Hearing Aid Noise Reduction Lowers the Sustained Listening Effort During Continuous Speech in Noise-A Combined Pupillometry and EEG Study

Lorenz Fiedler, Tirdad Seifi Ala, Carina Graversen, Emina Alickovic, Thomas Lunner and Dorothea Wendt

The self-archived postprint version of this journal article is available at Linköping University Institutional Repository (DiVA):
http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-180891

N.B.: When citing this work, cite the original publication.
https://doi.org/10.1097/AUD.0000000000001050

Original publication available at:
https://doi.org/10.1097/AUD.0000000000001050

Copyright: Lippincott, Williams & Wilkins
http://www.lww.com/
Hearing aid noise reduction lowers the sustained listening effort during continuous speech in noise — a combined pupillometry and EEG study

Lorenz Fiedler¹, Tirdad Seifi Ala¹,², Carina Graversen¹, Emina Alickovic¹,³, Thomas Lunner¹ & Dorothea Wendt¹,⁴

¹Eriksholm Research Centre, Oticon A/S, Denmark
²Hearing Sciences – Scottish Section, Division of Clinical Neuroscience, University of Nottingham, Glasgow, United Kingdom
³Department of Electrical Engineering, Linkoping University, Sweden
⁴Department of Health Technology, Technical University of Denmark, Denmark

Corresponding author:

Lorenz Fiedler
lfie@eriksholm.com
Eriksholm Research Centre
Rørtangvej 20
3070 Snekersten
Denmark

Conflicts of Interest and Source of Funding:

TSA has received funding from the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 765329.

Author contributions:

LF analyzed pupillometry data and drafted the manuscript, TSA designed the experiment, analyzed EEG data and edited the manuscript. CG, EALI, THLU and DW designed the experiment and edited the MS.
Abstract

Objectives: The investigation of auditory cognitive processes recently moved from strictly controlled, trial-based paradigms towards the presentation of continuous speech. This also allows the investigation of listening effort on larger time scales (i.e., sustained listening effort). Here we investigated the modulation of sustained listening effort by a noise reduction algorithm as applied in hearing aids in a listening scenario with noisy continuous speech. The investigated directional noise reduction algorithm mainly suppresses noise from the background.

Design: We recorded the pupil size and the electroencephalogram (EEG) in 22 hearing-impaired participants who listened to audio news clips in the presence of background multi-talker babble noise. We estimated how noise reduction (off, on) and signal-to-noise ratio (SNR; +3 dB, +8 dB) affect pupil size and the power in the parietal EEG alpha band (i.e., parietal alpha power) as well as the behavioral performance.

Results: Our results show that noise reduction reduces pupil size, while there was no significant effect of the SNR. Importantly, we found interactions of SNR and noise reduction, which suggested that noise reduction reduces pupil size predominantly under the lower SNR. Parietal alpha power showed a similar, yet non-significant pattern, with increased power under easier conditions. In line with the participants’ reports that one of the two presented talkers was more intelligible, we found a reduced pupil size, increased parietal alpha power and better performance when people listened to the more intelligible talker.
**Conclusions:** We show that the modulation of *sustained listening effort* (e.g. by hearing aid noise reduction) as indicated by *pupil size* and *parietal alpha power* can be studied under more ecologically valid conditions. Mainly concluded from *pupil size*, we demonstrate that hearing aid *noise reduction* lowers *sustained listening effort*. Our study approximates to real-world listening scenarios and evaluates the benefit of the signal processing as can be found in a modern hearing aid.
Introduction

Within the last decade, the investigation of auditory cognitive processes went beyond strictly repetitive paradigms, such as single words or sentences (i.e., trial-based designs), towards the presentation of non-repetitive, continuous stimuli such as speech (e.g., Lalor et al., 2009). While trial-based designs allow for maximal control, it was argued that more natural, non-repetitive stimuli allow more general conclusions (Hamilton and Huth, 2018). This also includes the presentation of multiple, concurrent speech streams or speech in noise, such as humans are faced with in everyday listening scenarios (Alexandrou et al., 2018). With the current study, we followed this line of research by investigating physiological measures of listening effort during the presentation of continuous, concurrent speech, which was recently proposed to lead to a stronger engagement into the listening task (Herrmann and Johnsrude., 2019). We investigated whether a noise reduction algorithm of a modern hearing aid leads to a reduction of listening effort during the presentation of continuous speech.

We investigated listening effort to evaluate a hearing aid noise reduction algorithm (further called noise reduction). In conventional speech-in-noise tests (e.g., Kalikow et al., 1977), the behavioral performance (i.e., listening performance; e.g. correctly repeated words) has been used to infer on the neural fate of the speech signals as well as the neural consequences of hearing impairment and its treatment. However, it was shown that it fails to represent the subjective performance of a listener or the benefit of a hearing aid (Plomp, 1977). Based on the argument that speech recognition in noise involves working memory resources, it was argued that other factors such as the cognitive effort spent to solve a listening task are important predictors for subjective
benefits (Rönnberg et al., 2008). For example, a hearing aid may not help to understand more words, but it may lower the cognitive effort of understanding the same amount of words. Henceforth, the cognitive effort spent for solving a listening task was called *listening effort* (Rönnberg et al., 2013; McGarrigle et al., 2014), which was later defined as “*the deliberate allocation of mental resources to overcome obstacles to goal pursuit when carrying out a [listening] task*” (Pichora-Fuller et al., 2016). Here we planned to investigate the modulation of *listening effort* by the *noise reduction* of a hearing aid. Note that we assumed that the manipulation of the task demand leads to a considerable modulation of listening effort as objectively assessed by physiological measures (Bruya and Tang, 2018). Hence, we did not measure subjective listening effort (cf., Hart and Staveland, 1988; Krueger et al., 2017). Objective quantification of *listening effort* was earlier investigated on neurophysiological measures (Peelle, 2018).

