to be low. Moreover, potential effects of arousal or emotional processes would be reflected in changes of the baseline pupil diameter as well. However, no changes in the baseline pupil value across conditions were found in the present study. In addition, when calculating the PPD relative to the pupil baseline value, one largely controls for those effects. Therefore, it is assumed that the observed differences in PPD are indeed reflecting changes in processing effort caused by NR processing while speech recognition in background noise.

To obtain a more complete picture of the processing effort involved in speech perception in an aided situation, subjective measures of effort, such as self-reported effort, must also be assessed. It has been shown that self-reported effort is not necessary reflected in more objective or physiologic measures of effort, indicating that those measures address different aspects of effort (Wendt et al. 2016). Future research can clarify how and to what extent pupillometry can be used as assessment tool for changes in processing effort resulting from signal processing technology in current HAs. Other acoustic scenarios, such as different types of background noise and a broader range of SNRs, should be evaluated in a more systematic study. In addition, more realistic communication situations can be evaluated by using a moving target or by changing the position of the target speaker. The present study used the VAC rationale as a first-fit approach, and no verification with a probe microphone was made, which is a limitation. We chose the VAC rationale (LeGoff 2015) in favor of the NAL prescription, since NAL may provide insufficient audibility below 4 kHz and may underestimate the importance of cognitive factors (Humes 2007). The VAC rationale has higher low-level gain that support increased audibility. However, it might be advisable to also include other prescriptive methods for future studies.

CONCLUSIONS

Although HA processing and NR algorithms often fail to improve speech intelligibility in situations with ecologic SNRs, specific signal processing settings and NR schemes are still often preferred by listeners. See Ohlenforst et al. (2017) for a review of the effect of HA amplification on processing effort. This preference may occur because the NR algorithm can free cognitive resources and thus reduce the effort required for successful speech communication. The results of the present study demonstrated that NR reduces the processing effort involved in speech recognition, as indicated by the pupil dilation. At positive SNRs where SRTs are no longer sensitive, NR processing can still help the hearing-impaired listener reduce the cognitive resources required for correct speech recognition.

REFERENCES

Akeroyd, M. A. (2008). Are individual differences in speech reception related to individual differences in cognitive ability? A survey of twenty experimental studies with normal and hearing-impaired adults. *Int J Audiol, 47*(Suppl 2), S53–S71.

Aston-Jones, G., & Cohen, J. D. (2005). An integrative theory of locus coeruleus-norepinephrine function: Adaptive gain and optimal performance. *Annu Rev Neurosci, 28*, 403–450.

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

Bentler, R., Wu, Y. H., Kettel, J., et al. (2008). Digital noise reduction: Outcomes from laboratory and field studies. *Int J Audiol, 47*, 447–460.

Boldt, J., Kjems, U., Pedersen, M. S., Lunner, T., & Wang, D. (2008). Estimation of the Ideal Binary Mask using Directional Systems. In *Proceedings of the 11th International Workshop on Acoustic Echo and Noise Control International Workshop on Acoustic Echo and Noise Control, University of Washington campus*. Seattle, Washington.

Brand, T., & Kollmeier, B. (2002). Efficient adaptive procedures for threshold and concurrent slope estimates for psychophysics and speech intelligibility tests. *J Acoust Soc Am, 111*, 2801–2810.

Brons, I., Houben, R., Dreschler, W. A. (2013). Perceptual effects of noise reduction with respect to personal preference, speech intelligibility, and listening effort. *Ear Hear*, 34, 29–41.

Brons, I., Houben, R., Dreschler, W. A. (2015). Acoustical and perceptual comparison of noise reduction and compression in hearing aids. *J Speech Lang Hear Res*, 58, 1363–1376.

Buus, S., & Florentine, M. (2002). Growth of loudness in listeners with cochlear hearing losses: Recruitment reconsidered. *J Assoc Res Otolaryngol*, 3, 120–139.

Daneman, M., & Carpenter, P. (1980). Individual differences in working memory and reading. *J. Verbal Learning Verbal Behav*, 19, 450–466.

Duquesnoy, A. J. (1983). Effect of a single interfering noise or speech source upon the binaural sentence intelligibility of aged persons. *J Acoust Soc Am*, 74, 739–743.

Einhäuser, W. (2017). The pupil as marker of cognitive processes. In Q. Zhao (Ed.), *Computational and Cognitive Neuroscience of Vision* (pp. 141–169). Singapore: Springer-Verlag.

Ephraim, Y., & Malah, D. (1984). Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator. *IEEE Trans Acoust Speech Signal Process, ASSP*, 32, 1109–1121.

