Memory & Cognition
https://doi.org/10.3758/s13421-024-01583-y

Noisy speech impairs retention of previously heard information only at short time scales

Violet A. Brown[1] · Katrina Sewell[1] · Jed Villanueva[1] · Julia F. Strand[1]

Accepted: 24 April 2024

Abstract

When speech is presented in noise, listeners must recruit cognitive resources to resolve the mismatch between the noisy input and representations in memory. A consequence of this effortful listening is impaired memory for content presented earlier. In the first study on effortful listening, Rabbitt, *The Quarterly Journal of Experimental Psychology*, *20*, 241–248 (1968; Experiment 2) found that recall for a list of digits was poorer when subsequent digits were presented with masking noise than without. Experiment 3 of that study extended this effect to more naturalistic, passage-length materials. Although the findings of Rabbitt's Experiment 2 have been replicated multiple times, no work has assessed the robustness of Experiment 3. We conducted a replication attempt of Rabbitt's Experiment 3 at three signal-to-noise ratios (SNRs). Results at one of the SNRs (Experiment 1a of the current study) were in the opposite direction from what Rabbitt, *The Quarterly Journal of Experimental Psychology*, *20*, 241–248, (1968) reported – that is, speech was recalled more accurately when it was followed by speech presented in noise rather than in the clear – and results at the other two SNRs showed no effect of noise (Experiments 1b and 1c). In addition, reanalysis of a replication of Rabbitt's seminal finding in his second experiment showed that the effect of effortful listening on previously presented information is transient. Thus, effortful listening caused by noise appears to only impair memory for information presented immediately before the noise, which may account for our finding that noise in the second-half of a long passage did not impair recall of information presented in the first half of the passage.

Keywords Speech · Recall · Listening effort · Replication

Introduction

In typical communicative settings, listeners aim not only to identify what was said, but also to retain that information for later. Relative to favorable listening conditions, presenting speech in background noise complicates both processes: Noise impairs speech intelligibility and it can also impair recall of what was heard (Mattys et al., 2012; Ward et al., 2016). One reason adding noise may lead to memory impairments for spoken material is simply that noisy speech is less intelligible; one cannot recall what was not heard. However, noise-induced memory impairments are often observed even when speech is identified correctly, indicating that poorer recall is not simply a result of misperception (Cousins et al., 2014; Ward et al., 2016).

There are at least two explanations for why intelligible speech may be recalled at lower rates when presented in masking noise than when presented without it. One explanation is that the sensory representations for speech presented in noise are degraded and encoded less effectively, which impairs downstream recall (Murphy et al., 2000). This explanation is akin to theories of visual memory positing that memories are not all-or-none. That is, the strengths of memory representations vary – as a result of either internal noise (e.g., perceptual, attentional, or memory processes) or external noise – and weaker memory representations are recalled less accurately (Brady et al., 2024).

However, these kinds of noise-induced differences in memory fidelity cannot wholly account for recall impairments in the presence of background noise. In Experiment 2 of a seminal paper in the auditory recall literature, Rabbitt (1968) demonstrated that recall of a block of digits was poorer when it was followed by a block of noise-masked digits than by a block of unmasked digits, regardless of whether the first block was noise-masked or unmasked. This study

✉ Violet A. Brown
violetbrownpsych@gmail.com

[1] Department of Psychology, Carleton College, Northfield, MN, USA

Published online: 17 May 2024

elegantly disentangled the detrimental effects of noise on speech identification from those on encoding: Because the noise manipulation happened *after* the to-be-remembered digits, any differences in recall cannot be attributed to the poor intelligibility of the target items or to degraded sensory representations. These noise-induced deficits in recall might instead be attributable to noise-induced changes in the allocation of cognitive resources.

A resource-based account (sometimes referred to in the literature as the "effortfulness hypothesis") claims that processing degraded speech requires cognitive resources that could otherwise be used for encoding information into memory (McCoy et al., 2005; Rabbitt, 1968). Specifically, according to the Ease of Language Understanding Model (Rönnberg et al., 2008), speech in noise leads to degraded acoustic input that does not cleanly map onto representations stored in the listener's mental lexicon. To resolve the mismatch and understand the spoken message, listeners must re-evaluate the impoverished input, which requires the recruitment of cognitive resources (namely, working memory capacity; Rönnberg et al., 2008) that exist in finite amounts. The recruitment of these additional resources leaves fewer resources available for encoding the speech into memory, therefore leading to poorer downstream recall (see Heinrich et al., 2008).[1] In line with the effortfulness hypothesis, individuals with hearing loss tend to recall spoken information less accurately than those with typical hearing – even when intelligibility is matched across groups (McCoy et al., 2005; Rabbitt, 1991) – presumably as a result of greater internal noise for individuals with hearing loss. Thus, although noise certainly degrades sensory representations and this may account in part for the poorer recall of speech in noise, there is also strong support in the literature for the resource-based account initially proposed by Rabbitt (1968).

The resource-based account as it is typically described focuses on the effects of noise on working memory processes, which operate on relatively short time scales (e.g., Bartsch & Oberauer, 2023). However, at longer timescales, processing noisy speech may also affect long-term memory processes. For example, long-term memory benefits from elaboration, whereby information in working memory is connected with representations in existing semantic networks to facilitate deeper encoding (Bartsch et al., 2018). Evidence that noise interferes with recall of information at longer timescales comes from Experiment 3 of Rabbitt (1968) in which participants listened to prose passages (approximately 5 min in length) and answered comprehension questions about their content. The first half of each passage was always presented without noise and the second half was presented either with or without noise. In line with Rabbitt's Experiment 2 results, second-half noise impaired recall of information presented in the first half of the passage. This may be because noisy speech in the second half interfered with elaboration; if second-half noise increases cognitive demand, participants may have been less readily able to make connections between incoming speech with information presented earlier in the passage, thereby leading to poorer recall of first-half content.

