Check for updates

Original Research Article

How Temporal Overlap in Task Processing and Acoustic Load Affect Listening Effort in Adults

Trends in Hearing
Volume 30: 1–15
© The Author(s) 2026
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/23312165261472354
journals.sagepub.com/home/tia
Sage

Julia Seitz[1], Jana R. Berger[1], Maria Anastasia Sinarso[1], and Janina Fels[1]

Abstract

Listening effort reflects the cognitive resources allocated to overcome obstacles in auditory processing. While acoustic degradation (e.g., noise, reverberation) is known to affect listening effort, little is known about how temporal overlap of task processing influences effortful listening in adults. This study examined behavioral and subjective listening effort in 26 adults using a sequential dual-task paradigm manipulating stimulus onset asynchrony (SOA; 0.05 *s*, 0.25 *s*, 1.5 *s*, single task), noise condition (silence, multi-talker babble at 0 dB SNR), and acoustic environment (anechoic, room effects of $T_{30} = 0.49$ s). Task difficulty via decreasing stimulus onset asynchrony (SOA) significantly increased word recognition response times and subjective effort ratings, even when speech intelligibility remained high (∼90%). Noise effects emerged only at longer SOAs, but vanished at very short SOAs where capacity was exhausted, indicating that SOA is a stronger predictor of listening effort than noise exposure. Interestingly, mild reverberation reduced behavioral listening effort compared to anechoic conditions in silence.

Keywords

listening effort, psychological refractory period, sequential dual-task paradigm, multi-talker babble, reverberation

Received: 2 March 2026; Revised: 30 June 2026; Accepted: 6 July 2026

1. Introduction

Listening effort has emerged as a key construct for understanding communication in challenging acoustic environments, particularly when speech remains intelligible, but processing becomes demanding. Within the Framework for Understanding Effortful Listening (FUEL), listening effort is defined as the deliberate allocation of limited cognitive resources to overcome obstacles in goal-directed auditory processing (Pichora-Fuller et al., 2016). This perspective emphasizes that listening effort is not solely determined by audibility, but rather, it is shaped by task demands, listener characteristics, and environmental conditions. To capture this multifaceted construct, research relies on behavioral, subjective, and physiological measures of listening effort (Pichora-Fuller et al., 2016). Behavioral measures quantify listening effort through observable performance changes, such as increased response times in speech recognition tasks, with slower responses reflecting greater cognitive engagement in the listening task (e.g., Gagné et al. (2017); Gatehouse and Gordon (1990); Seitz et al. (2024)). Subjective measures capture listeners' perceived effort through self-report questionnaires or rating scales reflecting the listener's conscious experience of task difficulty and cognitive load (e.g., Shields et al. (2023); Holube et al. (2016); Seitz and Fels (2025)). Physiological measures assess effort through neural or autonomic indicators such as pupil dilation or electroencephalography (e.g., Gómez-Merino et al., 2020, Peelle (2018); Richter et al. (2023)). Each measurement modality offers

[1]Institute for Hearing Technology and Acoustics, RWTH Aachen University, Aachen, Germany

Corresponding author:
Julia Seitz, Institute for Hearing Technology and Acoustics, RWTH Aachen University, Kopernikusstrasse 5, Aachen 52074, Germany.
Email: julia.seitz@akustik.rwth-aachen.de

Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (https://creativecommons.org/licenses/by-nc/4.0/) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (https://us.sagepub.com/en-us/nam/open-access-at-sage).

distinct advantages: behavioral measures quantify performance costs; subjective measures reveal perceptual experience; and physiological measures provide real-time biological reactions to cognitive engagement during task performance.

Common behavioral approaches to measure listening effort include dual-task and single-task paradigms. In the latter response times (RTs) in a speech perception task are interpreted as listening effort (e.g., Gatehouse and Gordon (1990); Seitz et al. (2024); Holube et al. (2016); Prodi and Visentin (2019)). The dual-task paradigm is based on the theoretical foundation that only limited resources are available for additional tasks if increased effort is spent on listening (Kahneman, 1973). There are two types of dual-task paradigms (Koch et al., 2018): one assessing dual-task performance relative to single-task performance to quantify dual-task interference, which is referred to in this work as the concurrent paradigm (for a review, see Gagné et al. (2017)), and a second examining dual-task interference as a function of the temporal overlap of task processing, known as the psychological refractory period (PRP) or sequential paradigm. In listening effort research, mostly concurrent dual-task paradigms were utilized (for an overview, see Gagné et al. (2017)). Only a few studies applied a sequential dual-task paradigm to investigate listening effort (Picou et al., 2011, 2013; Rakerd et al., 1996; Seitz and Fels, 2025). However, to the best of our knowledge, no studies have investigated listening effort in adults using a sequential dual-task paradigm that explicitly accounts for the PRP effect, even though this effect can substantially influence outcomes (Koch et al., 2018; Kuchinsky et al., 2024). The PRP effect refers to reduced processing efficiency of a second task when it closely follows a first task due to their simultaneous cognitive processing (Pashler, 1994). In the present study, the second task comprised speech perception, while the first task involved a visual decision-making task (visual stimulus detection). Simultaneous processing of the two tasks is induced by a short time interval between them, with the duration of this interval (SOAs) varying between tasks (Pashler, 1994). This sequential processing resembles everyday situations in which an ongoing conversation is interrupted by a competing visual event, such as a notification on a mobile phone. In line with the FUEL model (Pichora-Fuller et al., 2016), introducing the decision task requires intentional or evaluative resource allocation, thereby increasing task difficulty. Based on the results of Seitz and Fels (2025), who used a PRP paradigm to investigate listening effort in six- to ten-year-olds, it was hypothesized that shorter SOAs would lead to greater task difficulty due to the simultaneous processing of stimuli. In addition to these varying task difficulties, both tasks were performed individually in the experiment, which was assumed to be the easiest condition, and established a baseline. Furthermore, the listening experiment included a subjective measure based on the questionnaire by Seitz et al. (2024) and Seitz and Fels (2025) to validate the results on listening effort.

Not much is known about how structural bottlenecks and capacity limitations interact with acoustic demands in adult listeners. To address this gap, the present study aims to deliberately introduce a processing bottleneck and examine how this shapes behavioral and subjective listening effort. Structural models of attention (Broadbent, 1958; Deutsch and Deutsch, 1963; Treisman, 1960) primarily focus on identifying structural limitations on the flow of information within the cognitive processing chain. Implementing the PRP paradigm, Pashler (1994, 1984) identified response selection as the central bottleneck. Hence, response selection cannot be processed in parallel, but a second task must wait until critical processing of the first task is completed (Pashler, 1984, 1994). In contrast to structural models, capacity-limited models (Kahneman, 1973; Pichora-Fuller et al., 2016; Wickens, 1980, 1984) propose that performance depends on the amount and allocation of one or more pools of cognitive resources. Based on this, effort reflects the deployment of limited attentional capacity, which is influenced by factors such as task demands, arousal, and motivation. In their work, Kuchinsky et al. (2024) compared the FUEL (Pichora-Fuller et al., 2016), which is a single capacity model, with the multiple capacity model by Wickens (1980, 1984) that was extended by Horrey et al. (2006), including auditory, visual, peripheral, and cognitive capacities. Within this perspective, dual-task interference emerges when two tasks compete for overlapping resources in one or more of these domains (Kuchinsky et al., 2024). Depending on the nature of the tasks involved, this interference can occur across sensory, motor, or cognitive domains.