First, *pupil size* was assessed, since it was found to reflect cognitive effort facilitated to solve a task (Kahneman and Beatty, 1966). The sympathetic and the parasympathetic nervous system are linked to the dilator and sphincter muscles of the iris, respectively (McDougal and Gamlin, 2015). The activity in the sympathetic nervous system is associated with arousal in so called “*fight or flight*”-scenarios. Hence, increased activity in the sympathetic nervous system leads to pupil dilation. In contrast, the parasympathetic activity is associated with so called “*rest and digest*”-scenarios. Increased parasympathetic activity leads to constriction of the pupil. Hence, all other factors being equal (e.g. luminance), the pupil serves as an instant measure of facilitated cognitive effort where a larger pupil indicates increased effort (for a recent review: van der Wel and van Steenbergen, 2018).
Pupillometry has been adopted to auditory research and used to measure listening effort. Based on the pupil size, studies showed changes in listening effort related to changes in signal-to-noise ratio (SNR; Ohlenforst et al., 2017; Wendt et al., 2018). It was shown that lower speech intelligibility leads to an increased pupil size (Zekveld et al., 2010) and that different masker types vary pupil size (Koelewijn et al. 2014). Importantly, it was shown that a lower SNR leads to an increased pupil size but only up to the point where participants tend to give up. In sum, these studies suggest that the pupil size reflects the effort of selective listening in so-called cocktail-party scenarios (Cherry, 1953), where a target talker is presented during presence of one or multiple distractors.

As a second physiological measure for listening effort, we here consider parietal alpha power as obtained from electroencephalography (EEG; for a review: Peelle, 2018). Brain oscillations in the alpha frequency band (~8–13 Hz) are the most dominant marker of neural activity in EEG (Berger, 1929; Adrian and Matthews, 1934). The functional role of oscillatory power in this frequency band (i.e., alpha power) has been broadly discussed: One hypothesis is that increased alpha power leads to inhibition of brain regions (Klimesch et al., 2007). In theory, alpha power could either indicate the guidance of relevant signals by inhibiting irrelevant pathways (Jensen and Mazaheri, 2010), attenuate distracting signals (Kerlin et al., 2010) or, most crucially for listening effort, distribute cognitive resources across brain regions (Ocleser et al., 2012). Consequently, varying listening effort should result in a modulation of alpha power.
In contrast to pupil size, the direction (i.e., increase or decrease) of effort-related alpha power modulation found in the literature strongly depends on the task design and the investigated topographical regions. Evidence for its inhibitory functional role came from findings that parietal alpha power increases in the brain hemisphere where distracting speech is processed (i.e., contra-lateral to distractor) compared to relevant speech (contra-lateral to target; Kerlin et al., 2010; Wöstmann et al., 2016; Cartocci et al., 2019).

It was also argued that inhibitory alpha power works locally at specific cortical areas, such that auditory alpha power may show a different modulation than parietal or visual alpha power at the same time (Weisz et al., 2011; Keitel and Gross, 2016). Without a spatial separation of relevant and distracting auditory inputs, higher alpha power was associated with more adverse conditions and decreased speech intelligibility (Obleser and Weisz, 2012; Obleser et al., 2012; Becker et al., 2013; Strauß et al., 2014; Dimitrijevic et al., 2017; Wisniewski et al., 2017). It was argued that degraded speech leads to higher working memory load due to impeded lexical memory access. Controversially, higher working memory load and more adverse conditions were also found to lead to a decrease in alpha power (Miles et al., 2017; Hjortkjær et al., 2018). The reason for this discrepancy may occur because different study designs may step into different aspects of listening effort and thus reflect the modulation of different cognitive resources (Herrmann and Johnsrude, 2019). In a design similar to our current study, an increase of parietal alpha power under higher SNRs was found (Seifi Ala et al., 2020). Since noise reduction not only reduces the amount of energetic masking but may also free up cognitive resources due to decreased informational masking, in the current study we
will investigate the modulation of *parietal alpha power* while being agnostic about the direction of its modulation.

Previously it was shown that a hearing-aid, especially its noise-reduction algorithm, not only improved the SNR (i.e. reducing interfering background noise and enhancing the target speaker), but also reduced *listening effort* in dual-task paradigms (Sarampalis et al., 2009; Desjardins and Doherty, 2014) and pupillometry (Wendt et al., 2017; Ohlenforst et al., 2018). With the later studies, it was demonstrated that *noise reduction* leads to a decreased peak pupil dilation and thus, reduced listening effort for people with hearing loss. In cochlea implant patients it was shown that the ratio between EEG frontal theta power and parietal alpha power indicates reduced *listening effort* due to *noise reduction* (Cartocci, 2018). Wendt et al. 2017 reported a reduction in *listening effort* with active noise reduction scheme in situations with high speech intelligibility close to ceiling (i.e. 95% speech recognition), which indicates that pupillometry may provide additional information beyond behavioral performance. It was further suggested that pupil dilation is a useful tool to examine hearing aid signal processing in acoustic scenarios that are comparable to communication in everyday life, such as mostly positive SNRs (Smeds et al., 2015).

The goal of the current study was to evaluate the modulation of *listening effort* by a hearing-aid noise reduction algorithm under more ecologically valid listening conditions such as the presentation of continuous speech instead of single sentences. Recently, the cognitive resources spent to maintain attention to one of multiple continuous tone streams was studied (Zhao et al., 2019). The authors aimed to isolate the cognitive effort...
of maintaining attention (i.e., sustained attention) from other factors that could increase listening effort such as energetic masking. In contrast, our design does not dissociate between effort related to sustained attention and other possible factors for listening effort. With our current design, we studied how neurophysiological measures of listening effort unfold over time scales longer than a single sentence. Hence, we will further call our measure of interest sustained listening effort. Importantly, the dynamics of the pupil size have both tonic and phasic components (Beatty, 1982; Aston-Jones and Cohen, 2005). Phasic responses are transient components that occur in response to a stimulus, such as what has been found in classical sentence-in-noise tests or here directly after target onset in the form of a positive response (Wendt et al., 2016, 2018; Ohlenforst et al., 2017, 2018). In contrast, tonic activity is more sluggish and depends on the current state of arousal and/or memory load as well as fatigue (McGarrigle et al., 2017).

