

# The Presence of Another Individual Influences Listening Effort, But Not Performance

Hidde Pielage,<sup>1,2</sup> Adriana A. Zekveld,<sup>1</sup> Gabrielle H. Saunders,<sup>3</sup> Niek J. Versfeld,<sup>1</sup>  
Thomas Lunner,<sup>2</sup> and Sophia E. Kramer<sup>1</sup>

## INTRODUCTION

**Objectives:** The aim of this study was to modify a speech perception in noise test to assess whether the presence of another individual (copresence), relative to being alone, affected listening performance and effort expenditure. Furthermore, this study assessed if the effect of the other individual's presence on listening effort was influenced by the difficulty of the task and whether participants had to repeat the sentences they listened to or not.

**Design:** Thirty-four young, normal-hearing participants (mean age: 24.7 years) listened to spoken Dutch sentences that were masked with a stationary noise masker and presented through a loudspeaker. The participants alternated between repeating sentences (active condition) and not repeating sentences (passive condition). They did this either alone or together with another participant in the booth. When together, participants took turns repeating sentences. The speech-in-noise test was performed adaptively at three intelligibility levels (20%, 50%, and 80% sentences correct) in a block-wise fashion. During testing, pupil size was recorded as an objective outcome measure of listening effort.

**Results:** Lower speech intelligibility levels were associated with increased peak pupil dilation (PPDs) and doing the task in the presence of another individual (compared with doing it alone) significantly increased PPD. No interaction effect between intelligibility and copresence on PPD was found. The results suggested that the change of PPD between doing the task alone or together was especially apparent for people who started the experiment in the presence of another individual. Furthermore, PPD was significantly lower during passive listening, compared with active listening. Finally, it seemed that performance was unaffected by copresence.

**Conclusion:** The increased PPDs during listening in the presence of another participant suggest that more effort was invested during the task. However, it seems that the additional effort did not result in a change of performance. This study showed that at least one aspect of the social context of a listening situation (in this case copresence) can affect listening effort, indicating that social context might be important to consider in future cognitive hearing research.

**Key words:** Copresence, Listening effort, Pupil dilation response, Social context, Speech perception.

**Abbreviations:** BPS, baseline pupil size; HI, hearing impairment; PPD, peak pupil dilation; SFI, social facilitation and impairment; SNR, signal-to-noise ratio; SRT, speech reception threshold.

(Ear & Hearing 2021;42;1577–1589)

<sup>1</sup>Amsterdam Universitair Medisch Centrum, Vrije Universiteit Amsterdam, Otolaryngology Head and Neck Surgery, Ear & Hearing, Amsterdam Public Health Research Institute, Amsterdam, the Netherlands; <sup>2</sup>Eriksholm Research Centre, Sneekkersten, Denmark; and <sup>3</sup>Manchester Centre for Audiology and Deafness, University of Manchester, Manchester, United Kingdom.

Copyright © 2021 The Authors. Ear & Hearing is published on behalf of the American Auditory Society, by Wolters Kluwer Health, Inc. This is an open-access article distributed under the terms of the Creative Commons Attribution-Non Commercial-No Derivatives License 4.0 (CCBY-NC-ND), where it is permissible to download and share the work provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal.

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and text of this article on the journal's Web site ([www.ear-hearing.com](http://www.ear-hearing.com)).

Humans have a fundamental need to create and maintain social connections (Lieberman 2013). Because of this need, human behavior and underlying cognitive processes have often been found to be affected by the social context of a situation (Guerin & Innes 1984; Baumeister et al. 2002; Gere & MacDonald 2010). Listening often occurs when communicating with others; therefore, it is reasonable to assume that many listening scenarios have a social context. It is very well possible that the listening process is influenced by that context. However, social context is often overlooked in cognitive hearing research that looks at speech perception in noise and listening effort. Advancing our knowledge in this area could help us identify social context properties that might have a negative or positive influence on listening and listening effort. This could be useful when further examining the cognitive burden of HI. Social context could be important to consider when developing new interventions and technologies for those affected with hearing impairment.

In their consensus paper on listening effort, Pichora-Fuller et al. (2016) touched upon the subject of social context and noted that “some listeners find that the intellectual and social benefits of listening and conversing increase motivation and add value to expending effort” (page 6). This is true for people with hearing impairment, who report that social connectedness is an important reason for them to spend effort on listening (Matthen 2016; Hughes et al. 2018). Furthermore, it is in agreement with findings by Beechey et al. (2020a, 2020b). The authors observed that the benefits of conversation added value to expending effort when they elicited natural conversations between test participants. They also argued that a distinction can be made between effort spent to *maintain a conversation* and effort spent to *just listen*. According to the World Health Organization's International Classification of Functioning, Disability and Health, and the adaptation of this framework to hearing, these operate on a different level of human functioning (Kiessling et al. 2003). Listening is the process of hearing with intent and attention. It is an activity. Communication requires the bidirectional transfer of information, meaning, or intent between two or more people (Kiessling et al. 2003). Following this line of reasoning, listening effort can be regarded as an individual cognitive process, which is not necessarily tied to effort that is allocated to maintain a conversation. The current study aimed to explore if listening as an individual process can be affected by social context. Specifically, one aspect of social context was considered, namely the presence of another participant (copresence) during a speech perception task.

Research considering the combination of social and cognitive psychology has found that many cognitive processes can be influenced by the social context of a task. For example, Pickett et al. (2004) observed a positive relationship between a high

need for social connectedness and the accuracy with which participants decoded socially relevant cues, such as vocal tone and facial expression. Based on these findings, the authors theorized that persons with high belongingness needs would spend additional effort on the perception of social cues. Similarly, DeWall et al. (2008) falsely told participants that their personality test scores indicated they would likely end up alone later in life and found this increased the participants' performance on tasks that were framed as an indicator of social success. The authors suggested this indicates that the participants tried to compensate for news of future social unfulfillment by spending additional effort on a task that, they believed, measured their social capabilities. These studies show that when social wellbeing is threatened, it can result in an increase of effort and the resultant performance. Such effects are useful as they help to maintain or improve social wellbeing. Similar effects are found for social evaluative threat. Theories on social evaluative threat suggest that humans are fundamentally motivated to preserve the social self. When one's social status or value is threatened through the potential of negative evaluations by others (regardless of whether evaluation actually occurs or not), this is responded to by a state of heightened arousal (Rohleder et al. 2007). This state is thought to reflect a preallocation of cognitive resources and helps to overcome future challenges (Dickerson et al. 2004).

Copresence has been found to influence effort and task performance as well. For example, it has been shown that many animal species (including humans) tend to perform and spend effort differently in the presence of a conspecific (Belletier et al. 2019). Copresence may decrease performance on difficult tasks and increase performance on easy tasks compared with tasks performed in isolation. This is known as social facilitation-and-impairment (SFI). While the cognitive origins of these phenomena are still debated, they might relate to increased motivation (McFall et al. 2009) or to attentional mechanisms (Monfardini et al. 2016; Steinmetz & Pfattheicher 2017).

Copresence should thus be considered a relevant factor that may influence someone's listening activity in real life. Currently, little is known how copresence influences someone's performance on a standard speech perception in noise task administered in the laboratory. It is also unknown how copresence influences

effort invested during such a task. While speech perception in noise tasks seem to better reflect daily life listening than clinical tests like pure-tone and speech audiometry (Kramer et al. 1996; Spyridakou & Bamiou 2015), they still lack ecological validity (Keidser et al. 2020) in that characteristics of the social context are not taken into account. Even though the real-life social context of listening is much more complex than the mere presence of another person, a relatively simple manipulation was preferred in this study because of the novelty of this aspect. We aimed to investigate the effect of copresence on speech perception in noise performance and effort invested during the test. There is some early work describing SFI in a listening context (Beatty 1980; Beatty & Payne 1984), but it did not consider changes in effort as measured by physiological outcome measures.

In an attempt to gain insight into cognitive processes involved in listening, hearing research started considering the concept of listening effort. The Framework for Understanding Effortful Listening defines listening effort as "the deliberate allocation of mental resources to overcome obstacles in goal pursuit when carrying out a [listening] task" (Pichora-Fuller et al. 2016, p. 5). Several physiological measures have been used to assess listening effort, of which pupillometry is the most widely used (Zekveld et al. 2018). The autonomic pupil dilation response reflects the balance between sympathetic and parasympathetic nervous system activity (Kahneman 1973). Although the relative contribution of both nervous systems is complex, part of the dilation response is thought to be indicative of mental effort (Eckstein et al. 2017).

The aim of this study was to manipulate copresence during a standard listening task to examine if this affected performance and the investment of effort, as measured by pupillometry. This was done by having participants perform the task both alone and in the presence of another individual (in a dyad). The difficulty of the task was manipulated by changing the target proportion of correctly repeated sentences, resulting in three difficulty conditions. Additionally, participants were presented with sentences they had to repeat (dubbed active listening) and with sentences they did not have to repeat (dubbed passive listening). For an overview of the experimental design see Figure 1 (right side). We hypothesized that pupil dilations and performance would be impacted when a listening task was completed in the presence of another individual relative to when completed alone.

