






# Heard-text recall and listening effort under irrelevant speech and pseudo-speech in virtual reality

Cosima A. Ermert<sup>1,\*</sup>, Sabine J. Schlittmeier<sup>2</sup>, Andrea Bönsch<sup>3</sup>, Torsten W. Kuhlen<sup>3</sup>, and Janina Fels<sup>1</sup>

<sup>1</sup>Institute for Hearing Technology and Acoustics, RWTH Aachen University, Aachen, Germany

<sup>2</sup>Teaching & Research Area Work and Engineering Psychology, RWTH Aachen University, Aachen, Germany

<sup>3</sup>Visual Computing Institute, RWTH Aachen University, Aachen, Germany

Received 16 September 2025, Accepted 13 January 2026

**Abstract** – *Introduction:* Verbal communication depends on a listener’s ability to accurately comprehend and recall information conveyed in a conversation. The heard-text recall (HTR) paradigm can be used in a dual-task design to assess both memory performance and listening effort. The HTR paradigm uses running speech to simulate a conversation between two talkers. Thereby, it allows for talker visualization in virtual reality (VR), providing co-verbal visual cues like lip-movements, turn-taking cues, and gaze behavior. While the HTR in a dual-task design has been investigated under pink noise, the impact of more realistic irrelevant stimuli, such as speech, that provide temporal fluctuations and meaning compared to noise, remains unexplored.

*Methods:* In this study ( $N = 24$ ), the HTR task as primary task was administered in an immersive VR environment under three noise conditions: silence, pseudo-speech, and speech. Participants performed a vibrotactile secondary task to quantify listening effort in a dual-task design.

*Results:* The results indicate an effect of irrelevant speech on memory and speech comprehension as well as secondary task performance, with a stronger impact of speech relative to pseudo-speech.

*Discussion:* The study validates the sensitivity of the HTR in a dual-task design to background speech stimuli and highlights the relevance of linguistic interference-by-process for listening effort, speech comprehension, and memory.

**Keywords.** Listening effort, Virtual reality, Masking, Dual-task, Heard text recall

## 1 Introduction

Listening to and remembering speech is a demanding task, especially in listening scenarios with background noise. Background noise can give rise to listening effort, even if the target speech may still be perfectly intelligible. Listening effort generally refers to the cognitive resources allocated to understanding speech, especially under adverse conditions [1], and is defined as the “deliberate allocation of mental resources to overcome obstacles in goal pursuit when carrying out a task” (p. 10) [2]. Importantly, compared to, e.g., speech reception thresholds, listening effort can be measured at suprathreshold listening conditions – which is a lot more common for everyday listening scenarios [3], benefiting ecological validity [4].

Listening effort can be modulated by a multitude of factors [5, for a review], e.g., the intelligibility of the target signal [6–8], room acoustic reverberation [9],

target location changes [10], the listeners’ age [11, 12], or hearing impairment and hearing device algorithms [13–19]. But also different types of background noise impose qualitatively distinct challenges: from an acoustic perspective, a distinction is made between energetic masking (e.g., white noise, speech-shaped noise), where the target speech signal is obscured due to an energy overlap in relevant frequency bands [20], and information masking (e.g., speech, multi-talker babble), which occurs when listeners fail to successfully segregate target and masker streams or to allocate attention effectively to the target [21–23]. Informational masking, therefore, refers to perceptual and attentional challenges that go beyond spectrotemporal overlap. From a cognitive-psychological perspective, meaningful background speech as an informational masker poses yet another source of disruption: because it is processed up to the semantic level, it can interfere directly with comprehension and memory of the target material – a mechanism described as interference-by-process [24, 25].

\*Corresponding author: [cer@akustik.rwth-aachen.de](mailto:cer@akustik.rwth-aachen.de)

In listening effort experiments, meaningful background speech almost inevitably elicits larger listening effort than signals without linguistic meaning: this has been reported for irrelevant one-talker speech [26–32], two-talker babble [11, 33], as well as three- to four-talker babble [34–36]. When several speech maskers are compared directly, research suggests that, after controlling for the signal-to-noise ratio, effort increases as the number of competing voices decreases [28, 37]. The effect is likely two-fold: fewer voices might preserve more prosodic and semantic information and therefore induce stronger informational masking, but also allow for in-the-dip listening [23], which represents an energetic masking benefit. Even when speech is manipulated to be unintelligible, e.g., through time-reversal, it tends to evoke more effort than amplitude-modulated or stationary speech-shaped noise, underscoring the role of temporal and spectrotemporal variability [38]. Amplitude-modulated speech-shaped noise occupies an intermediate position: several studies observed greater listening effort for modulated than stationary speech-shaped noise, but less than for intelligible speech [32, 39–41]. As a general trend, listening effort seems to be modulated by the informational content and the time-variance of the background speech, e.g., allowing for catching parts of the target sentence in energy dips of the background speech [42–45].

There are multiple ways to quantify listening effort [46, 47]; it can be measured subjectively, e.g., via questionnaires [6, 48, 49], through physiological measures, e.g., via pupillometry [6, 18] and neural imaging [50, 51], or behaviorally, e.g., via dual-task paradigms [15, 52, 53]. Dual-task paradigms are based on the theory that the total cognitive processing resources are limited in capacity [54, 55], leading to performance limitations when two tasks demand the same processing capacities [1, 56]. Participants perform a primary listening task while at the same time attending to a secondary, concurrent task (e.g., reaction-time measurements or categorization tasks). Due to the limitation of cognitive resources, increased effort is evidenced by compromised secondary task performance as speech processing demands in the primary task increase [1, 56, 57]. To assess this decline in secondary task performance, both tasks are administered in isolation in single-tasking (i.e., only one task at a time; the primary task alone and the secondary task alone) and together as a dual-task (i.e., both tasks at the same time; the primary task and the secondary task together). Performance metrics are assessed for both tasks (primary task, secondary task) in both conditions (single-tasking, dual-tasking).

Dual-task designs with running speech are rare in the literature; most paradigms use sentence-, word-, or syllable-recognition tasks [1], which does not reflect speech processing in realistic listening situations. The recently developed heard-text recall (HTR) by Schlittmeier et al. [58, 59], in which participants listen to short, coherent texts and are asked content-related questions afterwards, assesses text comprehension and verbal memory. When implemented in a dual-task

design in combination with a secondary task, listening effort in running speech can be assessed in addition. The paradigm can also be embedded into virtual reality (VR) environments [60–62], where embodied conversational agents are animated to speak the HTR texts, effectively conveying co-verbal visual cues like lip movements, turn-taking cues, and gaze behavior. Because the primary HTR task explicitly requires semantic comprehension and memory, it should be especially susceptible to meaningful background speech that induces interference-by-process. In contrast, a non-linguistic secondary task (e.g., vibrotactile pattern discrimination) serving as the indicator of listening effort is not directly vulnerable to linguistic interference; decrements in performance are expected to reflect competition between limited cognitive resources [63] rather than semantic processing per se. Differences between single- and dual-tasking further reflect the listening effort exhibited in the primary listening task. In the present experiment, the HTR served as the primary task within a dual-task paradigm, combined with a non-linguistic vibrotactile pattern-discrimination secondary task.