We simulated a cocktail-party scenario (Cherry, 1953) consisting of two talkers in the foreground and multi-talker babble in the background. While the noise reduction mainly suppresses the background noise, selective auditory attention should dissociate between the two frontal talkers (Alickovic et al., 2020, under minor revision). To investigate the interaction between the noise reduction and the background noise, we varied the relative sound level (i.e., SNR) between the background and the frontal talkers.

We hypothesized that both higher SNR and activated noise reduction lead to a decrease in sustained listening effort. This should be reflected in a smaller pupil size and as well
as a modulation of *parietal alpha power*. In a design similar to the current one, an increase of *parietal alpha power* under easier conditions was found (Seifi Ala et al., 2020). Hence, here our working hypothesis is to find an increase in *parietal alpha power* with activated *noise reduction* compared to deactivated noise reduction.

**Methods**

**Participants**

This dataset has been used for another analysis that focused on the neural tracking of the speech signals (Alickovic et al., 2020). We recruited 22 hearing-impaired participants from the *Eriksholm Research Centre* database. The age span was between 40 and 80 (Mean = 67, SD = 11.2). Participants had at least 4 months of experience with their hearing aids. The maximal asymmetry in audiometric thresholds between the right and the left ear were 8 dB when averaged from 500 to 4000 Hz. The average audiometric threshold was 45 dB hearing level (HL) and the thresholds at 500, 1000, 2000, and 4000 Hz ranged from 33 to 58 dB HL (Figure 1B). The inclusion criteria were mild to moderate sensorineural hearing loss, normal or corrected-to-normal vision, and no history of neurological disorders, dyslexia or diabetes mellitus.

The study was approved by the ethics committee for the capital region of Denmark (Journal number H-1-2011-033). The study was conducted according to the Declaration of Helsinki, and all participants signed a written consent prior to the experiment.
Experimental setup

With the current experimental design, we sought to simulate an acoustically challenging cocktail-party listening scenario (Figure 1). The setup was inspired by the design of Das et al. (2018). The experiment took place in an acoustically and electrically shielded booth. Two loudspeakers (Genelec 8040A, Genelec Oy, lisalmi, Finland) were set up in front of the participant’s chair at an azimuthal angle of ±22 degrees to provide spatial separation of the frontal talkers. The resulting separation of 44 degrees is closest to the condition of ±30 degrees found in Das et al. (2018). Four loudspeakers of the same kind were set up in the back of the participants at an azimuthal angle of ±90 and ±150 degrees. Participants were positioned in the middle of the loudspeakers and the distance to each loudspeaker was 1.2 meters. The height of the chair was adjusted such that the ears of the participants were level with the center of the loudspeaker membrane. A screen was positioned between the frontal loudspeakers. Based on the experience from earlier studies, the light was dimmed to adjust the illuminance at the participants’ eye position to approximately 45 Lux.

Stimuli

We extracted 86 parts (i.e., trials) from publicly available news clips spoken by two different talkers. Half of the trials were spoken by a female and the other half by a male talker. Each part was 33 seconds long after pauses longer than 200 ms were removed. These news clips were used as target and distractor streams and presented at the frontal loudspeakers. The organization of the male and the female talker being target or distractor as well as their location (left or right loudspeaker) was randomized. Note that the factor talker was added to our analysis. Since we were not investigating gender
effects, we will refer to the male target talker as talker 1 and the female target talker as talker 2.

As background noise, four-talker babble was presented from each of the four loudspeakers in the back, resulting in a 16-talker babble noise. Each 4-talker babble-noise consisted of two female talkers and two male talkers, which were different from the talkers presented at the frontal loudspeakers. The set of four talkers was the same at each background loudspeaker, but it was assured that the identical news clip was not simultaneously presented at two different loudspeakers.

After the two frontal talkers were rms-equalized, the sound pressure level (SPL) of each of the two frontal talkers was set to 62 dB, respectively. The long-term average spectrum of the babble noise was matched to the long-term average spectrum of the target talkers. The SPL of each 4-talker babble noise was set to 53 dB or 48 dB. Since the summed SPL of the 16-talker babble noise is 6 dB above the SPL of the individual 4-talker babble noise, the signal-to-noise ratio (SNR) of the frontal talkers relative to the background noise was either +3 dB or +8 dB. Note that the effective SNR was even lower, since one of the frontal talkers in the role of a distractor will add to the noise. However, we will further use +3 dB and +8 dB as condition labels for the factor SNR.

Hearing aid and noise reduction

All participants were fit with identical hearing aids with non-individualized, closed soft tips. Two pairs of hearing aids were fit for each participant. In one pair, noise reduction was turned off and in the other pair, noise reduction was activated. The factor noise reduction with the condition labels on and off was used for statistical analysis. Other
than that, all signal processing features (e.g., feedback cancelling) were kept at default and did not vary between the conditions.

Both pairs of hearing aids amplified the sound based on each individual’s hearing threshold via the Voice Aligned Compression (VAC) rationale (Le Goff, 2015). The VAC amplification rationale is based on a wide dynamic range compression scheme with compression knee points between 20 and 50 dB SPL depending on the frequency range and the individual’s hearing thresholds. Similar to other standard fitting procedures (Scollie et al., 2005; Keidser et al., 2011), VAC was developed to fit hearing aids to individual needs to improve overall speech quality. It combines fast and slow compression in order to minimize distortion and to restore the amplitude modulation of speech. VAC is our standard fitting procedure for the hearing aids used in this study. The hearing aid was set to mimic the natural acoustic effect of the pinna by a microphone setting close to omnidirectional.