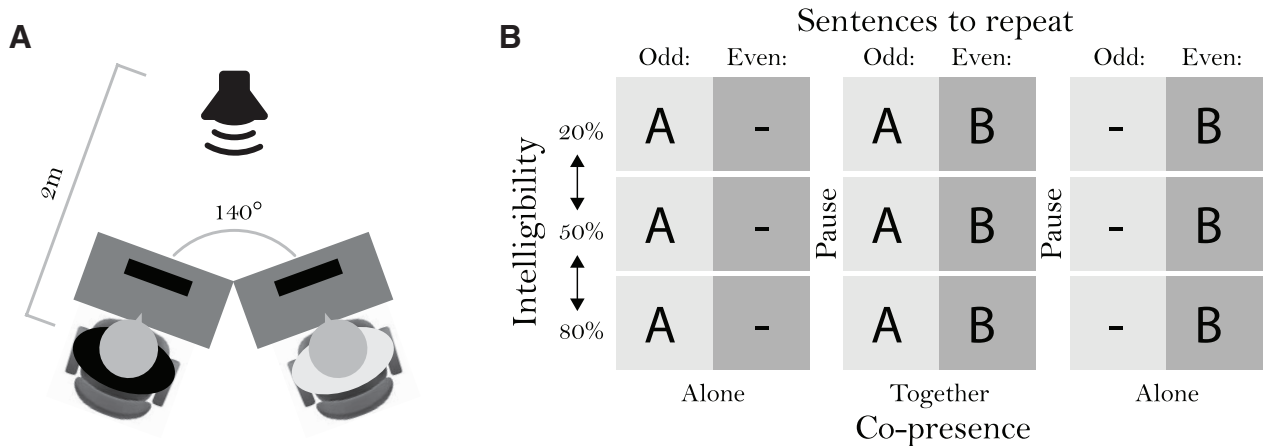


Fig. 1. A, Schematic display of the test setup when two participants were present. Participants were oriented at a 140° angle from each other and were placed so that the center of their head was 2-m away from the speaker. B, Schematic representation of a test session. Participant A always repeated odd-numbered sentences, while participant B always repeated even-numbered sentences. They did this both alone and together at three different intelligibility levels, the order of which was randomized.

## MATERIALS AND METHODS

### Participants

Thirty-four participants were recruited at the Vrije Universiteit and Amsterdam University Medical Center to participate in this study. Data from five participants had to be discarded (see Results section), leaving 29 participants (8 males) with ages ranging from 18 to 46 years (mean = 24.1, SD = 6.1). All participants had normal hearing, defined as pure-tone thresholds of  $\leq 20$  dB HL at all octave frequencies between 0.5 and 4 kHz, in both ears. They had no history of neurological or psychological disorders, diabetes, or eye-related diseases. After their participation, participants received 15 euros for their time. All participants provided written consent. Approval for this study was granted by the medical ethical research committee of the Amsterdam University Medical Center, location VUmc under reference number 2018.308.

The sample size was based on a power analysis accounting for a  $2 \times 3$  repeated measures ANOVA design and an effect size of partial  $\eta^2 = 0.30$ , which was found for pupil reactivity in response to social inclusion (Slegers et al. 2017). It revealed that 7 participants were needed to achieve sufficient experimental power. Additional participants were added for redundancy and to account for uncertainty of the effect size of the social presence manipulation. Because of multiple missing data points and the added fixed factor (whether or not participants had to repeat a sentence), it was later opted to use linear mixed effect modeling. Power was reassessed *post hoc* using the Simr package (Green & Macleod 2016) in R for data simulation. Parameters for simulating the data were derived from fitting the planned model to data from the first eight participants. Slopes of the fixed effects were reduced to estimate the required power to detect smaller effects. Running simulations revealed that at least 28 participants were required to reach sufficient power, justifying the number of recruited participants.

### Measurements

**Speech Task** • Speech performance was measured with an adaptive speech reception threshold (SRT) task (Plomp & Mimpen 1979). Participants were required to listen to and repeat Dutch sentences spoken by a female voice. The sentences were taken from earlier work by Versfeld et al. (2000) and varied in duration, from 1.4 to 2.0 seconds. One example of these sentences is “Ze moeten morgen weer vroeg op,” which translates to “They have to get up early again tomorrow.” To assess listening in noise performance, sentences were masked by a stationary noise masker, which started 3 seconds before sentence onset and ended 3 seconds after sentence offset. The masker, also taken from Versfeld et al. (2000), was designed such that its spectral shape was similar to the long-term average of the sentence material (Festen & Plomp 1990). Participants completed six blocks of 36 sentences, of which they had to repeat half of the sentences. Sentence lists were compiled before the start of the study and randomly assigned to experimental conditions. Participants heard no sentence more than once.

The difficulty within a block was manipulated using Kaernbach’s (1991) adaptive procedure to estimate 20% (step-size of 0.8 dB up and 3.2 dB down), 50% (step-size of 2 dB up and down), or 80% (step-size of 3.2 dB up and 0.8 dB down) sentence intelligibility. The signal-to-noise ratios (SNRs) of the first trial in each block were based on a rough estimation of the

SRTs expected for this population, as derived from Ohlenforst et al. (2017). For 20% intelligibility, the first trial started at  $-8$  dB SNR, for 50% this was at  $-6$  dB SNR, and for 80% this was at  $-4$  dB SNR. Trials were marked correct if the full sentence was repeated correctly. Compared with word scoring, full sentence scoring generally results in relatively high SNRs, which are more representative of realistic conditions (Smeds et al. 2015). Target sentence and masker intensity were changed depending on the SNR, while the overall sound pressure level remained fixed at 65 dB SPL. All participants completed one block of each intelligibility level once while doing the test alone, and once while doing it in the presence of another participant.

The task required participants to alternate between repeating and not repeating a target sentence to create both “active” (when participants were tasked to repeat a sentence) and “passive” (when they were not) listening situations. When two participants were present, they took turns repeating the sentences (each repeating 18 sentences per block). When a participant was alone, every other sentence was unrepeated. To avoid artifacts in the pupil data, no visual or auditory indicators were used to inform participants when they had to repeat a sentence and when not to. This means participants had to monitor for themselves which sentences to repeat. On the few occasions that participants lost track, the experimenter intervened. When participants were tested alone, the SNRs of the passive listening trials varied around the average SRT expected at that level of intelligibility for this population using step sizes like the ones described above. The exact pattern of SNRs for these sentences was identical for all participants and was derived from pilot data. Performance was assessed using SRTs, which were calculated by averaging the SNRs at which sentences 5 to 18 were presented. The first four SNRs were not included because the adaptive procedures had not stabilized yet in these trials.

**Pupil Dilation** • During each trial, pupil dilations were recorded from the onset of the masker until the offset of the three seconds of noise after the sentence. From the average of these recordings, baseline pupil size (BPS) and peak pupil dilation (PPD) were extracted. PPD in this context is considered to reflect a change in effort caused by listening to the target sentence in noise (Zekveld et al. 2018) and was therefore used as an objective measure of listening effort. The data acquired during the 3 seconds of noise after target sentence offset were included in the analysis to account for the slow pupil response (Winn et al. 2018). This interval did not include the window in which the participants responded.

**Subjective Rating** • To gain insight into the subjective experience of participants, four rating scales from Zekveld et al. (2010) were presented to participants after each block of sentences. Participants were asked to rate their perceived effort and performance, how often they gave-up listening, and how difficult the task was for them. Answers were given on a scale from 1 to 10, with one decimal precision. All scales were supplemented with labels corresponding to the numbers on the scale. For the full scales with accompanying labels, see Appendix A in Supplemental Digital Content 1, <http://links.lww.com/EANDH/A797>.

**Connectedness Questionnaires** • It was expected that some dyads might connect with one another more than others (e.g., through short conversations between blocks of sentences), adding additional variability to the copresence manipulation. To assess if such variability affected the data, a connectedness questionnaire was added, as well as a scale measuring overlap



between the self and the other. The used questionnaire was derived from the Connectedness for Groups questionnaire by Leach et al. (2008), which was initially intended to study feelings of connectedness within groups. Five of the original 14 items in the questionnaire were used. Items were chosen based on their applicability to dyads. Participants rated their experience of connectedness to the other participant by rating their agreement with each statement on a 7-point scale ranging from one (strongly disagree) to seven (strongly agree). The scales were labeled at the extremes and participants could choose any integer from 1 to 7. A lower score indicated less connectedness. An example statement is “Ik voel mij verbonden met de andere participant,” which translates to “I feel connected to the other participant.” A full list of the “connectedness for groups” questionnaire statements can be found in Appendix A in Supplemental Digital Content 1 (<http://links.lww.com/EANDH/A797>). Item scores on the questionnaire were averaged to produce a single score per participant. The Other-in-Self scale (Aron et al. 1992) was also used and had participants select which of five images best displayed their feeling of connectedness to the other participant, using images of two circles that overlapped in different amounts. The scale was coded by assigning ascending values (1 to 5) to each image where lower values corresponded to less overlap between the other and self.

## Procedure

Participants were tested in dyads in partly overlapping test sessions. Participants were paired based solely on availability and assurance that they were unacquainted with each other. After signing informed consent, participant A completed a pure-tone audiogram at 500, 1000, 2000, and 4000 Hz to check his/her hearing status and then completed the SRT task. For this task, the participant was seated behind the left of two desks in a large room. The room was treated to minimize reverberation and noise from outside and was lit by eight fluorescent bars at approximately 300 lx. The two desks were oriented at a 140° from each other and placed at an equal distance of 2 m from a loudspeaker (see Figure 1, left). The participant was instructed to repeat each *odd numbered* sentence (see Figure 1, right) after the masker noise ended while keeping their gaze fixed on the loudspeaker. The experimenter controlled the SRT task and monitored the pupil recordings from an adjacent room. A microphone allowed the experimenter to listen to the sentence repetitions. Participants were aware that the experimenter could hear but not see them. Participant A completed three blocks of sentences (20%, 50%, and 80% intelligibility). After each block, the participant was asked to complete the subjective rating scales.