In the present study, three variables are hypothesized to cause dual-task interference: the stimulus onset asynchrony (SOA), the noise condition, and the acoustic condition. Based on previous works (Koch et al., 2018; Pashler, 1994) using the PRP paradigm, SOAs of 0.05 s, 0.25 s, and 1.5 s were chosen, to create three stages of overlapping task processing. Listening effort in noise has been extensively studied in adults, with multi talker babble (MTB) emerging as a frequently used masking condition (see the review by Gagné et al. (2017)). MTB simulates realistic speech communication environments that adults encounter frequently in relevant everyday situations (Smeds et al., 2020). Furthermore, MTB has been found to elevate both response times and subjective effort ratings compared to quiet conditions (Holube et al., 2016; Huang et al., 2022; Picou et al., 2011; Seitz et al., 2024), making it ideal for examining the acoustic manipulations central to the FUEL model (Pichora-Fuller et al., 2016). The present study, therefore, compared silence to MTB at a fixed signal-to-noise ratio (SNR) of 0 dB. Huang et al. (2022) emphasized the importance of testing reverberation effects on listening effort to validate this part of the FUEL. Therefore, this study additionally included an anechoic condition as well as a listening condition with simulated room effects of $T_{30} = 0.49$ s.

According to the multiple capacity model (Horrey et al., 2006; Wickens, 1980, 1984), the present paradigm engages all four capacities: auditory resources for speech perception, visual resources for visual stimulus detection, peripheral resources

for the motor response, and cognitive resources for coordinating the multitasking situation. By manipulating the temporal overlap in task processing, the study targets the interplay between structural bottlenecks and shared resource limitations, linking these mechanisms directly to behavioral and subjective indices of listening effort.

Given these theoretical frameworks and design choices, three hypotheses guide the present investigation:

1. Adverse acoustics: Listening in adverse acoustic conditions involving MTB or room effects, compared to favorable conditions involving silence or anechoic conditions, increases both behavioral and subjective listening effort. This is evidenced by longer RTs on word recognition tasks and higher effort ratings.
2. Task difficulty: Increasing task difficulty, as measured by shorter SOA, requires more cognitive resources and thus increases behavioral and subjective listening effort.
3. Interaction: It is expected that there will be an interaction between adverse acoustics and task difficulty, with the strongest increase in listening effort occurring when there is a high level of task demand (i.e., short SOAs) and MTB and room effects are present.

2. Methods

2.1 Participants

26 adults between the ages of 20 and 30 participated in the study. All participants met the inclusion criteria: they had normal hearing, normal or corrected-to-normal vision, were fluent in German, and had no diagnosis of ADHD or epilepsy. Normal hearing was determined by pure tone audiometry: 20 dB HL - for the frequencies f = [125, 250, 500, 1k, 2k, 4k, 8k] Hz (World Health Organization, 2021). A visual acuity test determined if participants had normal or corrected-to-normal vision (International Organization for Standardization, 2020). The sample size was selected on the basis of previous studies on listening effort in similar experimental paradigms involving both children and adults (Seitz and Fels, 2025; Seitz et al., submitted for publication; Huang et al., 2022; Picou et al., 2016). The dataset of one participant had to be excluded due to a misunderstanding of the tasks. Therefore, the data of 25 participants (mean age 23.26 years, 13 female) were included in the analysis. The study was performed in compliance with the Declaration of Helsinki guidelines, and all participants were compensated with a 10 voucher from a local bookstore that also offers online dispatch. Ethical approval by the university's ethical committee was obtained before the recruitment of the participants (EK 24-070).

2.2 Experiment Task

In the experiment, the first task was a decision task and consisted of eight animal pictures that were arranged in a circle; four pictures were on the left and four pictures on the right side. The goal was to indicate on which side of the circle the target animal, a bear, was on by selecting the corresponding arrow on the screen. Participants could use either hand for selection. For each trial, the sequence of the animal pictures on the circle was created before the experiment, with two training lists and eight experimental lists. Within each list, the occurrence of the target animal on the right and left sides was balanced, ensuring that the target animal appeared on the same side no more than three times in succession, and no sequence was repeated.

The second task, a word-to-picture matching task (Klatte et al., 2010; Seitz et al., 2024), examined speech intelligibility by requiring participants to recognize spoken words. Behavioral listening effort was measured using the RTs. One trial consisted of a spoken word, followed by the presentation of four images - the image of the target word and three images of phonologically similar words - as response choices on the tablet screen. The response was given by selecting one of the images on the tablet screen. The task involved two training and eight different, equally difficult experiment lists. The list order was counterbalanced across participants, and within each list, the target word occurrence was randomized. In addition, the position of the target word image was randomized, with the position being changed after each trial. The words were recorded by a trained native German speaker using a Sennheiser MD 421-II-4 Dynamic Studio microphone in a sound-attenuated hearing booth at a sampling rate of 44.1 kHz and 16-bit resolution. The recordings were normalized according to EBU R-128 (European Broadcasting Union, 2011).

During the experiment, both tasks were performed in a single task condition and within a sequential dual-task paradigm. In the sequential dual-task paradigm (see Figure 1), the task complexity could be adapted. The word-to-picture matching task was executed after the decision task with varying SOA between the appearance of the decision task and the spoken word, thereby evoking the PRP effect. The three SOAs employed in this study were 0.05 s, 0.25 s and 1.5 s. The decision task measured the number of correct decisions and the time between the appearance of the picture until the response (RT1). The word-to-picture matching task evaluated the number of correct words and the time interval between the presentation of the acoustic stimulus and the response (RT2).

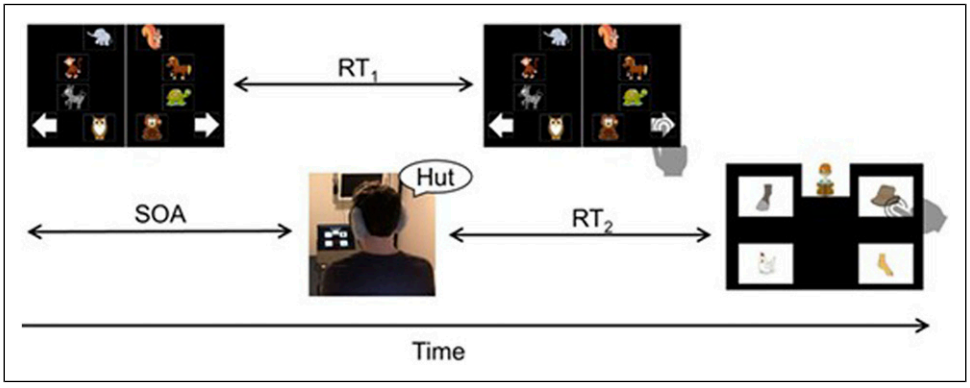

Figure 1. The paradigm structure with RT1 depicting the response time of the decision task and RT2 depicting the response time of the word recognition task. SOA shows the stimulus onset asynchrony between the first and the second task. For the word recognition trial, the German word "Hut" (Engl. hat) is depicted.

2.3 Questionnaire on Subjective Listening Effort

The present study utilized a questionnaire on subjective listening effort developed by Seitz et al. (2024) and applied by Seitz and Fels (2025) to investigate the subjective listening effort experienced by the participants. Both studies found reliable results regarding listening effort in children and adults, which is why the present study followed their approach. The questionnaire was developed for children aged six to ten years and adults. It comprised three questions, asking about perceived well-being, task performance, and listening effort. Responses were given on a five-point Likert scale, with the word description accompanied by a visual representation, ranging from a smiling face cartoon for very good (0), to a frown face for very bad (4). For analysis, the answer options were coded with numbers from zero to four, where zero was the best answer option for each question (e.g., "very good") and four was the subjectively worst (e.g., "very bad").

