UNIVERSITY of CALIFORNIA PRESS | Collabra: Psychology

Wennberg, J. W., Dillman-Hasso, N. H., Brown, V. A., & Strand, J. F. (2025). Impaired Performance in Noise: Disentangling Listening Effort from the Irrelevant Speech Effect. *Collabra: Psychology*, *11*(1). https://doi.org/10.1525/collabra.147319

Cognitive Psychology

Impaired Performance in Noise: Disentangling Listening Effort From the Irrelevant Speech Effect

Janna W. Wennberg[1], Naseem H. Dillman-Hasso[2], Violet A. Brown[3], Julia F. Strand[3]

[1] Department of Cognitive Science, University of California San Diego, CA, USA, [2] School of Environment and Natural Resources, The Ohio State University, Columbus, OH, USA, [3] Department of Psychology, Carleton College, Northfield, MN, USA

Keywords: irrelevant sound effect, speech perception, listening effort

https://doi.org/10.1525/collabra.147319

Collabra: Psychology

Vol. 11, Issue 1, 2025

Noise can reduce the intelligibility of spoken language and increase the effort necessary to understand speech. *Listening effort*, "the deliberate allocation of mental resources to overcome obstacles in goal pursuit when carrying out a [listening] task" (Pichora-Fuller et al., 2016), is commonly assessed by measuring response times to secondary tasks while listening to speech or by testing memory for the content of the speech. Increasing the level of background noise tends to slow responses and impair memory, and these effects are attributed to the resource-intensive process of reevaluating speech that was initially obscured or misheard. However, given that noise can impair performance on cognitive tasks that do not require processing auditory information, it is possible that noise-induced impairments typically ascribed to processing degraded speech may instead reflect increased cognitive load from the presence of noise itself. The current study assessed whether noise, in the absence of a speech task, can affect performance on tasks intended to measure listening effort. In Experiment 1 (positive control), target speech consisting of single words was presented aurally in background noise and we measured listening effort with three commonly-used paradigms. Experiment 2 was identical except that the target words were presented orthographically rather than aurally. Results showed that noise impaired performance on all three tasks when the target stimuli were presented aurally, consistent with a large body of work in the listening effort literature. Experiment 2 revealed that performance on some tasks was impaired by the presence of masking noise (particularly two-talker babble), indicating some domain-general interference. However, the magnitude of the noise-induced interference effects were markedly smaller in Experiment 2 than Experiment 1, suggesting that measures of listening effort capture variability attributable to the challenges associated with listening to speech in noise, and do not simply measure distraction or noise-induced cognitive interference.

Anyone who has attended a noisy sporting event or cocktail party is familiar with the challenges of listening to speech in noise. Not only does background noise lead to poorer performance on transcription or identification tasks (Pisoni, 1996), but it can also lead to increases in *listening effort*: "the deliberate allocation of mental resources to overcome obstacles in goal pursuit when carrying out a [listening] task" (Pichora-Fuller et al., 2016). Although the definition of listening effort outlined above applies to listening tasks broadly, the majority of research on listening effort has focused on the cognitive resources necessary to process speech. In the speech literature, listening effort is commonly measured using behavioral tasks in which participants listen to speech and either simultaneously perform another task (dual-task paradigms; see Gagné et al., 2017a for a review) or store the speech in memory for later recall (recall paradigms; e.g., McCoy et al., 2005; Rabbitt, 1968). These measures rely on the assumption that humans have a finite pool of cognitive resources (Kahneman, 1973), so when more resources are needed to process the speech, fewer remain available to efficiently perform other tasks.

According to the Ease of Language Understanding model (Rönnberg et al., 2013), any situation that elicits a mismatch between the incoming acoustic signal and representations of words in the listener's memory (e.g., background noise) will increase listening effort because additional cognitive resources such as working memory must be recruited to resolve these mismatches. Critically, the detrimental effects of noise or other types of signal degradation on listening effort are assumed to stem from the challenges of processing the speech in difficult listening conditions—that is, from the resource-intensive process of reevaluating phonemes or words that were initially obscured or misheard. A common method for degrading speech to experimentally induce listening effort is to add or increase the level of masking noise. Noise has been shown to affect mul-

tiple tasks intended to measure listening effort; it slows response times in dual-task paradigms (e.g., Picou & Ricketts, 2014), impairs performance on recall tasks (e.g., Brown & Strand, 2019a; Picou & Ricketts, 2014; Rabbitt, 1968), affects pupillary responses (see Van Engen & McLaughlin, 2018), and increases subjective ratings of listening effort (e.g., Johnson et al., 2015; Strand et al., 2018).

Noise-induced changes in tasks intended to measure listening effort are typically thought to reflect the cognitive challenges of parsing degraded speech. However, there is another possible explanation: These effects may reflect increased cognitive load from the presence of noise itself, rather than from the effect of noise on processing spoken words (see Francis, 2022). Indeed, even cognitive tasks that do not require processing acoustic speech can be negatively affected by the presence of noise (see Szalma & Hancock, 2011). For example, background noise impairs performance on the Raven's Progressive Matrices test (Dobbs et al., 2011), affects verbal reasoning (Dobbs et al., 2011), and hinders lipreading performance (Campbell et al., 2002; Myerson et al., 2016).

One area in which the negative effects of noise on cognitive performance are particularly well documented is in the memory literature: Noise impairs performance on memory tasks even when the noise is not relevant to the task (e.g., the task involves visual memory; Wais & Gazzaley, 2011; Weisz & Schlittmeier, 2006). This phenomenon—known as the irrelevant speech effect[1] (Baddeley & Salamé, 1986; Beaman et al., 1998; Beaman & Jones, 1997; Norris et al., 2004; Salamé & Baddeley, 1982)—may occur because noise (and speech in particular) is automatically processed in the phonological loop, which disrupts working memory for other information that is being rehearsed subvocally (e.g., Neath, 2000; Salamé & Baddeley, 1982). Given that verbal stimuli from both auditory and visual modalities are processed in the phonological loop, aurally presented speech stimuli can impair recall of to-be-remembered items, even when those items are presented visually. For example, Colle and Welsh (1976) demonstrated that recall of visually-presented digits was impaired when an unfamiliar language was played in the background. Critically, noise-induced interference in memory tasks does not occur for all types of noise: Physically changing sounds such as speech and music disrupt serial recall of visually-presented items, but steady-state noise typically does not (this is known as the changing state hypothesis; Jones & Macken, 1993; Jones & Morris, 1992). In sum, there is ample evidence that modulating noise can impair performance on memory tasks, even cross-modally.

To be clear, noise-induced cognitive interference cannot explain the entirety of the listening effort literature. First, many factors other than background noise have been shown to affect listening effort. For example, vocoded speech (Winn, 2016) and reverberant speech (Rennies et al., 2014) lead to greater listening effort than natural speech, and nonnative-accented speech requires greater listening effort than native-accented speech (Borghini & Hazan, 2018; Brown et al., 2020; McLaughlin & Van Engen, 2020). Second, steady-state noise increases listening effort without producing interference on other cognitive tasks (e.g., Brown & Strand, 2018), demonstrating that not all findings in the listening effort literature can be attributed to the irrelevant speech effect. Nevertheless, adding background noise is a common way to induce listening effort, and the mechanisms underlying noise-induced performance deficits on tasks intended to measure listening effort remain unclear. Slower response times and poorer recall for speech in noise may indeed reflect cognitive challenges associated with resolving mismatches between acoustic input and mental representations (as is typically assumed in the listening effort literature). Alternatively or in addition, these effects may be driven by the fact that noise impairs performance on cognitive tasks more generally. The existing work is not able to fully distinguish between these possibilities.

Prior work attempted to assess whether performance on a task commonly assumed to measure listening effort was affected by the presence of background noise in the absence of speech (Brown & Strand, 2018). In this task, participants made speeded judgments about visually-presented numbers while also listening to and repeating aurally-presented words (Picou & Ricketts, 2014; Sarampalis et al., 2009). In line with prior work, Brown and Strand (2018) demonstrated slower response times to the number judgment dual task as noise level increased. However, when the same task was performed without the aurally-presented speech, the noise-induced slowdowns on the number judgment dual task were not observed. This suggests that when the background noise was loud and speech was present, the slowed responses were a function of increased effort from processing degraded speech; the authors did not observe cognitive interference from the presence of noise itself.

However, the finding that noise alone does not impair performance on tasks intended to measure listening effort may not extend to other listening effort paradigms. Listening effort has been assessed using a variety of tasks, and there is growing doubt that those measures tap into the same underlying construct (Alhanbali et al., 2019; Strand et al., 2018, 2021). Indeed, measures of listening effort are often weakly intercorrelated (e.g., Johnson et al., 2015; Seeman & Sims, 2015; Strand et al., 2018), and they produce different patterns of results even when the same noise conditions and speech stimuli are used (Brown & Strand, 2019a). It is therefore possible that some tasks designed to measure listening effort are more affected by the presence of background noise than the dual-task paradigm used by

1 Note that the term *irrelevant speech effect* has also been broadened to include non-speech sounds (i.e., the *irrelevant sound effect*), reflecting the fact that modulating non-speech sounds such as changing tones can also interfere with working memory (Jones et al., 1999; Jones & Macken, 1993).

Brown and Strand (2018). Specifically, given that irrelevant sounds are particularly detrimental to memory tasks (e.g., Jones & Macken, 1993; Salamé & Baddeley, 1982), measures of listening effort that rely on encoding and recalling information may be more susceptible to noise-induced cognitive interference than dual-task paradigms.

Further, the results of Brown and Strand (2018) may not generalize to other types of noise. That study was conducted in steady-state background noise, which is commonly used to degrade speech by masking phonetic detail. However, steady-state noise is not as disruptive to cognitive, perceptual, and motor tasks as more complex forms of noise are (see Szalma & Hancock, 2011 for a meta-analysis), consistent with the changing-state hypothesis (Jones & Macken, 1993; Jones & Morris, 1992). For example, noise with temporal variation is especially detrimental to working memory performance (Jones et al., 1990; Salamé & Baddeley, 1989; Tremblay et al., 2001), and two-talker babble interferes with lipreading ability more than steady-state noise does (Lidestam et al., 2014; Myerson et al., 2016). Thus, although Brown and Strand (2018) provided evidence that steady-state noise in the absence of speech did not impair performance on a particular listening effort task, it is not clear whether that finding extends to other tasks or other types of masking noise.