Mohanathanasan et al. [64] investigated the influence of stationary pink noise with a challenging and a moderate signal-to-noise ratio on the primary HTR task and a secondary vibrotactile task. They demonstrated an effect of background noise on the primary HTR task only in the challenging signal-to-noise ratio and a sensitivity of the secondary vibrotactile task response times to both signal-to-noise ratios, while the performance in the secondary task in terms of error rates was unaffected. While studies employing a more realistic background noise (classroom noise) using the same dual-task exist [61, 62], these studies did not explicitly focus on the effect of noise variations. Thus, how different maskers (e.g., intelligible vs. unintelligible speech maskers) influence both performance and effort within the specified dual-task paradigm is still insufficiently explored. In the present study, building upon the VR implementation by Ehret et al. [60], we test whether maskers differing in semantic content and meaningfulness but matched in spectrotemporal structure (intelligible speech vs. pseudo-speech) [65], differentially affect primary HTR and secondary vibrotactile task performance.

We aim to disentangle effects of energetic masking (expected for both speech and pseudo-speech) from additional linguistic interference (expected only for intelligible speech) in the primary HTR task, and to test whether such additional interference indirectly taxes resources and thereby impairs secondary task performance in dual-tasking. Based on these considerations, we hypothesized that primary HTR performance would be best in silence, reduced under pseudo-speech due to energetic masking, and most strongly impaired under intelligible speech due to concurrent linguistic interference-by-process and increased informational masking.

For the dual-task condition, we further expected the pattern of performance decrements in the primary HTR task to be mirrored in the secondary task performance

(i.e., decreased accuracy and longer response times) as an indicator of listening effort. As more cognitive resources are drawn upon to maintain text comprehension and verbal memory in the primary HTR task under energetic and concurrent linguistic processing, fewer cognitive resources remain available for the secondary vibrotactile task. Even though the vibrotactile task is non-auditory and thus does not require semantic processing, background noise might influence the performance even in single-tasking [64] due to noise consuming mental resources and increasing stress [63]. To ensure that the performance decrements of the vibrotactile secondary task in dual-tasking can be attributed to increased listening effort in the primary task instead of general noise effects [63], and to obtain baseline performance, both tasks are also administered in single-tasking.

## 2 Method

### 2.1 Participants

We conducted a Bayesian design analysis for the present experiment design. Assuming a moderate main effect of the *masker condition*, a strong main effect of *number of tasks* [59, 64], and a small interaction, we performed a Bayesian assurance analysis in R (version 4.4.1) [66]. More details regarding the models can be found in the Data Analysis Section. The minimal sample size required to achieve at least 80% assurance was  $n = 21$ , to accommodate potential data loss and for a fully balanced design, we recruited  $n = 24$  participants. Participants were adults (9 female) and aged 22–33 years ( $M = 25.2$ ,  $SD = 2.6$ ). They had German as their first language, as the HTR task is only available in German at this point. Participants had to pass an audiometry (below 25 dBHL between 125 Hz and 8 kHz; pulsed pure-tone) and a Snellen Test (20/30) [67] to verify their visual acuity. They received a 10-euro voucher for a local bookstore as compensation for their time.

### 2.2 Dual-task paradigm

The dual-task paradigm consisted of a primary HTR and a secondary vibrotactile task. In the primary HTR [58] task, participants listened to and saw two embodied conversational agents narrate family stories in German consisting of ten sentences each. After the presentation of each text, nine content-related questions were displayed sequentially on a virtual television (see Fig. 1), and participants had to answer them vocally. An instructor listened to the participants' input and categorized the answer as correct or incorrect based on the ideal solution provided in the HTR database by Schlittmeier et al. [58], detailing all the texts, questions, and answers. By evaluating the number of correct responses, speech comprehension and memory could be examined [58].

The two agents took turns between sentences while narrating the stories. This turn-taking behavior was



**Figure 1.** Virtual reality setup with two embodied conversational agents narrating the HTR stories COMMA and the television displaying instructions and the HTR questions whenever appropriate.

implemented according to the suggestions by Ermert et al. [68]: successive sentences that form a unit of meaning were assigned to one embodied conversational agent before yielding the turn. Thus, the number of sentences spoken by each agent was pseudo-random, but the speaking time per text was approximately the same. Agents never interrupted each other. The texts had an average presentation duration of  $M = 43.4$  s,  $SD = 1.9$  s. The audio stimuli for the target sentences of the HTR were taken from the audio-visual speech and text database (AuViST) [68], which provides audio recordings and face tracking data for all sentences of the HTR.

The secondary task assessing listening effort was vibrotactile [64]. Participants held one HTC Vive controller (High Tech Computer Corporation, Taiwan) in each hand. The controllers vibrated in four distinct patterns: short–short, long–long, short–long, and long–short. If the second vibration was a repetition of the first vibration (i.e., short–short, long–long), participants had to click a designated button on the left controller. Otherwise, they had to click a designated button on the right controller [60].

Both tasks were administered together (dual-task) and also in isolation (single-task) to obtain a baseline performance.

### 2.3 Masker conditions

Three masker conditions were examined: silence (no masking), pseudo-speech (primarily energetic masking), and speech (energetic and informational masking). In the silence condition, only the target talkers were audible. In the pseudo-speech and speech conditions, a male and a female distractor talker spoke simultaneously and without breaks, in addition to the target talkers.

The stimuli in both the background speech conditions were obtained from female [69] and male [70] voice recordings of the German matrix sentences test [71–73]. In the pseudo-speech condition, the matrix sentences

were cut into syllables, one-syllable words were excluded, and the remaining syllables were rearranged in random order, ensuring that no meaningful words were formed. Crossfading of 10 ms was applied to smooth transitions (cf. [74]). In the speech condition, the original matrix sentences were played back in random order.

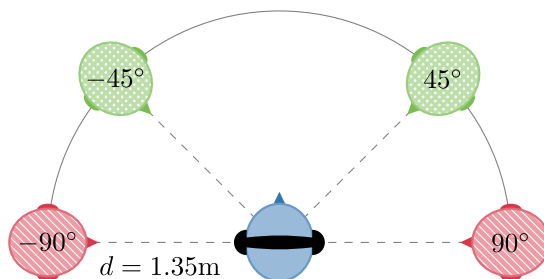
While other energetic maskers could have been possible, e.g., steady-state speech-shaped noise, time-modulated speech-shaped noise, foreign language recordings, or time-reversed target speech, none of them would have retained the same frequency spectrum of the speech condition while also preserving the temporal variations. Thus, differences in energetic masking compared to the speech condition could be possible [23]. Further, we aimed for a speech-like signal to increase realism of the background sounds compared to the steady-state noise signals used in the study by Mohanathasan et al. [64]. Therefore, the pseudo-speech signal described here was deemed a suitable solution. The stimuli were calibrated using an HMS III artificial head (HEAD acoustics GmbH, Germany) to 60 dB(A) for the target and 57 dB(A) for the background.

## 2.4 Audio-visual virtual reality setup

The experiment was conducted in a living room VR environment, which was presented via an HTC Vive Pro Eye head-mounted display (High Tech Computer Corporation, Taiwan) and created in Unreal Engine 5.3 (Epic Games Inc., United States of America). In the VR environment, participants saw two embodied conversational agents [75], one male and one female MetaHuman (Epic Games Inc., United States of America), narrating the HTR stories as a conversation. The embodied conversational agents were animated with gestures and gazing based on Ehret et al. [60] with synchronized face-tracked lip-movement from the audio-visual speech and text database [68]. All agent movement was implemented using the Character Plugin [76]. The study flow was controlled using the StudyFramework Plugin for Unreal Engine [77]. Further, the RWTH VR Toolkit Plugin was used to implement the participant interaction [78]. A video of an exemplary HTR text is provided in the publication by Ehret et al. [60] (supplementary video, condition “ $T_{full}$ ” at  $t = 3:23$  min).