This questionnaire was based on the literature review by Francis & Love (2020), who noted that subjective listening effort may not be a purely unitary construct, but might reflect multiple dimensions of the listening experience. Accordingly, Seitz et al. (2024) and Seitz and Fels (2025) analyzed the three questions as a composite listening experience score. As this approach may reflect a broader appraisal of the listening situation than subjective listening effort alone, the composite listening experience score was used as the primary measure in the present study, while the single question on listening effort was additionally analyzed.

2.4 Listening Experiment Procedure

The listening experiment was conducted in January 2024 in a sound-attenuated hearing booth. Before testing, the participant gave their informed consent. Initially, pure tone audiometry, visual acuity test, and headphone-related transfer function measurement were conducted. Then, the actual experiment task started. Within the introduction, the three tasks of the experiment - the decision task, word-to-picture matching task, and the sequential dual-task - were explained and trained. Both single tasks had four training trials, and the dual-task consisted of six training trials. After that, the participant could ask questions, and the questionnaire on subjective listening effort was introduced.

The main experiment was structured into eight blocks, which were counterbalanced across participants. The experiment involved four noise scenarios, each of which occurred in one single task block and one sequential task block. In the single task block, the decision task and the word-to-picture matching task were performed in succession, with 24 trials each, and the order was balanced across experiment blocks. The sequential task comprised 36 trials, with each SOA occurring randomly in 12 trials and with a maximum of three consecutive repetitions of the same SOA. Following each block, the participant had a break and was asked to fill out the questionnaire. The participant was guided through the experiment with a story, which also served as a progress indicator. The total duration of the experiment was approximately one hour.

2.5 Auralized Virtual Sound Environment

The listening experiment aimed to create a realistic acoustic setting that resembled a classroom or office environment. The virtual scene was reproduced binaurally using MATLAB 2022a (The MathWorks, Inc., Natick, MA, USA) and the MATLAB integration Virtual Acoustics 2022a (Institute for Hearing Technology and Acoustics, RWTH Aachen University

et al., 2022). To create binaural spatial signals, the experiment stimuli were convolved with an individualized head-related transfer function of the artificial head of the Institute of Hearing Technologies and Acoustics (Schmitz, 1995). For the individualization, the participants' head measures (width, depth, and height) were taken at the beginning of the experiment and subsequently employed to adjust the interaural time difference cues (Bomhardt and Fels, 2014). All sounds were played back via open headphones (HD 650 by Sennheiser, Wedemark, Germany) with individualized headphone equalization (Masiero and Fels, 2011). To this end, the headphone-related transfer function was measured at the beginning of the experiment.

For a realistic acoustic scene, a typical everyday scenario with multiple speakers was created using a four-talker multi-talker babble in addition to a silent condition. The MTB was recorded with four German girls between the ages of eight and nine who were each reading a German fairy tale in an anechoic chamber (Loh and Fels, 2023). Both noise conditions, alongside an anechoic condition, were tested with simulated room effects to create the feeling of sitting in a real room. Therefore, a room was designed with thirteen tables and dimensions of 9.1 m × 7.5 m × 3.22 m in SketchUp Make 2016 (Trimble Inc., Sunnyvale, CA, USA), which was also used by Seitz and Fels (2025). The simulated room is provided in the supplementary material. The reverberation time was computed to be $T_{30} = 0.49\,s$, averaged over the frequency bands $f = [500, 1000, 2000]\,Hz$, which is a typical office room (International Organization for Standardization, 2021).

In the experiment design, the listener was positioned midway through the virtual room at a height of 1.2 m. The virtual target speaker was positioned in front of the listener at a distance of 2 m. Surrounding the listener at a distance of 2 m, four virtual speakers - the distractors - were arranged in a circle: front-left (45°), back-left (135°), back-right (225°), and front-right (315°). The MTB playback from the four different distractor positions resulted in 16 talker streams arriving at the listeners' ears. The distances between the speakers and the listener were set so that the listener would be positioned outside the technical critical distance ($d_c = 0.057 \times \sqrt{1 \times 219.77 / 0.59}\ \text{m} = 1.1\ \text{m}$) (Weinzierl, 2009), at which direct and diffuse sound energy are equal.

Before conducting the listening experiment, the loudness of the speech material was calibrated and fixed at 60 dB SPL. The MTB was also set to 60 dB SPL, thereby establishing an SNR of 0 dB. This SNR was chosen to enable comparability to studies using the same paradigm in children (Seitz and Fels, 2025; Seitz et al., submitted for publication) as well as with adult studies investigating listening effort (Huang et al., 2022; Picou et al., 2016), where an SNR of 0 dB has been shown to create a sufficiently challenging yet feasible listening task, which leads to increased listening effort and decreased speech recognition. To conclude, two noise types (no noise and MTB) and two acoustic conditions (anechoic and with room effects) were investigated in a binaural virtual acoustic scene. The spectrograms of both noise types and selected stimuli for both acoustic conditions are included in the supplementary material.

2.6 Statistical Analysis

All statistical analyses were performed in R (v4.4.1; R Core Team (2024)), with the significance level set at $\alpha = .05$ for all analyses. The analyses of word recognition and decision task included only experiment trial data with RTs falling within three standard deviations of the mean for each participant (Lachaud and Renaud, 2011). Additionally, one trial of one participant in the decision task was excluded due to a software bug that produced a negative RT. In total, 109 trials (1.82%) from 25 participants were removed from the word recognition task, and 95 trials (1.58%) from 25 participants were removed from the decision task. Questionnaire scores were not filtered because only one data point per participant was collected for each experimental block.

The independent variables comprised acoustic condition (room effects, anechoic), noise condition (no noise, MTB), and task difficulty. The dependent variables were RT, measured in seconds, and error count in both the word recognition and decision tasks. Additionally, the questionnaire provided subjective ratings of well-being, perceived task performance, and listening effort, which were assessed for internal consistency using Cronbach's alpha (Cronbach, 1951). Then, for each experiment block and participant, the scores for the three questions were summed up into a single dependent variable referred to as the listening experience score, ranging from 0 (no effort) to 12 (maximum effort). In addition, the single question only asking about listening effort was analyzed.

RTs in both tasks were reciprocally transformed prior to analysis with linear mixed models (LMMs), fitted using restricted maximum likelihood estimation. Meanwhile, error counts were analyzed using generalized linear mixed models (GLMMs) with a binomial distribution and a logit link function, fitted by maximum likelihood estimation. Both model types were implemented using the lme4 package (Bates et al., 2015). For all analyses, the model with the lowest Akaike's Information Criterion was used. In all cases, this was the full model, which included noise condition, acoustic condition, and task difficulty as fixed effects, with all two- and three-way interactions, as well as random intercepts for participants. An LMM was used to

analyze the listening experience score and the single question on listening effort, employing the same fixed effects as in the word recognition and decision task models, also built based on a model selection procedure.

For all models, model diagnostics were carried out using the DHARMa package (Hartig, 2024), assessing residual distributions and random effects using QQ plots, as well as overdispersion, zero inflation, outliers, and homogeneity of variance. LRTs were applied to the GLMMs and LMMs, the latter refitted with maximum likelihood estimation. When significant main effects or interactions were detected, Bonferroni-adjusted pairwise comparisons of estimated marginal means (EMMs) were performed using the emmeans package (Lenth, 2025), applied to the final models.

Finally, Spearman's rank correlations were conducted using the cor.test function from the stats package in R (R Core Team, 2024) to explore associations between subjective (listening experience score) and behavioral (word recognition task's RT) listening effort. These correlations were calculated within each noise scenario and averaged over single task and PRP conditions to maintain a sufficient sample size.