The results of Rabbitt's Experiment 3 are consistent with other work showing poorer comprehension for passages presented in noise relative to in the clear (Wasiuk et al., 2021), for individuals with hearing loss relative to those with age-normal hearing (Piquado et al., 2012), and for some types of information in noise-vocoded passages (Ward et al., 2016). Unlike Rabbitt's work, however, these experiments involved recalling information from passages that were presented in noise (whether external or internal). A major benefit of Rabbitt's method of separating the to-be-recalled information from the noisy speech is that it disentangles the consequences of noise on speech identification from those on recall. Although the finding that effortful listening can impair recall of previously heard materials has been replicated for digits (Guang et al., 2021), it has never been replicated for longer, more naturalistic materials. Thus, in the current investigation, we attempt to replicate the finding that noise in the second half of a passage can impair retention of material presented in the first half of the passage (Rabbitt, 1968).

Experiment 1

Experiment 1a

Method

Participants After exclusions, participants included 150 native English speakers (ages 22–62 years, $M = 26.9$ years). Participants were recruited through Prolific (http://www.prolific.co) and received $4 for 22 min of participation. Prolific participants were limited to native English speakers between 22 and 68 years of age currently living in the USA to ensure that our population is a close match to Rabbitt's original sample (ages 22–68 years, $M = 42.6$ years). Individuals were required to have self-reported normal hearing and normal or corrected-to-normal vision. Carleton College's

[1] Note that the accounts described here are those that have been proposed to explain the effects of unintelligible noise (e.g., speech-shaped noise, white noise) on recall of speech; other explanations have been proposed when the masking consists of speech (e.g., multi-talker babble). For example, phonological and semantic information in the masker may interfere with encoding of the target speech, and segregating the target speech and masker streams may draw resources away from encoding (see Heinrich et al., 2008).

Institutional Review Board approved all research procedures. Rabbitt's participants were run in groups of 11–22, and ours completed the task individually in an online setting. We collected data from 165 participants to reach our preregistered sample size of 150. Participants were excluded for taking notes or using outside help during the task (N = 8) and failing to respond to more than 50% of the comprehension questions (N = 1). To achieve a fully balanced design, we did not analyze the data collected from the final six participants.

Rabbitt's Experiment 3 tested 124 participants using a between-subjects design: Participants were presented one of two passages and heard that passage with no noise at all or with noise in the second half only. To increase statistical power, we instead opted to manipulate second-half noise within-subjects (i.e., participants heard two passages, one in which the second half occurred in noise and one in which it occurred in the clear), and we increased the sample size to 150 to ensure that we could detect effects smaller than what Rabbitt observed. All data were collected online in August–September 2021.

Stimuli Rabbitt (1968) reported using two prose passages from *Scientific American* that were 682 and 712 words long. We used two passages from the free online psychology textbook NOBA (nobaproject.com) to match the stylistic content of Rabbitt's stimuli. The passages were 723 and 720 words long and described cross-cultural research on emotion and the psychology of aging.

Passages were recorded using a Yeti Blue microphone with a plosive screen by a native English speaker without a discernible regional accent. Recording was paced using a teleprompter to ensure a consistent speaking rate. The passages were 5 min 44 s and 5 min 45 s long, similar to the length of Rabbitt's passages (5 min 23 s; 5 min 8 s). The two passages were equated on root-mean-square amplitude using Adobe Audition (Version 22.0.2.61).

To generate masking noise, we extracted the long-term average spectrum for each passage to create speech-shaped noise using Praat (Winn, 2018). A Python script then modulated the masking noise based on the amplitude of the target speech, affording a constant SNR throughout the passage, in line with Rabbitt's method. Rabbitt (1968) used an SNR of 5, but pilot testing suggested that this SNR was too easy with our materials. We therefore presented the second half of passages at an SNR of -1. The masking noise and passages were mixed to generate two versions of each passage differing in whether noise was presented in the second half. Given that the first half was always presented without noise, the condition labels "clear" and "noise" indicate whether noise was present in the second half of the passage.

Next, we developed ten comprehension questions for each passage following the format described by Rabbitt (1968). Five questions assessed information that was only presented in the first half of the passage, and five questions assessed information exclusively from the second half. One "True or False" question and four "Fill in the Blank" questions were created for each passage half. For example, one of the questions from the passage about aging was "Older typists were found to compensate for age-related declines in speed by looking farther ahead at printed text" (true) and one of the fill-in-the-blank questions from the passage about emotion was "86% of US participants associated wrinkling of the nose with "disgust," but only 60% of _______ participants did" (Japanese). Comprehension questions were selected from a pilot study to ensure participants could not answer them without listening to the passages first, and to prevent floor and ceiling effects on recall performance. Comprehension questions were presented in black text on a white screen. We presented ten comprehension questions following the procedures of Rabbitt (1968); the main analyses described below were conducted on the questions about first-half content.

Procedure The experiment was programmed and presented in Gorilla Experiment Builder (Anwyl-Irvine et al., 2020). Participants were instructed to wear headphones, complete the task in a quiet space with minimal distractions, and refrain from taking notes while they listened to the passages. Before beginning the main task, a sound check was performed to ensure that participants had auto-play enabled on their browser, which was necessary for audio files to play without interruption. The sound check also allowed participants to adjust their volume to a comfortable level before beginning the main task. The task could be completed in any browser, but if the browser check failed, participants were redirected to a screen providing instructions for enabling auto-play in Google Chrome.