Next, participant A had a 15-minute break while participant B gave written informed consent and completed the audiogram, after which, he/she was instructed on the SRT task but was told to repeat only *even numbered* sentences. Participants A and B then completed three blocks of sentences and the subjective rating scales while seated in the testing room together. Following completion of the three blocks, participant B had a break while participant A completed the connectedness for groups and other-in-self questionnaires, after which his/her participation was complete. Participant B then completed three blocks of sentences and subjective rating questions while alone, still only repeating *even numbered* sentences. Finally, Participant B completed the connectedness for groups and other-in-self questionnaires, which concluded the experiment.

## Apparatus

The target sentences and masker (wave files: 44.1 Hz, 16 bit) were presented through an Optiplex 780 Analog Devices ADI 198× Integrated HD audio soundcard and amplified by a Samson Servo 4120 amplifier. Sound was played from a Tannoy Reveal speaker. The setup was calibrated such that the overall sound level was kept at a constant sound pressure of 65 dB SPL near the participants. Calibration was done by measuring sound pressure levels at both participants' positions. As shown by Aguirre et al. (2019), having two participants present is not expected to impact sound levels at the ears of each. For pupillometry, one SMI Red eye-tracker and one SMI Red Mobile eye-tracker were used (both produced by SensoMotoric Instruments, Germany). Pilot data revealed no systematic differences between the two trackers. Both trackers recorded pupil data at 60 Hz.

## Pupil Data Preprocessing

Pupil data were preprocessed in accordance with the pupillometry practices described by Zekveld et al. (2010) and Winn et al. (2018). Preprocessing was applied in a trial-by-trial fashion. Pupil data that corresponded to the period between the onset and offset of the masker will further be referred to as a pupil trace. The data were split into active and passive trials, resulting in blocks of 18 active listening and 18 passive listening traces per condition. The first four pupil traces of each block were removed for both active and passive listening. BPS in early trials is thought to be relatively unstable and their respective traces are therefore commonly excluded from further analysis (Wendt et al. 2018). We removed 4 traces so that it would coincide with the amount of excluded SNR values. From the remaining 14 traces, all zero values were coded as blinks and replaced through linear interpolation. Interpolation was performed between the fifth sample before a blink and the eighth successive nonzero sample after that blink. If more than 25 percent of a trace consisted of zeros, it was removed from further analysis. Less than one percent of all trials, both active and passive, were removed this way.

Next, the traces were smoothed using a five-point moving average filter. BPS was calculated by averaging all pupil data corresponding to the last second of noise before target sentence onset. The traces were cut from the onset of the target sentence to the offset of the masker, and a baseline correction was applied by subtracting BPS from all values in the remaining trace. Finally, traces within a block were averaged, and from that averaged trace, the maximum value was extracted as the PPD. By default, pupil traces were taken from the right eye, since some research suggests that dilation of the right pupil is more sensitive to cognitive effort (Liu et al. 2017; Wahn et al. 2017). One participant had an ocular abnormality in the right eye, so data from the left eye were used instead.

## Statistical Analysis

Pupil data, SRT scores, and the subjective rating scores were modeled as dependent variables using linear mixed effect modeling, as performed by the lme4 package (using default optimizers) (Bates et al. 2015) in R language (R Core Team 2019). The independent variables “intelligibility” (20%, 50%, or 80%) and “copresence” (alone or together with another participant) were always modeled as fixed effects. The fixed factor “task” was added to the pupil data, which discriminates between active and passive listening data. When interpreting the effects of task

one should keep in mind that SRTs of active and passive listening might differ slightly, as separate adaptive procedures were used to determine SNRs for each. Finally, the random effect structures used in the models were determined by using the “step” function from the lmerTest package (Kuznetsova et al. 2017). This function selects the most appropriate random effects structure based on the Akaike Information Criterion. Since no trial level data were available to the models, only slopes at the participant level could be included in the model. The exclusion of random slopes suggested that variance across participants for that variable was insufficient to justify the additional model complexity as a result of adding the slope to the random effects structure. To confirm the adequacy of the methods used, all outcome variables were checked for linearity and homogeneity. In case of nonlinearity, log-transformation was applied.

From the fitted models, fixed effect parameter estimates ( $\beta$ ) and their confidence intervals (CIs) were extracted and compiled into a table. Because the used intelligibility levels have a natural order, backward difference contrast coding was used to extract parameter estimates that compared each intelligibility level with their prior adjacent level (i.e., 50% compared with 20% and 80% compared with 50%). Copresence was coded so that the fixed effect estimate represented the change from the alone conditions to the together condition. The task estimate represented the change from active listening to passive listening.

While the fixed effect estimations and CIs were used to interpret the direction and size of the effects, a more formal test was used to assess statistical significance. More specifically, significance of the effects and interactions between them was assessed using a type III analysis of variance (ANOVA) together with Satterthwaite’s approximation for degrees of freedom. This test was performed using the lmerTest package’s functionality (Kuznetsova et al. 2017). Satterthwaite’s approximation is thought to yield comparatively conservative Type-I error rates for mixed-effect modeling (Luke 2017).

A correlation analysis was used to check if connectedness was significantly related to PPD. The difference in PPD between the alone and together conditions was calculated for all active listening conditions and averaged over intelligibility conditions. Then, Pearson’s correlation coefficient was used to evaluate if there was a significant relationship between the difference score and the connectedness for groups or other-in-self questionnaires.

**RESULTS**

The data from 5 out of 34 participants were excluded from all analyses due to equipment failing to save data or because the participant’s eyes were difficult to track, resulting in noisy pupil data. One additional participant was excluded because of prior experience with the sentence material, which was only discovered after the experiment had concluded. Furthermore, 7 conditions across 4 of the remaining participants were excluded due to poor pupil data quality. Among these conditions, 2 were performed alone at 20% intelligibility, 1 alone at 50%, 2 together at 20%, and 2 together at 50%.

To visualize the pupil traces, a grand trace mean was calculated and plotted for each condition. These traces can be found in Figure 2. A peak in pupil dilation is clearly present in the active listening traces approximately 2.5 second after sentence onset. While the passive listening traces followed a similar trend, they did not have a notable peak in the pupil dilation response. Furthermore, it seems that during active listening the together conditions generally produced larger responses than the alone conditions did for the same intelligibility.

**Baseline Pupil Size**

BPS was analyzed to check if baseline measures were affected by the task conditions. A model was fitted to BPS including intelligibility, copresence, and task as fixed effects (including interactions). Besides the random intercept term,

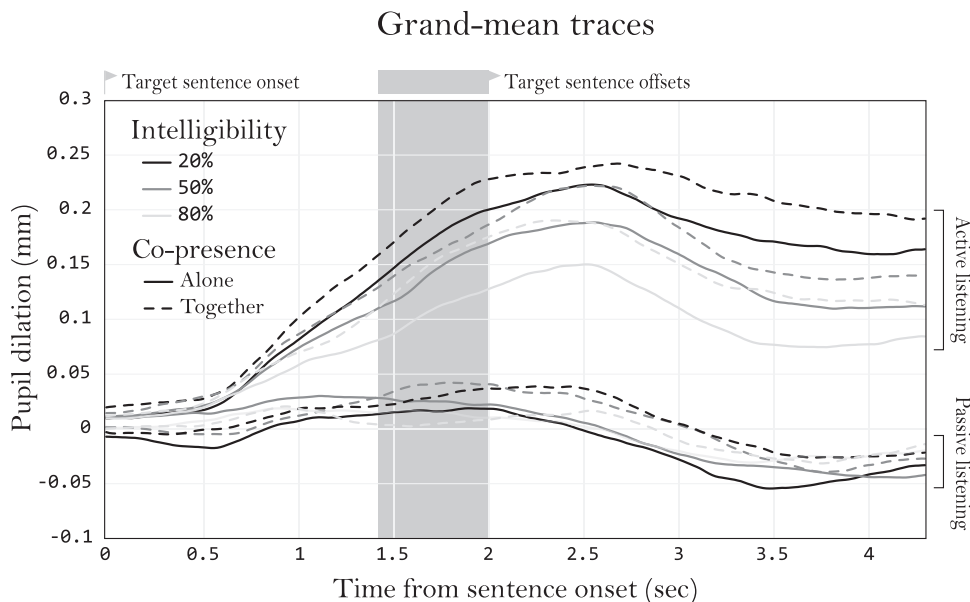


Fig. 2. Baseline corrected mean pupil traces of each condition averaged over participants. Zero seconds on the x-axis corresponds to target sentence onset. The highlighted area indicates the period during which the sentences (with variable durations) ended. Masker noise was playing over the whole duration of this graph. The plotted lines are cut at the duration of the shortest sentence (1.4 s) plus 3 s (postsentence noise). Participants were asked to respond after the offset of the noise, which ranged between 0.0 and 0.6 s after this plot, depending on the target sentence’s duration.