3. Results

Tables depicting all final models, including the variables, parameter estimates, and test statistics, are included in the supplementary material.

3.1 Listening Experiment Results

Although Levene's test indicated some heteroscedasticity in the LMMs for both tasks, examination of QQ plots for all GLMMs and LMMs revealed no violations of residual or random-effect normality. Levene's test is known to become more sensitive with increasing sample size (Field et al., 2012), so the minor deviations detected were expected and are unlikely to substantially bias the estimates (Barker and Shaw, 2015). Therefore, the fitted LMMs were retained, with these limitations acknowledged. Only significant results are reported in this section.

3.1.1 Response Time in the Word Recognition Task

In the LMM on word recognition task's RT, significant main effects were found for task difficulty ($\chi^2(12) = 359.57, p < .001$), noise condition ($\chi^2(8) = 26.03$, $p = .001$), and acoustic condition ($\chi^2(8) = 18.676$, $p = .017$) (see descriptives in Table 1). Overall, RTs were significantly slower in MTB than in silence, as well as with room effects compared to anechoic conditions. For task difficulty, RTs increased with decreasing SOA. Furthermore, the interactions between noise condition and acoustic condition ($\chi^2(4) = 10.03, p = .0399$) as well as noise condition and task difficulty ($\chi^2(6) = 12.61, p = .0497$) were significant. Post hoc tests revealed significantly longer RTs ($|z| = 2.717$, $p = .007$) in the MTB condition with room effects ($M = 1.52$ s, $SD = 0.50$ s) compared to the no noise condition with room effects ($M = 1.36$ s, $SD = 0.24$ s). Within the no noise condition, RTs were significantly longer ($|z| = 2.35, p = .019$) in the anechoic condition ($M = 1.40$ s, $SD = 0.31$ s) than with room effects ($M = 1.36$ s, $SD = 0.24$ s). This is illustrated in Figure 2.

The significant interaction of noise and task difficulty is depicted in Figure 3. RTs were significantly longer in MTB than in silence for the single task ($|z| = 2.92$, $p = .004$) and $SOA = 1.5$ s ($|z| = 2.72$, $p = .007$). For silence and MTB, RTs were significantly longer for $SOA = 0.05$ s and $SOA = 0.25$ s compared to the single task and $SOA = 1.5$ s. In addition, RTs were significantly longer for $SOA = 0.05$ s compared to $SOA = 0.25$ s in both noise types. The comparisons of the single task with $SOA = 1.5$ s for no noise and MTB were non-significant ($p > .05$, see Table 2).

Table 1. Descriptives for the Main Effects for the Response Time (RT) in the Word Recognition Task. *M* Denotes the Mean RT and *SD* the Standard Deviation.

Main Effect	Condition	*M*	*SD*
Noise	Silence	1.38 s	0.28 s
	MTB	1.49 s	0.41 s
Acoustic Condition	Anechoic	1.42 s	0.29 s
	Room Effects	1.44 s	0.40 s
Task Difficulty	Single Task	1.33 s	0.24 s
	SOA = 1.5 s	1.33 s	0.20 s
	SOA = 0.25 s	1.47 s	0.28 s
	SOA = 0.05 s	1.61 s	0.52 s

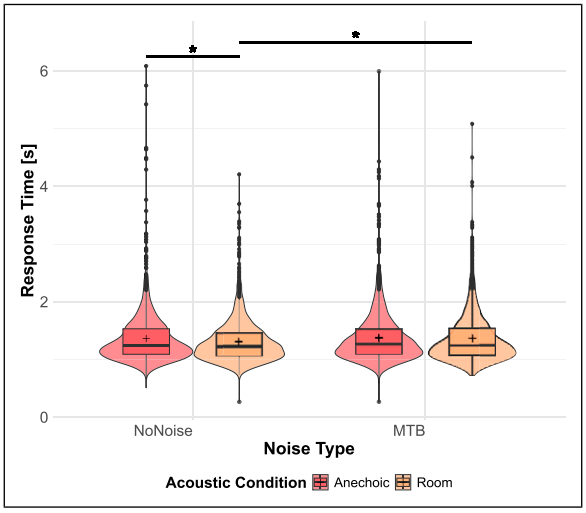

Figure 2. Word recognition response times across noise type and acoustic condition. Box plots display the median (center line), interquartile range, and mean (plus symbol), with noise type on the x-axis and the word recognition response time on the y-axis. The outliers are represented by circles, and the asterisks indicate $p < .05$ significance.

3.1.2 Performance in the Word Recognition Task

The GLMM on error count in the word recognition task resulted in significant main effects for noise condition ($\chi^2(8) = 195.07, p < .001$), acoustic condition ($\chi^2(8) = 20.06, p = .01$), and task difficulty ($\chi^2(12) = 21.49, p = .044$), as well as for the interaction between task difficulty and acoustic condition ($\chi^2(6) = 13.15, p = .041$) (see descriptives in Table 3). ER were significantly higher in MTB compared to in silence as well as with room effects compared to anechoic conditions. Bonferroni-corrected post hoc tests for the two-way interaction and the main effect of task difficulty revealed no significant comparisons ($p > .05$). The results of the word recognition tasks depicted as error rates (ERs) are shown in Figure 4.

3.1.3 Response Time in the Decision Task

The analysis of decision task RTs revealed a significant main effect of task difficulty ($\chi^2(12) = 200.78, p < .001$). Bonferroni-corrected post hoc tests revealed that RTs for the single task were significantly shorter than those with SOA (see Table 4). All other comparisons were non-significant ($p > .05$).

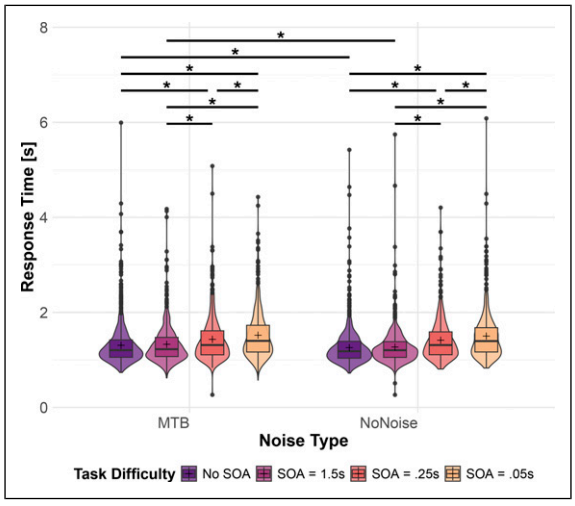

Figure 3. Word recognition response times across noise type and task difficulty. Box plots display the median (center line), interquartile range, and mean (plus symbol), with noise type on the x-axis and the word recognition response time on the y-axis. The outliers are represented by circles, and the asterisks indicate $p < .05$ significance.

Table 2. Pairwise Comparisons of the Interaction Between Noise Condition and Task Difficulty for the Linear Mixed Model (LMM) for the Response Time (RT) in the Word Recognition Task. M_{diff} Denotes the Mean RT Differences, $|z|$ the Absolute Value of the Standardized Test Statistic, and p the Significance Value.