Each participant heard both passages, one in the clear and one with second-half noise. Across participants, both passages were presented in the clear and second-half noise conditions an equal number of times. The order of the presentation of the passages and conditions were counterbalanced across participants. In line with Rabbitt's (1968) methods, participants were informed that background noise may or may not be present as they listen to the passages. Participants listened to each passage in full, then immediately answered the comprehension questions. Rabbitt did not specify the manner in which he presented the questions, so we elected to present the questions in a randomized order.[2]

[2] A potential concern is that by presenting the questions in a random order, the first questions may refer to content from the end of the second half of the passage, meaning they could conceivably be answered by drawing on information from working memory rather than long-term memory. However, our focus is on how well participants retain information from the first half of the passage – which ended more than 2 min before the questions were presented. Therefore, regardless of the order in which the questions are asked, participants must draw on long-term memory to answer questions about the first half of the passage.

After completing the experiment, participants completed a short demographic questionnaire.

Scoring typed responses Answers to the comprehension questions were labeled and scored as correct if the typed responses exactly matched the correct answer. We did not preregister criteria for correcting individual responses, but opted to count typos as correct if the error was a single-letter deletion (e.g., fals/false), addition (e.g., Japaneese/Japanese), or keystroke (e.g., univeralists/universalists) away from the correct word, or responses contained extraneous punctuation (e.g., false;/false). Responses that were common misspellings (e.g., embarrasment/embarrassment; openess/openness) or phonologically consistent with the target (e.g., narotisism/neuroticism) were also counted as correct. Pluralizations (e.g., universalist/universalists) and adding or removing suffixes (e.g., physiologic/physiological) to an otherwise correct response were also counted as correct (see R script for a full list of corrections).

Results and discussion

Data analysis was conducted in R (version 4.2.1; R Core Team, 2022) using the *lme4* package (version 1.1.30; Bates et al., 2015) and processed using the *tidyverse* package (2.0.0; Wickham et al., 2019). Given the binary nature of the dependent variable (accuracy for each comprehension question; 0 = incorrect, 1 = correct), we used generalized linear mixed models assuming a binomial distribution with a logit link function. We obtained *p*-values for the coefficients of interest using the *lmerTest* package (version 3.1.3; Kuznetsova et al., 2017).

Despite using different passages than Rabbitt (1968) did and collecting data online rather than in person, the retention scores for the clear condition we observed here were numerically similar to those reported in the earlier study: When the second half was presented in the clear, mean first-half recall was 40% in this study and 39% in Rabbitt's study, and when the second half was presented in noise, mean first-half recall was 46% in this study and 30% in Rabbitt's study. To assess whether noise in the second half of a passage affected recall of information from the first half, we built a model predicting first-half accuracy using second-half noise (clear vs. noise) and passage as fixed effects. The random effects structure included random intercepts for participants and items (questions), as well as by-participant random slopes for second-half noise. We attempted to model by-item random slopes for second-half noise, but this maximal model encountered unresolvable convergence and singularity issues.[3] We compared this model to a model lacking the fixed effect for second-half noise via a likelihood ratio test, which indicated that the full model provided a better fit for the data than the reduced model ($\chi^2_1 = 4.00$; $p = 0.045$). Examination of the summary output for the full model indicated that accuracy for first-half comprehension questions was slightly better when the second half was presented in noise than when it was presented in the clear ($B = 0.26$, $SE = 0.13$, $z = 2.04$, $p = 0.04$; Fig. 1).

The findings reported here are not likely to be driven by differences in the difficulty of the comprehension questions or restricted range, given the numerically similar recall rates between our study and Rabbitt's original study. Not only do these results not replicate Rabbitt's (1968) finding, but they actually provide some evidence for the opposite claim: Rabbitt found that noise in the second half of passages impaired recall of information presented in the first half, but we found that second-half noise actually *facilitated* recall of first-half information (see Table 1).

Although we attempted to replicate Rabbitt's methodology, methodological differences may account for the discrepant findings. For example, Rabbitt presented the second half of the passages at an SNR of 5, but pilot testing revealed that this noise level was too easy, so we opted to use an SNR of -1. However, the different SNRs may have had different consequences for retrieval of information from long-term memory. Relatively quiet background noise may induce moderate effort that interferes with elaborating upon first-half information, thereby impairing recall of first-half information (consistent with the results reported in Rabbitt's Experiment 3). In contrast, louder second-half background noise like what we used here may actually facilitate recall of first-half content by reducing retroactive interference from the second-half content (Wixted, 2010). That is, sufficiently loud background noise could render second-half content partially or wholly unintelligible, which could affect memory for first-half content in at least two ways.

First, when the entire passage is presented without noise, participants must encode 5 min of material on a given topic, and may therefore suffer from *cue-overload* upon retrieval – whereby an incorrect representation is accessed in response to a cue (Van Dyke, 2011; Watkins & Watkins, 1975). Given that first- and second-half content are semantically related, when participants are presented with a comprehension question about first-half content, they may inadvertently retrieve information about second-half content. In contrast, when second-half content is obscured by noise and therefore is not encoded as successfully, participants are less likely to experience this cue-overload because a given cue is tied to fewer representations, leading to better recall of first-half content.

[3] A likelihood ratio test indicated that excluding this random slope did not significantly affect model fit (see R script for more details).