**TABLE 1. Fixed effect estimates ( $\beta$ ) for BPS, together with corresponding CIs**

	BPS	
	B	CI (95%)
(Intercept)	+4.624	+4.360 to +4.889
Intelligibility (A)	-0.048	-0.156 to +0.060
Intelligibility (B)	-0.086	-0.179 to +0.006
Copresence	+0.025	-0.070 to +0.119
Task	+0.000	-0.039 to +0.039
Intelligibility (A) $\times$ Copresence	-0.062	-0.160 to +0.037
Intelligibility (B) $\times$ Copresence	-0.014	-0.110 to +0.082
Intelligibility (A) $\times$ Task	-0.011	-0.107 to +0.086
Intelligibility (B) $\times$ Task	-0.016	-0.111 to +0.079
Copresence $\times$ Task	-0.004	-0.060 to +0.051
Intelligibility (A) $\times$ Copresence $\times$ Task	-0.007	-0.144 to +0.131
Intelligibility (B) $\times$ Copresence $\times$ Task	+0.011	-0.124 to +0.146

The backward difference contrast coding resulted in two beta estimates for intelligibility, labeled contrast A (50% compared with 20%) and contrast B (80% compared with 50%). Copresence represents the difference in BPS when doing the task in the presence of another participant, compared with alone. The task estimate represents the change in BPS when actively listening, compared with passive listening. BPS, baseline pupil size; CIs, confidence intervals.

only subject-level random slopes for intelligibility and copresence were found to be appropriate and thus included in the random effects structure. Schematically the final model could be written as (lme4 notation):

$$\text{BPS} \sim \text{Intelligibility} * \text{Copresence} * \text{Task} + (1 + \text{Intelligibility} + \text{Copresence} | \text{Participants})$$

Visual inspection of the plotted residuals revealed no obvious deviations from the assumptions of linearity and homoscedasticity. Compared with the null model, the fitted model significantly improved predictions [ $\chi^2(11) = 45.67, p < 0.01$ ]. Fixed effect estimates and their CIs can be found in Table 1. Only intelligibility was found to have a significant main effect [ $F(2, 27) = 10.84, p < 0.01$ ], where higher intelligibility percentages resulted in lower BPS.

An exploratory analysis was done to check if the order in which the copresence conditions were performed influenced the data. A similar model was used to the one described above, but here the fixed factor “order” that coded for the copresence condition order was added. Therefore, order had two levels: “Alone First” and “Together First.” Task was removed from the model as a fixed factor since BPS seemed to be highly stable between active and passive listening. The model revealed a significant interaction between copresence and order conditions [ $F(1, 27) = 13.28, p < 0.01$ ], but no main effect of order. When visualizing this interaction (Figure 3, left) it seems that participants who started alone first had lower BPS in the together condition (compared with alone), and that participants who started together had lower BPS in the alone condition (compared with together). Even though no main effect of order was found, the plotted data indicated that BPS was generally higher

### Order effects

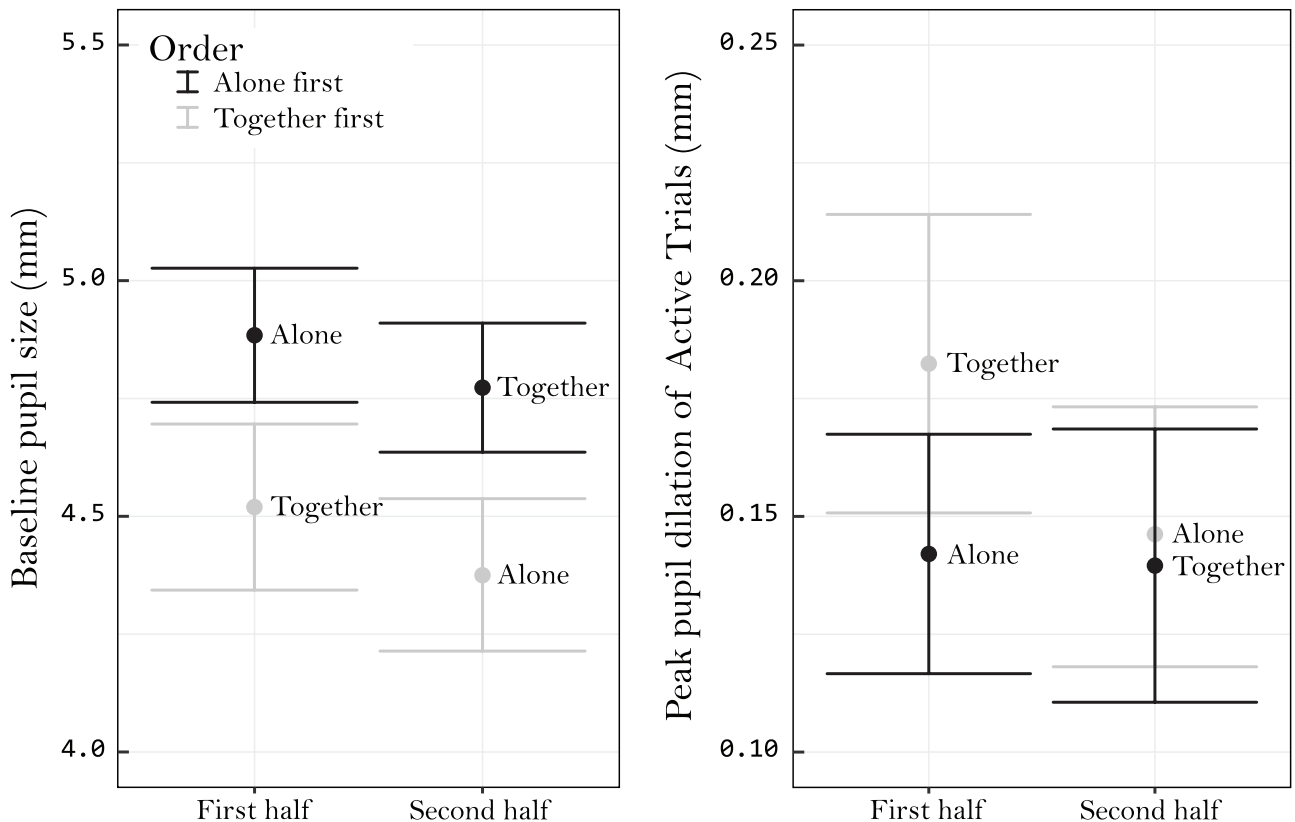


Fig. 3. Visualization of the effect caused by the order in which the copresence conditions were performed. BPS data have been averaged over intelligibility and task (passive and active listening). PPD data are averaged over intelligibility and only contains active data. Both graphs visualize the 95% confidence intervals around each average. BPS, baseline pupil size; CIs, confidence intervals; PPD, peak pupil dilation.

## Peak Pupil Dilation ±95% CI

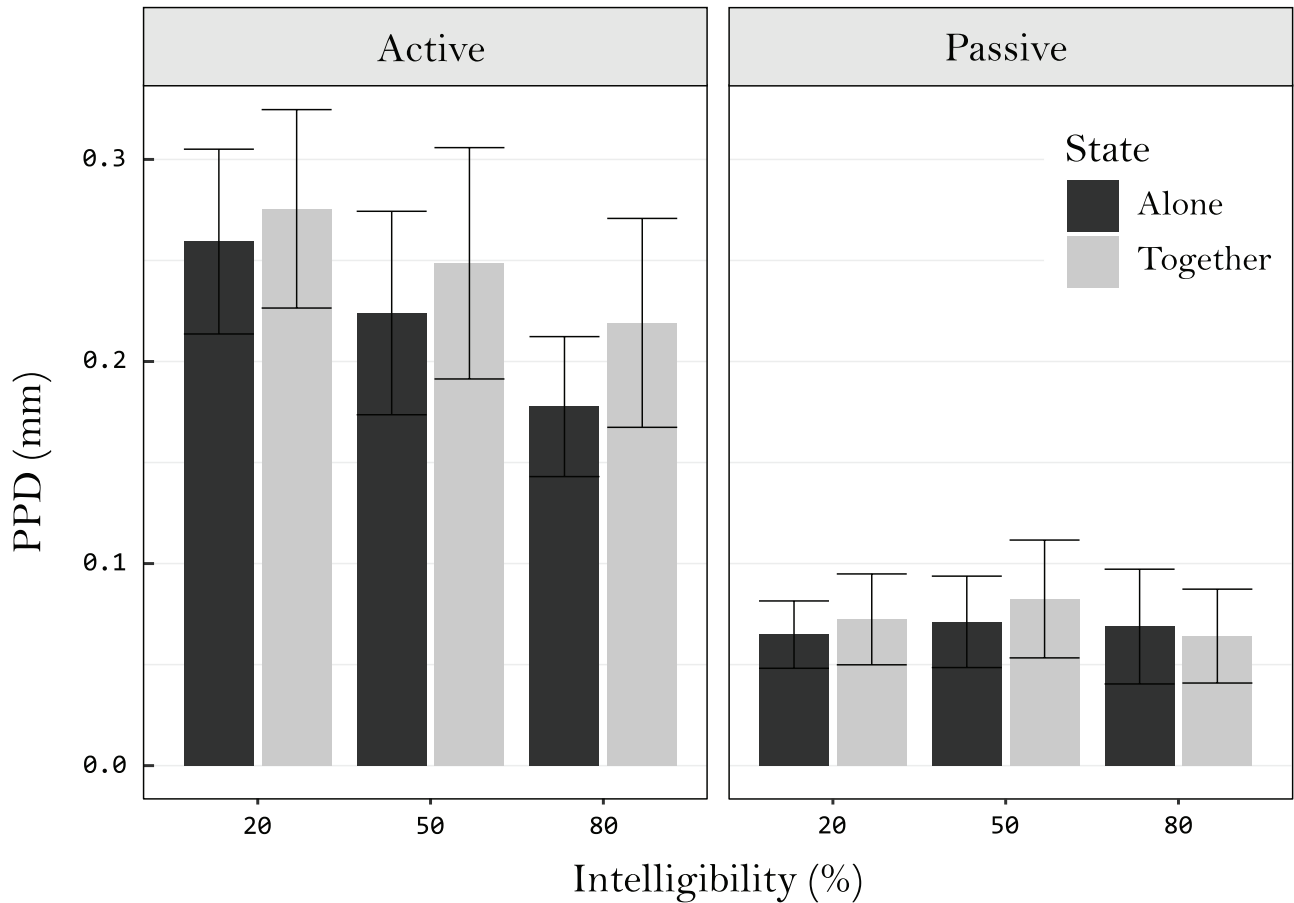


Fig. 4. Plotted means of peak pupil dilations (baseline corrected) in each condition, together with their 95% CI. CIs, confidence intervals.

for participants who began the experiment with the alone condition. Order did not interact with intelligibility.