The embodied conversational agents, as target talkers, were located at a distance of 1.35 m and a horizontal offset of  $\pm 45^\circ$  from the frontal direction. Invisible but audible distractor sound sources for pseudo-speech and speech conditions were located at  $\pm 90^\circ$  from the frontal direction, i.e., to the left and right (see Fig. 2). The gender of the target and background was matched, i.e., the distractor sound source at  $-90^\circ$  played irrelevant speech in a male voice, and the sound source at  $+90^\circ$  in a female voice.

The auditory scene was created in Virtual Acoustics v2022a [79] in combination with the Virtual Acoustics plugin for Unreal Engine [80] and reproduced via



**Figure 2.** Auditory virtual setup with two virtual target sources (green) and two invisible distractor sources (red).

Sennheiser HD650 headphones (Sennheiser electronic SE & Co. KG, Germany) using a generic head-related transfer function [81] and individual headphone equalization [82]. Using head tracking via the head-mounted display, dynamic scene rendering was employed.

## 2.5 Procedure

The experiment was performed in a sound-isolated booth. Participants sat in a chair. After passing the audiometry and vision test, participants were introduced to the head-mounted display and the virtual environment. Then, the training started. First, the secondary vibrotactile task was administered in silence. Participants completed at least 20 trials but could repeat the task as often as desired until they felt secure. Feedback was given automatically on the virtual television screen. Afterwards, the primary HTR task was trained in single tasking: one HTR text was administered without the secondary vibrotactile task. Participants answered the questions displayed on the virtual television screen. They practiced one more HTR text in the dual-tasking condition together with the secondary vibrotactile task. No background noise was played during training. Then, the main experiment started. No feedback was given in the main experiment. All participants performed both tasks under all combinations of the variables *number of tasks* (dual-task, single-task) and *masker condition* (silence, pseudo-speech, speech) in a within-subject design. The experiment consisted of three blocks: one per masker condition. The order of the noise blocks was counterbalanced using balanced Latin squares. At the beginning of each block, participants could listen to the noise to familiarize themselves with it. In each block, four HTR texts in dual-tasking (together with the secondary vibrotactile task), one HTR text in single-tasking, and 40 trials of the vibrotactile task in single-tasking were administered. The order of these tasks was balanced within the noise blocks. Texts 1–17 [58, 68] were used in random order. No texts were repeated. Participants could take breaks of no fixed length between noise blocks. At the end of the experiment, participants filled out a questionnaire regarding the perceived ecological validity of the experiment. As this aspect is not the focus of the present research

question, no further details about the questionnaire are reported here. The experiment took around 90 min, of which around 60 min were spent immersed in the VR environment.

## 2.6 Data analysis

The statistical analysis was carried out in R (version 4.4.1) [66]. The three outcome measures were the binary *performance* in the primary HTR task (1 = answer correct, 0 = answer incorrect), the binary *performance* in the secondary vibrotactile pattern categorization task (1 = answer correct, 0 = answer incorrect or missed), and the *response time* in the secondary vibrotactile task (cf. [64]). The independent variables were *number of tasks* (dual-task, single-task) and *masker condition* (silence, pseudo-speech, speech). Although participants were encouraged to prioritize the primary HTR task over the secondary vibrotactile task, some neglected the secondary task to such an extent that it raises concerns about the validity of the corresponding data. Therefore, a miss rate was calculated for each participant, and outliers were identified using the *is\_outlier()* function from the *rstatix* package (version 0.7.2) [83] and excluded from the analysis. This procedure resulted in the exclusion of three participants.

Separate Bayesian generalized mixed-effects models were fitted for each of the three outcome measures using the R package *brms* (version 2.18) [84]. The best-fitting model was determined in a step-wise procedure, starting with an intercept-only baseline model. In subsequent models, the random intercepts *participant* and *question* (combination of text and question number for the primary HTR task) and the independent variables *number of tasks* and *masker condition*, as well as their interaction, were incorporated. The prior distributions were sampled 16 000 times, employing No-U-Turn Sampling and using four separate chains of 5000 samples each, with the initial 1000 warm-up samples discarded for each independent variable and their interactions. A Bernoulli distribution on the logit scale was used for the *performance* and a Gamma distribution on the log scale for the *response time*. Priors were weakly informative, i.e., normal distributions each with a mean  $M = 0$  and a standard deviation  $SD = 2$  for the independent variables and heavily tailed Student- $t$  distributions with  $M = 0$  and  $SD = 2$  for the random intercepts. Using the leave-one-out cross-validation criterion, performance across models was compared based on the differences in expected log pointwise predictive density (elpd) and the standard error of the elpd of the models [85]. For all three independent variables, *performance*, *question* (for the primary HTR task), *number of tasks*, and *masker condition* improved the model fit. Models including the interaction between *number of tasks* and *masker condition* did not differ in a meaningful way from additive models for all three variables ( $\text{elpd}_{\text{diff}} < 2$ ). Thus, the interaction model was chosen to comprehensively encompass the experimental design.

Meaningful main effects or interactions detected in these models were explored further with post-hoc pairwise comparisons of estimated marginal means, conducted using the *emmeans* package (version 1.10.4) [86], and further summarized using the median and 95% Credible Intervals (CIs) calculated using the Highest Density Interval (HDI) of the Posterior Probability Distribution (PPD). The probability of direction (PD) of the PPD and the proportion of the 95% CI inside the region of practical equivalence (ROPE) (defined as  $\pm 0.1 \times SD$  [87]) were used to quantify effects. The PD of the PPD indicates whether an effect exists, with a PD roughly of 97.5% roughly corresponding to two-sided  $p$ -values of 0.05. As PD is not a clear index of a meaningful effect, the proportion of the 95% CI inside the ROPE was employed to signify an effect of negligible magnitude.

## 3 Results

The accumulated performances for the primary HTR task, secondary vibrotactile task, and response times in the secondary vibrotactile task are displayed in Figure 3. The pairwise comparisons for the final models are presented in Table 1. The pairwise comparisons for *number of tasks* indicated a meaningful effect for all three masker conditions and all three outcome measures with PDs > 99% and 0% in ROPE.