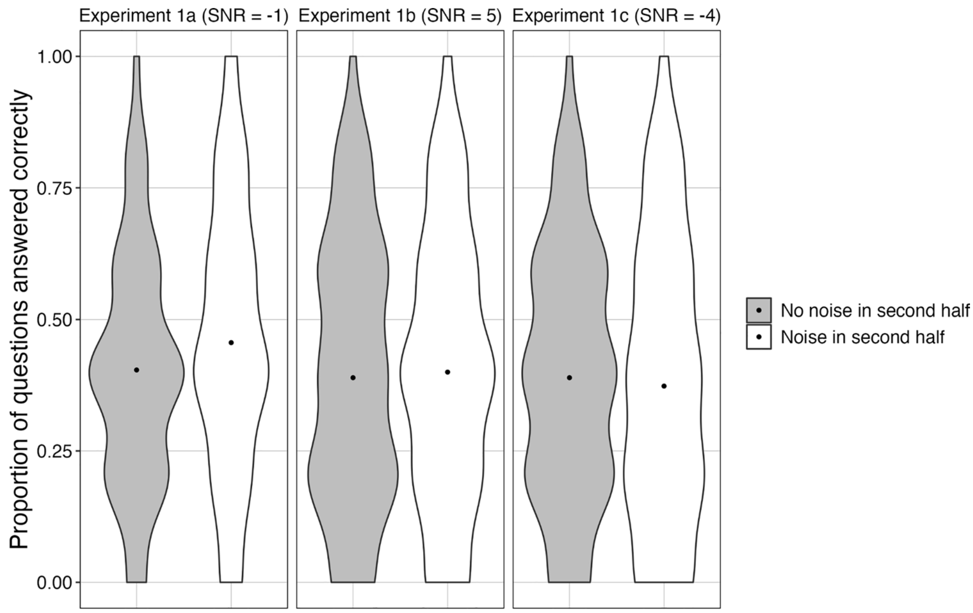

Fig. 1 By-participant comprehension accuracy for first-half material when the second-half of the passage was presented with and without noise. The dot represents the mean accuracy in each condition (gray = noise, white = no noise), the shape of each plot depicts the distribution of responses across participants

Table 1 Means and standard deviations for first-half comprehension question accuracy for Experiments 1a–1c

SNR	Clear in second half	Noise in second half
5	0.39 (0.49)	0.40 (0.49)
-1	0.40 (0.49)	0.46 (0.50)
-4	0.39 (0.49)	0.37 (0.48)

SNR signal-to-noise ratio

Alternatively or in addition, noise in the second half may lead to more degraded representations of that content in long-term memory (Brady et al., 2024). Given that low-fidelity representations capture attention to a lesser extent than high-fidelity representations (Williams et al., 2022), when participants answer comprehension questions, information about the first-half content may be more salient. Regardless of the mechanism, second-half noise may prevent that content from being encoded into long-term memory, thereby reducing the extent of retroactive interference on first-half content (Crouse, 1971; Dempster, 1988).

If the second-half noise was loud enough to substantially impair intelligibility – thereby reducing retroactive interference on first-half content – then performance on the comprehension questions should be worse for *second-half* content when that content was presented in noise. Indeed, an analysis mirroring those described above on second-half content (as opposed to first-half content) indicated that second-half noise significantly impaired accuracy for second-half comprehension questions (B = -0.30, SE = 0.12, z = -2.45, p = 0.01). Note, however, that participants may answer questions about content presented in noise less accurately than when it is not presented in noise either because noise impairs intelligibility (i.e., they could not hear the speech) or because noise increases cognitive demand (i.e., they identified the speech but did not successfully encode it). In the absence of a measure of intelligibility, there is no way to know whether poorer recall for the content of spoken passages presented in noise is driven by poor identification accuracy or the negative effects of noise on encoding (i.e., listening effort). To explicitly test whether the direction of the effect depends on the level of the background noise, and to more closely replicate Rabbitt's (1968) methods, Experiment 1b involved re-running the experiment at the easier SNR Rabbitt originally used (SNR = 5). Then, in Experiment 1c, we increased the level of the background noise to more explicitly test the retroactive interference hypothesis (SNR = -4).

Experiment 1b

Experiment 1b was identical to Experiment 1a, except that the SNR was 5 in the second-half noise condition.

Method

Participants We collected data from 160 participants to reach our preregistered sample size of 150. Participants were excluded for taking notes or using outside help during the task (N = 7). To achieve a fully balanced design, we did not analyze the data collected from the final three participants. Participants were native English speakers (ages 22–62 years, M = 42.3 years) recruited through Prolific (http://www.prolific.co). Data were collected online in February 2023.

Results and discussion

Our analytical approach was identical to that in Experiment 1a. The random effects structure included random intercepts for participants and items (questions), but no random slopes because including random slopes led to unresolvable convergence and singularity issues.[4] A likelihood ratio test indicated that the effect of second-half noise was not significant ($\chi^2_1 = 0.30$; $p = 0.58$; Fig. 1). These results suggest that the effect reported in Experiment 1a that second-half noise slightly improved recall of information presented in the first half of a passage (relative to presenting the second half in the clear) may be small and dependent on the difficulty of the speech identification task. In addition, analysis of second-half recall indicated that noise in the second half did not significantly affect recall of second-half content ($\chi^2_1 = 0.01$; $p = 0.99$). It is therefore unsurprising that we did not find an effect of second-half noise on comprehension of first-half content; the SNR of 5 may have been too easy to have a detectable negative effect on either intelligibility or encoding.

Importantly, regardless of whether we used a moderate SNR (Experiment 1a; SNR = -1) or an easy one (Experiment 1b; SNR = 5), we did not replicate Rabbitt's (1968) finding that second-half noise impairs first-half recall. If anything, we found that second-half noise slightly *improves* first-half recall (Experiment 1a), perhaps because loud second-half noise reduces the amount of information from the second half of the passage that can retroactively interfere with encoding of first-half information. If the effect observed in Experiment 1a is in fact driven by retroactive interference, then making the background noise even louder should increase the magnitude of the effect observed in Experiment 1a; that is, louder background noise should lead to even less retroactive interference because the second-half speech cannot be accurately heard, which should in turn lead to better recall of first-half information than we observed in either of the previous experiments. To explore this possibility, we ran the experiment a third time using a more difficult SNR.

