Dual-Task Accuracy and Response Time Index Effects of Spoken Sentence Predictability and Cognitive Load on Listening Effort

Cynthia R. Hunter [ID]

Abstract

A sequential dual-task design was used to assess the impacts of spoken sentence context and cognitive load on listening effort. Young adults with normal hearing listened to sentences masked by multitalker babble in which sentence-final words were either predictable or unpredictable. Each trial began with visual presentation of a short (low-load) or long (high-load) sequence of to-be-remembered digits. Words were identified more quickly and accurately in predictable than unpredictable sentence contexts. In addition, digits were recalled more quickly and accurately on trials on which the sentence was predictable, indicating reduced listening effort for predictable compared to unpredictable sentences. For word and digit recall response time but not for digit recall accuracy, the effect of predictability remained significant after exclusion of trials with incorrect word responses and was thus independent of speech intelligibility. In addition, under high cognitive load, words were identified more slowly and digits were recalled more slowly and less accurately than under low load. Participants' working memory and vocabulary were not correlated with the sentence context benefit in either word recognition or digit recall. Results indicate that listening effort is reduced when sentences are predictable and that cognitive load affects the processing of spoken words in sentence contexts.

Speech perception is one of many everyday activities that draws increasingly on central cognitive resources as task difficulty increases. Listening effort is defined as the allocation of cognitive resources that are required to comprehend a spoken or other auditory message (McGarrigle et al., 2014; Pichora-Fuller et al., 2016). Conversely, cognitive spare capacity refers to the amount of cognitive resources that remain available for allocation, and decreases in cognitive spare capacity can produce measurable decrements in performance on a wide range of tasks. The dual-task framework is among the most common designs used to investigate effortful listening (for a review, see Gagne et al., 2017). The use of the dual-task design to assess listening effort is based on the widely accepted basic theory from cognitive psychology that mental processes that require conscious control

and effort are executed more slowly than automatic processes and draw from a capacity-limited pool of cognitive resources, identified with working memory and/or attentional capacity (Engle, 2002; Kahneman, 1973). As such, performance declines on a secondary task as a function of increased difficulty of a concurrent speech recognition task indicate that there is less spare capacity

Speech Perception, Cognition, and Hearing Laboratory, Department of Speech-Language-Hearing: Sciences and Disorders, University of Kansas, Lawrence, United States

Corresponding author:
Cynthia R. Hunter, Department of Speech-Language-Hearing: Sciences and Disorders, University of Kansas, 1000 Sunnyside Avenue, Lawrence, KS 66045-7555, United States.
Email: c.hunter@ku.edu

available for allocation to secondary tasks when speech perception is difficult and as such provide an index of listening effort. For example, response times in secondary visuomotor tasks performed simultaneously with speech perception have been shown to increase as the signal-to-noise ratio (SNR) becomes less favorable (Gagne et al., 2017; Neher et al., 2014; Sarampalis et al., 2009; Seeman & Sims, 2015; Wu et al., 2016). Such results indicate that speech perception in adverse listening conditions demands working memory resources that would otherwise be used to support multitasking performance.

Performance trade-offs in dual-task designs occur when both tasks demand conscious control and effort at the same time. However, it is not necessary that the behavioral responses of both tasks be made concurrently. In a variant of the dual-task referred to as a sequential or memory load design, participants are tasked with remembering items presented prior to each stimulus of the primary task and reporting these items toward the end of each trial (Baddeley & Hitch, 1974). Here, the cognitive processes that are used to retain the memory load, such as conscious rehearsal, are theorized to involve working memory (e.g., see Doherty et al., 2019; Morey & Cowan, 2004). As such, performance on the memory task in the sequential dual-task design is taken as an index of the cognitive demand of the primary task. Using this design, reductions in recall for visually presented digits presented prior to a speech signal have been used to measure listening effort (Francis & Nusbaum, 2009; Hunter, 2020; Hunter & Pisoni, 2018; Luce et al., 1983; Rakerd et al., 1996). For example, Luce et al. (1983) showed that recall of digits presented prior to spoken words was reduced when the words were degraded by a speech synthesizer. A related paradigm known as a word recall design has also been used to investigate listening effort and is based on the reading span measure of working memory (Daneman & Carpenter, 1980). In this design, participants identify spoken words presented in quiet or various levels of background noise and are later asked to recall the words (Pichora-Fuller et al., 1995). This design differs from the memory load dual-task design in that it is the identified spoken words themselves that are to be recalled rather than unrelated items presented prior to the speech. Words that are presented in a noisy background tend to be recalled less accurately than words presented in quiet (McGarrigle et al., 2014; Sarampalis et al., 2009; Surprenant, 2007), even when listeners recognize both sets of words with equivalent accuracy (Pichora-Fuller et al., 1995; Rudner, 2016; Surprenant, 1999). Thus, there is evidence from both sequential dual-task designs and word recall designs that speech perception in adverse listening conditions demands central cognitive resources that would otherwise be used to support