### Peak Pupil Dilation

Average PPD values are plotted in Figure 4, together with their 95% CIs. The model fitted to PPD included intelligibility, copresence, and task as fixed effects (including interactions) together with the random intercept term and subject-level random slopes for copresence and task. Schematically the model could be written as (lme4 notation):

PPD ~ Intelligibility \* Copresence \* Task + (1 + Copresence + Task | Participants)

Again, no obvious deviations from homoscedasticity and normality were found when inspecting residual plots. The fixed effect estimations of this model can be found in Table 2. The fitted model significantly improved data predictions compared with the null model [ $\chi^2(11) = 87.16, p < 0.01$ ]. Main effects were found for intelligibility [ $F(2, 239) = 11.88, p < 0.01$ ], copresence [ $F(1, 27) = 4.42, p = 0.05$ ], and task [ $F(1, 28) = 98.33, p < 0.01$ ]. Higher intelligibility was associated with lower PPD, which can be confirmed by visually inspecting the data (Figures 2 and 4) and the parameter estimates (Table 2). The fixed effect estimation for copresence suggests that doing the task together with another participant, relative to alone, was associated with

increased PPDs. The task estimate suggests that passive listening was associated with a reduction in PPD as compared with active listening. Besides its main effect, task was also found to interact with intelligibility [ $F(2, 239) = 10.19, p < 0.01$ ], suggesting that the effects of intelligibility on PPD were greater during active listening than during passive listening, which was further tested in the analysis below. Note that passive listening did not produce a notable peak in pupil dilations (Figure 2). While this differs from active listening, PPD was extracted using the same method and was still regarded to represent maximum task-evoked processing load during a trial (Zekveld et al. 2011).

The interaction effect between active and passive listening and intelligibility indicated smaller intelligibility effects on PPD during the passive listening conditions. To explore this further, separate models were fitted to the active and passive pupil data. These models included fixed factors for intelligibility and copresence, together with their interaction. Random slopes for intelligibility and copresence were included in the active data model, but no random slopes were found to be appropriate for the passive data model. For the active data, the model was found to significantly improve data predictions [ $\chi^2(5) = 21.56, p < 0.01$ ] beyond the null model. Similar to using the full data set, significant effects were found for intelligibility [ $F(2, 34) = 13.17, p < 0.01$ ] and copresence [ $F(1, 28) = 4.36, p = 0.05$ ]. Fixed effect estimations for the active data (Table 3)



**TABLE 2. Fixed effect estimates ( $\beta$ ) for PPD, together with corresponding CIs**

	PPD	
	$\beta$	CI (95%)
(Intercept)	+0.220	+0.181 to +0.259
Intelligibility (A)	-0.038	-0.068 to -0.007
Intelligibility (B)	-0.044	-0.074 to -0.015
Copresence	+0.028	+0.008 to +0.049
Task	-0.152	-0.187 to -0.118
Intelligibility (A) $\times$ Copresence	+0.003	-0.040 to +0.046
Intelligibility (B) $\times$ Copresence	+0.018	-0.024 to +0.060
Intelligibility (A) $\times$ Task	+0.044	+0.001 to +0.087
Intelligibility (B) $\times$ Task	+0.043	+0.001 to +0.085
Copresence $\times$ Task	-0.023	-0.047 to +0.002
Intelligibility (A) $\times$ Copresence $\times$ Task	-0.002	-0.063 to +0.059
Intelligibility (B) $\times$ Copresence $\times$ Task	-0.034	-0.094 to +0.026

The backward difference contrast coding resulted in two beta estimates for intelligibility, labeled contrast A (50% compared with 20%) and contrast B (80% compared with 50%). Copresence represents the difference in BPS when doing the task in the presence of another participant, compared with alone. The task estimate represents the change in PPD when actively listening, compared with passive listening. CIs, confidence intervals; PPD, peak pupil dilation.

**TABLE 3. Fixed effect estimates ( $\beta$ ) for PPDs corresponding to active listening, together with CIs**

	Active PPD	
	B	CI (95%)
(Intercept)	+0.220	+0.182 to +0.259
Intelligibility (A)	-0.038	-0.072 to -0.003
Intelligibility (B)	-0.045	-0.074 to -0.015
Copresence	+0.029	+0.002 to +0.055
Intelligibility (A) $\times$ Copresence	-0.001	-0.041 to +0.040
Intelligibility (B) $\times$ Copresence	+0.020	-0.020 to +0.059

The backward difference contrast coding resulted in two beta estimates for intelligibility, labeled contrast A (50% compared with 20%) and contrast B (80% compared with 50%). Copresence represents the difference in PPD when doing the task in the presence of another participant, compared with alone. CIs, confidence intervals; PPD, peak pupil dilation.

were comparable to those of the model fitted to the full dataset. Conversely, for the passive data, the model did not significantly improve upon the null model [ $\chi^2(5) = 2.55, p = 0.77$ ], suggesting that there is no evidence that intelligibility and copresence affected the pupil during passive listening. To assess if the passive pupil data reflected any dilations beyond baseline, a t-test was used to compare absolute PPDs to BPS during passive listening. The test revealed that passive PPDs differed significantly from the BPS values [ $t(642) = -2.61, p < 0.01$ ], indicating that on average, the PPD of passive trials was larger than the BPS.

**TABLE 4. Connectedness questionnaire scores median and IQR**

Questions	Median	IQR
I feel connected to the other participant <i>Ik voel mij verbonden met de andere participant</i>	4	3
I feel a solidarity with the other participant <i>Ik voel mij solidair met de andere participant</i>	4	3
I feel involved with the other participant <i>Ik voel mij betrokken met de andere participant</i>	4	3
I share a lot of similarities with the other participant <i>Ik heb veel overeenkomsten met de andere participant</i>	4	2.25
Me and the other participant look alike <i>Ik lijk veel op de andere participant</i>	3	2
Other-in-self score	2.5	1

Top: connectedness for groups questions. Bottom: other-in-self scores. IQR, interquartile range.

However, this finding should be interpreted with care since the amplitudes of the pupil responses during passive listening were relatively small.

Similar to the analysis of BPS, another model was used to assess if the order in which the copresence conditions were performed affected the data. Only data from the active listening conditions were used since copresence did not affect the pupil during passive listening. Again, a significant interaction was found between copresence and the order of the copresence conditions [ $F(2, 34) = 13.47, p < 0.01$ ]. Visualization of this interaction (Figure 3, right), suggested that the change in PPD between the alone and together conditions was more apparent for participants who started the experiment together with another participant, compared with participants who started the experiment in the alone condition. Indeed, a post hoc two-sample t-test revealed a significant difference in PPD during the first half of the experiment between participants who started the experiment alone and participants who started together [ $t(73) = -2.28, p = 0.03$ ]. No difference was found between these two groups during the second half of the experiment [ $t(80) = 0.13, p = 0.89$ ], adding to the idea that copresence only had an effect if the second participant joined at the start of the experiment. Following the advice of Armstrong (2014), no multiple testing correction was applied to these t-tests. No other effects of order were found as it did not have a significant main effect and did not interact with intelligibility.

### Connectedness Questionnaire

Descriptive statistics for the modified Connectedness for Groups questionnaire and the Other-in-Self scale are shown in Table 4. The median scores indicate that participants felt somewhat connected to the other participant. The interquartile ranges suggest there was some variation in mean connectedness scores for the modified Connectedness for Groups questionnaire, but very little for the Other-in-Self Scale. Pearson's correlation coefficient was used to assess the relationship between the two scores and the difference in PPD between the together and alone conditions, as averaged over the intelligibility conditions. Neither the Connectedness for Groups questionnaire [ $r(21) = 0.39, p = 0.06$ ], nor the Other-in-Self scale [ $r(21) = 0.23, p = 0.29$ ] correlated significantly with the differences in PPD between the copresence conditions.