The Current Study

The goal of the current study was to distinguish between two explanations for noise-induced performance decrements on tasks intended to measure listening effort. One explanation is that noise leads to mismatches between the incoming speech and representations of words stored in memory, and cognitive resources (i.e., listening effort) must be recruited to resolve these mismatches, consistent with the Ease of Language Understanding model (Rönnberg et al., 2013). Another explanation is that noise may increase cognitive load and therefore may interfere with working memory and processing speed, consistent with research on the irrelevant speech effect (Baddeley & Salamé, 1986). Either mechanism alone or both explanations jointly could explain the finding that performance on listening effort tasks is impaired by noise. The current research aimed to disentangle these explanations and provide insights about the mechanism by which noise affects performance on tasks intended to measure listening effort.

The present study employed three widely-used listening effort tasks in which participants listened to and repeated words in silence, steady-state speech-shaped noise, and two-talker babble. Noise was always presented aurally, and the target stimuli were presented either aurally (Experiment 1) as they are in traditional listening effort tasks, or orthographically (Experiment 2) such that the tasks closely mirrored listening effort tasks but noise effects could not be attributable to mismatches between speech input and representations in memory. Experiment 1 served as a positive control to ensure that our lab can demonstrate effects typically attributable to listening effort using each of the three paradigms, and Experiment 2 enabled us to assess the extent to which performance impairments on the three tasks were due to noise-related cognitive interference generally rather than listening effort resulting from processing degraded speech.

The three tasks we selected are well established in the listening effort literature: a vibrotactile dual-task paradigm that requires making judgments about the duration of vibrations presented to the index finger while listening to and repeating words (Brown & Strand, 2019a; Fraser et al., 2010; Gosselin & Gagné, 2011), a verbal dual-task paradigm that involves judging whether isolated words are nouns (Picou & Ricketts, 2014; Strand et al., 2018), and a running memory task that requires storing and later recalling lists of words (McCoy et al., 2005; Sommers & Phelps, 2016; Strand et al., 2018). We selected these tasks for three reasons. First, we wanted to include both a dual-task and a recall paradigm, as they are the most commonly used behavioral methods of assessing listening effort. Second, we wanted to include a dual-task paradigm that requires verbal processing (making a judgment about a word) to enable a more direct comparison to the recall task, which also requires verbal processing. Had we only included the non-verbal dual-task paradigm (the vibrotactile dual task) and found that background noise affected performance on the recall and dual-task paradigms differently, it would be unclear whether the observed differences were the result of a difference between recall and dual-task paradigms or between verbal and non-verbal tasks. Third, given that successfully performing the noun judgment dual task relies on accurately perceiving the speech (i.e., a participant cannot judge a word as a noun if they did not hear it), we also wanted to include a dual-task paradigm in which the speech task and listening effort task could be completed independently. Each task was completed in silence, steady-state noise (to induce masking at the level of the auditory periphery, i.e., energetic masking), and two-talker babble (to induce masking attributable to higher-level cognitive processing, i.e., informational masking; Freyman et al., 1999; see Figure 1).

Finally, to obtain a measure of self-reported listening difficulty, participants completed the four questions on the NASA Task Load Index (NASA-TLX; Hart & Staveland, 1988) intended to measure mental demand, performance, effort, and frustration. We excluded the questions regarding physical and temporal demand because they are not relevant to our research question and may confuse participants. These ratings were obtained throughout each of the three tasks in both experiments.

Hypotheses

Experiment 1

In Experiment 1, the target stimuli were presented aurally, as is always the case with listening effort research. Experiment 1 therefore serves as a positive control to ensure that our lab can demonstrate results consistent with prior work on listening effort using these particular tasks, types of noise, and stimuli.

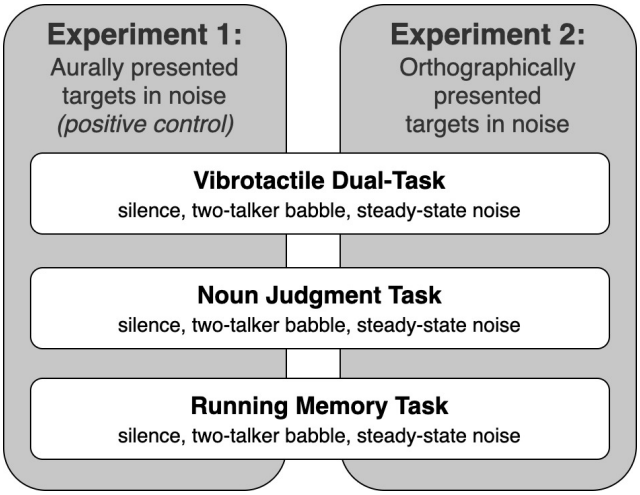

Figure 1. Schematic of experimental procedure. Participants completed all three tasks and all three noise conditions in either Experiment 1 or Experiment 2.

> ***Hypothesis 1:*** *We expected to replicate the well-established finding that both steady-state noise and two-talker babble adversely affect performance on all three tasks relative to listening in quiet.* We did not have specific predictions about whether two-talker babble or steady-state noise would lead to greater increases in listening effort because the extent of effort depends on idiosyncrasies of the stimuli such as characteristics of the talkers and signal-to-noise ratio. The magnitude of the effects from Experiment 1 will serve as a point of comparison for Experiment 2.

Experiment 2

The purpose of Experiment 2 was to assess whether the performance deficits observed in Experiment 1 might be driven by noise-induced cognitive interference rather than listening effort associated with processing degraded speech. Experiment 2 was identical to Experiment 1 except that the target words were presented orthographically rather than aurally. If background noise impairs performance on the tasks in Experiment 2, it would suggest that findings typically attributed to difficulty with listening to speech may stem from cognitive challenges associated with noise. Our hypotheses for Experiment 2 were as follows:

> ***Hypothesis 2a:*** *We expected that steady-state noise would not impair performance relative to performance on the same tasks in silence.* Brown and Strand (2018) found that response times to a non-verbal dual-task paradigm were unaffected by the level of steady-state noise, so we expected that performance on the vibrotactile dual task (another non-verbal dual-task paradigm) is unlikely to be affected by steady-state noise. Furthermore, the changing state hypothesis predicts that modulating noise (such as speech) taxes the phonological loop, but steady-state noise does not (Jones et al., 1990; Salamé & Baddeley, 1987; Weisz & Schlittmeier, 2006). We therefore expected that performance on the noun judgment and recall tasks would also be unaffected by steady-state noise. If we find that steady-state noise impairs performance on the recall task, this would be inconsistent with the changing state hypothesis (Jones & Macken, 1993; Jones & Morris, 1992) but would be consistent with the general finding that irrelevant sounds can impair performance on cognitive tasks. Although we do not anticipate this effect, this would indicate that tasks intended to measure listening effort may also be affected by cognitive challenges associated with the presence of noise.
>
> ***Hypothesis 2b:*** *We expected that performance on all three tasks would be negatively affected by the presence of two-talker babble relative to performance in silence.* Baddeley's model of working memory would predict that two-talker babble automatically enters the phonological loop and interferes with processing and remembering words (see Baddeley, 1992). This leaves fewer resources available for completing the other tasks, leading to slower response times and poorer recall of orthographic words when the task is completed in two-talker babble relative to silence. Given research on the irrelevant speech effect, finding no difference between the two-talker babble and silence conditions for any task (and the recall task in particular) would be surprising and would indicate that the two-talker babble did not place sufficient strain on the phonological loop to impair performance. This would strengthen the argument that these listening effort tasks are assessing the effort that results from processing degraded speech rather than from the mere presence of background noise.
>
> ***Hypothesis 2c:*** *We expected that the magnitude of the interference from the two-talker babble would be largest for the recall task, smaller for the noun judgment dual task, and smallest for the vibrotactile dual-task paradigm.* We anticipated that two-talker babble would be particularly taxing for tasks that require greater recruitment of working memory (the recall task) and those that require more verbal processing (the noun judgment dual task). These experiments will clarify which tasks assess the listening effort associated with processing degraded speech and which tasks assess more general noise-induced cognitive interference.

Experiment 1: Auditory Presentation (Positive Control)

All stimuli, raw data, code for statistical analysis, and the Stage 1 Registered Report are available at https://osf.io/asnqj/.

Method

Participants

Participants consisted of 55 young adults (ages 18–28; see Supplementary Materials for full demographic information) from the Carleton College community with self-reported normal or corrected-to-normal vision and no known hearing impairment. We recruited participants via posted advertisements, word of mouth, and email. Sample size was pre-determined via power analysis (see Supplementary Materials). All procedures were approved by the Carleton College Institutional Review Board, and participants gave

written consent prior to participating. Participants in both experiments received $15 for 80 minutes of participation. To reach our intended sample size, we collected data from 65 participants. Each task used data from the first 55 usable participants for that task (see below). Additional data was discarded. Thus, the sample size was 55 participants for each task, and although most of the participants overlapped across tasks, the participants are not identical across the three tasks.

Speech Stimuli

Speech stimuli consisted of 810 words selected from the English Lexicon Project (Balota et al., 2007), and each word's dominant part of speech was determined from the SUBTLEX-US database (Brysbaert et al., 2012). Following the conventions of Strand, Brown, and Barbour (2020), we subsetted the full database to only include words with log-frequencies of three or higher (Brysbaert & New, 2009), two to five phonemes, and one or two syllables. We also excluded proper nouns, articles, conjunctions, interjections, profane or emotionally evocative words, and homographs (e.g. "lead"). We then selected words at random from this set, replacing multiple instances of homophones and words with multiple forms (e.g. "bad" and "badly"). Finally, we replaced words as needed until 55% of words were classified predominantly as nouns to follow the norms of previous studies that have used the noun judgment dual task (Picou & Ricketts, 2014; Strand et al., 2018). Of the 810 words, we randomly assigned 180 to appear in the vibrotactile dual task, 270 to appear in the noun judgment dual task, and 360 to appear in the recall task, maintaining the 55% noun composition in each of the three tasks. The number of stimuli presented in each task was determined to ensure that the analyses associated with each task were sufficiently powered. The three sets of words were matched on length, frequency, number of orthographic and phonological neighbors, number of syllables, and number of phonemes.