With activated noise reduction, a fast-acting combination of minimum variance distortion-less response (MVDR) beam-former (Kjems and Jensen, 2012) and a single-channel Wiener post-filter (Jensen and Pedersen, 2015) was applied before the VAC. The noise reduction algorithm is based on the finding that a multi-channel Wiener filter can be decomposed into a beam-former and a single-channel Wiener filter, which is better suited for implementation into hearing aids. The noise reduction mainly attenuates interfering sounds originating behind the listener, which should mainly affect the background noise in the current study. To confirm this, we measured the output SNR-improvement of the noise reduction in the hearing aids, which was here defined as the
difference between the two frontal talkers and four background talkers. The hearing aid output was measured on a Head and Torso Simulator (HATS). A pair of hearing aids were put on the HATS and the output SNRs of the hearing aids were derived using the phase-inversion technique (Hagerman and Oloffson, 2004). The articulation-index weighted SNR improvements were 6.24 dB and 5.17 dB at +3 dB SNR and +8 dB SNR for noise reduction on compared to off (see also Alickovic et al., 2020).

Procedure and task

The order of conditions followed a blocked design with the order of blocks being randomized across participants. For each participant, the experiment consisted of four major blocks (20 trials per block). In order to blind the randomization of noise reduction to the participant, the hearing aids were always taken out of the booth and inserted again between the blocks. In two blocks, the noise reduction was switched on and in the other two blocks it was switched off. Within each block, the SNR was fixed at either +3 or +8 dB. The order was randomized such that all four possible combinations of SNRs (+3, +8 dB) and noise reduction (off, on) occurred once per participant. This randomization was balanced across participants. Over the 20 trials within each block, the target talker and its location changed every fifth trial, such that all four combinations of talker (T1, T2) and position (left, right) occurred within each block in a random succession. To indicate the target talker and its position, before every fifth trial, a 5-second snippet from the talker’s voice was presented at the to-be-attended loudspeaker. Before and during each trial, the to-be-attended loudspeaker was also indicated by an arrow at the screen. To acclimatize to the hearing aids, before the start
of the four major blocks, the participants listened to six training trials with noise reduction on.

Participants were instructed to visually fixate a cross in the middle of the screen during listening. The presentation of the sound started with background babble-noise of five seconds. This period mainly served as baseline for the acquired physiological measures. Subsequently, the two news clips were presented at the frontal speakers in the presence of the ongoing background babble-noise (Figure 1A). After each trial, a statement about the content of the to-be-attended news clip was displayed on the screen, e.g. “Der kommer et stigende antal krydstogtturister til København” ("An increasing number of cruise tourists come to Copenhagen"). Participants were asked to indicate whether this statement was correct or wrong. Consequently, the chance level was 50%.

Without being asked explicitly, 15 of the 22 participants reported the male talker being harder to understand (or the female talker being harder to ignore). None of the participants reported the opposite. Hence, we included the factor talker in our analysis. To avoid the impression that we are investigating gender effects on intelligibility, we will refer to the talkers as talker 1 (male) and talker 2 (female).

<<<Figure 1 about here>>>

Pupillometry data acquisition and preprocessing

We used an SMI Red eye tracker (SensoMotoric Instruments, Teltow, Germany) for recording of the pupil size. The eye tracker was put approximately 75 centimeters in front of the participant at an angle of approximately 35°. We recorded data from both eyes. The following analysis was done with MATLAB R2018b (The MathWorks Inc.,
Nattick, Massachusetts). The pupil size from both eyes was extracted in the time range of the noise presentation, which started 5 seconds before the speech signal.

As found in the recorded pupil size, eye blinks appear either as missing data samples or as zeros. We detected such samples and removed samples within the time range 35 ms before and 100 ms after their occurrence. In contrast to earlier studies, we decided against an outlier-based removal criterion (e.g., ±2 standard deviations), since we had to deal with a non-stationary signal which expressed in a negative drift across the 33 seconds (Figure 2A). For each participant, we further processed the pupil size of the eye that showed fewer missing samples (left eye: 8 participants, right eye: 14 participants). If less than 60% of data were available, a trial was rejected from further analysis. We only included the pupil size of participants to further analysis if at least 60% of the trials were available. This resulted in a total of 16 participants with a mean of 95% of available trials. For 10 participants, 100% of the trials were available. The minimum of available trials for a participant were 71%. For the remaining trials, missing samples were linearly interpolated and a convolution with a hamming window of two seconds (121 points) was used for smoothing. As a baseline for each trial, the mean pupil size of the noise-only time range between –4 and 0 seconds relative to target speech onset was subtracted from the pupil size of the whole trial. The mean pupil size of the time range of target speech between 0 and 33 seconds was calculated, which served as a measure for sustained listening effort. Consequently, we derived one value for the pupil size per trial and per participant.
EEG data acquisition and preprocessing

EEG data were recorded using a BioSemi ActiveTwo system (Biosemi, Netherlands) with a standard cap including 64 surface electrodes mounted on the scalp according to the international 10-20 system and external electrodes on both mastoids. The cap included driven right leg (DRL) and common mode sense (CMS) electrodes. The later serves as reference for all other electrodes during recording. All electrodes were mounted by applying conductive gel. The BioSemi LabView software was used to monitor the electrode offset voltage and to record the EEG data. To ensure good connectivity, electrodes were prepared (and if necessary, supplied with additional gel) such that the offset voltage was stable and below 50 mV. The EEG signals were recorded with a sampling rate of 1024 Hz.

The EEG data were preprocessed with MATLAB R2018b (The MathWorks Inc., Nattick, Massachusetts) and the Fieldtrip toolbox (Oostenveld et al., 2011). The continuous EEG data were re-referenced to the averaged signal of the two mastoid electrodes, high- and lowpass-filtered between 0.5 and 70 Hz, respectively (Hamming-window finite impulse response). The filter order was set to $3f_s/f_c$, where $f_s$ is the sample-rate and $f_c$ is the cut-off frequency. The filter was applied both forward and backward to avoid phase shifts. A 50 Hz notch band-stop filter was applied as well. The EEG data were down-sampled to 256 Hz. Subsequently, the EEG data were epoched in trials between the onset of the babble-noise and the offset of the mixture of target and babble-noise.