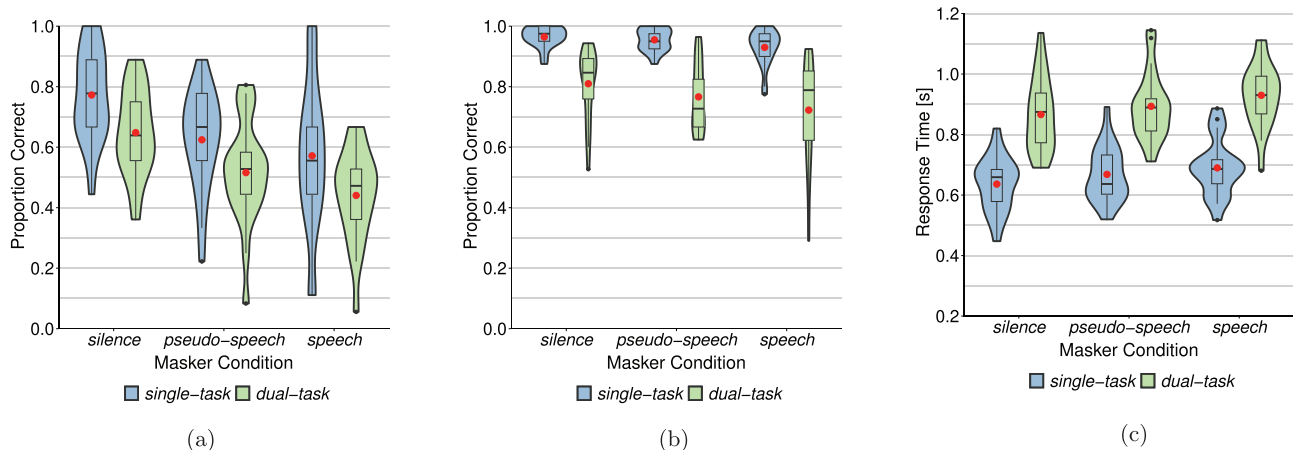
Regarding the comparisons of *masker condition* in the primary HTR task, PDs were generally quite high (> 97.5%), indicating the existence of an effect. This was supported by the low percentage within ROPE across conditions. An exception was the comparison of pseudo-speech and speech in the *single-task* condition with a PD of 78.08%.

In the secondary vibrotactile task, PDs for all comparisons of *masker condition* were high, and the percentage of PPD within ROPE was low across conditions; except for the comparisons between speech and pseudo-speech in the *single-task* condition with a PD of 84.80% and a high percentage within ROPE, and the comparison between pseudo-speech and speech, in which the 95% CI encompasses the null. Thus, no clear evidence for an effect between these conditions could be gathered.

Lastly, most PD of the response times in the secondary vibrotactile task were over 95.7% combined with low percentages within ROPE, signifying an effect. Only the comparison between pseudo-speech and speech in the *single-task* evoked a PD below 97.5%, and the CI encompasses the null, indicating the absence of an effect.

## 4 Discussion

As evident in other studies employing the HTR in a dual-task design [59, 62, 64], performance in dual-tasking was worse than in single-tasking in all three outcome measures (primary HTR task performance, secondary vibrotactile task performance, secondary vibrotactile task response time), reflecting a dual-task cost.



**Figure 3.** Results of the (a) primary HTR task, (b) secondary vibrotactile task, and (c) response times in the secondary vibrotactile task. Proportion correct (for the primary and secondary vibrotactile task) and accumulated response times in seconds (for the secondary vibrotactile task) are plotted as a function of the *masker condition* (silence, pseudo-speech, speech) and the *number of tasks* (single-task, dual-task), averaged per participant. The boxes represent the interquartile ranges, while the median is indicated with a horizontal line and the mean with a red circle.

Our hypothesis that primary HTR performance would be best in silence, reduced under pseudo-speech, and most strongly impaired under intelligible speech in both single- and dual-tasking, was supported by the results: a high dependency on the masker condition could be detected in the primary HTR task. Short-term memory of verbal information, which the primary HTR task depends on, has been demonstrated to be impacted more strongly by meaningful than foreign speech or pseudo-speech that does not convey meaning [74, 88]. Similar results can be found for speech comprehension [1, 37]. The comparison between pseudo-speech and speech was not meaningful only in single-tasking. It is possible that the difference between these two maskers was not as pronounced in single-tasking as in dual-tasking, as more cognitive resources were available for processing. Further, the limited number of trials in the single-task condition compared to the dual-task condition might not have been sensitive enough to account for the difference between pseudo-speech and speech. The current design was chosen to minimize variance within the dual-task condition, as these conditions were the central focus of our research hypotheses, and in accordance with previous HTR studies [59, 62, 64]. For future studies, a more balanced number of trials should be aimed for.

We further hypothesized that if more cognitive resources are drawn upon to maintain primary task performance, the performance in the secondary vibrotactile task would be affected in dual-tasking. The results generally suggested that listening in the speech condition elicited the worst secondary task performance and the highest response times, while performance in silence was best. This is supported by literature, where listening effort has been shown to be higher if the target speech is degraded by speech than by non-speech or pseudo-speech stimuli [1, 11, 30]. It should be noted, however,

that this difference is rather subtle, as seen in the median in Figure 3.

Further, performance in single-tasking was comparable to dual-tasking, hinting towards a general noise effect within the vibrotactile secondary task, potentially due to the maskers drawing upon mental resources [63]. This is further supported by the results of Mohanathan et al. [64], who also found a sensitivity of the secondary task towards noise in single-tasking. Listening effort should be primarily apparent in the differences of the masker conditions between single-tasking and dual-tasking. There were some comparisons that did not yield meaningfulness in single-tasking, but were significant in dual-tasking, indicating increased listening effort. For example, we could not detect a meaningful difference between silence and pseudo-speech, as well as between pseudo-speech and speech, in the secondary vibrotactile task performance in single-tasking, contrary to dual-tasking, while the response times mirrored the pattern of the primary HTR task. In dual-tasking, all three comparisons were meaningful in both behavioral metrics. This suggests that the added cost of processing the meaningful speech signal compared to the pseudo-speech signal leads to more errors and longer response times in dual-tasking and thus increased listening effort.

These results should, however, be taken with a grain of salt. Generally, rather few mistakes were made in the secondary task in single-tasking in general, as visible in Figure 3b, indicating a ceiling effect in the secondary task performance. This is comparable to participants' performance in Mohanathan et al. [64]: albeit in a non-VR scenario, they report average performances of close to one in single-tasking. Schiller et al. [62] report a similar high secondary vibrotactile task performance in single-tasking in VR. Further, as discussed before, the number of single-tasking trials in the primary task was limited.

**Table 1.** Summary of pairwise comparisons of the combinations of *masker condition* and *number of tasks* in the primary HTR and the secondary vibrotactile task. Meaningful effects are indicated in bold.

	Masker condition	Number of tasks	Median	95% CI	PD	% in ROPE
Primary Task	Silence – pseudo-speech	Single-task	0.18	[0.08, 0.28]	<b>99.99%</b>	<b>0%</b>
	Silence – speech		0.22	[0.11, 0.32]	<b>100%</b>	<b>0%</b>
	Pseudo-speech – speech		0.04	[-0.06, 0.15]	78.07%	14.09%
	Silence – pseudo-speech	Dual-task	0.17	[0.11, 0.22]	<b>100%</b>	<b>0%</b>
	Silence – speech		0.25	[0.19, 0.30]	<b>100%</b>	<b>0%</b>
	Pseudo-speech – speech		0.07	[0.02, 0.13]	<b>99.56%</b>	<b>0%</b>
	Silence	Single-task – dual-task	0.13	[0.05, 0.20]	<b>99.92%</b>	<b>0%</b>
	Pseudo-speech		0.12	[0.03, 0.21]	<b>99.55%</b>	<b>0%</b>
Speech	0.15		[0.06, 0.24]	<b>99.98%</b>	<b>0%</b>	
Secondary Task	Silence – pseudo-speech	Single-task	8.76e-03	[-0.01, 0.03]	84.80%	39.59%
	Silence – speech		0.03	[0.01, 0.05]	<b>99.96%</b>	<b>0%</b>
	Pseudo-speech – speech		0.02	[0.00, 0.05]	98.47%	3.91%
	Silence – pseudo-speech	Dual-task	0.04	[0.01, 0.08]	<b>99.64%</b>	<b>0%</b>
	Silence – speech		0.09	[0.05, 0.13]	<b>100%</b>	<b>0%</b>
	Pseudo-speech – speech		0.04	[0.01, 0.08]	<b>99.22%</b>	<b>0%</b>
	Silence	Single-task – dual-task	0.15	[0.11, 0.19]	<b>100%</b>	<b>0%</b>
	Pseudo-speech		0.18	[0.14, 0.23]	<b>100%</b>	<b>0%</b>
Speech	0.20		[0.16, 0.25]	<b>100%</b>	<b>0%</b>	
Secondary Task Response Time	Silence – pseudo-speech	Single-task	-31.97	[-54.57, -11.05]	<b>99.83%</b>	<b>0%</b>
	Silence – speech		-51.92	[-74.50, -30.41]	<b>100%</b>	<b>0%</b>
	Pseudo-speech – speech		-19.95	[-41.43, 3.39]	96.08%	11.63%
	Silence – pseudo-speech	Dual-task	-32.85	[-57.57, -5.70]	<b>99.29%</b>	<b>0.41%</b>
	Silence – speech		-65.21	[-92.51, -37.85]	<b>100%</b>	<b>0%</b>
	Pseudo-speech – speech		-32.39	[-60.84, -4.87]	<b>99.01%</b>	<b>1.45%</b>
	Silence	Single-task – dual-task	-225.75	[-250.90, -200.42]	<b>100%</b>	<b>0%</b>
	Pseudo-speech		-226.68	[-252.50, -200.16]	<b>100%</b>	<b>0%</b>
Speech	-238.89		[-267.10, -211.45]	<b>100%</b>	<b>0%</b>	