Speech stimuli were recorded with a Blue Yeti microphone by a female speaker without a strong regional accent and edited and equated on root-mean-square amplitude using Adobe Audition. Speech was presented in noise at a signal-to-noise ratio (SNR) of -3 dB for the steady-state noise condition and 0 dB for the two-talker babble condition. These SNRs were chosen via pilot testing to determine a level of noise that resulted in intelligibility levels of approximately 70% correct in each noise condition. This level of difficulty was chosen to make the task difficult enough for noise effects to emerge but not so difficult that participants can not hear the speech and therefore can not complete the tasks. Speech files were played at approximately 68 dB SPL throughout the experiment, and noise levels were set to attain the desired SNR.[2]

Noise Stimuli

The steady-state noise consisted of speech-shaped noise generated in Praat (version 6.0.36) to match the long-term average spectrum of the target stimuli (Winn, 2018). The two-talker babble also matched the long-term average spectrum of the target stimuli (see Brouwer et al., 2012), and consisted of two different female speakers producing simple, meaningful sentences from the Bamford-Kowal-Bench sentence list (Bench et al., 1979; e.g., "The clown had a funny face"). We obtained recordings of these sentences that had been equalized on RMS amplitude from Van Engen (2010). To generate the two-talker babble, we combined the audio files for each speaker into one continuous stream, then overlaid the two speakers streams. Natural fluctuations in speaking speed ensure that sentences do not consistently start and stop at the same time, and the original stimuli are created such that there are no pauses between sentences.

For all tasks, noise played continuously during blocks with background noise present, but paused when participants were completing the NASA-TLX questionnaire. We chose continuous noise presentation for the vibrotactile and noun judgment dual tasks to increase temporal uncertainty, thus making the listening task more challenging. Although running memory tasks typically do not have noise present during the recall portion, we chose to include it here because continuous noise presentation for some tasks but not for others could result in different degrees of noise habituation across tasks, further limiting our ability to compare effects of noise across these tasks.

NASA-TLX

Participants completed four of the six questions from the NASA-TLX throughout the three tasks. Participants responded by clicking a location along an unnumbered 21-point scale ranging from "Very Low" (or "Failure" in the case of the performance question) to "Very High" (or "Perfect" in the case of the performance question).[3] Each question was presented in the following order:

1. "How mentally demanding was the task?" [mental demand]
2. "How successful were you in accomplishing what you were asked to do?" [performance]
3. "How hard did you have to work to accomplish your level of performance?" [effort]
4. "How insecure, discouraged, irritated, stressed, and annoyed were you?" [frustration]

We only analyzed data for the effort question, but we included the other questions to isolate subjective effort from other factors (such as beliefs about performance).

2 We had originally planned to hold the level of the noise constant and change the amplitude of the speech, but realized that did not specify the level at which to present the speech in silence. Therefore, we set the level of the speech and manipulated the noise level.

3 Note that the original survey presented "Perfect" on the left end of the performance scale and "Failure" on the right end, but we switched the scale limits to be consistent with the other questions on the NASA-TLX.

Procedure

Participants sat a comfortable distance from a 21.5-inch iMac computer in a sound-attenuating booth and wore sound-isolating headphones to attenuate sounds from the vibrotactile apparatus as well as those outside the testing environment. All participants completed three tasks: a vibrotactile pulse length classification task, a noun judgment dual task, and a running memory task (see Figure 1). Each task was conducted in three noise conditions: silence, steady-state noise, and two-talker babble. The three tasks were blocked and the order of the blocks was pseudorandomized. Within each task, words were randomly divided into three lists to be presented in each of the noise conditions to ensure that within a task, words appeared in all conditions approximately the same number of times across participants. Within each task, the noise conditions were blocked, and the order in which they were presented was randomized. For all tasks, participants repeated words aloud as they heard them, and responses were recorded in Audacity (version 3.2) and coded for identification accuracy offline by research assistants.

Each participant completed the NASA-TLX four times per noise condition to provide multiple observations per participant per condition. Thus, the NASA-TLX was completed every 15 words for the vibrotactile dual task, every 22 words for the noun judgment dual task (with the exception of the last presentation, which occurred after 24 words), and after every four lists for the recall task.

Vibrotactile dual task. The vibrotactile dual task was identical to the one used in two previous studies by our lab (Brown & Strand, 2019a, 2019b). Vibrotactile stimuli were presented via a custom-made apparatus consisting of a 3D-printed finger rest and a direct current vibrating motor that delivers pulse trains of various lengths. Vibrations were delivered to the index finger of the participant's nondominant hand. During the task, participants were presented with short (100 ms), medium (150 ms), and long (250 ms) pulses from the vibrotactile device. The apparatus and the participant's hand were placed inside a box lined with noise-attenuating foam to reduce any sounds generated from the vibrating apparatus. Prior to the main task, all participants completed a familiarization block. Participants were first presented with two short pulses, two medium pulses, and two long pulses. During familiarization, participants identified 18 randomly ordered pulses (six of each length) by pressing the appropriate button on the box. In the event of an incorrect response, the correct answer was immediately displayed on the screen. In order to pass the familiarization phase, participants must obtain at least 75% accuracy (14 out of the 18 pulses). If this threshold was not met, the entire familiarization block was repeated. Following successful completion of the familiarization block, participants completed the remaining three experimental blocks.

Participants completed a total of 180 trials in the main task (60 per noise condition). Each trial consisted of a vibrotactile pulse and an auditorily presented word. The pulse began somewhere between 100 ms before the onset of the word and 150 ms after the onset of the word, randomly selected from 50 ms intervals. We chose these onset times because they ensure that the cognitive processing required to perform the speech task and the vibrotactile pulse classification task coincide. That is, even if the shortest pulse is presented at the earliest onset time and the pulse itself does not coincide with presentation of the word, making a judgment about the pulse will coincide with presentation of the word. After the word was presented, there was a variable interstimulus interval ranging from 2,500 to 3,500 ms, randomly selected from 250 ms intervals. Participants responded to the vibrotactile stimulus by pressing the appropriate button on a button box and then repeating the word aloud. The outcome measure of interest was response time to make the vibration length judgment, measured from the onset of the vibration.

Noun judgment dual task. The procedure for the noun judgment dual task followed the conventions of previous work implementing this task (Picou & Ricketts, 2014; Strand et al., 2018). At the start of each trial, a word was presented through headphones, and participants were asked to press a button on a button box as quickly as possible if the word can ever be classified as a noun. After making this noun judgment, participants repeated the word they perceived aloud, regardless of whether they had classified it as a noun. The interstimulus interval again ranged from 2,500 to 3,500 ms in random 250 ms intervals. The outcome measure of interest was response time on trials during which the participant indicated that the word was a noun. Following the conventions of prior work, we analyzed all noun responses as opposed to all "correct" responses because many nouns can be categorized as other parts of speech (see Picou & Ricketts, 2014; Strand et al., 2018). Participants completed a total of 270 trials in the main task (90 per noise condition).

Running memory task. The procedures for this task were similar to those we employed in our previous work (Brown & Strand, 2019a). Participants completed three blocks of the running memory task (McCoy et al., 2005; Morris & Jones, 1990; Sommers et al., 2015; Sommers & Phelps, 2016; Strand et al., 2018), one per noise condition. In each block, participants were presented 16 lists of words ranging in length from five to ten words, with 1,000 ms between each word. Each list length was presented three times per noise condition, with the exception of the shortest and longest list lengths (5 and 10 words), which were each presented twice, for a total of 120 words per condition. Words were assigned to lists randomly, but each list was manually checked to ensure that none of the words within a list were semantically related. Participants were instructed to repeat each word aloud immediately after presentation, and at the end of each list, verbally recall the last four words in each list in any order. The next trial was initiated after a button press or after eight seconds elapsed. We included all words when calculating intelligibility scores for exclusion purposes (see below), but the outcome measure of interest is recall of the words in the 3- and 4- back positions of each list. With 16 lists of words per noise condition, this results in 32 critical items per noise condition. Words

were counted as being recalled correctly if they were recalled as they had been perceived—that is, if the participant initially misperceived a word but then later recalled that misperception, this is counted as correct recall (e.g., Brown & Strand, 2019a; Johnson et al., 2015; Pichora-Fuller et al., 1995).

Results

Model Fitting

We used both frequentist and Bayesian multilevel modeling to assess evidence for our hypotheses, and used the following packages in R version 4.2 (R Core Team, 2022) for data cleaning and analysis: the *tidyverse* suite of packages (Wickham et al., 2019), *lme4* (version 1.1.23; Bates et al., 2014), *lmerTest* (Kuznetsova et al., 2017), *brms* (Bürkner, 2017), *data.table* (Barrett et al., 2023), *bayestestR* (Makowski et al., 2019), and *effsize* (Torchiano & Torchiano, 2020). Of primary interest are the outcomes of the Bayesian models, primarily because Bayesian approaches enable us to accept the null hypothesis. However, because we created a sampling plan with frequentist techniques, we wanted to include frequentist models as well (see Supplementary Materials). Furthermore, given the widespread use of frequentist models in psychological science, including these analyses increases the accessibility and future replicability of our work.

Each Bayesian model had four MCMC chains of 6,000 iterations total, 2,000 of which were for warm-up. This resulted in 16,000 post-warm up samples per model. Divergent transitions during sampling can lead to inaccurate posterior sampling, so when we encountered divergent transitions, we increased the "adapt delta" parameter (a sampler control parameter) up to .99. This slows the sampler but ensures more accurate transitions. If divergent transitions were still identified, we continued to increase the "adapt delta" value until there were no longer divergent transitions. To ensure that estimation went smoothly and no convergence issues were encountered, we checked that all R-hat values were equal to 1.00 after sampling (or at least below 1.10). We used default priors for all models, which included a uniform distribution over the coefficient estimates of fixed effects, a half-student t-distribution (df = 3, μ = 0) over group-level effects, and an LKJ(1) correlation prior for the correlations between group-level effects. Although concerns have been raised about the use of default priors for studies with small samples (Smid & Winter, 2020), all experiments reported here employ large sample sizes to avoid these and other issues related to studies with small samples. We performed posterior predictive checks on each model, which are available in Supplementary Figures 2-4, and model objects are available online with all data and code.