The EEG data were visually inspected in order to detect noisy channels. Noisy channels were removed and interpolated from neighbor channels. On average, 2.6 channels were
removed (sd: 2.2, range: 0–7). To remove artifacts such as eye blinks, eye movements, heartbeat, muscle artifacts and single channel noise, independent component analysis was applied to the data (Makeig et al., 2004). Based on their topographical distribution as well as their time domain signals and their power spectrum, components that could be clearly associated with artifacts were removed from the data. On average, 35.3 components were removed (sd: 8.2, range: 17–52). At the parietal channels of interest it resulted in a removal of 33% of the variance (sd: 17%, range: 10–62%). Subsequently, the EEG data were re-referenced to the average of all channels in order to be consistent with an earlier study (Seifi Ala et al., 2020). During the described data cleaning, we identified two participants where technical problems must have led to noisy recordings such that these participants had to be excluded from further analysis.

**Extraction of parietal alpha power from EEG**

The absolute power (spectral density) for both the baseline (–4 to 0 seconds relative to target onset) and the target period (0 to 33 seconds) was calculated with the P-welch method at each EEG channel (Hann-window length of 1 second with 50% overlap). To assess the changes of EEG alpha power that are related to the listening task, event-related spectral perturbation (ERSP) was calculated (Pfurtscheller, 2001). The outcome of the measure is the relative change of power compared to the baseline in percent. The formula to calculate ERSP is as follows:

$$\text{ERSP} = \frac{A - R}{R} \times 100\%$$  

where $A$ is the absolute alpha power (8–13 Hz) of the target period and $R$ is the absolute alpha power of the baseline. We hypothesized to find a modulation of *parietal alpha*
power. Hence, parietal channels were selected as channels of interest ERSP was calculated for each channel and subsequently averaged across the channels of interest (CPz, CP1–CP4, Pz, P1–P8, POz, PO3, PO4; see Figure 2B). Consequently, we derived one ERSP value per trial and per participant, which was submitted to the statistical analysis as the measure for parietal alpha power.

Statistical analysis

We used (generalized) linear mixed-effects models in MATLAB to model if the fixed effects of noise reduction, SNR and talker predict the response variables behavioral performance, pupil size and parietal alpha power. We chose the approach of mixed models since it allowed us to submit single trial responses to the statistical analysis, where the model accounts for the fact that varying number of trials were available per participant and per condition. To allow for dependent testing, a random intercept was modelled for every participant. We coded the hypothetically harder levels with −0.5 (i.e., noise reduction off, SNR +3 dB, talker 1) and the easier levels with 0.5 (i.e., noise reduction on, SNR +8 dB, talker 2). With this coding the β-estimates directly quantify the change related to the manipulation of the factor from hard to easy. Consequently, we expected to find positive β-estimates for behavioral performance, negative β-estimates for mean pupil size (i.e., smaller pupil for easier condition) and positive β-estimates for mean parietal alpha power (i.e., higher parietal alpha power for easier condition). For the behavioral data, we applied generalized mixed models with a logit link function to account for its binomial distribution. Since the β-estimates of the behavioral performance are in logit-space, we transformed them back to the percentage of correctly answered questions relative to the mean intercept of the behavioral
performance (i.e., 87%). To exploit maximal statistical power, the separate models for pupil size and parietal alpha power were fit on all available participants for each measure, respectively (pupil size: \( N = 16 \), parietal alpha power: \( N = 20 \)). Note that there were only 15 participants for whom both pupil size and parietal alpha power were available. For these 15 participants, the average amount of trials where both pupil size and parietal alpha power were available was 92% (sd: 11%, range: 75–100%). To ensure that the selection of participants did not affect the results, as a control analysis we fit the models to the 15 participants for whom both pupil size and parietal alpha power were available (see supplements).

Since we hypothesized that both pupil size and parietal alpha power are neurophysiological measures for sustained listening effort, we decided to additionally analyze whether any trial-to-trial relationship between pupil size and parietal alpha power can be found in the 15 participants for whom both measures were available. We added parietal alpha power as a predictor to the model of pupil size and vice versa. With this approach, we estimated if one measure of sustained listening effort may explain some variance of the other measure. We only inspected a main effect in the models and not the interaction with all the other factors.

**Results**

**Behavioral performance**

After each trial, a two-choice question was presented regarding the content of the to-be-attended news clip, such that the chance level was 50%. All but one of the participants performed significantly above chance level, indicating that they attended
the news clips as instructed. This also reflects in significantly positive intercept of the behavioral performance ($\beta = 1.9 \approx 87\%$, $SE = 0.11$, $t_{1671} = 17.43$, $p = 0.001*10^{-62}$). The one participant who did not properly perform the task was excluded from further analysis.

We inspected how both hearing aid noise reduction and the signal-to-noise ratio (SNR) affected the performance. To account for the participants’ reports of one of the two talkers being more intelligible, we also investigated the effect of talker on the behavioral performance. Main effects and two-way interactions on the behavioral performance are depicted in Figure 2.

Even though the mean behavioral performance was higher with noise reduction on compared to off, we did not find a significant effect of noise reduction on behavioral performance ($\beta = 0.22 \approx 2\%$, $SE = 0.15$, $t_{1671} = 1.5$, $p = 0.14$). We found a slightly higher but non-significant behavioral performance under the better SNR ($\beta = 0.27 \approx 3.1\%$, $SE = 0.27$, $t_{1671} = 3.4$, $p = 0.067$). Confirming the subjective reports, we found an effect of talker on behavioral performance ($\beta = 0.79 \approx 9\%$, $SE = 0.15$, $t_{1671} = 29.5$, $p = 0.014*10^{-5}$), which indicates that participants were more likely giving a correct answer after attending to talker 2 compared to talker 1. Even though visual inspection may suggest interactions of SNR and noise reduction as well as talker and SNR, we did not find any significant interactions that affect behavioral performance (Noise reduction & SNR, $\beta = -0.34$, $SE = 0.30$, $t_{1671} = -1.14$, $p = 0.253$; Noise reduction & Talker, $\beta = 0.004$, $SE = 0.30$, $t_{1671} = -1.14$, $p = 0.98$; SNR & Talker, $\beta = -0.47$, $SE = 0.30$, $t_{1671} = -1.57$, $p = 0.116$).
The covariate trial number did not have a significant effect on the behavioral performance ($\beta = 0.4$, $SE = 0.24$, $t_{1671} = 1.6$, $p = 0.105$), which indicates that participants did not significantly become better or worse at the task.