retention of information in memory and that this can be measured as an effect on memory performance sometime after the listening effort occurs.

In addition to speech signal quality, listening effort may be affected by internal factors, such as one's linguistic knowledge or the amount of cognitive resources allocated to other, concurrent tasks (Mattys et al., 2012; Rudner, 2016). Although it is well known that spoken word recognition accuracy improves when words are embedded in a predictable sentence context (Bilger et al., 1984; Kalikow et al., 1977), a less well-examined question is whether sentence contexts increase accuracy and reduce listening effort, or instead increase accuracy through a cognitively demanding process that increases listening effort. Two studies using the sequential (memory load) dual-task design with spectrally degraded speech have observed effects of sentence predictability on performance in a secondary memory task (Hunter, 2020; Hunter & Pisoni, 2018). Specifically, recall of memory load stimuli was more accurate on trials in which the spoken sentence was predictable rather than unpredictable. Such downstream effects of sentence predictability on later recall of unrelated digits indicate that listening is less effortful when sentences are predictable such that more cognitive resources are available to rehearse and remember other information.

Understanding the impact of sentence predictability on cognitive spare capacity during speech perception in adverse listening conditions is important because this is a potentially modifiable factor that could reduce the strain of listening effort on a listener's cognitive capacity. For people with hearing loss, the near-constant exertion of listening effort when participating in spoken communication can lead to fatigue (Alhanbali et al., 2017, 2018; Hornsby, 2013; Hornsby & Kipp, 2016; Hornsby et al., 2016) and potentially chronic stress (Pichora-Fuller, 2016). In addition, listening effort by definition will reduce one's capacity to mentally process the informational content of spoken communication. If it can be shown that sentence predictability is protective against reduced cognitive spare capacity, this may have potential applications for habilitation strategies for hearing loss. For example, people with hearing loss could be counseled or trained as part of aural rehabilitation to take advantage of sentence contexts for the purpose of preventing listening fatigue.

As described earlier, two prior studies using a dual-task memory load design with spectrally degraded spoken sentences have observed downstream effects of sentence predictability on recall accuracy in the secondary task, indicating that sentence context preserves cognitive spare capacity compared to a lack of context (Hunter, 2020; Hunter & Pisoni, 2018). These results suggest that contextual facilitation in adverse listening conditions is a relatively automatic process compared to

the listening effort demanded by low-context sentences. In addition to behavioral data, Hunter (2020) presented electrophysiological evidence that sentence predictability supports ease of listening. Specifically, both alpha-band oscillatory power and the amplitude of the P300 or late positive complex event-related potential tracked with memory load and sentence predictability, indicating that predictable sentence context increases cognitive spare capacity (Hunter, 2020). Similar findings using a pupillometry measure were reported by (Winn, 2016), who found that pupil size was smaller in the seconds after sentence offset for predictable sentences relative to unpredictable sentences for both young normal hearing participants listening to spectrally degraded speech and for older cochlear implant users. In addition, Obleser and Kotz (2009), also using spectrally degraded speech, found that fewer neural resources were allocated to speech processing for predictable compared to unpredictable sentences in brain areas associated with spoken word recognition and working memory. Each of these findings indicates that in adverse listening conditions, less effort is required for sentence recognition when sentences are predictable.