Behavioral Measures. The vibrotactile and noun judgment dual tasks used response time (in ms) as the outcome, whereas the recall task used accuracy (0 or 1) as the outcome. We assumed a Gaussian distribution and an identity link function for response time data. Although response times tend to be skewed, linear mixed effects models tend to be robust to violations of the normality assumption, and we trimmed extreme response time values before beginning data analysis (see Exclusion Criteria).[4] For recall data, we assumed a Bernoulli distribution with a logit link function due to the binary nature of the data (i.e., the outcome of each trial is either 0 or 1).

Participants and words were included as random effects, and following the recommendations of Barr and colleagues (2013) and Brown (2021), we attempted to fit the model with the maximal theoretically-motivated random effects structure justified by the design. Specifically, we included by-participant and by-item random intercepts, as well as by-participant and by-item random slopes for noise type for all tasks except the noun judgment dual task. In this task, responses were only included in the analysis if the word was classified as a noun. Given that some words may never or very rarely be classified as nouns, some levels of the random effect for words have very few observations. This sort of imbalance in the data can cause convergence issues, and random slopes estimates for levels with only a few observations are uninformative, so we did not include by-word random slopes in any of the analyses for the noun judgment dual task.

NASA-TLX. The primary goal of this study was to assess whether the three tasks described above that are commonly used to behaviorally assess listening effort are also sensitive to changes in background noise in the absence of speech. However, as a supplemental exploratory measure, we assessed the effect of background noise on the subjective effort reported for each task separately. We used linear mixed effects models (assuming a Gaussian distribution with an identity link function) with random intercepts for participants and by-participant random slopes for noise. Random effects for items were not included in this analysis because participants responded to a block of stimuli rather than a particular item, and only one item on the NASA-TLX (the question related to perceived effort) was analyzed. Given that each participant responded to the effort question four times per noise condition in each task, we analyzed 12 responses per participant for each task. Model fitting and comparison criteria for all subjective effort analyses were identical to those for our behavioral listening effort analyses.

4 To ensure that our results were not affected by this modeling decision, we ran posterior predictive checks to examine whether the Gaussian model captures the skew of the response time distribution. Although a lognormal model better captures this skew, the fixed effects of interest were qualitatively similar between Gaussian and lognormal models. For ease of interpretation and to be consistent with our preregistered plan, we report results assuming normally distributed residuals.

Exclusion Criteria

For each task, we determined each participant's accuracy at repeating the target words aloud to ensure that we only included participants who attended to the primary speech task. Particularly for the dual-task paradigms, if participants stopped attending to the speech task, this may result in performance improvements on the secondary tasks. If, for any task and any noise condition, a participant's word identification accuracy was more than three standard deviations below the mean accuracy for that condition, that participant's data was excluded from all analyses for that task. However, if a participant met this exclusion criterion in a condition but still achieved at least 90% accuracy for that condition, they were not excluded. We made this decision because if performance is at ceiling level in a given condition, this renders high means and small standard deviations, which can result in an extremely conservative performance cutoff. We wanted to avoid unnecessarily discarding data from participants who were performing the task reasonably well, so we removed participants only if their performance was below 90% in addition to being three standard deviations below the mean. Finally, participants were excluded from a response time task if their mean response time in any condition was more than three standard deviations above or below the mean for that condition. Note that if a participant met an exclusion criterion in any of the three noise conditions for a particular task, their data was excluded from all noise conditions, but only for that task.

Individual trials were excluded if response times to the the secondary tasks in the two dual-task paradigms were more than three median absolute deviations above or below that participant's median response time. Following the recommendations of Leys and colleagues (2013), we used medians and median absolute deviations in lieu of means and standard deviations here because raw response times tend to be skewed. During data analysis, we discovered some trials in which participants pushed the noun button multiple times per trial, particularly in the two-talker babble condition. This was unexpected and was possibly the result of participants responding to words in the background noise rather than (or in addition to) the target speech. Although this was not part of our analysis plan, we opted to remove all trials in which the noun button was pressed more than once because responses to the two-talker babble are uninformative, and it is unclear which response we should use if a trial was responded to multiple times. All multiple-press trials came from three participants who pushed the noun button multiple times per trial on more than 60% of trials. Two of these participants would have been replaced for the preregistered criterion of having low accuracy at word identification, and we opted to replace the third participant as well out of concern for data quality. We made these decisions before conducting any statistical analyses.

Critical Tests

We assessed the effects of background noise on response time or recall separately for each task. We used a dummy coding scheme in which the silence condition served as the reference level. Although we did not have a specific prediction about whether listening effort would differ between two-talker babble and steady-state noise, we conducted that pairwise comparison as well by re-leveling the full model. The Bayesian analyses reported below involved examining the 95% highest density interval (HDI) around the fixed effect of noise. For this study, we adopted the decision rule that values outside the HDI are rejected; thus, we interpreted the effect of noise as significant if the HDI did not contain zero (Nicenboim & Vasishth, 2016). If the HDI contains zero, that implies that zero is a plausible estimated value for the coefficient, suggesting that there is not good evidence that the coefficient differs from zero.

Given that the focus of this study was on the behavioral measures of listening effort rather than speech intelligibility, we did not statistically analyze the intelligibility data, but present descriptive statistics in Table 1. The results of the Bayesian mixed effects analyses are reported below, and the results of the corresponding frequentist analyses are reported in the Supplementary Materials (the two sets of analyses revealed identical patterns of results). Note that all references to response times below refer to response times to the secondary task (either identifying the length of the pulses in the vibrotactile dual task or classifying the word as a noun in the noun judgment dual task).

Vibrotactile Dual Task

Mean accuracy at identifying the pulses as short, medium, or long was 73.49%. Response times to the pulses were on average an estimated 106 ms slower in steady-state noise than silence (B = 105.82; SE = 16.95; HDI: [71.68, 138.27]) and 181 ms slower in two-talker babble than silence (B = 180.80; SE = 22.56; HDI: [135.89, 225.06]). Releveling the noise variable revealed that response times were estimated to be 75 ms slower in two-talker babble than in steady-state noise (B = 74.82; SE = 20.97; HDI: [34.48, 116.84]). None of these HDIs included zero, suggesting that the differences in response times between all three noise levels were reliable (see Figure 2).

Subjective Effort. Subjective effort ratings were on average 3.6 points higher in steady-state noise than in silence (B = 3.60, SE = 0.47, HDI: [2.68, 4.54]) and 5.3 points higher in two-talker babble than in silence (B = 5.34, SE = 0.56, HDI: [4.23, 6.44]). Effort ratings were estimated to be 1.7 points higher in two-talker babble than in steady-state noise (B = 1.73, SE = 0.35, HDI: [1.03, 2.41]). None of the HDIs included zero, indicating reliable evidence for response time differences across the three noise conditions.

Noun Judgment Dual Task

Relative to the silence condition, response times to the noun judgment dual task were an estimated 93 ms slower in steady-state noise (B = 93.46; SE = 18.50; HDI: [56.89, 129.30]) and 90 ms slower in two-talker babble (B = 89.86; SE = 14.86; HDI: [61.18, 118.98]). Unlike in the vibrotactile dual task, the HDI for the comparison between response times in steady-state noise and two-talker babble contained

Table 1. By-participant means and standard deviations for speech identification accuracy in the three tasks in Experiment 1.

	Silence	Steady-state noise	Two-talker babble
Vibrotactile dual task	99% (2%)	68% (7%)	67% (13%)
Noun judgment dual task	98% (2%)	73% (5%)	72% (8%)
Running memory task	99% (1%)	72% (5%)	73% (9%)

Figure 2. By-participant average response times to vibrotactile stimuli for each noise type in Experiment 1 (auditory). Dots represent means.

zero (*B* = -4.01, *SE* = 19.34, *HDI*: [-41.87, 33.56]), suggesting that we did not find evidence that response times differed systematically between the two types of noise for the noun judgment dual task (Figure 3).

Subjective Effort. Participants rated the speech identification task as 5.6 points more effortful in steady-state noise (*B* = 5.62; *SE* = 0.55; *HDI*: [4.52, 6.69]) and 8.2 points more effortful in two-talker babble (*B* = 8.23; *SE* = 0.62; *HDI*: [7.01, 9.44]) than in silence. Additionally, participants reported 2.6 points more effort in two-talker babble than in steady-state noise (*B* = 2.62; *SE* = 0.40; *HDI*: [1.83, 3.40]). Note that the latter pairwise comparison differs from the response time data: The difference between steady-state noise and two-talker babble was not reliable for the response time data, yet participants reported more subjective effort in two-talker babble than in steady-state noise (i.e., the HDI did *not* contain zero in the subjective effort analysis).

Running Memory Task

A Bayesian mixed model assuming a Bernoulli data-generating process using a logit link function indicated that participants recalled fewer 3- and 4-back words in steady-state noise (*B* = -0.35; *SE* = 0.09; *HDI*: [-0.53, -0.18]) and two-talker babble (*B* = -1.23; *SE* = 0.11; *HDI*: [-1.45, -1.02]) than in silence. Recall was also worse in two-talker babble than in steady-state noise (*B* = -0.88; *SE* = 0.10; *HDI*: [-1.07, -0.68]). None of the HDIs included zero, suggesting that we have reliable evidence for differences in recall across the three noise levels (Figure 4).

Subjective Effort. Participants rated the speech identification task as 3.2 points more effortful in steady-state noise (*B* = 3.24; *SE* = 0.40; *HDI*: [2.48, 4.04]) and 4.8 points more effortful in two-talker babble (*B* = 4.78; *SE* = 0.54; *HDI*: [3.68, 5.80]) than in silence. Additionally, participants reported that the listening task was 1.5 points more effort in two-talker babble than in steady-state noise (*B* = 1.54; *SE* = 0.36; *HDI*: [0.82, 2.23]). None of the HDIs included zero.

Taken together, the results of Experiment 1 are consistent with our **Hypothesis 1** outlined above: Background noise slowed secondary task response times and impaired recall of previously-heard speech. These results are consistent with previous work in the listening effort literature, and this experiment therefore serves as a positive control demonstrating that we can reproduce robust findings in the listening effort literature when the target stimuli are pre-

Figure 3. By-participant average response times to the noun judgment dual task for each noise type in Experiment 1 (auditory). Dots represent means.