Listening-effort effects may be more pronounced when the primary task is strongly prioritized. In canonical dual-task paradigms, participants are instructed to maintain performance on the primary listening task across conditions, so that any decrements are expressed in the secondary task [1]. We did not observe this pattern: performance on the primary HTR task was affected by both the masker condition and the number of concurrent tasks. This likely attenuated listening-effort effects in the secondary task. We can only hypothesize why this was the case. One possibility is a timing difference between the tasks: information in the primary HTR task had to be retained until the end of the text (with no predefined memory target), whereas the vibrotactile task required immediate responses. This may have made it easier to prioritize the secondary task over the primary HTR task, with participants only realizing insufficient prioritization when answering the content-related questions after each trial. Notably, three participants exhibited exceptionally high proportions of missed responses in the vibrotactile task, suggesting that some may have abandoned the secondary task in most trials. The choice of a suitable secondary vibrotactile task is an intricate pursuit, balancing factors such as task difficulty, timing, and modality [1]. While the secondary vibrotactile task is advantageous in the sense that it is decoupled from the audio-visual VR scene (cf. [64]), an improved bal-

ance between the difficulty of the primary and secondary vibrotactile task could be aimed for.

The HTR task together with a vibrotactile secondary task has been examined by Mohanathan et al. [64] under pink stationary noise at two noise conditions: a moderate and challenging signal-to-noise ratio. The primary HTR task was sensitive only to the challenging signal-to-noise ratio, and the response times in the secondary vibrotactile task were sensitive to both signal-to-noise ratios. The secondary vibrotactile task performance in terms of errors was, in contrast, unaffected. It is possible that the stationary noise conditions in Mohanathan et al. [64] were not strong enough to elicit clear differences in all three outcome measures. In the present study, with the pseudo-speech and speech as more challenging auditory markers, all three outcome measures were affected. Speech comprehension and memory of running speech were worse under intelligible speech than under pseudo-speech and best in silence. While differences in vibrotactile secondary task performance, both with respect to the accuracy and response times between masker conditions could be shown, these indicate listening effort only to a limited extent and rather show that also in the secondary task, noise consumes cognitive resources and triggers compensatory efforts [63]. Still, small differences between masker effects when comparing single-tasking

and dual-tasking indicate that listening effort rises for speech compared to pseudo-speech.

## 5 Limitations

While an effort was made to provide a realistic listening scenario with spatialized audio and co-verbal visual cues, there are still some design factors that provide room for improvement when aiming for complex, realistic scenarios. As has been shown in literature, the processing of audio-visual speech conveyed by virtual agents is different than that of real-talkers [49, 89, 90], and the naturalness of the animation of the agents could still be improved [60]. The background talkers were not visualized. This was done to investigate the effect of background speech in isolation from possible confounding effects of visual distractors. This approach may have presented certain plausibility and ecological validity challenges, potentially limiting the transferability of the results to realistic scenarios. For example, it might have increased task demands due to the added burden of resolving sensory mismatches [91]. However, Ehret et al. [61] investigated the importance of embodied background noise sources with a dual-task paradigm using the HTR as primary task and a vibrotactile dual-task and found that listening effort and memory were not influenced by the visualization or absence of embodied distractors.

Furthermore, room acoustic effects were not included to avoid distinct signal-to-noise ratios for each target source due to differential effects of room reflections, e.g., by varying distance to the walls or room modes. As reverberation can be, however, a relevant modulator of listening effort [9, 12, 92–94], they should be included in further studies.

As discussed, some comparisons might not have yielded meaningfulness due to the limited number of datapoints (i.e., only nine questions in the HTR in single-tasking). This decision was made to reduce variance in the dual-task condition while balancing time- and resource-demands of the participants. More trials, or a more balanced distribution of HTR texts between single- and dual-tasking, are recommended for future studies.

## 6 Conclusion

The HTR paradigm by Schlittmeier et al. [58] in which participants listen to short stories and have to remember the content, measuring speech comprehension, memory, and – when combined with a secondary task – listening effort, was examined under three types of background noise: silence, pseudo-speech without meaning, and intelligible meaningful speech. A plausible scenario was created by presenting an audio-visual VR scene, in which embodied conversational agents spoke the HTR sentences. Memory of running speech in the primary HTR task was best in silence and worse under meaningful speech, replicating established findings from paradigms using simpler stimuli. An increased listening effort for speech over pseudo-speech

is indicated in the vibrotactile secondary task. These findings validate the HTR dual-task in VR within the broader scope of listening effort literature and highlight the relevance of linguistic competition in speech background noises for memory and listening effort. To further enhance the realism of the experimental environment, room acoustics, embodied distractor talkers, and more diverse sound environments should be included.

### Acknowledgments

The authors would like to thank Chinthusa Mohanathanasan and Jonathan Ehret for their valuable project collaboration.

### Funding

This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation): SPP2236 – Project number 444724862: Listening to and remembering conversations between two talkers: cognitive research using embodied conversational agents in audiovisual virtual environments. Sabine Schlittmeier’s contribution was supported by a grant from the HEAD-Genuit-Stiftung (Head-Genuit Foundation; P-16/10-W).

### Conflicts of interest

The authors declare that there is no conflict of interest.

### Data availability statement

Data are available on request from the authors.

### Author contribution statement

C.E., S.S., and J.F. conceptualized the experiment. C.E. and S.S. developed the methodology. C.E. and A.B. developed the experiment software. C.E. curated the data and analyzed the results. C.E. wrote the manuscript; A.B., S.S., T.K., and J.F. reviewed and edited the manuscript. J.F., S.F., and T.K. provided resources and acquired funding. J.F. supervised the experiment and was responsible for project administration.

### Ethics approval

The study was pre-approved by the ethics committee of the medical faculty of RWTH Aachen University (EK396-19) and conducted in accordance with the Declaration of Helsinki.

### Informed consent

Participants gave informed written consent prior to participating in the experiment.