In brief, we showed that the behavioral performance well above chance indicates that the participants followed the task as instructed. The behavioral performance was most strongly modulated by the talker, which indicates that the subjective reports of talker 2 being more intelligible also reflect in the quantitative measure of the behavioral performance.

Pupil size and parietal alpha power deflect from baseline during listening

To depict how pupil size and parietal alpha power unfold during sustained attention, we averaged the time courses of these measures across conditions (Figure 3, upper panels) and per conditions across participants (Figure 3, lower panels). Importantly, a change from baseline emerged both in pupil size and parietal alpha power over the course of the 33 seconds. This change is highly consistent across participants as well as across conditions. This consistency is statistically confirmed by significant estimates of the intercepts: While pupil size shows a significantly negative change from baseline ($\beta = -0.16$, $SE = 0.037$, $t_{1172} = -4.24$, $p = 0.023\times10^{-3}$), parietal alpha power shows a significantly positive change from baseline ($\beta = 50.2$, $SE = 9.1$, $t_{1507} = 5.52$, $p = 0.039\times10^{-5}$). This confirms that we can observe that listening to the news clips in background noise leads to a highly consistent change from baseline of pupil size and parietal alpha power. Note
that baseline values as well as trial number were added to the models as predictors and their effects are described in the supplements.

In brief, we showed that both pupil size and parietal alpha power show a deflection from baseline during listening.

<<Figure 3 about here>>>

Reduction of sustained listening effort due to noise reduction mainly reflects in decreased pupil size and under lower SNR.

We investigated both pupil size and parietal alpha power as neurophysiological measures of sustained listening effort during sustained attention to 33-second news clips. We hypothesized that pupil size and parietal alpha power as measures for sustained listening effort are modulated by hearing aid noise reduction and SNR. To account for the participants’ reports that one of the two talkers is more intelligible, we also investigated the effect of talker on the measures for sustained listening effort. All main effects and two-way interactions of pupil size are depicted in Figure 4 and parietal alpha power in Figure 5.

Confirming our hypothesis, we found that noise reduction leads to a significantly reduced pupil size ($\beta = -0.046$, $SE = 0.008$, $t_{1172} = -5.66$, $p = 0.018*10^{-6}$). In contrast, we did not find a significant effect of noise reduction on parietal alpha power ($\beta = 5.85$, $SE = 3.92$, $t_{1507} = 1.4$, $p = 0.136$). Thus, a reduction of sustained listening effort during noise reduction is only suggested by the pupil size.

We found that the modulation of the SNR had neither a significant effect on pupil size ($\beta = -0.013$, $SE = 0.008$, $t_{1172} = -1.6$, $p = 0.107$) nor on parietal alpha power ($\beta = 5.9$, $SE =$
3.91, t_{1507} = 1.53, p = 0.127). Thus, a reduction of sustained listening effort due to a higher SNR was not evident in any of the two measures.

Interestingly, we found a significantly decreased pupil size for talker 2 compared to talker 1 (β = -0.037, SE = 0.008, t_{1172}, p = 0.0056*10^{-3}) and significantly increased parietal alpha power (β = 9.2, SE = 3.90, t_{1507} = 2.35, p = 0.019). In sum, both measures suggest that the better intelligible talker 2 led to decreased sustained listening effort compared to talker 1.

Most importantly, we found a significant interaction of SNR and noise reduction on pupil size (β = 0.038, SE = 0.0164, t_{1172} = 2.31, p = 0.021), which suggests that the noise-reduction related decrease of pupil size is stronger under the lower SNR of +3 dB compared to +8 dB. We observed a similar but non-significant mirrored pattern in parietal alpha power (β = -14.46, SE = 7.82, t_{1507} = 1.85, p = 0.065), which would suggest the noise-reduction related increase of parietal alpha power is stronger under the lower SNR of +3 dB compared to +8 dB. In sum, these two findings suggest that noise reduction leads to a decreased sustained listening effort predominantly under the worse SNR of +3 dB compared to +8 dB.

Furthermore, we found a significant interaction of SNR and talker on pupil size (β = 0.033, SE = 0.016, t_{1172} = 2.05, p = 0.041), which suggests that a smaller pupil size due to a better SNR (+8 vs +3) can only be observed in response to the less intelligible talker 1. Other than that, we did not find any two-way or three-way interactions.
Noteworthy, our blocked design introduced some predictability of task demands, because the same condition was presented multiple times in a row. We did an exploratory analysis to check if this predictability had any effect on the baseline. We found significant differences in the baseline pupil size (see supplements, Figure S1 & S2). In brief, we found that mainly pupil size was sensitive to the manipulation of noise reduction, none of the measures was sensitive to the manipulation of the SNR and both pupil size and parietal alpha power were sensitive to talker. Additionally, we found and interaction of SNR and noise reduction on pupil size, which suggest that the effect of noise reduction is stronger under the worse SNR.

Pupil size and parietal alpha power are independent measures of sustained listening effort

Since we hypothesized that both pupil size and parietal alpha power are neurophysiological measures for sustained listening effort, we exploratorily analyzed whether any trial-to-trial relationship between pupil size and parietal alpha power is evident.

We neither found a main effect of parietal alpha power on pupil size ($\beta = -0.005, SE = 0.004, t_{1090} = -1.23, p = 0.22$) nor a main effect of pupil size on parietal alpha power ($\beta = -2.8, SE = 2.45, t_{1089} = -1.15, p = 0.25$). This indicates that the two measures of sustained listening effort are uncorrelated. Even ignoring potential confounds by removing all the other factors did not lead to a detection of any relationship. We thus conclude that the two measures for sustained listening effort are independent.
Discussion

We conducted a combined pupillometry and EEG study to investigate whether both pupil size and parietal alpha power in the EEG reflect sustained listening effort during continuous speech. We hypothesized that a hearing aid noise reduction algorithm lowers the sustained listening effort, which was tested in two different signal-to-noise ratios (SNR) during the presentation of continuous speech in background noise. We simulated a cocktail-party scenario where hearing impaired participants were asked to listen to one and ignore another talker in the presence of background noise. We found that noise reduction leads to decreased sustained listening effort predominantly under the lower SNR, which mainly manifested in a smaller pupil size. Confirming the participants’ reports that one talker was more intelligible, we also found a smaller pupil size and increased parietal alpha power when participants listened to this talker compared to the other talker. Even though the modulation of pupil size and parietal alpha power had similar patterns, a single trial analysis did not reveal any relationship between the two.