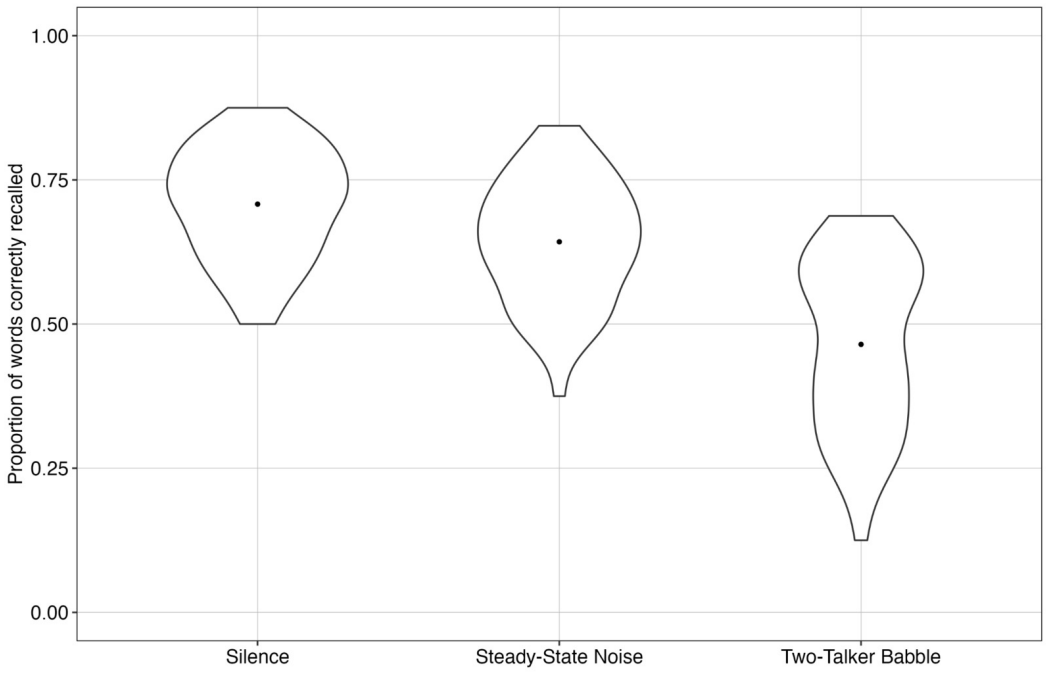

Figure 4. By-participant average proportion of words recalled to three- and four-back words recalled for each noise type in Experiment 1 (auditory). Dots represent means.

sented aurally. Thus, any null effects observed in Experiment 2 cannot be attributable to methodological limitations that preclude us from observing effects of noise on response time and recall.

Experiment 2: Orthographic Presentation

Experiment 2 followed the conventions of Experiment 1, but rather than presenting target words *aurally*, we presented them *orthographically* on the screen for participants to read. Any deviations from Experiment 1 are explicitly outlined below.

Method

Participants

Experiment 2 included data from 160 participants who did not participate in Experiment 1. To reach this sample size, we collected data from 187 participants and used the first 160 eligible datasets for each task after applying exclusion criteria, as we did in Experiment 1 (see Supplementary Materials for demographic information).

Sample Size Justification

The sample size of 160 participants is considerably larger than the sample size in Experiment 1 to account for the possibility that the effects in Experiment 2 are smaller than those in Experiment 1 (see Supplementary Materials for power analyses).

Procedure

Orthographic stimuli were presented in uppercase black text at the center of a gray screen in a sans serif font. Words were presented on the screen for a duration that matched the average duration of the target speech stimuli from Experiment 1 (627 ms), and the interstimulus intervals also matched those in Experiment 1.

Results

Participants read the isolated words presented on the screen with high levels of accuracy (99%–100% accuracy in all conditions). Posterior predictive checks are shown in Supplementary Figures 6-8.

Vibrotactile Dual Task

Mean accuracy at identifying the pulses as short, medium, or long was 77.9%. Response times were on average an estimated 40 ms slower in steady-state noise than silence (B = 40.48; SE = 6.64; *HDI*: [27.75, 53.75]) and 34 ms slower in two-talker babble than silence (B = 33.55; SE = 5.43; *HDI*: [22.81, 44.08]).[5] Re-leveling the noise variable revealed that response times did not systematically differ in two-talker babble and steady-state noise (B = -7.03; SE = 6.48; *HDI*: [-19.65, 5.71]; Figure 5). Although the HDIs comparing steady-state noise and two-talker babble to silence both excluded zero, these differences were substantially smaller than those in Experiment 1 (see "Comparison Across Experiments" section below).

Subjective Effort. Subjective effort ratings were estimated to be 2.4 points higher on average in steady-state noise than in silence (B = 2.37, SE = 0.27, *HDI*: [1.84, 2.90]), 3.9 points higher in two-talker babble than in silence (B = 3.93, SE = 0.30, *HDI*: [3.35, 4.52]) and 1.6 points higher in two-talker babble than in steady-state noise (B = 1.55, SE = 0.21, *HDI*: [1.15, 1.95]). None of the HDIs included zero.

Noun Judgment Dual Task

Response times did not systematically differ between silence and steady-state noise (B = -4.31; SE = 8.45; *HDI*: [-20.91, 11.57]), silence and two-talker babble (B = -6.23; SE = 7.40; *HDI*: [-21.17, 7.77]), or steady-state noise and two-talker babble (B = -1.98; SE = 9.47; *HDI*: [-20.63, 16.36]). Thus, we did not find evidence that noise affected response times to the noun judgment dual task when the primary task involved reading rather than listening to words (see Figure 6).

Subjective Effort. Participants rated the task as 1.4 point more effortful in steady-state noise than silence (B = 1.40; SE = 0.25; *HDI*: [0.89, 1.86]), 3.0 points more effortful in two-talker babble than silence (B = 2.95; SE = 0.27; *HDI*: [2.43, 3.49]), and 1.6 points more effortful in two-talker babble than in steady-state noise (B = 1.55; SE = 0.24; *HDI*: [1.10, 2.03]). None of the HDIs included zero, suggesting reliable differences in subjective effort between all noise conditions.

Running Memory Task

A Bayesian mixed model assuming a Bernoulli data-generating process and using a logit link function indicated that recall of 3- and 4-back words did not reliably differ between the steady-state noise and silence conditions (B = -0.09; SE = 0.21; *HDI*: [-0.50, 0.33]), but was worse in two-talker babble than in silence (B = -0.75; SE = 0.21; *HDI*: [-1.15, -0.34]). Recall was also worse in two-talker babble than in steady-state noise (B = -0.65; SE = 0.07; *HDI*: [-0.79, -0.51]; see Figure 7).

Subjective Effort. Participants rated the text identification task as 1 point more effortful in steady-state noise than silence (B = 1.12; SE = 0.21; HDI: [0.72, 1.53]), 3 points more effortful in two-talker babble than in silence (B = 3.22; SE = 0.27; HDI: [2.68, 3.74]), and 2 points more effortful in two-talker babble than in steady-state noise (B = 2.10; SE = 0.22; HDI: [1.67, 2.52]). None of the HDIs included zero.

Comparisons Across Experiments

It is possible that noise-induced impairments on listening effort tasks are the result of both listening effort from processing degraded speech and more general noise-induced cognitive interference. In that case, performance impairments should be larger in Experiment 1 than Experiment 2, as Experiment 2 cannot induce listening effort associated with processing degraded speech. We could not

5 We had originally specified that we would run separate analyses comparing steady-state noise and two-talker babble to silence. However, during data analysis we realized that those pairwise comparisons are easily accessible if silence is used as the reference level, so we opted not to conduct those unnecessary analyses.

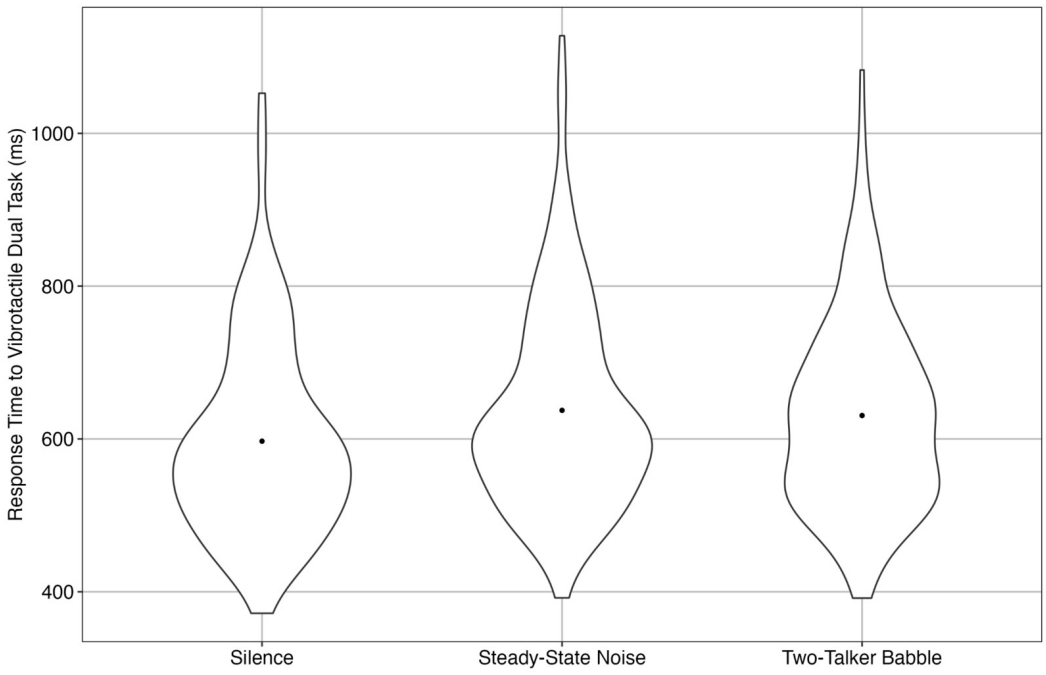

Figure 5. By-participant average response times to vibrotactile stimuli for each noise type in Experiment 2 (orthographic). Dots represent means.