### References

1. J.-P. Gagné, J. Besser, U. Lemke: Behavioral assessment of listening effort using a dual-task paradigm: a review. *Trends in Hearing* 21 (2017) 2331216516687287.
2. M.K. Pichora-Fuller, S.E. Kramer, M.A. Eckert, B. Edwards, B.W.Y. Hornsby, L.E. Humes, U. Lemke, T. Lunner, M. Matthen, C.L. Mackersie, G. Naylor, N.A. Phillips, M. Richter, M. Rudner, M.S. Sommers, K.L. Tremblay, A. Wingfield: Hearing impairment and cognitive energy: the framework for understanding effortful listening (FUEL). *Ear and Hearing* 37, Suppl 1.1 (2016) 5–27.
3. F. Wolters, K. Smeds, E. Schmidt, E.K. Christensen, C. Norup: Common sound scenarios: a context-driven categorization of everyday sound environments for application in hearing-device research. *Journal of the American Academy of Audiology* 27, 7 (2016) 527–540.

4. G. Keidser, G. Naylor, D.S. Brungart, A. Caduff, J. Campos, S. Carlile, M.G. Carpenter, G. Grimm, V. Hohmann, I. Holube, S. Launer, T. Lunner, R. Mehra, F. Rapport, M. Slaney, K. Smeds: The quest for ecological validity in hearing science: What it is, Why it matters, and How to advance it. *Ear and Hearing* 41 (2020) 5–19.
5. U. Lemke, J. Besser: Cognitive load and listening effort: concepts and age-related considerations. *Ear and Hearing* 37 (2016) 77–84.
6. A.A. Zekveld, S.E. Kramer: Cognitive processing load across a wide range of listening conditions: insights from pupillometry. *Psychophysiology* 51, 3 (2014) 277–284.
7. D. Wendt, R.K. Hietkamp, T. Lunner: Impact of noise and noise reduction on processing effort: a pupillometry study. *Ear and Hearing* 38, 6 (2017) 690.
8. M. Krueger, M. Schulte, M.A. Zokoll, K.C. Wagener, M. Meis, T. Brand, I. Holube: Relation between listening effort and speech intelligibility in noise. *American Journal of Audiology* 26, 3 (2017) 378–392.
9. J. Rennie, H. Schepker, I. Holube, B. Kollmeier: Listening effort and speech intelligibility in listening situations affected by noise and reverberation. *The Journal of the Acoustical Society of America* 136, 5 (2014) 2642–2653.
10. T. Koelewijn, H. de Kluiver, B.G. Shinn-Cunningham, A.A. Zekveld, S.E. Kramer: The pupil response reveals increased listening effort when it is difficult to focus attention. *Hearing Research* 323 (2015) 81–90.
11. J.L. Desjardins, K.A. Doherty: Age-related changes in listening effort for various types of masker noises. *Ear and Hearing* 34, 3 (2013) 261.
12. J. Seitz, K. Loh, J. Fels: Listening effort in children and adults in classroom noise. *Scientific Reports* 14, 1 (2024) 25200.
13. D.J. Strauss, F.I. Corona-Strauss, C. Trenado, C. Bernarding, W. Reith, M. Latzel, M. Froehlich: Electrophysiological correlates of listening effort: neurodynamical modeling and measurement. *Cognitive Neurodynamics* 4, 2 (2010) 119–131.
14. M.B. Winn, J.R. Edwards, R.Y. Litovsky: The impact of auditory spectral resolution on listening effort revealed by pupil dilation. *Ear and Hearing* 36, 4 (2015) e153.
15. E.M. Picou, T.A. Ricketts, B.W.Y. Hornsby: How hearing aids, background noise, visual cues influence objective listening effort. *Ear and Hearing* 34, 5 (2013) e52.
16. B.W.Y. Hornsby: The effects of hearing aid use on listening effort and mental fatigue associated with sustained speech processing demands. *Ear and Hearing* 34, 5 (2013) 523–534.
17. S. Alhanbali, P. Dawes, S. Lloyd, K.J. Munro: Self-reported listening-related effort and fatigue in hearing-impaired adults. *Ear and Hearing* 38, 1 (2017) e39.
18. S.E. Kuchinsky, J.B. Ahlstrom, K.I. Vaden Jr., S.L. Cute, L.E. Humes, J.R. Dubno, M.A. Eckert: Pupil size varies with word listening and response selection difficulty in older adults with hearing loss. *Psychophysiology* 50, 1 (2013) 23–34.
19. B. Ohlenforst, A.A. Zekveld, E.P. Jansma, Y. Wang, G. Naylor, A. Lorens, T. Lunner, S.E. Kramer: Effects of hearing impairment and hearing aid amplification on listening effort: a systematic review. *Ear and Hearing* 38, 3 (2017) 267–281.
20. D.S. Brungart: Informational and energetic masking effects in the perception of two simultaneous talkers. *The Journal of the Acoustical Society of America* 109, 3 (2001) 1101–1109.
21. B.G. Shinn-Cunningham: Object-based auditory and visual attention. *Trends in Cognitive Sciences* 12, 5 (2008) 182–186.
22. M. Cooke, M.L. Garcia Lecumberri, J. Barker: The foreign language cocktail party problem: energetic and informational masking effects in non-native speech perception. *The Journal of the Acoustical Society of America* 123, 1 (2008) 414–427.
23. G. Kidd, H.S. Colburn: Informational masking in speech recognition, in: *The Auditory System at the Cocktail Party*. Springer International Publishing, 2017, pp. 75–109.
24. R.W. Hughes, F. Vachon, D.M. Jones: Disruption of short-term memory by changing and deviant sounds: support for a duplex-mechanism account of auditory distraction. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 33, 6 (2007) 1050–1061.
25. J.E. Marsh, R.W. Hughes, D.M. Jones: Interference by process, not content, determines semantic auditory distraction. *Cognition* 110, 1 (2009) 23–38.
26. T. Koelewijn, A.A. Zekveld, J.M. Festen, S.E. Kramer: Pupil dilation uncovers extra listening effort in the presence of a single-talker masker. *Ear and Hearing* 33, 2 (2012) 291.
27. B. Ohlenforst, A.A. Zekveld, T. Lunner, D. Wendt, G. Naylor, Y. Wang, N.J. Versfeld, S.E. Kramer: Impact of stimulus-related factors and hearing impairment on listening effort as indicated by pupil dilation. *Hearing Research* 351 (2017) 68–79.
28. D. Wendt, T. Koelewijn, P. Książek, S.E. Kramer, T. Lunner: 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. *Hearing Research* 369 (2018) 67–78.
29. S. Villard, T. Perrachione, S.-J. Lim, A. Alam, G. Kidd: Listening effort elicited by energetic versus informational masking. *Proceedings of Meetings on Acoustics* 45, 1 (2022) 050002.
30. S. Villard, T.K. Perrachione, S.-J. Lim, A. Alam, G. Kidd Jr.: Energetic and informational masking place dissociable demands on listening effort: evidence from simultaneous electroencephalography and pupillometry. *The Journal of the Acoustical Society of America* 154, 2 (2023) 1152–1167.
31. V. Stenbäck, E. Marsja, M. Hällgren, B. Lyxell, B. Larsby: Informational masking and listening effort in speech recognition in noise: the role of working memory capacity and inhibitory control in older adults with and without hearing impairment. *Journal of Speech, Language, and Hearing Research* 65, 11 (2022) 4417–4428.
32. T. Koelewijn, A.A. Zekveld, J.M. Festen, S.E. Kramer: The influence of informational masking on speech perception and pupil response in adults with hearing impairment. *The Journal of the Acoustical Society of America* 135, 3 (2014) 1596–1606.
33. A.L. Francis, M.K. MacPherson, B. Chandrasekaran, A.M. Alvar: Autonomic nervous system responses during perception of masked speech may reflect constructs other than subjective listening effort. *Frontiers in Psychology* 7 (2016) 170291.
34. B. Ohlenforst, D. Wendt, S.E. Kramer, G. Naylor, A.A. Zekveld, T. Lunner: Impact of SNR, masker type and