With the current study, we assessed pupil size as an established measure of sustained listening effort. In contrast to earlier sentence-based studies (Ohlenforst et al., 2017, 2018; Wendt et al., 2017, 2018), we presented continuous speech stimuli to simulate more ecologically valid listening conditions. This revealed that the time course of the pupil size was essentially different from conventional pupil responses. In previous studies, an increase of pupil size after sentence onset was typically shown, of which the maximum (i.e., peak pupil dilation) was extracted as a marker for listening effort. In contrast, after an initial, comparably small positive dilation, we here found a strong and
highly consistent negative-going deflection from baseline. While we assume that the initial positive response is comparable to what is found in sentence-based studies, in the current experiment negative-going deflection was much more prominent. A similar morphology of the pupil response was shown before (McGarrigle et al., 2017; Zhao et al., 2019). It was argued that the negative-going deflection is a release from initial arousal and that a stronger decrease reflects less listening effort. However, one might argue that there should be a positive deflection because the listening effort to be higher during the presentation of target speech compared to noise-only in the baseline. Since the observed negativity slowly establishes over the course of the 33 seconds, we assume that it rather describes the release from initial baseline arousal and the transition into a constant, tonic listening state, rather than a phasic response to the target stimulus onset. Stimuli longer than 33 seconds may lead to a full transition into the listening state such that the pupil size reach a constant level which solely depends on the current listening condition.

We showed that pupil size indicates a general reduction of sustained listening effort. This reduction was enhanced under the lower SNR. To our knowledge, this is the first study to show that listening effort can be reduced by a hearing aid noise reduction scheme in hearing impaired participants listening to continuous speech. We consider such a design to be closer to real-world listening scenarios than sentence-based hearing-in-noise tests (cf. Wendt et al., 2017; Ohlenforst et al., 2018). Continuous speech stimuli have been extensively used to study auditory selective attention in previous studies (Ding and Simon, 2011; Mesgarani and Chang, 2012; O’Sullivan et al., 2014; Fiedler et al., 2019; Alickovic et al., 2020). Due to their higher ecological validity, continuous stimuli were
suggested to lead to more generalizable results (Alexandrou et al., 2018; Hamilton and Huth, 2018) and an enhanced engagement into the task (Herrmann and Johnsrude., 2019). However, from a clinical point of view, such an ecological validity may cost some specificity: Even though we found lower listening effort with noise reduction, it can be challenging to trace down what aspect of the noise reduction led to lowered listening effort. Especially the fact that we did not find an effect of SNR raises the question to which extend noise reduction differs from the manipulation of the SNR. Another question should be whether the individual measurement of listening effort based on pupil size indicates which condition is subjectively preferred. For example, the participants that show higher listening effort with noise reduction switched on (compared to switched off) may also prefer the noise reduction to be switched off. In our current study, we did not assess subjective ratings of listening effort. Various scales exist to assess workload (Hart and Staveland, 1988) or listening effort (Krueger et al., 2017). One common problem with such subjective ratings is that the concept of listening effort is complex and that participants tend to evaluate the task demand and not the spent resources (Zekveld et al., 2010; Bruya and Tang, 2018). Here we assumed that a manipulation of the task demand by noise reduction and SNR leads to a considerable modulation of listening effort. Upcoming studies should go back and forth between designs with more control (e.g. single sentences) versus more ecological validity (e.g., continuous speech). The latter can be achieved by the simulation of conversations in real or virtual environments (Shavit-Cohen and Zion Golumbic, 2019).

The direction of alpha power modulation remains controversial. On average, in our study parietal alpha power showed a pattern with inversed similarity to the pupil size,
which means that in conditions with increased pupil size we found lower parietal alpha power. Since the modulation is very consistently pointing towards a larger pupil reflecting increased effort, from our current study we can only conclude that less listening effort leads to increased parietal alpha power. This is in conflict with some of the earlier findings (Obleser and Weisz, 2012; Obleser et al., 2012; Strauß et al., 2014), where it was argued that parietal alpha power reflects the inhibition of brain areas unrelated to the current task such that cognitive resources can be allocated to the listening task. One explanation why we have found an effect in the opposite direction could be that alpha does not have a linear but an inverted U-shape relation to attentional gain (Rajagovindan and Ding, 2010; Decruy et al., 2019). Depending on the operational point (i.e., the average difficulty) being on one or the other side of the inverted U’s maximum, a modulation of cognitive resources related to attention would result in one or the other direction of alpha power modulation. Another explanation of the current controversy might be the fact that alpha power is a dominant oscillation present in the several brain areas (Keitel and Gross, 2016). Its functional role might vary from one brain area to the other, such that its modulation may strongly depend on the task. Since listening to speech involves brain areas far beyond auditory cortex (Hickok and Poeppel, 2007), with the manipulation of hearing aid noise reduction and the SNR we might have tapped into the modulation of several cognitive resources. This would also explain why the topographic maps do not show a clear pattern of parietal modulation. Recently, the modulation of alpha power in frontal and temporal areas was connected to speech intelligibility (Dimitrijevic et al., 2017) and listening effort.
(Dimitrijevic et al., 2019). Here we restricted our hypothesis to a clearly defined parietal area.

Even though we hypothesized that both pupil size and parietal alpha power are measures for listening effort and we found inversely similar patterns on the average level, we did not find any trial-to-trial relationship between the two measures. However, this is in line with earlier studies with combined pupillometry and EEG (McMahon et al., 2016; Miles et al., 2017; Hjortkjær et al., 2018). Again, it brings up the question which aspect the two measures of listening effort reflect. As it was recently pointed out, the engagement into a listening task may depend on various cognitive resources and the individual limit of each single resource may limit the overall engagement of a listener (Herrmann and Johnsrude., 2019).