Figure 6. By-participant average response times to the noun judgment dual task for each noise type in Experiment 2 (orthographic). Dots represent means.

analyze the data from Experiments 1 and 2 in the same model because the time course of hearing and reading speech are quite different, leading to response time differences between experiments for reasons other than effects of background noise. We therefore calculated effect sizes (Cohen's *d)* to assess the extent to which each noise type impaired performance relative to silence for each task in Experiment 1 compared to Experiment 2. The difference in effect sizes between Experiment 1 and Experiment 2 (referring to the difference between the silence and steady-state noise conditions as well as the difference between the silence and two-talker babble conditions, in each task separately) provides an indication of the extent to which findings typically ascribed to listening effort were in fact due

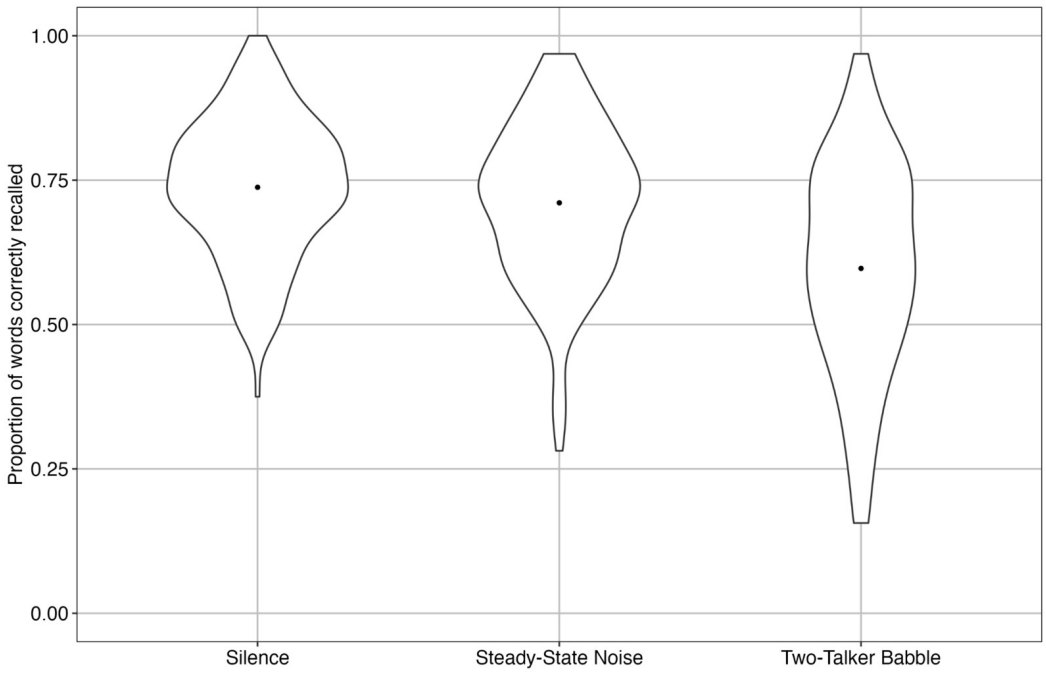

Figure 7. By-participant average proportion of three- and four-back words recalled for each noise type in Experiment 2 (orthographic). Dots represent means.

Table 2. Effect sizes (Cohen's *d*) for the comparisons between silence and either steady-state noise or two-talker babble across experiments

	Silence vs. steady-state noise		Silence vs. two-talker babble	
	Exp 1	Exp 2	Exp 1	Exp 2
Vibrotactile dual task	0.40	0.24	0.63	0.19
Noun judgment dual task	0.31	-0.01	0.31	-0.02
Running memory task[a]	0.19	0.05	0.68	0.41

[a]Note that Cohen's *d* values for the recall task were calculated by converting log-odds (extracted from the model summary outputs) to Cohen's *d* via the following formula: Cohen's *d* = log odds * ($\sqrt{3}$ / π).

to listening effort from processing degraded speech rather than to noise-induced cognitive interference. These values are reported in Table 2.

Both steady-state noise and two-talker babble impaired performance more when the stimuli were presented auditorily (Experiment 1) than when they were presented orthographically (Experiment 2) for all three behavioral tasks as well as the subjective measure of effort (NASA-TLX) collected during all three tasks. These findings suggest that at least some of the commonly-observed noise-induced decrements in performance on listening effort tasks are indeed attributable to the challenge that noise poses for listening to speech, rather than cognitive impairments generated by the noise alone.

Discussion

According to the Ease of Language Understanding model, processing speech in noise requires cognitive resources that exist in finite amounts, leaving fewer resources available to complete concurrent cognitive tasks and therefore poorer performance on these secondary tasks (Rönnberg et al., 2013). However, if background noise interferes with cognitive processing generally, it may be that some of these noise-induced impairments would emerge regardless of whether the task involved listening, questioning the interpretation that noise increases listening effort. The current study therefore tested whether noise similarly impairs performance on several tasks that are commonly used to assess listening effort when the primary task involves reading rather than listening to words. Experiment 1 served as a positive control to demonstrate that we can replicate the robust finding that noise impairs task performance when the primary task involves listening to speech, and Experiment 2 was identical but used an orthographic primary task rather than an auditory speech identification task.

Our first hypothesis (**Hypothesis 1**) was that we would replicate the well-established finding that both steady-state noise and two-talker babble adversely affect performance on all three tasks relative to listening in silence. As hypothesized, both types of noise slowed response times to

classify pulses as short, medium, or long; slowed response times to classify words as nouns; and impaired recall of verbally-presented word lists (see Table 2). These results replicate previous findings in the literature (see for example Gagné et al., 2017b) and confirm that our stimuli and procedures can produce the noise-induced performance changes that are typically attributed to listening effort.

Experiment 2 involved assessing the effects of noise on performance on the same tasks, but we presented words orthographically rather than aurally to assess whether these effects persist when the noise does not directly interfere with the perceptual clarity of the target stimuli. Our first specific hypothesis was that steady-state noise would *not* affect task performance relative to silence for any of the three tasks involving orthographic presentation (**Hypothesis 2a**). Consistent with this hypothesis, when the primary task involved reading rather than listening, steady-state noise had no effect on either response times to the noun judgment dual task or performance on the recall task. In other words, the steady-state noise-induced performance deficits for the auditory versions of these two tasks disappeared when the primary task involved reading rather than listening. This suggests that when listening effort researchers use this particular running memory or noun judgment dual task and present speech in steady-state noise, observed differences in "listening effort" indeed appear to reflect differences in the effort necessary to identify speech rather than general noise-induced cognitive interference or distraction.

For the vibrotactile dual task, however, we found a reliable difference in secondary task response times between the silence and steady-state noise conditions when the primary task involved reading, inconsistent with Hypothesis 2a. However, the coefficient for the estimated effect of steady-state noise on response times was more than 2.5 times larger when words were presented auditorily (B = 106 ms in Experiment 1) than when they were presented orthographically (B = 40 ms in Experiment 2). Thus, this cross-modal interference leads us to conclude that a small though non-negligible portion of the negative effect of steady-state noise on response times to a secondary vibrotactile dual task—which is typically attributed to listening effort—may instead be attributable to general noise-induced cognitive interference or distraction. Given that the effect of noise was much larger in Experiment 1 than in Experiment 2 (106 ms vs. 40 ms, Cohen's d = 0.40 vs. 0.24), the vibrotactile dual task also appears to measure the dual-task costs associated specifically with listening.

Hypothesis 2b predicted that performance on all three tasks would be negatively affected by two-talker babble relative to performance in silence. This hypothesis was supported for the vibrotactile dual task and the running memory task, but not the noun judgment dual task. The finding that two-talker babble can impair performance even for non-auditory tasks is consistent with irrelevant speech effect: Meaningful speech in the two-talker babble is processed automatically by the phonological loop (Jones & Macken, 1993; Salamé & Baddeley, 1982), which diverts resources away from responding to the vibrotactile pulses or encoding information into long term memory.

Somewhat surprisingly, there were no negative effects of two-talker babble in the noun judgment dual task when stimuli were presented orthographically. It is not clear why the detrimental effects of two-talker babble may have emerged for the non-verbal response time task (the vibrotactile dual task) but not the verbal one (the noun judgment dual task); indeed, we had expected that the babble interference effect would be larger for verbal relative to non-verbal tasks (consistent with the irrelevant speech effect). One key difference between the two response time tasks that may account for this finding is that the vibrotactile dual task involves dividing attention between two completely unrelated tasks (identifying the word and responding to the vibrotactile pulses). For the noun judgment dual task, however, the secondary task involves making a judgment about the primary task stimulus. Thus, processing the primary and secondary task stimuli in the vibrotactile dual task involves processing two streams of *concurrent* information, whereas identifying a word and then classifying it as a noun is a *sequential* task (see Gagné et al., 2017b for more on the distinction between concurrent and sequential dual-task paradigms). Because sequential tasks do not require processing temporally coincident information, they may be easier for participants to complete than concurrent paradigms; that is, reading a word and then making a judgment about it is likely to be a less demanding task than dividing attention between two concurrent tasks.[6] If the noun judgment dual task is less cognitively demanding overall, effects of background noise might be less pronounced for this task because there are sufficient cognitive resources available to quickly and accurately complete both the identification and judgment portions of the task.

Our final hypothesis was that the magnitude of interference from two-talker babble relative to silence would be largest for the recall task and smallest for the vibrotactile dual task, with the effect size for the noun judgment dual task falling somewhere in between (**Hypothesis 2c**). Results revealed that the effect of two-talker babble was indeed largest for the recall task. The finding that recall tasks are quite susceptible to two-talker babble—even when the target stimuli are presented orthographically rather than auditorily—suggests some caution when interpreting results from studies using two-talker babble along with memory-based measures of listening effort. The irrelevant speech effect implies that verbal information, like the speech in two-talker babble, interferes with the rehearsal

6 Another piece of evidence suggesting that the noun judgment dual task may be easier is that it is more akin to a go/no-go task whereas the vibrotactile dual task is more akin to a three-alternative forced-choice task, and the former often produce lower error rates and higher accuracy than corresponding forced-choice tasks (Moret-Tatay & Perea, 2011).

process; effects typically attributed to listening effort in studies using recall paradigms may therefore be largely driven by this verbal interference rather than effortful listening.