Ephraim, Y., & Malah, D. (1985). Speech enhancement using a minimum mean-square error log-spectral amplitude estimator. *IEEE Trans Acoust Speech Signal Process, ASSP*, 33, 443–445.

Fredelake, S., Holube, I., Schlueter, A., et al. (2012). Measurement and prediction of the acceptable noise level for single-microphone noise reduction algorithms. *Int J Audiol*, 51, 299–308.

Gatehouse, S., & Gordon, J. (1990). Response times to speech stimuli as measures of benefit from amplification. *Br J Audiol*, 24, 63–68.

Anderson Gosselin, P., & Gagné, J. P. (2011). Older adults expend more listening effort than young adults recognizing speech in noise. *J Speech Lang Hear Res*, 54, 944–958.

Gosselin, P. A., & Gagné, J. P. (2011). Older adults expend more listening effort than young adults recognizing audiovisual speech in noise. *Int J Audiol*, 50, 786–792.

Hagerman, B., & Kinnefors, C. (1995). Efficient adaptive methods for measuring speech reception threshold in quiet and in noise. *Scand Audiol*, 24, 71–77.

Humes, L. E. (2007). The contributions of audibility and cognitive factors to the benefit provided by amplified speech to older adults. *J Am Acad Audiol*, 18, 590–603.

Hick, C. B., & Tharpe, A. M. (2002). Listening effort and fatigue in school-age children with and without hearing loss. *J Speech Lang Hear Res*, 45, 573–584.

Hornsby, B. W. (2013). The effects of hearing aid use on listening effort and mental fatigue associated with sustained speech processing demands. *Ear Hear*, 34, 523–534.

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

Jensen, J., & Pedersen, M. S. (2015). Analysis of beamformer directed single-channel noise reduction system for hearing aid applications. *IEEE Int Conf Acoust, Speech Signal Process (ICASSP)*, 5728–5732.

Johnsrude, I., & Rodd, J. M. (2015). "Factors that increase processing demands when listening to speech." In G. Hickok & S. L. Small (Eds.), *Neurobiology of Language* (pp. 491–502). London, UK: Academic Press.

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

Kahneman, D. (1973). *Attention and Effort*. Englewood Cliffs, NJ: Prentice-Hall, Inc.

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

Kjems, U., & Jensen, J. (2012). Maximum likelihood based noise covariance matrix estimation for multi-microphone speech enhancement. *Proc Eur Signal Process Conf (EUSIPCO)*, 295–299.

Kjems, U., Boldt, J. B., Pedersen, M. S., et al. (2009). Role of mask pattern in intelligibility of ideal binary-masked noisy speech. *J Acoust Soc Am*, 126, 1415–1426.

Koelewijn, T., Zekveld, A. A., Festen, J. M., et al. (2014). The influence of informational masking on speech perception and pupil response in adults with hearing impairment. *J Acoust Soc Am*, 135, 1596–1606.

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

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

Le Goff, N. (2015). *Amplifying Soft Sounds - A Personal Matter. Oticon Whitepaper*. Retrieved July 4, 2016 from http://www.oticon.global/evidence.

Le Goff, N., Jensen, J., Perdersen, M.S., et al. (2016). *An Introduction to OpenSound Navigator, Oticon Whitepaper*. Retrieved July 4, 2016 from http://oticon.global/evidence.

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

Lunner, T. (2003). Cognitive function in relation to hearing aid use. *Int J Audiol*, 42(Suppl 1), S49–S58.

Lunner, T., Rudner, M., Rosenbom, T., et al. (2016). Using speech recall in hearing aid fitting and outcome evaluation under ecological test conditions. *Ear Hear*, 37(Suppl 1), 145S–154S.

Mattys, S. L., Davis, M. H., Bradlow, A. R., et al. (2012). Speech recognition in adverse conditions: A review. *Lang Cogn Process*, 27, 953–978.

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 A*, 58, 22–33.

McGarrigle, R., Munro, K. J., Dawes, P., et al. (2014). Listening effort and fatigue: What exactly are we measuring? A British Society of Audiology Cognition in Hearing Special Interest Group 'white paper.' *Int J Audiol*, 53, 433–440.

Nachtegaal, J., Kuik, D. J., Anema, J. R., Goverts, S. T., Festen, J. M., & Kramer, S. E. (2009). Hearing status, need for recovery after work, and psychosocial work characteristics: Results from an Internet-based national survey on hearing. *Int J Audiol*, 48, 684–691.