SOA	Noise	M_{diff}	$\|z\|$	p
0.05 s vs. Single Task	No noise	.26 s	12.97	< .001
0.05 s vs. 1.5 s	No noise	.25 s	10.78	< .001
0.05 s vs. 0.25 s	No noise	.10 s	3.45	.003
0.25 s vs. Single Task	No noise	.16 s	9.00	< .001
0.25 s vs. 1.5 s	No noise	.15 s	7.34	< .001
1.5 s vs. Single Task	No noise	.01 s	0.53	>.05
0.05 s vs. Single Task	MTB	.31 s	10.84	< .001
0.05 s vs. 1.5 s	MTB	.32 s	8.33	< .001
0.05 s vs. 0.25 s	MTB	.18 s	3.74	.001
0.25 s vs. Single Task	MTB	.13 s	6.53	< .001
0.25 s vs. 1.5 s	MTB	.14 s	4.58	< .001
Single Task vs. 1.5 s	MTB	.01 s	1.27	>.05

3.1.4 Performance in the Decision Task

No main effects or interactions were found to be significant ($p > .05$) in the analysis of error count for the decision task. The average overall error rate was $M = .02$, $SD = 0.03$.

3.2 Subjective Listening Effort

Two analyses were conducted to assess subjective listening effort. The first analyzed only the single question on listening effort, while the second evaluated the joint listening experience score based on the approach by Seitz et al. (2024) and Seitz and Fels (2025). For the listening experience score, the three questionnaire variables were assessed for internal consistency. Cronbach's alpha was $\alpha = .758$, which is considered very good regarding internal consistency (Janssens, 2008). No model assumptions were violated in the LMMs.

For the listening experience score, the LMM revealed significant main effects of task difficulty ($\chi^2(4) = 17.22$, $p = .002$) and noise type ($\chi^2(4) = 52.03$, $p < .001$). Post hoc tests revealed higher listening experience score values for all SOAs compared to the single task condition ($|z| = 3.67$, $M_{diff} = 0.69$, $SE = .19$, $p < .001$), and in the MTB noise condition compared to the no-noise condition ($|z| = 7.40$, $M_{diff} = 1.39$, $SE = .19$, $p < .001$). All other comparisons were non-significant ($p > .05$). Figure 5 depicts both significant main effects for the listening experience score.

For the analysis of the single effort question, the LMM indicated significant main effects of task difficulty ($\chi^2(4) = 24.43$, $p < .001$) and noise type ($\chi^2(4) = 73.53$, $p < .001$). Post hoc tests indicated that participants reported higher perceived effort in the PRP task compared to the single task condition ($|z| = 4.79$, $M_{diff} = 0.43$, $SE = .09$, $p < .001$), and in the MTB noise condition compared to the no-noise condition ($|z| = 9.25$, $M_{diff} = 0.83$, $SE = .09$, $p < .001$). All other comparisons were non-

Table 3. Descriptives for the Main Effects for the Error Rate (ER) in the Word Recognition Task. *M* Denotes the Mean RT and *SD* the Standard Deviation.

Main Effect	Condition	*M*	*SD*
Noise	Silence	0.01	0.03
	MTB	0.01	0.09
Acoustic Condition	Anechoic	0.05	0.08
	Room Effects	0.06	0.09
Task Difficulty	Single Task	0.05	0.07
	SOA = 1.5 s	0.05	0.09
	SOA = 0.25 s	0.05	0.08
	SOA = 0.05 s	0.07	0.09

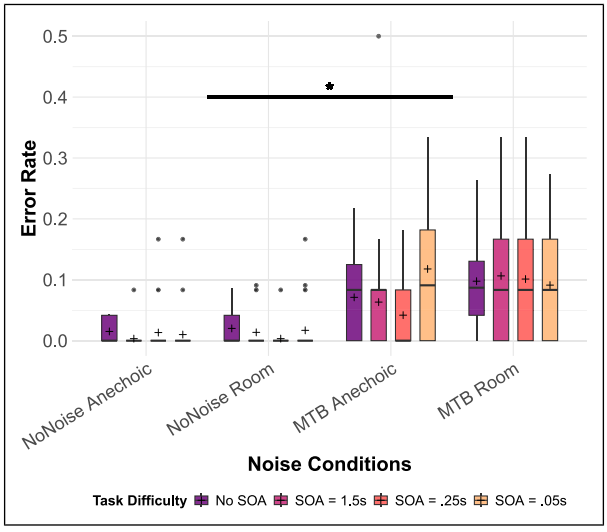

Figure 4. Word recognition error rates across noise type, acoustic conditions, and task difficulty. Box plots display the median (center line), interquartile range, and mean (plus symbol), with the combination of noise type and acoustic conditions on the x-axis and the word recognition error rate on the y-axis. The outliers are represented by circles. The asterisk indicates $p <$ .05 significance for the comparison of no noise and multi-talker babble.

significant ($p > .05$). Taken together, the analyses based on the composite listening experience score and the single effort item showed the same significant results. Accordingly, the following discussion refers to them collectively as subjective listening effort.

3.2.1 Correlation Between Subjective and Behavioral Listening Effort

Spearman correlations were calculated for each noise condition between subjective and behavioral listening effort. The listening experience score was positively correlated with the word recognition task's RT in the MTB condition with ($r_s = .32$, $p = .024$) and without room effects ($r_s = .34, p = .015$). The correlation for no noise with ($r_s = .16, p = .28$) and without room effects ($r_s = .24$, $p = .097$) was non-significant.

4. Discussion

4.1 Influence of Adverse Acoustics on Listening Effort

For this study, it was hypothesized that listening in adverse acoustic conditions involving MTB or room effects, compared to favorable conditions involving silence or anechoic conditions, increases both behavioral and subjective listening effort. The significant interaction of noise and acoustic condition for the RT of the word recognition task revealed significantly longer RTs for MTB compared to silence in the condition with room effects. This effect was expected and is in line with results for behavioral listening effort by Holube et al. (2016), Seitz et al. (2024), and Gatehouse and Gordon (1990) and supports the FUEL (Pichora-Fuller et al., 2016). That was underlined by the significant main effect of noise on subjective listening effort. Interestingly, for the anechoic MTB condition and that with room effects, behavioral and subjective listening effort were significantly correlated. This contradicts the conclusion of the systematic review by Shields et al. (2023) that listening effort

Table 4. Results of Bonferroni-Corrected Post Hoc Tests Comparing Different Levels of Task Difficulty for the Response Time (RT) in the Decision Task. $M_1 - M_2$ Denotes the Mean RT Differences and SD_1 and SD_2 the Respective Standard Deviation, Whereas the Order of the Values in the First Column Denotes 1 and 2. $|z|$ Shows the Absolute Value of the Standardized Test Statistic and p the Significance Value.

Task Difficulty Comparison	$M_1 - M_2$	SD_1	SD_2	$\|z\|$	p
SOA = 1.5 s vs. Single Task	1.27 s − 1.14 s	0.28 s	0.14 s	15.68	< .001
SOA = 0.25 s vs. Single Task	1.26 s − 1.14 s	0.22 s	0.14 s	15.03	< .001
SOA = 0.05 s vs. Single Task	1.26 s − 1.14 s	0.21 s	0.14 s	13.88	< .001

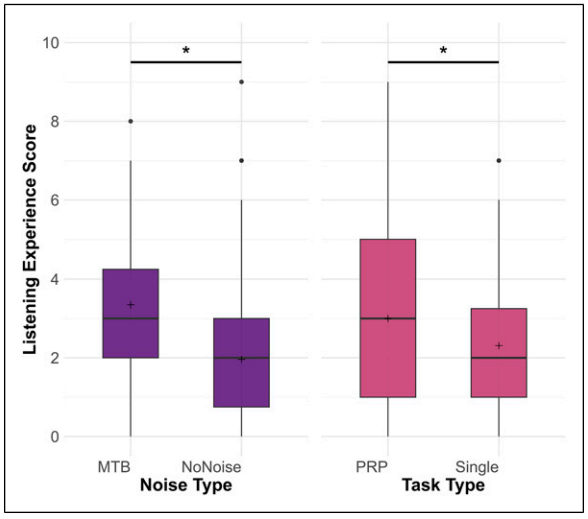

Figure 5. Box plots display the median (center line), interquartile range, and mean (plus symbol), with the listening experience score on the y-axis and noise type as well as task difficulty on the x-axis. The outliers are represented by circles, and the asterisks indicate $p < .05$ significance.

measures were overall poorly correlated. In line with this, it might be that other confounding factors, such as an individual's cognitive capacity and motivation, influence listening effort measures differently, leading to divergent outcomes. This could also explain why there was just a correlation for the MTB conditions, where it could have been that an increased task demand had more influence than those individual differences. Furthermore, each measure may describe a distinct cognitive stage or process involved in sound perception that leads to listening effort (Shields et al., 2023). Shields et al. (2023) motivated future work to integrate different participant groups while ensuring that the study is sufficiently powered. It is additionally important for future work to consider the participants' individual differences, cognitive capacity, and motivation.