Hypothesis 2c also predicted that the magnitude of the interference from two-talker babble would be larger for the noun judgment dual task than the vibrotactile dual task. Instead, we found the opposite pattern: Two-talker babble did not affect response times to the noun judgment dual task, but led to reliably slower response times in the vibrotactile dual task. These findings can be accounted for by the theoretical framework outlined above regarding the relative difficulties of concurrent (e.g., vibrotactile dual task) and sequential (e.g., noun judgment dual task) dual-task paradigms.

Finally, although not the primary outcome in the current study, an interesting pattern of results emerged in the NASA-TLX data. All analyses revealed that subjective effort ratings were higher in steady-state noise than in silence, and higher in two-talker babble than steady-state noise. Thus, it appears that even when speech identification accuracy is approximately matched across these two types of noise, performing cognitive tasks is subjectively more demanding in speech-like noise. This stepwise pattern of results emerges regardless of whether the two-talker babble provides sensory interference with the cognitive task in question, and emerged even when objective performance was unaffected by noise (in Experiment 2). It is also worth noting that the differences in subjective effort between both noise types and silence were larger in Experiment 1 than in Experiment 2 for all three tasks. Thus, consistent with the behavioral data, the subjective effort data suggest that such measures can reliably detect the effect of noise on listening effort above and beyond effects of noise on cognitive processing generally.

Conclusions

The current study tested whether tasks that are typically used to measure listening effort may inadvertently be measuring noise-related cognitive load instead of or in addition to the effort associated with processing speech specifically. If we had shown noise-induced performance decrements of the same magnitude for orthographically- and aurally-presented words, that would have suggested that some effects typically ascribed to the effort needed to process degraded speech may instead be attributable to other forms of cognitive interference. We found that although some tasks designed to measure listening effort can be moderately sensitive to noise even when words are presented orthographically, the magnitude of the effects of noise were smaller (and often unreliable) with orthographically-presented words than aurally-presented words *in every condition tested*. This suggests that effects typically attributed to listening effort are indeed driven at least in part by the cognitive challenges of listening to speech in noise and not to noise-induced cognitive interference more generally. This is encouraging news for listening effort researchers, as it suggests that our tasks are actually tapping into the challenges of listening. However, future research—especially work that involves measures of listening effort not included here—should consider including orthographic negative controls to explicitly test whether the effects observed with speech-in-noise are also present without speech.

Acknowledgements

This work was supported by the National Science Foundation through a Graduate Research Fellowship awarded to Violet Brown (DGE-1745038), a National Institutes of Health grant via the National Institute on Deafness and Communication Disorders awarded to Julia Strand (R15-DC018114), and Carleton College. We are grateful to Laurie Avila, Nicholas Berry, Grace Farwell, Helen Hu, Yuxin Lin, Gigi Paulig, Caroline Saksena, and Jed Villanueva for running participants

Competing Interests

The authors have no conflicts of interest.

Editors: Don van Ravenzwaaij (Editor-in-Chief)

Submitted: December 01, 2020 PST. Accepted: August 28, 2025 PST. Published: December 16, 2025 PST.

References

Alhanbali, S., Dawes, P., Millman, R. E., & Munro, K. J. (2019). Measures of listening effort are multidimensional. *Ear and Hearing*. https://doi.org/10.1097/AUD.0000000000000697

Baddeley, A. (1992). Working memory. *Science*, *255*(5044), 556–559.

Baddeley, A., & Salamé, P. (1986). The unattended speech effect: perception or memory? *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *12*(4), 525–529. https://doi.org/10.1037/0278-7393.12.4.525

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., Neely, J. H., Nelson, D. L., Simpson, G. B., & Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods*, *39*(3), 445–459. https://doi.org/10.3758/BF03193014

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3). https://doi.org/10.1016/j.jml.2012.11.001

Barrett, T., Dowle, M., Srinivasan, A., Gorecki, J., Chirico, M., Hocking, T., & Schwendinger, B. (2023). *data.table: Extension of "data.frame"* (1.14.8). https://Rdatatable.gitlab.io/data.table

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R., Singmann, H., Dai, B., Scheipl, F., Grothendieck, G., & Green, P. (2014). *Package "lme4" (Versions 1.1-15)*. R foundation for statistical computing. https://github.com/lme4/lme4/

Beaman, C. P., & Jones, D. M. (1997). Role of serial order in the irrelevant speech effect: Tests of the changing-state hypothesis. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *23*(2), 459. https://doi.org/10.1037/0278-7393.23.2.459

Beaman, C. P., Philip Beaman, C., & Jones, D. M. (1998). Irrelevant Sound Disrupts Order Information in Free Recall as in Serial Recall. *The Quarterly Journal of Experimental Psychology Section A*, *51*(3), 615–636. https://doi.org/10.1080/713755774

Bench, J., Kowal, A., & Bamford, J. (1979). The BKB (Bamford-Kowal-Bench) sentence lists for partially-hearing children. *British Journal of Audiology*, *13*(3), 108–112. https://doi.org/10.3109/03005367909078884

Borghini, G., & Hazan, V. (2018). Listening Effort During Sentence Processing Is Increased for Non-native Listeners: A Pupillometry Study. *Frontiers in Neuroscience*, *12*, 152. https://doi.org/10.3389/fnins.2018.00152

Brouwer, S., Van Engen, K. J., Calandruccio, L., & Bradlow, A. R. (2012). Linguistic contributions to speech-on-speech masking for native and non-native listeners: language familiarity and semantic content. *The Journal of the Acoustical Society of America*, *131*(2), 1449–1464. https://doi.org/10.1121/1.3675943

Brown, V. A. (2021). An Introduction to Linear Mixed-Effects Modeling in R. *Advances in Methods and Practices in Psychological Science*, *4*(1), 2515245920960351. https://doi.org/10.1177/2515245920960351

Brown, V. A., McLaughlin, D. J., Strand, J. F., & Van Engen, K. J. (2020). Rapid adaptation to fully intelligible nonnative-accented speech reduces listening effort. *The Quarterly Journal of Experimental Psychology*. https://doi.org/10.1177/1747021820916726

Brown, V. A., & Strand, J. F. (2018). Noise increases listening effort in normal-hearing young adults, regardless of working memory capacity. *Language, Cognition and Neuroscience*, *34*, 628–640.

Brown, V. A., & Strand, J. F. (2019a). About face: Seeing the talker improves spoken word recognition but increases listening effort. *Journal of Cognition*, *2*(1). https://doi.org/10.5334/joc.89

Brown, V. A., & Strand, J. F. (2019b). "Paying" attention to audiovisual speech: Do incongruent stimuli incur greater costs? *Attention, Perception & Psychophysics*, *81*(6), 1743–1756. https://doi.org/10.3758/s13414-019-01772-x

Brysbaert, M., & New, B. (2009). Moving beyond Kucera and Francis: a critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, *41*(4), 977–990. https://doi.org/10.3758/BRM.41.4.977

Brysbaert, M., New, B., & Keuleers, E. (2012). Adding part-of-speech information to the SUBTLEX-US word frequencies. *Behavior Research Methods*, *44*(4), 991–997. https://doi.org/10.3758/s13428-012-0190-4

Bürkner, P.-C. (2017). brms: An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical Software, Articles*, *80*(1), 1–28. https://doi.org/10.18637/jss.v080.i01

Campbell, T., Beaman, C. P., & Berry, D. C. (2002). Changing-state disruption of lip-reading by irrelevant sound in perceptual and memory tasks. *The European Journal of Cognitive Psychology*, *14*(4), 461–474. https://doi.org/10.1080/09541440143000168

Colle, H. A., & Welsh, A. (1976). Acoustic masking in primary memory. *Journal of Verbal Learning and Verbal Behavior*, *15*(1), 17–31. https://doi.org/10.1016/S0022-5371(76)90003-7

Dobbs, S., Furnham, A., & McClelland, A. (2011). The effect of background music and noise on the cognitive test performance of introverts and extraverts. *Applied Cognitive Psychology*, *25*(2), 307–313. https://doi.org/10.1002/acp.1692

Francis, A. L. (2022). Adding noise is a confounded nuisance. *The Journal of the Acoustical Society of America*, *152*(3), 1375. https://doi.org/10.1121/10.0013874

Fraser, S., Gagné, J.-P., Alepins, M., & Dubois, P. (2010). Evaluating the effort expended to understand speech in noise using a dual-task paradigm: The effects of providing visual speech cues. *Journal of Speech, Language, and Hearing Research: JSLHR*, *53*(1), 18–33. https://doi.org/10.1044/1092-4388(2009/08-0140)

Freyman, R. L., Helfer, K. S., McCall, D. D., & Clifton, R. K. (1999). The role of perceived spatial separation in the unmasking of speech. *The Journal of the Acoustical Society of America*, *106*(6), 3578–3588. https://doi.org/10.1121/1.428211

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

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

Gosselin, P. A., & Gagné, J.-P. (2011). Older adults expend more listening effort than young adults recognizing audiovisual speech in noise. *International Journal of Audiology*, *50*(11), 786–792. https://doi.org/10.3109/14992027.2011.599870

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (task load index): Results of empirical and theoretical research. *Advances in Psychology*, *52*, 139–183. https://doi.org/10.1016/S0166-4115(08)62386-9

Johnson, J., Xu, J., Cox, R., & Pendergraft, P. (2015). A comparison of two methods for measuring listening effort as part of an audiologic test battery. *American Journal of Audiology*, *24*(3), 419–431. https://doi.org/10.1044/2015_AJA-14-0058

Jones, D. M., Alford, D., Bridges, A., Tremblay, S., & Macken, B. (1999). Organizational factors in selective attention: The interplay of acoustic distinctiveness and auditory streaming in the irrelevant sound effect. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *25*(2), 464–473. https://doi.org/10.1037/0278-7393.25.2.464

Jones, D. M., & Macken, W. J. (1993). Irrelevant tones produce an irrelevant speech effect: Implications for phonological coding in working memory. In *Journal of Experimental Psychology: Learning, Memory, and Cognition* (Vol. 19, Issue 2, pp. 369–381). https://doi.org/10.1037//0278-7393.19.2.369

Jones, D. M., Miles, C., & Page, J. (1990). Disruption of proofreading by irrelevant speech: Effects of attention, arousal or memory? In *Applied Cognitive Psychology* (Vol. 4, Issue 2, pp. 89–108). https://doi.org/10.1002/acp.2350040203

Jones, D. M., & Morris, N. (1992). Irrelevant speech and serial recall: implications for theories of attention and working memory. *Scandinavian Journal of Psychology*, *33*(3), 212–229. https://doi.org/10.1111/j.1467-9450.1992.tb00911.x

Kahneman, D. (1973). *Attention and effort*. Prentice-Hall.