Evidence from word recall designs is also consistent with reduced listening effort when sentences are predictable. Multiple studies have observed greater accuracy in both sentence-final word recognition and recall for predictable than unpredictable sentences (Johnson et al., 2015; Pichora-Fuller et al., 1995; Sarampalis et al., 2009; Strand et al., 2018). For example, as part of a larger examination of the effects of noise reduction algorithms on listening effort, Sarampalis et al. (2009) observed that both word identification and later recall of words presented in background noise were more accurate in high-context sentences, indicating a release of cognitive resources for high-context sentences. Although effects of sentence predictability from the word recall design have been taken to indicate reduced listening effort for high predictability sentences, it is also possible that in the word recall design predictable contexts could exert their effect on recall of sentence-final words by acting as a memory retrieval cue. That is, words from predictable sentences may be recalled more accurately in this design because predictable context itself is more memorable and serves to cue recall of the sentence-final words. In contrast, in the sequential dual-task (memory load) design, the items to be remembered are unrelated to the sentence contexts, and thus differences in recall performance across high- and low-context sentences can be more clearly attributed to listening effort per se.

In contrast to prior results with the sequential dual-task and word recall design, other studies that have examined the effect of context on listening effort using a simultaneous dual-task design observed no significant effects of sentence context on secondary task performance, despite clear context benefits in word recognition accuracy (Desjardins & Doherty, 2013, 2014; Feuerstein, 1992). Each of these studies presented stimuli in background noise and used a simultaneous visual-motor secondary task in which response speed to a visual target was used to quantify listening effort. In addition, correlational studies have provided evidence that listening effort increases when sentences are predictable, rather than the converse. These studies examined correlations of working memory capacity with the context benefit in word recognition to investigate the hypothesis that taking advantage of sentence context to support spoken word recognition is effortful. At times, contextual facilitation must involve holding the acoustic-phonetic forms of unrecognized spoken words in phonological working memory while continuing to process incoming words. By definition, this should involve working memory, which is distinguished from a passive memory store by consisting not only of memory storage, but also the simultaneous processing of other information (Baddeley & Hitch, 1974). From this perspective, contextual facilitation should demand cognitive resources, potentially increasing listening effort (Pichora-Fuller et al., 1995). If so, it follows that individuals with greater working memory capacity should experience a larger boost in word recognition accuracy from context given that they have greater baseline cognitive capacity that could be allocated to an effortful form of contextual facilitation. Support for this hypothesis comes from studies that have found verbal working memory capacity to be correlated with contextual facilitation. Two studies observed that working memory capacity (reading span) was correlated with the ability to benefit from semantically related text cues presented prior to a spoken sentence in young adults with normal hearing (Zekveld et al., 2012, 2013). In both studies, spoken sentences were degraded by background noise. Similar findings were reported by a study with older adult participants that examined the sentential context benefit for a phoneme-monitoring task (Janse & Jesse, 2014). The relevant findings from this study were that verbal working memory was related to context benefit for the response time to identify target phonemes embedded in words in the sentence. In addition, a recent study with adult post-lingually deaf cochlear implant users observed a correlation of verbal working memory with the use of contextual information in sentences (Dingemanse & Goedegebure, 2019). Correlations of context benefit with working memory capacity suggest that the context benefit for word accuracy draws on working memory resources.

In sum, it is not yet clear whether the well-known context benefit in word recognition accuracy when words occur in predictable sentences is achieved by a