Importantly, we have found an effect of the talker on both measures which we interpret as a modulation of sustained listening effort. However, we cannot infer what caused this difference. Even though the stimuli were normalized, the subjectively reported difference in intelligibility could have been caused by a difference in energetic masking due to spectral differences. This would lead to increased effort at a comparably early stage of the neural encoding of the speech signal. However, if not fully restored, a noise degraded speech signal can also require more effort at later stages due to more demanding semantic processing (Rönnberg et al., 2013). Alternatively, the reported difference could also emerge from a difference in semantic complexity (Wendt et al., 2016), such that not earlier but only later stages of the neural encoding would demand more resources. Furthermore, differences in speech rate between the talkers may have
had an influence on the listening effort (Krause and Braida, 2002; Müller et al., 2019). An exploratory analysis of the modulation spectrum of the talkers indicated that talker 2 showed stronger modulation around 4 Hz, in the range where the rate of syllables is usually found (see Figure S3). As stated earlier, we did not infer the effect of gender from the effect of the talker, since we only presented one example of each gender. To infer on general effects of gender, a representative sample of various female and male voices should be presented, which would have gone far off the scope of this study. With our design, we could not provide further insight into these aspects, but we could show that both measures were modulated after participants consistently expressed intelligibility concerns.

One considerable limitation of our study is the generalizability of the hearing aid fitting strategy to recommended fitting strategies. Here we used non-individualized, closed tips for better control of the signal that enters the auditory pathway. However, in practice, patients with mild to moderate hearing loss (and intact hearing in the low frequency range) would have been fitted with open domes. Therefore, noise reduction would only affect higher frequencies, such that the overall effect of noise reduction might be reduced for patients with open fit. Future studies should address this to claim higher ecological validity. Furthermore, we only quantified the effect of noise reduction once on a dummy head. In future studies, in-situ measurements should be conducted to quantify the effect of noise reduction in the individual patient.
Conclusion

We show that the modulation of sustained listening effort (e.g. by hearing aid noise reduction) as indicated by pupil size and parietal alpha power can be studied under more ecologically valid conditions. We here show that hearing aid noise reduction lowers sustained listening effort as mainly indicated by the pupil size. While the interpretation of pupil size modulation is in line with the literature, the direction of alpha power modulation due to a varying listening effort stays controversial. In sum, our study further approximates real-world listening scenarios and supports the evaluation of the benefit of signal processing as applied in a modern hearing aid.
References


Available at: https://doi.org/10.1111/ejn.13855.


Keitel A, Gross J (2016) Individual Human Brain Areas Can Be Identified from Their Characteristic Spectral Activation Fingerprints. PLOS Biol 14:e1002498 Available at: https://doi.org/10.1371/journal.pbio.1002498.


Listening: The Influence of Sentence Intelligibility. Ear Hear 31 Available at: https://journals.lww.com/ear-hearing/Fulltext/2010/08000/Pupil_Response_as_an_Indication_of_Effortful.4.aspx.

Figure 1: Experimental design and audiograms. A) Loudspeaker setup: Six loudspeakers were used to present target and distractor speech signals in front of the participant (±22°, orange & blue) and background noise (±90° & ±150°, grey). The target talker and their locations were fully counterbalanced. The SNR was defined as sound pressure level at one of the frontal loudspeakers relative to the summed sound pressure level of the background loudspeakers. Speech signals: The background noise was presented for 5 seconds before the target and distractor speech signal were added for the following 33 seconds. Afterwards, participants were asked to answer a comprehension question. B) Audiometric thresholds (mean of both ears) of single subjects (grey lines) and averaged across participants (thick black line).

Figure 2: Behavioral performance of correctly answered questions. Black lines and grey shadows depict the model’s slope estimate and confidence interval. Dots & grey lines depict single participant means. From left to right, the first three panels depict the main effects of the factors noise reduction, signal-to-noise ratio (SNR) and talker. The three panels on the right depict the 2-way interactions.

Figure 3: Time courses of pupil size and parietal alpha power. Panels show the baselined time courses of pupil size (left) and alpha power (right). Upper panels show the thin-lined single participant time courses (averaged across conditions) and the bold-lined grand average time course. Lower panels show the across-participants averages per condition. Topographic map shows grand average alpha power change from baseline and the highlighted parietal channels that were selected for further investigation.

Figure 4: Pupil size change from baseline. Black lines and grey shadows depict the model’s slope estimate and confidence interval. Dots & grey lines depict single participant means. From left to right, the first three panels depict the main effects of the factors noise reduction, signal-to-noise ratio (SNR) and talker. The three panels on the right depict the 2-way interactions.

Figure 5: Alpha power change from baseline. Black lines and grey shadows depict the model’s slope estimate and confidence interval. Dots & grey lines depict single participant means. From left to right, the first three panels depict the main effects of the factors noise reduction, signal-to-noise ratio (SNR) and talker. The three panels on the right depict the 2-way interactions. Topographic maps depict the contrasts between the levels and interactions, respectively.
Fig 1

A

Loudspeaker setup

Talker 1
Talker 2
Babble noise

Level

SNR

Speech signals

Time [s]

B

Hearing threshold [dB HL]

Frequency [Hz]
Main effects

Noise reduction  
\( p = 0.136 \)

SNR  
\( p = 0.067 \)

Talker  
\( p < 0.001 \)

Interactions

Noise reduction & SNR  
\( p = 0.253 \)

Noise reduction & Talker  
\( p = 0.988 \)

SNR & Talker  
\( p = 0.116 \)

Fig 2
A Pupil size

B Parietal alpha power

Fig 3
Main effects

Noise reduction
p < 0.001

SNR
p = 0.108

Talker
p < 0.001

Interactions

Noise reduction & SNR
p = 0.021

Noise reduction & Talker
p = 0.848

SNR & Talker
p = 0.041

Fig 4