Neher, T. (2014). Relating hearing loss and executive functions to hearing aid users' preference for, and speech recognition with, different combinations of binaural noise reduction and microphone directionality. *Front Neurosci*, 8, 391.

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

Ng, E. H., Rudner, M., Lunner, T., et al. (2015). Noise reduction improves memory for target language speech in competing native but not foreign language speech. *Ear Hear*, 36, 82–91.

Nilsson, M., Soli, S. D., Sullivan, J. A. (1994). Development of the hearing in noise test for the measurement of speech reception thresholds in quiet and in noise. *J Acoust Soc Am*, 95, 1085–1099.

Nielsen, J. B., & Dau, T. (2011). The Danish hearing in noise test. *Int J Audiol*, 50, 202–208.

Ohlenforst, B., Zekveld, A. A., Jansma, E. P., et al. (2017). Effects of hearing impairment and hearing aid amplification on listening effort: A systematic review. *Ear Hear*, 38, 267–281.

Petersen, E. B., Lunner, T., Vestergaard, M. D., et al. (2016). Danish reading span data from hearing-aid users, including a sub-group analysis of their relationship to speech-in-noise performance. *Int J Audiol*, 55, 254–261.

Pichora-Fuller, M. K., Kramer, S. E., Eckert, M., et al. (2016). Hearing impairment and cognitive energy: A framework for understanding effortful listening (FUEL). *Ear Hear*, 37, S5–S37.

Picou, E. M., Ricketts, T. A., Hornsby, B. W. (2013). How hearing aids, background noise, and visual cues influence objective listening effort. *Ear Hear*, 34, e52–e64.

Piquado, T., Isaacowitz, D., Wingfield, A. (2010). Pupillometry as a measure of cognitive effort in younger and older adults. *Psychophysiology*, 47, 560–569.

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.

Rabbitt, P. (1990). Mild hearing loss can cause apparent memory failures which increase with age and reduce with IQ. *Acta Otolaryngol Suppl*, 476, 167–175; discussion 176.

Rönnberg, J., Arlinger, S., Lyxell, B., et al. (1989). Visual evoked potentials: Relation to adult speechreading and cognitive function. *J Speech Hear Res*, 32, 725–735.

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(Suppl 1), S68–S76.

Rönnberg, J., Lunner, T., Zekveld, A., et al. (2013) The Ease of Language Understanding (ELU) model: theoretical, empirical, and clinical advances. *Front Syst Neurosci*, 7, 31.

Sarampalis, A., Kalluri, S., Edwards, B., et al. (2009). Objective measures of listening effort: Effects of background noise and noise reduction. *J Speech Lang Hear Res*, 52, 1230–1240.

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. *Neuroimage*, 20, 114–124.

Smeds, K., Wolters, F., Rung, M. (2015). Estimation of signal-to-noise ratios in realistic sound scenarios. *J Am Acad Audiol*, 26, 183–196.

Souza, P., Arehart, K., Neher, T. (2015) Working memory and hearing aid processing: Literature findings, future directions, and clinical applications. *Front Psychol*, 6, 1894.

Studebaker, G. A. (1985). A "rationalized" arcsine transform. *J Speech Hear Res*, 28, 455–462.

Wang, D., Kjems, U., Pedersen, M. S., et al. (2008). Speech perception of noise with binary gains. *J Acoust Soc Am*, 124, 2303–2307.

Wendt, D., Kollmeier, B., Brand, T. (2015). How hearing impairment affects sentence comprehension: Using eye fixations to investigate the duration of speech processing. *Trends Hear*, 19, 1–18.

Wendt, D., Dau, T., Hjortkjær, J. (2016). Impact of background noise and sentence complexity on processing demands during sentence comprehension. *Front Psychol*, 7, 345.

Zekveld, A. A., Kramer, S. E., Festen, J. M. (2010). Pupil response as an indication of effortful listening: The influence of sentence intelligibility. *Ear Hear*, 31, 480–490.

Zekveld, A. A., Kramer, S. E., Festen, J. M. (2011). Cognitive load during speech perception in noise: The influence of age, hearing loss, and cognition on the pupil response. *Ear Hear*, 32, 498–510.

Zekveld, A. A., & Kramer, S. E. (2014). Cognitive processing load across a wide range of listening conditions: Insights from pupillometry. *Psychophysiology*, 51, 277–284.

REFERENCE NOTE

Haverkamp, L. (2015). Erfassung von Alltagssituationen mithilfe von Echtzeitaufnahmen und subjektiven Bewertungen [In German]. Unpublished Bachleor Thesis, 2015, Jade Hochschule, Oldenburg, Germany.