[4] A likelihood ratio test indicated that excluding these random slopes did not significantly affect model fit (see R script for details).

Experiment 1c

Experiment 1c was identical to Experiments 1a and b, except that the SNR was -4 in the second-half noise condition.

Method

Participants We collected data from 160 participants to reach our preregistered sample size of 150. Participants were excluded for taking notes or using outside help during the task (N = 10). Participants included 150 native English speakers (ages 22–62 years, M = 40.6 years) recruited through Prolific (http://www.prolific.co). Data were collected in September 2023.

Results and discussion

Our analytical approach was identical to that in Experiments 1a and 1b. The random effects structure included random intercepts for participants and items (questions), as well as by-participant random slopes for second-half noise. Unlike in Experiments 1a and 1b, by-item random slopes for second-half noise were included because they did not lead to convergence and singularity issues (see R script for details). A likelihood ratio test indicated that the effect of second-half noise on comprehension of first-half content was not significant ($\chi^2_1 = 1.07$; $p = .30$; Fig. 1), despite a significant effect of second-half noise on comprehension of second-half content ($B = -0.94$, $SE = 0.15$, $z = -6.25$, $p < .001$). Thus, consistent with the previous two experiments, this experiment did not replicate Rabbitt's finding that noise in the second half of a passage impairs recall of information presented in the first half of the passage, despite impairing recall of second-half content.

The retroactive interference hypothesis predicts that louder second-half noise should produce less retroactive interference and therefore better recall of first-half content (see the *Results and discussion* section for Experiment 1a). If this is the case, then the beneficial effect of second-half noise on first-half recall should be largest in this experiment, given that this SNR was the most difficult of the three. However, we did not observe an effect of second-half noise on first-half recall in this experiment; indeed, the only experiment that was consistent with the retroactive interference hypothesis was Experiment 1a, which used an intermediate SNR.

Finally, to evaluate whether evidence for either a resource-based or retroactive interference hypothesis emerges with larger sample sizes, we conducted an exploratory analysis on the combined data from the three sub-experiments within Experiment 1 – which differed only in SNR – to determine whether second-half noise affected first-half recall across the 450 participants tested (controlling for SNR). The full model in this analysis included fixed effects for second-half noise, SNR, and passage; random intercepts for participants and items; and by-participant and by-item random slopes for second-half noise (a model including by-item random slopes for SNR did not converge; see R script for more details). A likelihood ratio test comparing this model to a model lacking the fixed effect for second-half noise indicated that noise in the second half did not significantly affect recall of information from the first half of the passage ($\chi^2_1 = 0.54$; $p = .46$).

Experiment 1: Discussion

In this series of three experiments, we attempted to replicate Rabbitt's (1968) finding that when the second half of a spoken passage is presented in background noise, retention of information from the first half of the passage is worse than when the second half is presented in the clear. Regardless of whether the level of the background noise was easy (Experiment 1b; SNR = 5), moderate (Experiment 1a; SNR = -1), or hard (Experiment 1c; SNR = -4), we did not replicate Rabbitt's finding. Indeed, in Experiment 1a, second-half noise actually *improved* recall of previous material (despite impairing recall of second-half content), and the effect was non-significant in Experiments 1b and 1c. It is unclear why noise in the second half improved first-half recall in Experiment 1a, whereas it had no effect in the other two experiments; it may be that the effects of effortful listening and retroactive interference on spoken passage retention are small and only appear in very specific listening conditions, or perhaps the negative effects of noise on encoding and the beneficial effects of noise on retroactive interference offset one another. Regardless of the reason for the unexpected, significant effect in Experiment 1a, none of these three experiments replicated Rabbitt's finding that second-half noise impairs retention of first-half information.

A previous investigation from our lab successfully replicated Rabbitt's finding that second-half noise impairs recall of the first half of short lists of digits (Guang et al., 2021). It is therefore unexpected that the three experiments reported here – which were well powered to detect effects smaller than what Rabbitt originally observed – did not replicate the finding with spoken passages. One reason that subsequent noisy speech might differently affect recalling lists of digits versus 5-min passages is the timescale at which these processes operate; indeed, digit lists are much shorter and recalling them therefore relies more heavily on working memory, whereas recalling information presented 5 min earlier in a spoken passage requires that the information be encoded into long-term memory. It may be that the negative effects of effortful listening on downstream recall are transient and primarily disrupt working memory. If this is the case, then noise should primarily impair recall of information presented immediately before the onset of noise. Although the current study was not designed to test this prediction, the list construction of our previous replication study (Guang et al., 2021) – which included digits in four discrete positions in each half of the list – enables us to test the time course of the effects of effortful listening on recall of previously-presented items.

Experiment 2: Reanalysis of existing data

Method

Guang et al. (2021) reported a direct replication of Rabbitt's second experiment (Rabbitt, 1968) in an online setting. Participants heard lists of eight digits, and the first half (i.e., items 1–4) and second half (i.e., items 5–8) were presented either in the clear or in the same modulating noise as described above. After each list, participants were instructed to recall either the first-half or the second-half digits. That replication study found that the presence of masking noise in the second half impaired recall of digits in the first half, in line with the findings of Rabbitt's (1968) second experiment.

The analyses reported by Guang and colleagues (Guang et al., 2021) and Rabbitt (1968) involved aggregating data from all positions in the first half of the list of digits. That is, those studies assessed whether the presence of noise in the second half affected recall of first-half digits regardless of the particular position in the list at which a digit occurred. However, if effortful listening impairs working memory, recall should be poorest for items near the end of the first half (i.e., items in position 4) because these items have not yet been encoded by the time the noise begins. Thus, the digit in the final position of the first half is likely to show larger effects of second-half noise than those in earlier first-half list positions.