Kuznetsova, A., Brockhoff, P., & Christensen, R. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software, Articles*, *82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology*, *49*(4), 764–766. https://doi.org/10.1016/j.jesp.2013.03.013

Lidestam, B., Holgersson, J., & Moradi, S. (2014). Comparison of informational vs. energetic masking effects on speechreading performance. *Frontiers in Psychology*, *5*, 639.

Makowski, D., Ben-Shachar, M., & Lüdecke, D. (2019). BayestestR: Describing effects and their uncertainty, existence and significance within the Bayesian framework. *Journal of Open Source Software*, *4*(40), 1541. https://doi.org/10.21105/joss.01541

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. https://doi.org/10.1080/02724980443000151

McLaughlin, D. J., & Van Engen, K. J. (2020). Task-evoked pupillary response for intelligible accented speech. *The Journal of the Acoustical Society of America*, *147*. https://doi.org/10.1121/10.0000718

Moret-Tatay, C., & Perea, M. (2011). Is the go/no-go lexical decision task preferable to the yes/no task with developing readers? *Journal of Experimental Child Psychology*, *110*(1), 125–132. https://doi.org/10.1016/j.jecp.2011.04.005

Morris, N., & Jones, D. M. (1990). Memory updating in working memory: The role of the central executive. *British Journal of Psychology*, *81*(2), 111–121. https://doi.org/10.1111/j.2044-8295.1990.tb02349.x

Myerson, J., Spehar, B., Tye-Murray, N., Van Engen, K., Hale, S., & Sommers, M. S. (2016). Cross-modal informational masking of lipreading by babble. *Attention, Perception & Psychophysics*, *78*(1), 346–354. https://doi.org/10.3758/s13414-015-0990-6

Neath, I. (2000). Modeling the effects of irrelevant speech on memory. *Psychonomic Bulletin & Review*, *7*(3), 403–423. https://doi.org/10.3758/BF03214356

Nicenboim, B., & Vasishth, S. (2016). Statistical methods for linguistic research: Foundational Ideas—Part II: Statistical methods for linguistics--Part II. *Language and Linguistics Compass*, *10*(11), 591–613. https://doi.org/10.1111/lnc3.12207

Norris, D., Baddeley, A. D., & Page, M. P. A. (2004). Retroactive effects of irrelevant speech on serial recall from short-term memory. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *30*(5), 1093–1105. https://doi.org/10.1037/0278-7393.30.5.1093

Pichora-Fuller, M. K., Kramer, S. E., Eckert, M. A., Edwards, B., Hornsby, B. W. Y., Humes, L. E., Lemke, U., Lunner, T., Matthen, M., Mackersie, C. L., Naylor, G., Phillips, N. A., Richter, M., Rudner, M., Sommers, M. S., Tremblay, K. L., & Wingfield, A. (2016). Hearing impairment and cognitive energy: The Framework for Understanding Effortful Listening (FUEL). *Ear and Hearing*, *37 Suppl 1*, 5S-27S. https://doi.org/10.1097/AUD.0000000000000312

Pichora-Fuller, M. K., Schneider, B. A., & Daneman, M. (1995). How young and old adults listen to and remember speech in noise. *The Journal of the Acoustical Society of America*, *97*(1), 593–608. https://doi.org/10.1121/1.412282

Picou, E. M., & Ricketts, T. A. (2014). The effect of changing the secondary task in dual-task paradigms for measuring listening effort. *Ear and Hearing*, *35*(6), 611–622. https://doi.org/10.1097/AUD.0000000000000055

Pisoni, D. B. (1996). Word Identification in Noise. *Language and Cognitive Processes*, *11*(6), 681–687. https://doi.org/10.1080/016909696387097

R Core Team. (2022). *R 4.2.2*. 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. https://doi.org/10.1080/14640746808400158

Rennies, J., Schepker, H., Holube, I., & Kollmeier, B. (2014). Listening effort and speech intelligibility in listening situations affected by noise and reverberation. *The Journal of the Acoustical Society of America*, *136*(5), 2642–2653. https://doi.org/10.1121/1.4897398

Rönnberg, J., Lunner, T., Zekveld, A., Sörqvist, P., Danielsson, H., Lyxell, B., Dahlström, O., Signoret, C., Stenfelt, S., Pichora-Fuller, M. K., & Rudner, M. (2013). The Ease of Language Understanding (ELU) model: Theoretical, empirical, and clinical advances. *Frontiers in Systems Neuroscience*, *7*, 31. https://doi.org/10.3389/fnsys.2013.00031

Salamé, P., & Baddeley, A. (1982). Disruption of short-term memory by unattended speech: Implications for the structure of working memory. *Journal of Verbal Learning and Verbal Behavior*, *21*(2), 150–164. https://doi.org/10.1016/S0022-5371(82)90521-7

Salamé, P., & Baddeley, A. (1987). Noise, unattended speech and short-term memory. *Ergonomics*, *30*(8), 1185–1194. https://doi.org/10.1080/00140138708966007

Salamé, P., & Baddeley, A. (1989). Effects of Background Music on Phonological Short-Term Memory. In *The Quarterly Journal of Experimental Psychology Section A* (Vol. 41, Issue 1, pp. 107–122). https://doi.org/10.1080/14640748908402355

Sarampalis, A., Kalluri, S., Edwards, B., & Hafter, E. (2009). Objective measures of listening effort: Effects of background noise and noise reduction. *Journal of Speech, Language, and Hearing Research: JSLHR*, *52*(5), 1230–1240.

Seeman, S., & Sims, R. (2015). Comparison of psychophysiological and dual-task measures of listening effort. *Journal of Speech, Language, and Hearing Research: JSLHR*, *58*(6), 1781–1792. https://doi.org/10.1044/2015_JSLHR-H-14-0180

Smid, S. C., & Winter, S. D. (2020). Dangers of the defaults: A tutorial on the impact of default priors when using Bayesian SEM with small samples. *Frontiers in Psychology*, *11*, 611963. https://doi.org/10.3389/fpsyg.2020.611963

Sommers, M. S., & Phelps, D. (2016). Listening effort in younger and older adults: A comparison of auditory-only and auditory-visual presentations. *Ear and Hearing*, *37 Suppl 1*, 62S-8S. https://doi.org/10.1097/AUD.0000000000000322

Sommers, M. S., Tye-Murray, N., Barcroft, J., & Spehar, B. P. (2015). The effects of meaning-based auditory training on behavioral measures of perceptual effort in individuals with impaired hearing. *Seminars in Hearing*, *36*(4), 263–272. https://doi.org/10.1055/s-0035-1564454

Strand, J. F., Brown, V. A., & Barbour, D. L. (2020). Talking points: A modulating circle increases listening effort without improving speech recognition in young adults. *Psychonomic Bulletin & Review*. https://doi.org/10.3758/s13423-020-01713-y

Strand, J. F., Brown, V. A., Merchant, M. B., Brown, H. E., & Smith, J. (2018). Measuring listening effort: Convergent validity, sensitivity, and links with cognitive and personality measures. *Journal of Speech, Language, and Hearing Research: JSLHR*, *61*, 1463–1486.

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.

Szalma, J. L., & Hancock, P. A. (2011). Noise effects on human performance: a meta-analytic synthesis. *Psychological Bulletin*, *137*(4), 682–707. https://doi.org/10.1037/a0023987

Torchiano, M., & Torchiano, M. M. (2020). *Effsize*. r.meteo.uni.wroc.pl. http://r.meteo.uni.wroc.pl/web/packages/effsize/effsize.pdf

Tremblay, S., MacKen, W. J., & Jones, D. M. (2001). The impact of broadband noise on serial memory: Changes in band-pass frequency increase disruption. *Memory*, *9*(4), 323–331. https://doi.org/10.1080/09658210143000010

Van Engen, K. J. (2010). Similarity and familiarity: Second language sentence recognition in first- and second-language multi-talker babble. *Speech Communication*, *52*(11–12), 943–953. https://doi.org/10.1016/j.specom.2010.05.002

Van Engen, K. J., & McLaughlin, D. J. (2018). Eyes and ears: Using eye tracking and pupillometry to understand challenges to speech recognition. *Hearing Research*, *369*, 56–66. https://doi.org/10.1016/j.heares.2018.04.013

Wais, P. E., & Gazzaley, A. (2011). The impact of auditory distraction on retrieval of visual memories. *Psychonomic Bulletin & Review*, *18*(6), 1090–1097. https://doi.org/10.3758/s13423-011-0169-7

Weisz, N., & Schlittmeier, S. J. (2006). Detrimental effects of irrelevant speech on serial recall of visual items are reflected in reduced visual N1 and reduced theta activity. *Cerebral Cortex*, *16*(8), 1097–1105. https://doi.org/10.1093/cercor/bhj051

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. https://doi.org/10.21105/joss.01686

Winn, M. B. (2016). Rapid release from listening effort resulting from semantic context, and effects of spectral degradation and cochlear implants. *Trends in Hearing*, *20*. https://doi.org/10.1177/2331216516669723

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

Supplementary Materials

Supplemental Material

Download: https://collabra.scholasticahq.com/article/147319-impaired-performance-in-noise-disentangling-listening-effort-from-the-irrelevant-speech-effect/attachment/310328.docx?auth_token=SwFf3ijpqK6Rio9OiNsN

Peer Review Communication

Download: https://collabra.scholasticahq.com/article/147319-impaired-performance-in-noise-disentangling-listening-effort-from-the-irrelevant-speech-effect/attachment/310329.docx?auth_token=SwFf3ijpqK6Rio9OiNsN