Interestingly, the RTs in the word recognition task in silence were significantly higher in the anechoic than in the condition with room effects. Similar trends were found by Picou et al. (2016): They measured behavioral listening effort with word categorization RTs in a dual-task paradigm and found non-significantly longer RTs for a low-reverberant ($T_{30} < 0.1$ s) compared to a moderately-reverberant ($T_{30} = .475$ s), and highly-reverberant ($T_{30} = .834$ s) room in quiet. Holube et al., 2016 investigated subjective and physiological listening effort comparing hard ($T_{60} = 4$ s) and easy ($T_{60} = 0.5$ s) reverberant conditions in quiet and found significantly higher subjective and physiological effort for the hard condition. Notably, their chosen reverberation times were much higher than those in the present study and in Picou et al. (2016), and different listening effort measures were applied.

Further research investigated the effect of room effects on behavioral listening effort only in the presence of noise. For example, Prodi and Visentin (2019) found an increase in behavioral listening effort, measured using RTs during sentence recognition for pink and speech-shaped noise when comparing $T_{30} = 0$ s, $T_{30} = .3$ s, and $T_{30} = .6$ s at 90% speech intelligibility. Also, Huang et al. (2022) found that longer reverberation times ($T_{30} = 1.223$ s compared to $T_{30} = .469$ s) negatively affect behavioral and subjective listening effort in four-talker babble. Interestingly, the speaker setup in the study by Huang et al. (2022) was based on the critical distance beyond which word recognition performance does not further degrade, whereas the present study's speaker setup was designed based on the technical critical distance at which direct sound and diffuse sound energy density are equal (Weinzierl, 2009). It is important to choose the distance between target and receiver larger than the technical critical distance to ensure that the diffuse sound energy (including the reverberation) is larger than the direct sound energy.

If one considers the present findings of this study, they suggest that a certain degree of room effects in quiet might benefit listening effort. This would somewhat contradict the FUEL (Pichora-Fuller et al., 2016), based on which one would expect increased listening effort with increased reverberation. As in our everyday life, we experience reverberation everywhere; a certain degree of reverberation may be necessary for effortless speech processing, because anechoic scenarios are artificial. Interestingly, the present study conducted with five- and six-year-olds (Seitz et al., submitted for publication) found the same results of increased behavioral listening effort for anechoic conditions compared to with room effects.

The results of the present study indicate that further research with a range of reverberation times—from the technical critical distance (Weinzierl, 2009) up to the distance at which word recognition performance does not further degrade (Huang

et al., 2022)–is needed. Especially, as it became evident that literature on the influence of room effects on listening effort is scarce, as only a few studies have investigated this (Holube et al., 2016; Picou et al., 2016; Huang et al., 2022; Prodi and Visentin, 2019). If future research is in line with the present study's findings and that by Picou et al. (2016), the FUEL (Pichora-Fuller et al., 2016) would need to be adapted regarding the impact of reverberation on listening effort.

Overall, hypothesis one receives partial support: MTB led as expected to increased exerted behavioral and subjective effort, yet room effects indicated an opposite trend in quiet.

4.2 Influence of Task Difficulty on Listening Effort

It was hypothesized that increasing task difficulty as induced by shorter SOAs requires more cognitive resources and thus increases behavioral and subjective listening effort. For behavioral listening effort, decreasing SOAs significantly increased word recognition task RTs for both noise types, indicating the psychological refractory period (PRP) effect (Pashler, 1984, 1994). In addition, the decision task's RT was significantly longer in trials with any SOA than in trials without, a finding similarly reported in five- to ten-year-olds by Seitz et al. (submitted for publication) and Seitz and Fels (2025). Furthermore, subjectively, higher listening effort was reported for experiment blocks with SOA than for those without. These three converging measures—behavioral word recognition RTs, behavioral decision task RTs, and subjective effort ratings—demonstrate that the SOAs increased task difficulty, resulting in increased behavioral and subjective listening effort.

As expected, a significant main effect of noise was found for the word recognition task's ERs. Interestingly, the different task difficulties did not significantly influence the error rate with overall speech intelligibility remaining high (~90%) across all conditions, including MTB. The fact that increased effort is exerted even when speech intelligibility is unaffected by task difficulty aligns with the FUEL Pichora-Fuller et al. (2016), which declares that listening effort reflects cognitive resource allocation independent of audibility. Thus, hypothesis two is accepted for behavioral and subjective listening effort.

4.3 Interaction Effect of Adverse Acoustics and Task Difficulty on Listening Effort

The third hypothesis stated that there would be an interaction between adverse acoustics and task difficulty, with the strongest increase in listening effort occurring when there is a high level of task demand (e.g., short SOAs) and a challenging acoustic load (e.g., MTB and room effects are present).

An interaction between noise and task difficulty was found revealing significantly higher word recognition task's RTs in both no noise and MTB for shorter SOAs (SOA = .05 s and SOA = .25 s) compared to longer (SOA = 1.5 s) and the single task. However, MTB led to significantly increased RTs only in the SOA = 1.5 s and in the absent SOA conditions, but not for shorter SOAs. This suggests that the task difficulty in the form of decreasing SOAs was a stronger predictor for behavioral listening effort than noise type or acoustic condition.

When interpreting these results in the context of structural bottlenecks and capacity limitations (Koch et al., 2018), the increased RTs in both tasks with decreasing SOA provide evidence that processing resources are shared between those two tasks, thus supporting the models of limited cognitive capacity. The central bottleneck model would have been supported if different levels of task difficulty had influenced only word recognition RTs, indicating strict sequential processing. This interpretation is in line with the results of Seitz and Fels (2025) in six- to ten-year-olds and Seitz et al. (submitted for publication) in five- to six-year-olds. Seitz and Fels (2025) motivated future work to use shorter SOAs (which was done in the present study) as they just tested SOAs >1 s and did not find any interaction between SOA and noise. However, they also found significant differences in both tasks' RTs for the comparison of any SOA and the single task, supporting the limited capacity model (Seitz and Fels, 2025).

As noise only significantly impacted the RTs in conditions with the long and the absent SOA, the results seem to be more easily explained with a single capacity model, such as the FUEL (Pichora-Fuller et al., 2016) than a multi-capacity model (Horrey et al., 2006; Wickens, 1980, 1984). Under a straightforward multi-capacity interpretation, one might have expected further RT increases at short SOAs in MTB, as noise draws from partially different capacities (auditory) than the visual stimulus detection (visual), or the multi-tasking situation in general (cognitive). The present results indicate that the induced multi-tasking situation at the shortest SOAs already approached the participants' available capacity and aligns better with the single-capacity model, though multi-capacity effects might emerge under different task configurations.