Results and discussion

To test this possibility, we conducted an exploratory reanalysis of Guang et al.'s (2021) data. Although this experiment varied whether noise was present in the first half as well as the second half, to more closely match the conditions of Experiments 1a–c (and Rabbitt's experiment using passages), our reanalysis only included trials in which lists were presented in the clear in the first half. To test whether second-half noise had a different effect on recall depending on list position, we built a model including fixed effects for position, second-half noise, and the interaction, and compared it to a model lacking the interaction term. The random effects structure included random intercepts for both participants and items, but no random slopes (see R script for details). We implemented a dummy coding scheme such that "clear" was coded as 0 and "noise" was coded as 1, and position 4 was the reference level for the

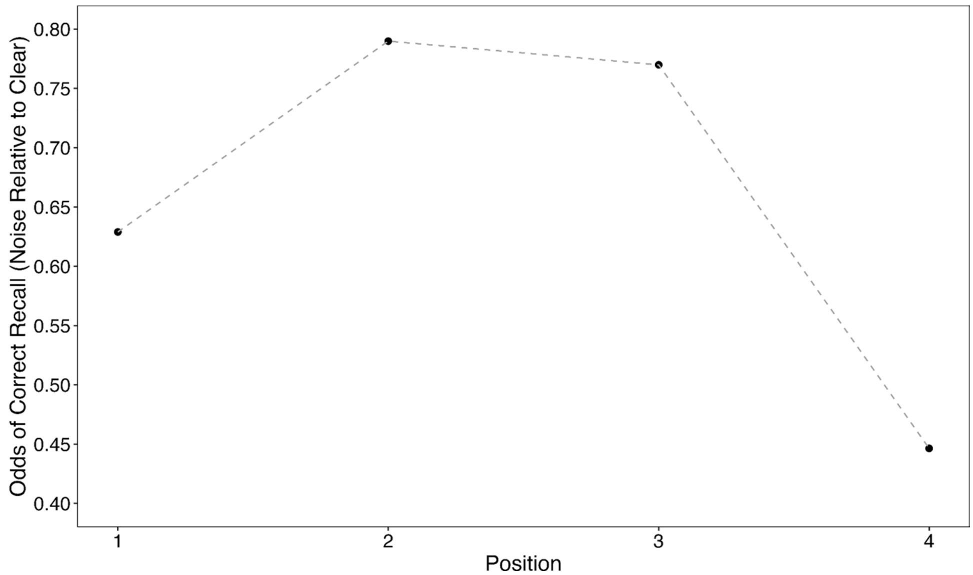

Fig. 2 Odds of correctly recalling a word in noise relative to in the clear for each list position in Guang et al. (2021). Each point represents the exponentiated coefficient from a model predicting recall from second-half noise for each of the four possible list positions

list position variable (because this is the position we have a specific hypothesis about). A likelihood ratio test indicated that the interaction was significant ($\chi^2_3 = 15.91$; $p = .001$), and examination of the summary output for the model including the interaction term showed a significant effect of noise at position 4 ($B = -0.81$, $SE = 0.17$, $z = -4.80$, $p < 0.001$). Relative to position 4, the effect of noise on recall was significantly weaker in position 1 ($B = 0.70$, $SE = 0.22$, $z = 3.24$, $p = 0.001$), position 2 ($B = 0.58$, $SE = 0.20$, $z = 2.98$, $p = 0.003$), and position 3 ($B = 0.65$, $SE = 0.19$, $z = 3.43$, $p < .001$).

Given the significant interaction between list position and second-half noise, we subsetted the data four times such that each subset contained data from each of the four possible first-half list positions (1–4) to assess the effect of noise at each list position. We built four separate generalized linear mixed models in which the dependent variable was recall of digits in a particular list position, and the fixed effect of interest was second-half noise. All models included random intercepts for both participants and items, and the models for positions 1 and 3 additionally included by-participant random slopes for second-half noise (the models for positions 2 and 4 did not include random slopes because we encountered unresolvable convergence issues; see R script for details). The effect of second-half noise was not significant at position 1 ($B = -0.46$, $SE = 0.44$, $z = -1.06$, $p = 0.29$), position 2 ($B = -0.24$, $SE = 0.20$, $z = -1.18$, $p = 0.24$), or position 3 ($B = -0.26$, $SE = 0.28$, $z = -0.95$, $p = 0.34$). However, the effect of noise was significant at position 4 ($B = -0.81$, $SE = 0.17$, $z = -4.87$, $p < .001$); this was expected given the simple effect of noise in the interaction model reported above (in which position 4 was the reference level), but the main effect of noise also exists in the subsetted data for position 4.

Given that the fixed-effects estimates from these models represent the log-odds of recalling the digit in that position when the second half is presented with noise relative to without, exponentiating the coefficient produces the odds ratio for recall in noise relative to in the clear. These odds ratios are plotted in Fig. 2.[5] As expected based on the nature of the interaction reported above, the magnitude of the effect of noise on recall was largest for the digit that occurred immediately before the noise (position 4). These findings are consistent with previous work showing that adding masking noise to a single item in a word list impairs retention for the masked word as well as the word presented immediately before, but not the other words in the list (Cousins et al., 2014; Piquado et al., 2010). Together, this reanalysis of our previous data supports the claim that memory impairments from noise are driven by disruptions to working memory that prevent information from being encoded into long-term memory: The effect of noise in the second half was larger for the item that was presented immediately before the noise than for the items that were presented earlier.

[5] Note that although the odds ratio for the effect of noise appears to be numerically lower in position 1 than in positions 2 and 3 in Fig. 2, re-leveling the interaction model such that position 1 is coded as the reference level reveals that that the effect of noise is not significantly stronger in position 1 than in the middle two positions.