### Speech Reception Thresholds

Mean SRTs across conditions are plotted in Figure 5. A model including intelligibility and copresence as fixed effects (including interaction), together with the random intercept, was fitted to the SRTs. Subject-level random slopes for intelligibility and copresence were not found to be appropriate for this model. Schematically the model is described as (lme4 notation):



Speech Reception  
Thresholds ±95% CI

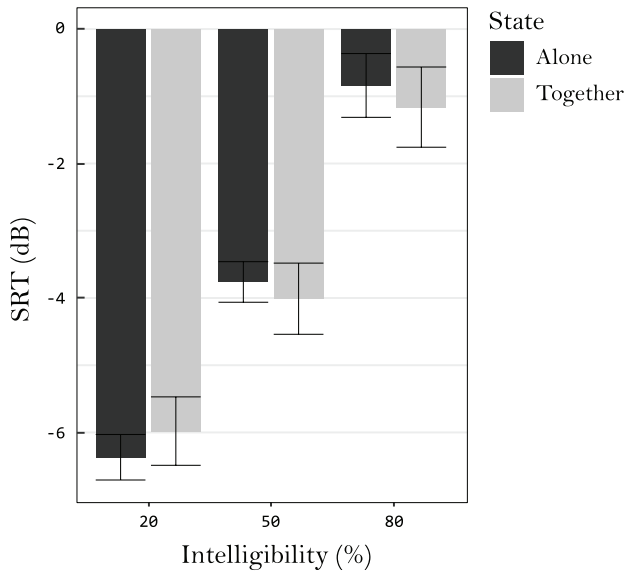


Fig. 5. SRT means for each intelligibility and copresence condition, together with their 95% confidence intervals. CIs, confidence intervals; SRTs, speech reception thresholds.

$$\text{SRT} \sim \text{Intelligibility} * \text{Co-presence} + (1 | \text{Participants})$$

Task was not included as a fixed factor since SRTs were not available for the passive listening condition. Fixed effect estimations of this model can be found in Table 5. The model significantly improved data predictions compared with the null model [ $\chi^2(5) = 224.48, p < 0.01$ ]. However, only intelligibility was found to hold statistical significance [ $F(2, 136) = 247.90, p < 0.01$ ]. Copresence did not affect the SRTs. There was also no significant interaction between copresence and intelligibility.

**Subjective Ratings**

The subjective effort, performance, giving-up, and difficulty ratings were each modelled as dependent variables. The models fitted to these ratings hosted intelligibility and copresence as fixed factors together with their interaction term. Task was not included since no distinction was made between active and passive listening when acquiring the subjective ratings. Random slopes for intelligibility and copresence were added to the model for all subjective ratings except giving-up. The step function determined that the giving-up scale was invariant across subjects for both copresence and intelligibility; therefore, only the random intercept was included in the giving-up model. When fitted to the effort [ $\chi^2(5) = 67.00, p < 0.01$ ], performance [ $\chi^2(5) = 87.31, p < 0.01$ ], and difficulty [ $\chi^2(5) = 66.79, p < 0.01$ ] ratings, the models significantly improved predictions compared with their respective null models. For these subjective ratings, intelligibility had a significant effect on the data [effort:  $F(2, 45) = 89.98, p < 0.01$ ; performance:  $F(2, 38) = 161.78, p < 0.01$ ; and difficulty:  $F(2, 83) = 104.27, p < 0.01$ ], but copresence did not. Fixed effect estimations of the models can be found in Table 6. The estimations suggest that participants perceived less effort, a higher performance, less giving-up, and a lower

**TABLE 5. Fixed effect estimates ( $\beta$ ) for SRTs, together with confidence intervals (CIs)**

	SRT	
	$\beta$	CI (95%)
(Intercept)	-3.658	-3.931 to -3.385
Intelligibility (A)	+2.606	+1.953 to +3.258
Intelligibility (B)	+2.926	+2.285 to +3.566
Copresence	-0.062	-0.436 to +0.313
Intelligibility (A) × Copresence	-0.637	-1.564 to +0.290
Intelligibility (B) × Copresence	-0.073	-0.983 to +0.838

The backward difference contrast coding resulted in two beta estimates for intelligibility, labeled contrast A (50% compared with 20%) and contrast B (80% compared with 50%). Copresence represents the difference in SRT when doing the task in the presence of another participant, compared with alone. CIs, confidence intervals; SRTs, speech reception thresholds.

difficulty when intelligibility was increased. Inspection of residual plots revealed no deviations from the assumptions of linearity and homoscedasticity.

Inspection of the subjective rating data revealed that the giving-up rating scale data were nonnormally distributed. To account for this nonnormality, the giving-up data were log-transformed. When fitting the previously described model to the log-transformed data, it significantly improved predictions compared with the null model [ $\chi^2(5) = 153.45, p < 0.01$ ]. A significant effect for both intelligibility [ $F(2, 170) = 122.35, p < 0.01$ ] and copresence [ $F(1, 170) = 3.93, p < 0.05$ ] was found. The parameter estimates, which can be found in Table 6, suggest that as intelligibility increased, participants perceived they gave-up less on listening. Furthermore, the estimate for copresence suggests that participants experienced more giving-up when the task was performed in the presence of the other participant. However, this effect is relatively small and should be interpreted with care as its lower confidence interval approaches zero.

**DISCUSSION**

This study assessed whether performing an adaptive listening task together with another participant, compared with doing it alone (copresence), influenced effort (as measured using PPDs), and listening performance (as measured using SRTs). For both PPDs and SRTs, it was also assessed if there was an interaction between copresence and sentence intelligibility (20%, 50%, and 80%). Both active listening (in which participants were required to repeat sentences aloud) and passive listening (in which no response had to be given) were examined. The results showed a significantly larger PPD during listening when another participant was present, compared with listening when alone, suggesting that more effort was exerted. However, SRTs did not differ significantly between these two conditions, suggesting that the additional effort did not lead to a change in performance. No interaction was found between copresence and sentence intelligibility for either measure. Furthermore, no correlation was observed between the Connectedness for Groups questionnaire or Other-in-Self scale scores (which were intended to measure connectedness within a dyad) and the change in PPD between the copresence conditions.

The finding that PPD increased when the task was conducted in the presence of another participant suggested that more effort was exerted while another participant was present compared with when participants did the task alone. Interestingly, this

**TABLE 6.** Fixed effect estimates ( $\beta$ ) for the subjective rating scales, together with CIs

	Effort		Performance	
	$\beta$	CI (95%)	$\beta$	CI (95%)
(Intercept)	+6.351	+6.087 to +6.615	+6.237	+5.922 to +6.552
Intelligibility (A)	-1.726	-2.136 to -1.317	+2.350	+1.882 to +2.818
Intelligibility (B)	-1.662	-2.212 to -1.112	+1.685	+1.142 to +2.228
Copresence	+0.025	-0.317 to +0.368	-0.179	-0.521 to +0.162
Intelligibility (A) $\times$ Copresence	+0.447	-0.109 to +1.003	-0.485	-1.133 to +0.162
Intelligibility (B) $\times$ Copresence	-0.491	-1.048 to +0.065	+0.044	-0.603 to +0.691
	Giving-up		Difficulty	
	B	CI (95%)	$\beta$	CI (95%)
(Intercept)	+1.050	+0.939 to +1.162	+5.384	+5.032 to +5.736
Intelligibility (A)	-0.461	-0.615 to -0.307	-1.841	-2.328 to -1.354
Intelligibility (B)	-0.402	-0.556 to -0.247	-1.703	-2.305 to -1.101
Copresence	+0.090	+0.001 to +0.179	+0.081	-0.316 to +0.478
Intelligibility (A) $\times$ Copresence	+0.067	-0.151 to +0.285	+0.376	-0.309 to +1.062
Intelligibility (B) $\times$ Copresence	-0.082	-0.300 to +0.136	-0.218	-0.903 to +0.467

The backward difference contrast coding resulted in two beta estimates for intelligibility, labeled contrast A (50% compared with 20%) and contrast B (80% compared with 50%). Copresence represents the difference in the ratings when doing the task in the presence of another participant, compared with alone. Log-transformed data were used for the giving-up estimates. CIs, confidence intervals.

effect was not clearly reflected in the self-rating scores, suggesting that participants did not experience a change in cognitive demands caused by the other participant's presence. This might indicate that the increase in PPD did not necessarily reflect a change in effort. However, it could also be that the subjective ratings did not accurately measure subjective effort. The latter notion is supported by the finding that, when prompted to rate their exerted effort, participants tend to rate their performance instead (Moore & Picou 2018). Indeed, just like the subjective ratings, SRTs (performance) were stable over the copresence conditions. Models of listening effort (Pichora-Fuller et al. 2016) suggest that a change in effort, as was found in this study, could be explained by both a change in task demand as well as a change in motivation. An interesting parallel can be drawn between these two predictors of listening effort and the ongoing discussion on the cognitive processes that underlie SFI.

One hypothesis of SFI proposes that the presence of another individual causes an attentional conflict, eventually leading to attention focusing. When considered in the context of the current study, an attentional conflict could have caused the speech-in-noise task to be more demanding on cognitive resources, and thus have increased effort. Indeed, attention mechanisms are an important part of listening effort (Koelewijn et al. 2015). Interestingly, it seemed that the effect of copresence on PPD only occurred when participants started together with another participant, not when they first completed the alone condition and were then accompanied by the other participant. Attentional focusing could account for this finding when considering task experience. Participants who started alone first would already have had experience with the task when doing the together blocks. They would then require less attention to complete them successfully, leading to reduced PPDs.

Alternatively, SFI effects are explained through an increase of motivation caused by the presence of a social other (McFall et al. 2009). This account of SFI suggests that a change in motivation could have occurred when participants were threatened by the potential of social evaluation (caused by the presence of the other participant). This then induced a state of heightened arousal, signaling the preemptive allocation of effort. The order

effect of copresence conditions could be explained through participants who started together first having less confidence in their abilities to successfully complete the task, leading to an increased experience of social evaluative threat. This could have been further amplified by the knowledge that the other participant did have experience with the task. However, the current study did not find evidence of an increase in baseline arousal during the together conditions, since BPS was constant between copresence conditions. Such an increase would be expected if the allocation of resources was preemptive, as proposed by the motivational account. Future studies could use a more sophisticated design to examine the origin of the copresence effect.