- noise reduction processing on sentence recognition performance and listening effort as indicated by the pupil dilation response. *Hearing Research* 365 (2018) 90–99.
35. O. Tuomainen, S. Rosen, L. Taschenberger, V. Hazan: The effects of informational and energetic/modulation masking on the efficiency and ease of speech communication across the lifespan. *Speech Communication* 162 (2024) 103101.
  36. A. Kuusinen, K. Kondraciuk, T. Lokki: Effects of masker type and reverberation on speech-in-noise recognition thresholds and listening effort as indexed by pupil dilation responses. *Journal of the Audio Engineering Society* (2023) 10650.
  37. A.B. Jagadeesh, A.K. Uppunda: Speech-on-speech masking: effect of maskers with different degrees of linguistic information. *Canadian Journal of Speech-Language Pathology & Audiology* 45, 2 (2021) 143–156.
  38. D. Brungart, N. Iyer, E.R. Thompson, B.D. Simpson, S. Gordon-Salant, J. Schurman, C. Vogel, K. Grant: Interactions between listening effort and masker type on the energetic and informational masking of speech stimuli. *Proceedings of Meetings of Acoustics* 19, 1 (2013) 060146.
  39. H. Meister, S. Rählmann, U. Lemke, J. Besser: Verbal response times as a potential indicator of cognitive load during conventional speech audiometry with matrix sentences. *Trends in Hearing* 22 (2018) 2331216518793255.
  40. M. Krueger, M. Schulte, T. Brand, I. Holube: Development of an adaptive scaling method for subjective listening effort. *The Journal of the Acoustical Society of America* 141, 6 (2017) 4680–4693.
  41. I. Holube, S. Taesler, S. Ibelings, M. Hansen, J. Ooster: Automated measurement of speech recognition, reaction time, speech rate and their relation to self-reported listening effort for normal-hearing and hearing-impaired listeners using various maskers. *Trends in Hearing* 28 (2024) 23312165241276435.
  42. B. Larsby, M. Hällgren, B. Lyxell: The interference of different background noises on speech processing in elderly hearing impaired subjects. *International Journal of Audiology* 47 (2008) 83–90.
  43. T. Francart, A. van Wieringen, J. Wouters: Comparison of fluctuating maskers for speech recognition tests. *International Journal of Audiology* 50, 1 (2011) 2–13.
  44. S. Rosen, P. Souza, C. Ekelund, A.A. Majeed: Listening to speech in a background of other talkers: effects of talker number and noise vocoding. *The Journal of the Acoustical Society of America* 133, 4 (2013) 2431–2443.
  45. J. Rennie, V. Best, E. Roverud, G. Kidd: Energetic and informational components of speech-on-speech masking in binaural speech intelligibility and perceived listening effort. *Trends in Hearing* 23 (2019) 2331216519854597.
  46. R. McGarrigle, K.J. Munro, P. Dawes, A.J. Stewart, D.R. Moore, J.G. Barry, S. Amitay: Listening effort and fatigue: What exactly are we measuring? A British society of audiology cognition in hearing special interest group “White Paper”. *International Journal Of Audiology* 53, 7 (2014) 433–440.
  47. J.E. Peelle: Listening effort: How the cognitive consequences of acoustic challenge are reflected in brain and behavior. *Ear and Hearing* 39, 2 (2018) 204–214.
  48. K. Sewell, V. Brown, G. Farwell, M. Rogers, X. Zhang, J. Strand: The effects of temporal cues, point-light displays, and faces on speech identification and listening effort. *PLoS One* 18 (2023) e0290826.
  49. J. Nirme, B. Sahlén, V.L. Åhlander, J. Brännström, M. Haake: Audio-visual speech comprehension in noise with real and virtual speakers. *Speech Communication* 116 (2020) 44–55.
  50. C. Bernarding, D.J. Strauss, R. Hannemann, H. Seidler, F.I. Corona-Strauss: Objective assessment of listening effort in the oscillatory EEG: comparison of different hearing aid configurations, in: 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Chicago, IL, USA 2014, pp. 2653–2656.
  51. A. Dimitrijevic, M.L. Smith, D.S. Kadis, D.R. Moore: Neural indices of listening effort in noisy environments. *Scientific Reports* 9, 1 (2019) 11278.
  52. A. Devesse, A. van Wieringen, J. Wouters: AVATAR assesses speech understanding and multitask costs in ecologically relevant listening situations. *Ear and Hearing* 41, 3 (2020) 521–531.
  53. Y.-H. Wu, N. Aksan, M. Rizzo, E. Stangl, X. Zhang, R. Bentler: Measuring listening effort: driving simulator versus simple dual-task paradigm. *Ear and Hearing* 35, 6 (2014) 623–632.
  54. D.E. Broadbent: *Perception and Communication*. Pergamon Press, Elmsford, NY, US, 1958.
  55. D. Kahneman: *Attention and Effort*. Prentice-Hall, 1973.
  56. S.E. Kuchinsky, F.J. Gallun, A.K.C. Lee: Note on the dual-task paradigm and its use to measure listening effort. *Trends in Hearing* 28 (2024) 23312165241292215.
  57. P.A. Gosselin, J.-P. Gagné: Older adults expend more listening effort than young adults recognizing audiovisual speech in noise. *International Journal of Audiology* 50, 11 (2011) 786–792.
  58. S.J. Schlittmeier, I.S. Schiller, C. Mohanathasan, A. Liebl: Measuring text comprehension and memory: a comprehensive database for Heard Text Recall: (HTR) and Read Text Recall: (RTR) paradigms, with optional note-taking and graphical displays, Lehr- und Forschungsgebiet Arbeits- und Ingenieurpsychologie, RWTH Aachen University, RWTH-2023-05285, 2023.
  59. E. Fintor, L. Aspöck, J. Fels, S.J. Schlittmeier: The role of spatial separation of two talkers’ auditory stimuli in the listener’s memory of running speech: listening effort in a non-noisy conversational setting. *International Journal of Audiology* 61, 5 (2022) 371–379.
  60. J. Ehret, A. Bönsch, P. Nossol, C.A. Ermert, C. Mohanathasan, S.J. Schlittmeier, J. Fels, T.W. Kuhlen: Who’s next? Integrating non-verbal turn-taking cues for embodied conversational agents, in: Proceedings of the 23rd ACM International Conference on Intelligent Virtual Agents: (IVA ’23), 2023, pp. 1–8.
  61. J. Ehret, I.S. Schiller, C. Breuer, T.W. Kuhlen: Audio-visual coherence: is embodiment of background noise sources a necessity?, in: *IEEE Virtual Reality*. IEEE, 2024.
  62. I.S. Schiller, C. Breuer, L. Aspöck, J. Ehret, A. Bönsch, T.W. Kuhlen, J. Fels, S.J. Schlittmeier: A lecturer’s voice quality and its effect on memory, listening effort, and perception in a VR environment. *Scientific Reports* 14, 1 (2024) 12407.
  63. J.L. Szalma, P.A. Hancock: Noise effects on human performance: a meta-analytic synthesis. *Psychological Bulletin* 137, 4 (2011) 682–707.
  64. C. Mohanathasan, C.A. Ermert, J. Fels, T.W. Kuhlen, S.J. Schlittmeier: Exploring short-term memory and listening effort in two-talker conversations: the influence of soft and moderate background noise. *PLoS One* 20, 2 (2025) e0318821.