General discussion

Across three sub-experiments differing in the SNR at which second-half noise was presented, we did not replicate the finding that subsequent noise has a negative effect on first-half recall. Indeed, in Experiment 1a (SNR = -1) noise slightly *improved* first-half recall, and in Experiments 1b (SNR = 5) and 1c (SNR = -4) noise had no effect on first-half recall. The finding that subsequent noise has a different effect on recall of lists versus passages may be due to differences in the time course of the two types of materials: If the negative effect of effortful listening on recall of previously heard information is transient, noise should only affect recall of recently heard information and not speech presented minutes earlier. Indeed, Rabbitt himself argued that his findings were attributable to the negative effects of noise on the *rehearsal* process (1968), and rehearsal is expected to play a larger role in list recall (which relies more on working memory) than passage retention (which relies on long-term memory). Thus, the effect of effortful listening on recall is likely to be stronger at shorter timescales (see Strand et al., 2021, page 10, for more on this argument).

To test this hypothesis, we conducted a reanalysis of data from our previous replication study of Rabbitt (1968; Experiment 2). We showed that noise presented in the second half of a spoken list of eight digits impairs recall of digits presented in the first half of the list (Guang et al., 2021), consistent with Rabbitt's original finding. In the reanalysis presented here, we demonstrated that the negative effect of noise on recall was more pronounced for the digit in the fourth position – that is, the digit that was spoken immediately before the onset of the noise – than for the digits presented in the initial three positions. This reanalysis provides some evidence that the effect of effortful listening on recall of previously heard information is transient, supporting the claim that retention of information from longer passages is unlikely to be affected by noise-induced effort several minutes later.

This is not to say that subsequent noise cannot affect retention of previously heard information from longer passages. Indeed, in Experiment 1a we demonstrated that second-half noise slightly improved first-half retention. One explanation for this finding is that noise in the second half of a passage leads to poor speech identification accuracy, which means that less verbal information is available to retroactively interfere with encoding of first-half information than when the second half of the passage is presented in the clear, leading to better retention of first-half information. Although this explanation can account for our finding in Experiment 1a that second-half noise slightly improved first-half recall, this cannot be the only mechanism through which second-half noise affects first-half recall of longer passages: Under the retroactive interference hypothesis, increasing the level of background noise should increase the first-half retention benefit, but we found no effect of second-noise on first-half recall when we used a more difficult level of background noise (Experiment 1c).

Together, these experiments provide evidence that although subsequent noise may impair recall of previously heard information, this effect of effortful listening is transient and does not appear to influence retention of information presented minutes earlier. It is possible that differences in the patterns of results observed in Experiments 1a–c and Rabbitt (1968) are driven by the fact that changing the SNR influences the balance between retroactive interference and effortful encoding. That is, retroactive interference models and resource models make opposite predictions regarding the effects of subsequent noise on recall of previously heard information. However, this explanation seems unlikely given the transient nature of the effect of effortful listening and the weak evidence for retroactive interference we observed in Experiments 1a–1c.

It is unclear why Rabbitt (1968) found that second-half noise impairs recall of first-half information, regardless of whether the materials consisted of spoken lists or longer passages. The magnitude of the effect of second-half noise on first-half passage retention was relatively small and differed across the two passages (we calculated Cohen's *d* values of approximately 0.25 and 0.44 based on Rabbitt's Table 3), suggesting that particulars of the passages and perhaps statistical noise contributed to the observed effect. Although we attempted to match characteristics of our sample to those in the original study, we conducted these experiments online, and it is conceivable that participant characteristics and the experimental setting account for the differences between our results and Rabbitt's original finding.

Another difference between the methods we employed in our replication attempt and those in the original study is that second-half noise was manipulated between-subjects in the original study, but here we manipulated this variable within-subjects to increase statistical power. This design choice introduces the possibility that carryover effects might affect results. Specifically, if participants complete the no-noise condition first, by the time they get to the second half of the second passage (which would be presented in noise given our design), familiarity with the talker's voice may lead to smaller effects of noise than if the order of the conditions was reversed. However, this is unlikely to account for our findings because even if participants listened to the passage with noise first, the first half of the passage (approximately 2.5 min) was presented in the clear in all conditions, which provides listeners with ample time to adapt to the talker's voice before any noise is presented. Regardless, it appears that any potential negative effects of subsequent noise on retention are not robust to minor variations in experimental methodology and therefore are unlikely to have practical significance in everyday listening situations.

This investigation highlights the value of conducting both direct (or as close to direct as possible) and conceptual replications (see Crandall & Sherman, 2016, and Simons, 2014, for arguments about the benefits of each approach). Rabbitt's original study (Rabbitt, 1968) included multiple experiments that all sought to test the claim that listening to noisy speech impairs recall of previously heard information. Experiment 2 (words) and Experiment 3 (passages) of that study might therefore be considered conceptual replications of one another, because they tested the same underlying theoretical claim. Given our successful direct replication of Rabbitt's Experiment 2 (Guang et al., 2021), we expected Rabbitt's Experiment 3 to replicate here. The unexpected failure to replicate demonstrates the value of conducting direct replications even if a conceptual replication has already been successful. The studies taken together shed light on the temporal bounds of the effect, such that noisy speech impairs recall of previously presented information only at short time scales.

Funding This work was supported by Carleton College and the National Institutes of Health via the National Institute on Deafness and Communication Disorders (R15-DC018114; R15-DC018114-02). We are grateful to Claire Guang, Naseem Dillman-Hasso, Helen Hu, and Gigi Paulig for technical assistance and research support.

References

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods, 52*(1), 388–407.

Bartsch, L. M., & Oberauer, K. (2023). The contribution of episodic long-term memory to working memory for bindings. *Cognition, 231*, 105330.