Theories about SFI primarily try to explain the contrasting findings of increased and decreased task performance when accompanied by a social other. In this study, the presence of another participant was not found to affect performance on the speech-in-noise task, suggesting that there was no actual facilitation or impairment of performance. This finding differs from some earlier work that linked listening in the presence of strangers (compared with listening in isolation) to reduced speech comprehension (Beatty 1980; Beatty & Payne 1984). The opposite was found in a more recent study by Zekveld et al. (2019), who showed an increase of performance for participants that received explicit social evaluative feedback, compared with controls who did not receive feedback. This effect was more prominent in easy listening conditions (71% sentence intelligibility), compared with harder ones. For the current study, the stationary noise mask might not have allowed for much improvement of performance, even when participants tried harder (Kidd & Colburn 2017).

PPDs during passive listening were found to be considerably lower than PPDs during active listening. While still significantly higher than BPS, passive PPDs only reflected a small change in cognitive demand and could be explained by the pupil merely reacting to the onset of the target sentence. Regardless, the difference between active and passive PPDs suggests that considerably less effort was exerted during passive listening. This finding is in line with a study by Zekveld et al. (2014), who found a reduced pupil response during passive listening tasks

as well. One explanation for these results is that participants stopped listening since they were not asked to repeat the target sentences. However, this seems unlikely because monitoring which sentences to repeat required some level of engagement in the present study. Furthermore, participants might have been driven by competitiveness to listen to the other participant's responses.

If it were the case that participants were listening during the passive conditions, the decreased PPDs would indicate that less effort was used. Recent research on listening effort suggests that it is influenced by a variety of cognitive processes, for example memory and attention (Francis & Love 2020; Seifi Ala et al. 2020). During passive listening, the reliance on these processes was likely reduced, which would explain the reduced pupil response. The involvement of memory might have been especially salient as participants had to retain the target sentence in memory during the 3 seconds of noise after target sentence offset. Participants could have employed different strategies depending on whether they had to listen actively or passively. For example, participants might have employed active rehearsal to correctly repeat a sentence during active listening, or they might have tried to fill in any gaps in the sentence in this period using the context of the sentence. During passive listening they might not have attempted to hear the sentence and therefore might not have had any information to rehearse, or they may not have tried to actively remember or complete the target sentence, leading to a reduction of effort. These interpretations should be examined in future research using methods that allow testing of speech perception during passive listening.

No relationship was found between the outcomes of the connectedness questionnaires and the pupil measurements. This might be because the questionnaires were insensitive to variations in the perceived connectedness between participants in a dyad, in part because the dyads did not have an opportunity to socially connect and form meaningful opinions about one another. Alternatively, it might be that connectedness does not impact PPD. Future studies could study in more detail if and how the social dynamics of dyads (or larger groups) influence listening effort.

A main effect of intelligibility was found on BPS, where lower intelligibility levels corresponded with higher BPS. This might be because pupil dilations had not fully returned to baseline between trials or that baseline arousal was increased in response to the difficulty of the block. This difference in BPS might have confounded PPD as pupil traces were baseline corrected using BPS. There is some research that suggests that PPD scales linearly regardless of BPS, as long as it approximates the middle of the dynamic range (Reilly et al. 2019). However, any influence of BPS on PPD cannot be ruled out completely. Furthermore, as BPS might have varied over trials this also raises the concern that PPD could have been confounded by trial-to-trial changes in BPS. As pupil traces were averaged within conditions, this information was lost. While analyses to account for such effects are beyond the scope of the current paper, in future studies they should be considered. BPS was also influenced by the combination of copresence and order of copresence conditions such that BPS was generally lower in the second half of the experiment. We suggest this is a result of reduced arousal or fatigue (Winn et al. 2018).

Most of the random slopes that were excluded corresponded to a fixed factor that did not affect the data significantly. Since

there is no systematic change caused by this factor, the model is unlikely to improve by trying to account for between-subject variance of that factor. For example, since active and passive trials were interleaved within a block, it is likely that there was little change of BPS between the two. As this was the case for all participants, the between-subject variance was relatively small, thus the addition of a slope for task is not justifiable. However, for both the PPD and SRT data the random slope for intelligibility was excluded, even though intelligibility significantly affected the data. This could reflect that the changes caused by intelligibility were relatively similar across participants (at least not variant enough to justify the random slope). Possibly, by setting certain criteria for hearing thresholds, a normally hearing sample was included with very similar hearing abilities. Therefore, performance and effort changes did not vary enough between participants to consider a random slope for intelligibility.

By pairing participants to form a dyad, differences in characteristics of the participants (e.g., sex, ethnicity) or their performance on the task might have introduced additional variance in the copresence manipulation, posing as a limitation. For example, an opposite sex dyad could have resulted in a stronger (or weaker) effect, compared with same sex dyads. However, because effects caused by observers are generally fairly robust (Guerin & Innes 1984), we do not consider it to be problematic for this study's purposes. Another limitation is that the trials in which participants lost track whether they had to repeat a sentence or not were not logged and were thus not accounted for during analysis. Finally, pupil size has been found to convey social information like emotion and listening engagement (Kang & Wheatley 2017; Kret 2018). It is unclear whether such changes in pupil dilation are caused by a change in effort expenditure, or by a more direct process (unrelated to effort). In case of the latter, the presence of another participant might have confounded pupil dilation as a measure of effort. However, since participants could only see each other in the periphery of their vision and since the speech material did not contain particularly emotional material, such influences should have been minimal.

In summary, copresence during testing was found to increase the pupil dilation response when performing a speech-in-noise task in the absence of an effect on performance, suggesting that copresence did indeed lead to a change in effort. This change could be explained by motivation originating from social evaluative stress or attentional conflicts caused by the other participant's presence. The effect of copresence was especially apparent for participants that first completed the together blocks. The finding that pupil size was reduced during passive listening either suggests disengagement, or that some cognitive processes were present during active listening but not during passive listening.

## CONCLUSION

This study is the first to show that a minimal social context manipulation—here the mere presence of another individual— influences the pupil dilation response indexing listening effort. Even though ecologically valid social dynamics in realistic listening situations are much more complex, this at least shows that one element of it (copresence) can be manipulated and should be considered when studying ecologically valid settings. Research on speech-in-noise perception and listening effort



often aims to improve quality of life for those with hearing disabilities. While much progress has been made, experimental research often neglects to consider social factors that play a role in real-life listening situations.

### ACKNOWLEDGEMENTS

All authors had significant contributions to this work. H.P. designed and performed the experiment, analyzed the data and wrote the manuscript, A.A.Z., G.H.S., N.J.V., T.L. and S.E.K. all discussed and provided valuable feedback on the design, analyses and the article at all stages of the research. S.E.K. was the principle investigator of the project. The authors would like to express their gratitude to J.H.M van Beek for his technical support during the experiment's setup.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie-Sklodowska-Curie grant agreement No 765329. There are no conflicts of interest financially or otherwise.

Address for correspondence: Hidde Pielage, Otolaryngology Head and Neck Surgery, Ear & Hearing, Amsterdam Public Health Research Institute, Amsterdam UMC, Vrije Universiteit Amsterdam, De Boelelaan, 1117 Amsterdam, the Netherlands. E-mail: h.pielage@amsterdamumc.nl.

Received May 26, 2020; accepted February 2, 2021; published online ahead of print April 1, 2021.