In view of the FUEL (Pichora-Fuller et al., 2016), this suggests that some task demands—such as multi-tasking situations potentially activating automatic attention mechanisms in the case of very short SOAs—exhaust available cognitive capacity, potentially acting as a stronger predictor than input-related demands like noise. This could mean that the multi-tasking process may be prioritized compared to acoustic degradation. In the FUEL, no prioritization between automatic attention and input-related demands is present. The present findings raise the possibility of an innate prioritization of automatic attention over input-related demands, which requires future validation and may warrant model extensions.

Beechey and Naylor (2025) adapted the FUEL by more broadly conceptualizing adaptive capacity as encompassing cognitive, physical, and linguistic resources, integrated with environmental factors. They recognize listening effort as just one way of adapting to a listening situation (Beechey and Naylor, 2025). They emphasize that listening effort is not only a cost but also a functional investment that can yield perceptual benefits (Beechey and Naylor, 2025). This model includes a specific type of voluntary prioritization based on the functional investment the listener is willing to make. Future work could further elaborate the FUEL, as done by Beechey and Naylor (2025), to not only include voluntary prioritization but also involuntary prioritization—particularly if subsequent research supports the present suggestion of an underlying innate prioritization of automatic attention mechanisms.

Overall, an interaction between noise and task difficulty on subjective and behavioral listening effort was suspected. The results led to only partial support for hypothesis three. The strongest increase in behavioral listening effort occurred under high task demand (SOA = .05 s), but without being further influenced by noise. Although the hypothesis anticipated the greatest effort under combined adverse acoustics and high task demand, the findings suggest that task difficulty (short SOAs) was the more critical determinant, overshadowing additive effects of acoustic manipulation. For subjective listening effort, no significant interaction occurred. Future research could examine how noise and task difficulty interact under more adverse acoustic conditions. For example, a more challenging SNR could be used, or a more diverse population, e.g., older or hearing-impaired adults, could be investigated.

5. Conclusion

This study examined behavioral and subjective listening effort in adults using a sequential dual-task paradigm provoking the PRP effect. By systematically manipulating temporal overlap in task processing in the form of SOAs, noise condition, and acoustic environment, the study targeted the interplay between listening effort and structural bottlenecks as well as capacity models of cognitive resources.

Four key findings emerged. First, MTB significantly impacted both behavioral listening effort, reflected in longer word recognition response times, and increased subjective effort ratings. Additionally, behavioral and subjective measures were significantly correlated in MTB conditions. Second, contrary to FUEL predictions, mild reverberation ($T_{30} = 0.49$ s) in quiet conditions led to significantly lower behavioral listening effort than in anechoic environments. This finding aligns with Picou et al. (2016) and contradicts Prodi and Visentin (2019), emphasizing that future research is needed to examine the effect of multiple reverberation times on listening effort. Third, task difficulty in the form of decreasing SOAs significantly elevated both behavioral and subjective listening effort, even when speech intelligibility remained high and unaffected (~90%), supporting FUEL's dissociation between audibility and effort. As manipulating task difficulty affected both the word recognition and decision tasks' RTs, the present data support the limited capacity model more than the central bottleneck theory. Fourth, a significant interaction between noise and task difficulty emerged for behavioral listening effort, revealing that SOA was a stronger predictor of behavioral listening effort than acoustic load. Future work should further investigate this, exploring the prioritization of temporal task demands over acoustic adversity.

Acknowledgements

The authors are grateful to all participants. The authors thank Prof. Maria Klatte and her team for developing the adjusted word lists and providing the stimulus material. The authors also thank Iring Koch for the helpful discussions about the psychological refractory period effect, and Carolin Breuer for the helpful comments on an earlier version of this manuscript.

ORCID iDs

Julia Seitz https://orcid.org/0009-0003-1822-0781
Jana R. Berger https://orcid.org/0009-0001-8249-4053
Maria Anastasia Sinarso https://orcid.org/0009-0004-2484-5283
Janina Fels https://orcid.org/0000-0002-8694-7750

Author Contributions

The concept of this study was conceived and designed by J.S. and J.F. The experimental design and analysis plan were developed by J.S. J.R.B. carried out implementation and data collection under the supervision of J.S. Statistical analyses were conducted by M.A.S. in support of J.S. Results interpretation and literature search were conducted by J.S. J.S. took the lead in writing the manuscript and was supported by J.R.B and M.A.S. All authors provided critical feedback and helped finalize the research, analyses, and manuscript. J.F. supervised the project and provided all necessary resources.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This study was funded by the European Union's Horizon 2020 Research and Innovation Program under grant agreement No. 874724 (Equal-Life).

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Data Availability Statement

The datasets collected and analyzed during the current study are available from the corresponding author, J.S., on reasonable request.

Supplemental Material

Supplemental material for this article is available online.

References

Barker, L. E., & Shaw, K. M. (2015). Best (but oft-forgotten) practices: Checking assumptions concerning regression residuals. *The American Journal of Clinical Nutrition*, *102*(3), 533–539. https://doi.org/10.3945/ajcn.115.113498

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Beechey, T., & Naylor, G. (2025). How purposeful adaptive responses to adverse conditions facilitate successful auditory functioning: A conceptual model. *Trends in Hearing*, *29*, 23312165251317010. https://doi.org/10.1177/23312165251317010

Bomhardt, R., & Fels, J. (2014). Analytical interaural time difference model for the individualization of arbitrary head-related impulse responses. In *Audio Eng. Soc. Convention 137*. Audio Engineering Society.

Broadbent, D. E. (1958). *Perception and communication*. Pergamon Press.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*(3), 297–334. https://doi.org/10.1007/BF02310555

Deutsch, J. A., & Deutsch, D. (1963). Attention: Some theoretical considerations. *Psychological Review*, *70*(1), 80–90. https://doi.org/10.1037/h0039115

European Broadcasting Union. (2011). *EBU Recommendation R 128 - Loudness normalisation and permitted maximum level of audio signals*. European Broadcasting Union.

Field, A., Miles, J., & Field, Z. (2012). *Discovering statistics using R*. Sage Publications.

Francis, A. L., & Love, J. (2020). Listening effort: Are we measuring cognition or affect, or both? *WIREs Cognitive Science*, *11*(1), e1514. https://doi.org/10.1002/wcs.1514

Gagné, J.-P., Besser, J., & Lemke, U. (2017). Behavioral assessment of listening effort using a dual-task paradigm: A review. *Trends in Hearing*, *21*, 2331216516687287. https://doi.org/10.1177/2331216516687287

Gatehouse, S., & Gordon, J. (1990). Response times to speech stimuli as measures of benefit from amplification. *British Journal of Audiology*, *24*(1), 63–68. https://doi.org/10.3109/03005369009077843

Gómez-Merino, N., Gheller, F., Spicciarelli, G., & Trevisi, P. (2020). Pupillometry as a measure for listening effort in children: A review. *Hearing, Balance and Communication*, *18*(3), 152–158. https://doi.org/10.1080/21695717.2020.1807256

Hartig, F. (2024). DHARMa: Residual diagnostics for hierarchical (multi-level / mixed) regression models. *R package version 4.7) [Computer software]*.