65. C.A. Ermert, J. Ehret, C. Mohanathanasan, A. Bönsch, T.W. Kuhlen, S.J. Schlittmeier, J. Fels: Influence of (non-) intelligible background speech on memory and listening effort in conversational situations, in: Proceedings of DAS/DAGA 2025: 51st Annual Meeting on Acoustics, 2025, pp. 159–160.
66. R Core Team R: A Language and Environment for Statistical Computing, Vienna, 2023.
67. H. Snellen: Probebuchstaben zur Bestimmung der Sehschärfe. H. Peters, 1873.
68. C.A. Ermert, C. Mohanathanasan, J. Ehret, S.J. Schlittmeier, T. Kuhlen, J. Fels: AuViST – An Audio-Visual Speech and Text Database for the Heard-Text-Recall Paradigm. Institute for Hearing Technology and Acoustics, RWTH Aachen University, 2023.
69. K.C. Wagener, S. Hochmuth, M. Ahrlich, M.A. Zokoll, B. Kollmeier: Der weibliche Oldenburger Satztest, 17. Jahrestagung der Deutschen Gesellschaft für Audiologie, 2014, 4p.
70. K.C. Wagener, T. Brand: Sentence intelligibility in noise for listeners with normal hearing and hearing impairment: influence of measurement procedure and masking parameters. *International Journal of Audiology* 44, 3 (2005) 144–156.
71. K. Wagener, V. Kuehnel, B. Kollmeier: Entwicklung und Evaluation eines Satztests für die deutsche Sprache I: Design des Oldenburger Satztests. *Zeitschrift für Audiologie* 38, 1 (1999) 4–15.
72. K. Wagener, T. Brand, B. Kollmeier: Entwicklung und Evaluation eines Satztests für die Deutsche Sprache II: Optimierung des Oldenburger Satztests. *Zeitschrift für Audiologie* 38, 2 (1999) 44–56.
73. K. Wagener, T. Brand, B. Kollmeier: Entwicklung und Evaluation eines Satztests für die Deutsche Sprache III: Evaluation des Oldenburger Satztests. *Zeitschrift für Audiologie* 38, 3 (1999) 86–95.
74. C.A. Ermert, M. Yadav, J.E. Marsh, S.J. Schlittmeier, T.W. Kuhlen, J. Fels: Serial recall in spatial acoustic environments: irrelevant sound effect and spatial source alternations. *Scientific Reports* 15, 1 (2025) 32473.
75. J. Cassell: Embodied conversational interface agents. *Communications of the ACM* 43, 4 (2000) 70–78.
76. Virtual Reality & Immersive Visualization Group, RWTH Aachen University: Virtual Acoustics Plugin, 2022. <https://git-ce.rwth-aachen.de/vr-vis/VR-Group/unreal-development/plugins/character-plugin>.
77. J. Ehret, A. Bönsch, J. Fels, S.J. Schlittmeier, T.W. Kuhlen: StudyFramework: comfortably setting up and conducting factorial-design studies using the unreal engine, in: 2024 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops: (VRW), 2024, pp. 442–449.
78. D. Gilbert, J. Ehret, M. Krüger, D. Rupp, S. Pape, K. Helwig, T. Römer, S. Oehrl, A.C. Demiralp, F. Qurabi, V. Wolf, K. Karwacki, L. Schröder: Virtual Reality, Immersive Visualization Group: RWTH VR Group Unreal Engine Toolkit, 2024.
79. Institute for Hearing Technology and Acoustics, RWTH Aachen University, Philipp Schäfer, Jonas Stienen, Lukas Aspöck, Michael Vorländer: Virtual Acoustics – A Real-Time Auralization Framework for Scientific Research, 2021. <https://doi.org/10.5281/zenodo.13744493>.
80. Virtual Reality & Immersive Visualization Group and Institute for Hearing Technology and Acoustics, RWTH Aachen University: Virtual Acoustics Plugin, 2023. <https://git-ce.rwth-aachen.de/vr-vis/VR-Group/unreal-development/plugins/unreal-va-plugin>.
81. A. Schmitz: Ein neues digitales Kunstkopfmesssystem. *Acustica: International Journal on Acoustics* 81, 4 (1995) 416–420.
82. B. Masiero, J. Fels: Perceptually robust headphone equalization for binaural reproduction, in: Audio Engineering Society Convention. Vol. 130. Audio Engineering Society, 2011, pp. 1–7.
83. A. Kassambara: Rstatix: Pipe-friendly framework for basic statistical tests, 2023.
84. P.-C. Bürkner: BRMS: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software* 80 (2017) 1–28.
85. T. Sivula, M. Magnusson, A.A. Matamoros, A. Vehtari: Uncertainty in Bayesian leave-one-out cross-validation based model comparison. *Bayesian Analysis* 1 (2025) 1–31.
86. R.V. Lenth, B. Banfai, B. Bolker, P. Buerkner, I. Giné-Vázquez, M. Herve, M. Jung, J. Love, F. Miguez, J. Piaskowski, H. Riebl, H. Singmann: Emmeans: Estimated Marginal Means, Aka Least-Squares Means, 2025.
87. J.K. Kruschke, T.M. Liddell: The Bayesian new statistics: hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective. *Psychonomic Bulletin & Review* 25, 1 (2018) 178–206.
88. M. Yadav, M. Georgi, L. Leist, M. Klatte, S.J. Schlittmeier, J. Fels: Cognitive performance in open-plan office acoustic simulations: effects of room acoustics and semantics but not spatial separation of sound sources. *Applied Acoustics* 211 (2023) 109559.
89. A. Devesse, A. Dudek, A. van Wieringen, J. Wouters: Speech intelligibility of virtual humans. *International Journal of Audiology* 57, 12 (2018) 914–922.
90. M.M.E. Hendrikse, G. Llorach, G. Grimm, V. Hohmann: Influence of visual cues on head and eye movements during listening tasks in multi-talker audiovisual environments with animated characters. *Speech Communication* 101 (2018) 70–84.
91. K. Shavit-Cohen, E.Z. Golumbic: The dynamics of attention shifts among concurrent speech in a naturalistic multi-speaker virtual environment. *Frontiers in Human Neuroscience* 13 (2019) 386.
92. E.M. Picou, J. Gordon, T.A. Ricketts: The effects of noise and reverberation on listening effort in adults with normal hearing. *Ear and Hearing* 37, 1 (2016) 1–13.
93. E.M. Picou, B. Bean, S.C. Marcum, T.A. Ricketts, B.W.Y. Hornsby: Moderate reverberation does not increase subjective fatigue, subjective listening effort, or behavioral listening effort in school-aged children. *Frontiers in Psychology* 10 (2019) 1749.
94. N. Prodi, C. Visentin: A slight increase in reverberation time in the classroom affects performance and behavioral listening effort. *Ear and Hearing* 43, 2 (2022) 460.