### REFERENCES

- Aguirre, S. L., Bramsløw, L., Lunner, T., Whitmer, W. M. (2019). Spatial cue distortions within a virtualized sound field caused by an additional listener. Paper presented at the International Congress on Acoustics. <http://dx.doi.org/10.18154/RWTH-CONV-238997>
- Armstrong, R. A. (2014, September 1). When to use the Bonferroni correction. *Ophthalmic Physiol Opt*, *34*, 502–508.
- Aron, A., Aron, E. N., Smollan, D. (1992). Inclusion of other in the self scale and the structure of interpersonal closeness. *J Pers Soc Psychol*, *63*, 596–612.
- Bates, D., Mächler, M., Bolker, B. M., Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *J Stat Softw*, *67*, 1–51.
- Baumeister, R. F., Twenge, J. M., Nuss, C. K. (2002). Effects of social exclusion on cognitive processes: Anticipated aloneness reduces intelligent thought. *J Pers Soc Psychol*, *83*, 817–827.
- Beatty, M. J. (1980). Social facilitation and listening comprehension. *Percept Mot Skills*, *51*(3 Pt 2), 1222.
- Beatty, M. J., & Payne, S. K. (1984). Effects of social facilitation on listening comprehension. *Communication Quarterly*, *32*, 37–40.
- Beechey, T., Buchholz, J. M., Keidser, G. (2020a). Hearing aid amplification reduces communication effort of people with hearing impairment and their conversation partners. *J Speech Lang Hear Res*, *63*, 1299–1311.
- Beechey, T., Buchholz, J. M., Keidser, G. (2020b). hearing impairment increases communication effort during conversations in noise. *J Speech Lang Hear Res*, *63*, 305–320.
- Belletier, C., Normand, A., Huguet, P. (2019). Social-facilitation-and-impairment effects: From motivation to cognition and the social brain. *Current Directions in Psychological Science*, *28*, 260–265.
- DeWall, C. N., Baumeister, R. F., Vohs, K. D. (2008). Satiated with belongingness? Effects of acceptance, rejection, and task framing on self-regulatory performance. *J Pers Soc Psychol*, *95*, 1367–1382.
- Dickerson, S. S., Gruenewald, T. L., Kemeny, M. E. (2004). When the social self is threatened: Shame, physiology, and health. *J Pers*, *72*, 1191–1216.
- Eckstein, M. K., Guerra-Carrillo, B., Miller Singley, A. T., Bunge, S. A. (2017). Beyond eye gaze: What else can eyetracking reveal about cognition and cognitive development? *Dev Cogn Neurosci*, *25*, 69–91.
- Festen, J. M., & Plomp, R. (1990). Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing. *J Acoust Soc Am*, *88*, 1725–1736.
- Francis, A. L., & Love, J. (2020). Listening effort: Are we measuring cognition or affect, or both? *Wiley Interdiscip Rev Cogn Sci*, *11*, e1514.
- Gere, J., & MacDonald, G. (2010). An update of the empirical case for the need to belong. *J Individ Psychol*, *66*, 93–115.
- Green, P., & Macleod, C. J. (2016). SIMR: An R package for power analysis of generalized linear mixed models by simulation. *Methods Ecol Evol*, *7*, 493–498.
- Guerin, B., & Innes, J. M. (1984, June 1). Explanations of social facilitation: A review. *Curr Psychol*, *3*, 32–52.
- Hughes, S. E., Hutchings, H. A., Rapport, F. L., McMahon, C. M., Boisvert, I. (2018). Social connectedness and perceived listening effort in adult cochlear implant users: A grounded theory to establish content validity for a new patient-reported outcome measure. *Ear Hear*, *39*, 922–934.
- Kaernbach, C. (1991). Simple adaptive testing with the weighted up-down method. *Percept Psychophys*, *49*, 227–229.
- Kahneman, D. (1973). *Attention and Effort*.
- Kang, O., & Wheatley, T. (2017). Pupil dilation patterns spontaneously synchronize across individuals during shared attention. *J Exp Psychol Gen*, *146*, 569–576.
- Keidser, G., Naylor, G., Brungart, D. S., Caduff, A., Campos, J., Carlile, S., Carpenter, M. G., Grimm, G., Hohmann, V., Holube, I., Launer, S., Lunner, T., Mehra, R., Rapport, F., Slaney, M., Smeds, K. (2020). The quest for ecological validity in hearing science: What it is, why it matters, and how to advance it. *Ear Hear*, *41*(Suppl 1), 5S–19S.
- Kidd, G., Jr., & Colburn, H. S. (2017). Informational masking in speech recognition. In *The Auditory System at The Auditory System at the Cocktail Party* (pp. 75–109). Springer.
- Kiessling, J., Pichora-Fuller, M. K., Gatehouse, S., Stephens, D., Arlinger, S., Chisolm, T., Davis, A. C., Erber, N. P., Hickson, L., Holmes, A., Rosenhall, U., von Wedel, H. (2003). Candidature for and delivery of audiological services: Special needs of older people. *Int J Audiol*, *42*(Suppl 2), 2S92–2101.
- Koelwijn, T., de Kluiver, H., Shinn-Cunningham, B. G., Zekveld, A. A., Kramer, S. E. (2015). The pupil response reveals increased listening effort when it is difficult to focus attention. *Hear Res*, *323*, 81–90.
- Kramer, S. E., Kapteyn, T. S., Festen, J. M., Tobi, H. (1996). The relationships between self-reported hearing disability and measures of auditory disability. *Audiology*, *35*, 277–287.
- Kret, M. E. (2018). The role of pupil size in communication. Is there room for learning? *Cogn Emot*, *32*, 1139–1145.
- Kuznetsova, A., Brockhoff, P. B., Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *J Stat Softw*, *82*, 1–26.
- Leach, C. W., van Zomeren, M., Zebe, S., Vliek, M. L., Pennekamp, S. F., Doosje, B., Ouwerkerk, J. W., Spears, R. (2008). Group-level self-definition and self-investment: A hierarchical (multicomponent) model of in-group identification. *J Pers Soc Psychol*, *95*, 144–165.
- Lieberman, M. D. (2013). *Social: Why Our Brains Are Wired To Connect*. Crown Publishers.
- Liu, Y., Rodenkirch, C., Moskowicz, N., Schriver, B., Wang, Q. (2017). Dynamic lateralization of pupil dilation evoked by locus coeruleus activation results from sympathetic, not parasympathetic, contributions. *Cell Rep*, *20*, 3099–3112.
- Luke, S. G. (2017). Evaluating significance in linear mixed-effects models in R. *Behav Res Methods*, *49*, 1494–1502.
- Matthen, M. (2016). Effort and displeasure in people who are hard of hearing. *Ear Hear*, *37*(Suppl 1), 28S–34S.
- McFall, S. R., Jamieson, J. P., Harkins, S. G. (2009). Testing the mere effort account of the evaluation-performance relationship. *J Pers Soc Psychol*, *96*, 135–154.
- Monfardini, E., Redouté, J., Hadj-Bouziane, F., Hynaux, C., Fradin, J., Huguet, P., Costes, N., Meunier, M. (2016). Others' sheer presence boosts brain activity in the attention (but not the motivation) network. *Cereb Cortex*, *26*, 2427–2439.
- Moore, T. M., & Picou, E. M. (2018). A potential bias in subjective ratings of mental effort. *J Speech Lang Hear Res*, *61*, 2405–2421.
- Ohlenforst, B., Zekveld, A. A., Lunner, T., Wendt, D., Naylor, G., Wang, Y., Versfeld, N. J., Kramer, S. E. (2017). Impact of stimulus-related factors and hearing impairment on listening effort as indicated by pupil dilation. *Hear Res*, *351*, 68–79.
- Pichora-Fuller, M. K., Kramer, S. E., Eckert, M. A., Edwards, B., Hornsby, B. W. Y., Humes, L. E., Lemke, U., Lunner, T., Matthen, M., Mackersie, C. L., Naylor, G., Phillips, N. A., Richter, M., Rudner, M., Sommers, M. S., Tremblay, K. L., Wingfield, A. (2016). Hearing impairment and cognitive energy: The framework for understanding effortful listening (FUEL). *Ear and Hearing*, *37*, 5S–27S.



- Pickett, C. L., Gardner, W. L., Knowles, M. (2004). Getting a cue: The need to belong and enhanced sensitivity to social cues. *Pers Soc Psychol Bull*, 30, 1095–1107.
- Plomp, R., & Mimpen, A. M. (1979). Improving the reliability of testing the speech reception threshold for sentences. *Audiology*, 18, 43–52.
- R Core Team. (2019). *R: A Language and Environment for Statistical Computing* (pp. 1–1731). <https://www.yumpu.com/en/document/view/6853895/r-a-language-and-environment-for-statistical-computing>.
- Reilly, J., Kelly, A., Kim, S. H., Jett, S., Zuckerman, B. (2019). The human task-evoked pupillary response function is linear: Implications for baseline response scaling in pupillometry. *Behav Res Methods*, 51, 865–878.
- Rohleder, N., Beulen, S. E., Chen, E., Wolf, J. M., Kirschbaum, C. (2007). Stress on the dance floor: The cortisol stress response to social-evaluative threat in competitive ballroom dancers. *Pers Soc Psychol Bull*, 33, 69–84.
- Seifi Ala, T., Graverson, C., Whitmer, W., Hadley, L., Lunner, T. (2020). Motivation modifies effort when the task demands it: EEG evidence of an interaction between reward and demand in a speech-in-noise task. *Manuscript in Preparation*.
- Sleegers, W. W. A., Proulx, T., van Beest, I. (2017). The social pain of Cyberball: Decreased pupillary reactivity to exclusion cues. *J Exp Soc Psychol*, 69, 187–200.
- Smeds, K., Wolters, F., Rung, M. (2015). estimation of signal-to-noise ratios in realistic sound scenarios. *J Am Acad Audiol*, 26, 183–196.
- Spyridakou, C., & Bamiou, D.-E. (2015). Need of speech-in-noise testing to assess listening difficulties in older adults. *Hear Balance Commun*, 13, 65–76.
- Steinmetz, J., & Pfattheicher, S. (2017). Beyond social facilitation: A review of the far-reaching effects of social attention. *Social Cognition*, 35, 585–599.
- Versfeld, N. J., Daalder, L., Festen, J. M., Houtgast, T. (2000). Method for the selection of sentence materials for efficient measurement of the speech reception threshold. *J Acoust Soc Am*, 107, 1671–1684.
- Wahn, B., Ferris, D. P., Hairston, W. D., Peter, K. (2017). Pupil size asymmetries are modulated by an interaction between attentional load and task experience. *BioRxiv*, 1–8. <https://doi.org/10.1101/137893>
- Wendt, D., Koelewijn, T., Książek, P., Kramer, S. E., Lunner, T. (2018). Toward a more comprehensive understanding of the impact of masker type and signal-to-noise ratio on the pupillary response while performing a speech-in-noise test. *Hear Res*, 369, 67–78.
- Winn, M. B., Wendt, D., Koelewijn, T., Kuchinsky, S. E. (2018). best practices and advice for using pupillometry to measure listening effort: An introduction for those who want to get started. *Trends Hear*, 22, 2331216518800869.
- Zekveld, A. A., Heslenfeld, D. J., Johnsrude, I. S., Versfeld, N. J., Kramer, S. E. (2014). The eye as a window to the listening brain: Neural correlates of pupil size as a measure of cognitive listening load. *Neuroimage*, 101, 76–86.
- Zekveld, A. A., Koelewijn, T., Kramer, S. E. (2018). The pupil dilation response to auditory stimuli: Current state of knowledge. *Trends Hear*, 22, 2331216518777174.
- 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., van Scheepen, J. A. M., Versfeld, N. J., Veerman, E. C. I., Kramer, S. E. (2019). Please try harder! The influence of hearing status and evaluative feedback during listening on the pupil dilation response, saliva-cortisol and saliva alpha-amylase levels. *Hear Res*, 381, 107768.