Holube, I., Haeder, K., Imbery, C., & Weber, R. (2016). Subjective Listening Effort and Electrodermal Activity in Listening Situations with Reverberation and Noise. *Trends in Hearing*, *20*, 233121651666773. https://doi.org/10.1177/2331216516667734

Horrey, W. J., Wickens, C. D., & Consalus, K. P. (2006). Modeling drivers' visual attention allocation while interacting with in-vehicle technologies. *Journal of Experimental Psychology: Applied*, *12*(2), 67–78. https://doi.org/10.1037/1076-898X.12.2.67

Huang, H., Ricketts, T. A., Hornsby, B. W. Y., & Picou, E. M. (2022). Effects of critical distance and reverberation on listening effort in adults. *Journal of Speech, Language, and Hearing Research*, *65*(12), 4837–4851. https://doi.org/10.1044/2022_JSLHR-22-00109

Institute for Hearing Technology and Acoustics, RWTH Aachen University. Schäfer, P., Palenda, P., Aspöck, L., & Vorländer, M. (2022). *Virtual Acoustics - A real-time auralization framework for scientific research (Version 2022a) [Computer software]*. Zenodo. https://doi.org/10.5281/zenodo.13744523

International Organization for Standardization. (2020). **Ophthalmic optics - Visual acuity testing - Standard and clinical optotypes and their presentation* (DIN EN ISO Standard No. 8596:2020)*. https://doi.org/10.31030/3144393

International Organization for Standardization. (2021). **Acoustics — Acoustic quality of open office spaces.* (DIN EN ISO Standard No. 22955)*.

Janssens, W. (2008). *Marketing research with SPSS*. Pearson Education.

Kahneman, D. (1973). *Attention and Effort* (*1063*, pp. 218–226). Prentice-Hall.

Klatte, M., Lachmann, T., & Meis, M. (2010). Effects of noise and reverberation on speech perception and listening comprehension of children and adults in a classroom-like setting. *Noise Health*, *12*(49), 270–282. https://doi.org/10.4103/1463-1741.70506

Koch, I., Poljac, E., Müller, H., & Kiesel, A. (2018). Cognitive structure, flexibility, and plasticity in human multitasking—An integrative review of dual-task and task-switching research. *Psychological Bulletin*, *144*(6), 557–599. https://doi.org/10.1037/bul0000144

Kuchinsky, S. E., Gallun, F. J., & Lee, A. K. C. (2024). Note on the dual-task paradigm and its use to measure listening effort. *Trends in Hearing*, *28*, 23312165241292215. https://doi.org/10.1177/23312165241292215

Lachaud, C. M., & Renaud, O. (2011). A tutorial for analyzing human reaction times: How to filter data, manage missing values, and choose a statistical model. *Applied Psycholinguistics*, *32*(2), 389–416. https://doi.org/10.1017/S0142716410000457

Lenth, R. V. (2025). emmeans: Estimated Marginal Means, aka Least-Squares Means. *R package version 1.10.5) [Computer software]*, https://rvlenth.github.io/emmeans/

Loh, K., & Fels, J. (2023). ChildASA dataset: Speech and noise material for child-appropriate paradigms on auditory selective attention. *Technical Report*, https://doi.org/10.18154/RWTH-2023-00740

Masiero, B., & Fels, J. (2011). Perceptually robust headphone equalization for binaural reproduction. In Audio Engineering Society Convention (*130*). Audio Engineering Society.

Pashler, H. (1984). Processing stages in overlapping tasks: Evidence for a central bottleneck. *Journal of Experimental Psychology: Human Perception and Performance*, *10*(3), 358–377. https://doi.org/10.1037/0096-1523.10.3.358

Pashler, H. (1994). Dual-task interference in simple tasks: Data and theory. *Psychological Bulletin*, *116*(2), 220–244. https://doi.org/10.1037/0033-2909.116.2.220

Peelle, J. E. (2018). Listening effort: How the cognitive consequences of acoustic challenge are reflected in brain and behavior. *Ear & Hearing*, *39*(2), 204–214. https://doi.org/10.1097/AUD.0000000000000494

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 & Hearing*, *37*(1), 5S–27S. https://doi.org/10.1097/AUD.0000000000000312

Picou, E. M., Gordon, J., & Ricketts, T. A. (2016). The effects of noise and reverberation on listening effort in adults with normal hearing. *Ear and Hearing*, *37*(1), 1–13. https://doi.org/10.1097/AUD.0000000000000222

Picou, E. M., Ricketts, T. A., & Hornsby, B. W. Y. (2011). Visual cues and listening effort: Individual variability. *Journal of Speech, Language, and Hearing Research*, *54*(5), 1416–1430. https://doi.org/10.1044/1092-4388(2011/10-0154

Picou, E. M., Ricketts, T. A., & Hornsby, B. W. Y. (2013). How hearing aids, background noise, and visual cues influence objective listening effort. *Ear and Hearing*, *34*(5), e52–e64. https://doi.org/10.1097/AUD.0b013e31827f0431

Prodi, N., & Visentin, C. (2019). Impact of background noise fluctuation and reverberation on response time in a speech reception task. *Journal of Speech, Language, and Hearing Research*, *62*(11), 4179–4195. https://doi.org/10.1044/2019_JSLHR-H-19-0180

R Core Team. (2024). *R: A language and environment for statistical computing (Version 4.4.1)*. R Foundation for Statistical Computing. [Computer software]. https://www.R-project.org/

Rakerd, B., Seitz, P. F., & Whearty, M. (1996). Assessing the cognitive demands of speech listening for people with hearing losses. *Ear and Hearing*, *17*(2), 97–106. https://doi.org/10.1097/00003446-199604000-00002

Richter, M., Buhiyan, T., Bramsløw, L., Innes-Brown, H., Fiedler, L., Hadley, L. V., Naylor, G., Saunders, G. H., Wendt, D., Whitmer, W. M., Zekveld, A. A., & Kramer, S. E. (2023). Combining multiple psychophysiological measures of listening effort: Challenges and recommendations. *Seminars in Hearing*, *44*(2), 95–105. https://doi.org/10.1055/s-0043-1767669

Schmitz, A. (1995). Ein neues digitales kunstkopfmeßsystem. *Acta Acust. Acust*, *81*, 416–420.

Seitz, J., & Fels, J. (2025). *Differences between subjective and behavioral listening effort in children across age and task complexity*. Ear & Hearing. https://doi.org/10.1097/AUD.0000000000001759

Seitz, J., Loh, K., & Fels, J. (2024). Listening effort in children and adults in classroom noise. *Scientific Reports*, *14*(1), 25200. https://doi.org/10.1038/s41598-024-76932-7

Seitz, J., Westphal, V., & Fels, J. (2026). Subjective and Behavioral Listening Effort Across Task Difficulty and Listening Scenarios in Five- and Six-Year-Old Children. *Hearing Research*, (submitted for publication).

Shields, C., Sladen, M., Bruce, I. A., Kluk, K., & Nichani, J. (2023). Exploring the correlations between measures of listening effort in adults and children: A systematic review with narrative synthesis. *Trends in Hearing*, *27*, 23312165221137116. https://doi.org/10.1177/23312165221137116

Smeds, K., Gotowiec, S., Wolters, F., Herrlin, P., Larsson, J., & Dahlquist, M. (2020). Selecting scenarios for hearing-related laboratory testing. *Ear and Hearing*, *41*(Suppl 1), 20S–30S. https://doi.org/10.1097/AUD.0000000000000930

Treisman, A. M. (1960). Contextual cues in selective listening. *Quarterly Journal of Experimental Psychology*, *12*(4), 242–248. https://doi.org/10.1080/17470216008416714

Weinzierl, S. (Ed.), (2009). *Handbuch der Audiotechnik*. Springer. https://doi.org/10.1007/978-3-540-34301-1

Wickens, C. D. (1980). The structure of attentional resources. In R. S. Nickerson (Ed.), *Attention and performance VIII* (pp. 239–257). Erlbaum.

Wickens, C. D. (1984). Processing resources in attention. In R. Parasuraman & R. Davies (Eds.), *Varieties of attention* (pp. 63–101). Academic Press.

World Health Organization. (2021). **World report on hearing**. World Health Organization.