Bartsch, L. M., Singmann, H., & Oberauer, K. (2018). The effects of refreshing and elaboration on working memory performance, and their contributions to long-term memory formation. *Memory & Cognition, 46*(5), 796–808.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, Articles, 67*(1), 1–48.

Brady, T. F., Robinson, M. M., & Williams, J. R. (2024). Noisy and hierarchical visual memory across timescales. *Nature Reviews Psychology, 3*, 1–17.

Cousins, K. A. Q., Dar, H., Wingfield, A., & Miller, P. (2014). Acoustic masking disrupts time-dependent mechanisms of memory encoding in word-list recall. *Memory & Cognition, 42*(4), 622–638.

Crandall, C. S., & Sherman, J. W. (2016). On the scientific superiority of conceptual replications for scientific progress. *Journal of Experimental Social Psychology, 66*, 93–99.

Crouse, J. H. (1971). Retroactive interference in reading prose materials. *Journal of Educational Psychology, 62*(1), 39–44.

Dempster, F. N. (1988). Retroactive interference in the retention of prose: A reconsideration and new evidence. *Applied Cognitive Psychology, 2*(2), 97–113.

Guang, C., Lefkowitz, E., Dillman-Hasso, N., Brown, V. A., & Strand, J. F. (2021). Recall of speech is impaired by subsequent masking noise: A replication of Rabbitt (1968) Experiment 2. *Auditory Perception & Cognition*. https://doi.org/10.1080/25742442.2021.1896908

Heinrich, A., Schneider, B. A., & Craik, F. I. M. (2008). Investigating the influence of continuous babble on auditory short-term memory performance. *Quarterly Journal of Experimental Psychology, 61*(5), 735–751.

Kuznetsova, A., Brockhoff, P., & Christensen, R. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software, Articles, 82*(13), 1–26.

Mattys, S. L., Davis, M. H., Bradlow, A. R., & Scott, S. K. (2012). Speech recognition in adverse conditions: A review. *Language and Cognitive Processes, 27*(7–8), 953–978.

McCoy, S. L., Tun, P. A., Cox, L. C., Colangelo, M., Stewart, R. A., & Wingfield, A. (2005). Hearing loss and perceptual effort: Downstream effects on older adults' memory for speech. *The Quarterly Journal of Experimental Psychology A, Human Experimental Psychology, 58*(1), 22–33.

Murphy, D. R., Craik, F. I. M., Li, K. Z. H., & Schneider, B. A. (2000). Comparing the effects of aging and background noise on short-term memory performance. *Psychology and Aging, 15*(2), 323–334.

Piquado, T., Benichov, J. I., Brownell, H., & Wingfield, A. (2012). The hidden effect of hearing acuity on speech recall, and compensatory effects of self-paced listening. *International Journal of Audiology, 51*(8), 576–583.

Piquado, T., Cousins, K. A. Q., Wingfield, A., & Miller, P. (2010). Effects of degraded sensory input on memory for speech: Behavioral data and a test of biologically constrained computational models. *Brain Research, 1365*, 48–65.

R Core Team. (2022). *R 4.2.1*. R Foundation for Statistical Computing.

Rabbitt, P. M. (1968). Channel-capacity, intelligibility and immediate memory. *The Quarterly Journal of Experimental Psychology, 20*(3), 241–248.

Rabbitt, P. M. (1991). Mild hearing loss can cause apparent memory failures which increase with age and reduce with IQ. *Acta Oto-Laryngologica. Supplementum, 476*, 167–175. discussion 176.

Rönnberg, J., Rudner, M., Foo, C., & Lunner, T. (2008). Cognition counts: A working memory system for ease of language understanding (ELU). *International Journal of Audiology, 47*(sup2), S99–S105.

Simons, D. J. (2014). The value of direct replication. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science, 9*(1), 76–80.

Strand, J. F., Ray, L., Dillman-Hasso, N. H., Villanueva, J., & Brown, V. A. (2021). Understanding speech amid the jingle and jangle: Recommendations for improving measurement practices in listening effort research. *Auditory Perception & Cognition, 3*(4), 1–20.

Van Dyke, J. A. (2011). Cue-dependent interference in comprehension. *Journal of Memory and Language, 65*(3), 247–263.

Ward, C. M., Rogers, C. S., Van Engen, K. J., & Peelle, J. E. (2016). Effects of age, acoustic challenge, and verbal working memory on recall of narrative speech. *Experimental Aging Research, 42*(1), 97–111.

Wasiuk, P. A., Radvansky, G. A., Greene, R. L., & Calandruccio, L. (2021). Spoken narrative comprehension for young adult listeners: effects of competing voices and noise. *International Journal of Audiology, 60*(9), 711–722.

Watkins, O. C., & Watkins, M. J. (1975). Buildup of proactive inhibition as a cue-overload effect. *Journal of Experimental Psychology Human Learning and Memory, 1*(4), 442–452.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., …, & Yutani, H. (2019). Welcome to the Tidyverse. *Journal of Open Source Software*, *4*(43), 1686.

Williams, J. R., Brady, T. F., & Störmer, V. S. (2022). Guidance of attention by working memory is a matter of representational fidelity. *Journal of Experimental Psychology Human Perception and Performance, 48*(3), 202–231.

Winn, M. B. (2018). *Praat script for creating speech-shaped noise [software] version 12*. http://www.mattwinn.com/praat.html

Wixted, J. T. (2010). The role of retroactive interference and consolidation in everyday forgetting. *Forgetting., 338*, 285–312.

Open Practices Statement All data, code for analysis, and stimuli are available via the Open Science Framework at: https://osf.io/znv7a/, and the preregistration document for Experiment 1 at: https://osf.io/bj3tw. Experiment 2 was a reanalysis of existing data from our lab and was not preregistered.