relatively automatic process and hence reduces listening effort, or instead via an effortful process that demands cognitive resources. Broadly speaking, both automatic and effortful routes to prediction during language comprehension have been proposed (for review, see Huettig, 2015). Automatic linguistic prediction has been theorized to involve spreading activation in lexical networks. For example, predictable sentences typically contain words that are semantically related, which in many models of the lexicon would result in spreading, associative activation among words with overlapping meanings (e.g., Collins & Loftus, 1975). Associative activation is predictive in the sense that words that follow a semantically related word in a sentence would be preactivated and hence able to reach a threshold level of activation for recognition with less bottom-up activation needed from acoustic-phonetic input. Effortless prediction may occur within language networks, without involving central cognitive resources. By contrast, more effortful linguistic prediction is thought to recruit central cognitive resources to combine multiple types of information, including syntax, semantics, and other relevant information, into higher-order representations (for a review, see Kuperberg, 2007). An automatic associative route and an effortful combinatorial route are likely both involved in making linguistic predictions of different types. However, given that the bulk of research on linguistic prediction has involved reading rather than listening and used stimuli that are not degraded, less is known about how and under what conditions these mechanisms apply to linguistic prediction in adverse listening conditions.

The overall goal of the current study was to expand the evidence base on this question using the dual-task memory load design. To this end, the behavioral design used in prior sequential dual-task studies (Hunter, 2020; Hunter & Pisoni, 2018) was modified as detailed later to more closely relate results with the memory load design, which have to date all indicated that predictable sentence contexts reduce listening effort, to results from the prior simultaneous dual-task and correlational studies discussed earlier. Additional improvements to the sequential dual-task design for examining listening effort were also implemented, as described later.

First, prior simultaneous dual-task studies that did not observe secondary task benefits for predictable sentences (Desjardins & Doherty, 2013, 2014; Feuerstein, 1992) as well as studies that observed a correlation of context benefit with working memory (Dingemanse & Goedegebure, 2019; Janse & Jesse, 2014; Zekveld et al., 2012, 2013) all used background noise to degrade the spoken sentences. This contrasts with the sequential dual-task studies (Hunter, 2020; Hunter & Pisoni, 2018) and physiological studies (Hunter, 2020; Obleser and Kotz, 2009; Winn, 2016) conducted to date in which stimuli were degraded with vocoding to simulate

listening with a cochlear implant. It may be that background noise is a more cognitively demanding form of degradation than vocoding. Speech recognition in background noise is thought to require selective attention to the target speech instead of the masker, as well as potentially other types of informational masking, particularly when the masker is speech (Kidd et al., 2008; Rosen et al., 2013). In the current study, stimuli were degraded with a multitalker babble masker. In addition, the sentences were presented at an individualized SNR that was set to approximate 50% accuracy for word recognition, similar to the intelligibility level used by prior studies (Zekveld et al., 2011). Further, correlations of context benefit in both the primary and secondary tasks with working memory capacity and vocabulary (as a proxy measure of lexical-semantic knowledge) were examined. If the context benefit for sentences presented in noise is effortful and increases demand for cognitive resources (in contrast to the context benefit for sentences that are spectrally degraded), then the current study should fail to replicate the prior finding of decreased listening effort for predictable sentences in the dual-task memory load design, and may also observe correlations of working memory with context benefit.

Second, unlike prior studies that have used the sequential dual-task design to investigate contextual facilitation, response time in both tasks was examined in addition to response accuracy. Given that controlled processes require greater processing time, reduced cognitive spare capacity in a dual-task experiment may present as slowed response time, reduced accuracy, or both. There is no clear indication of whether accuracy or response time in a dual-task design is a more appropriate measure for listening effort (see Gagne et al., 2017). However, an extensive literature in cognitive psychology documents ubiquitous trade-offs between response speed and accuracy (Bogacz et al., 2010; Wagenmakers et al., 2007; Wickelgren, 1977). For example, participants may maintain high accuracy as task difficulty increases by responding more carefully, but at the cost of slowed response times. As such, including both response time and accuracy in analysis of dual-task responses should yield a more complete picture of cognitive resource allocation than either measure alone. In the current study, response time was examined for both word identification and digit recall accuracy. With respect to the word identification response, response times during speech audiometry have been taken to index listening effort (Houben et al., 2013; Meister et al., 2018; Pals et al., 2015). Following these prior studies, word identification response times in the current study were used as a measure of listening effort. However, given that in the current design these were typed responses and as such differ from the spoken word identification response times validated by prior studies, the inclusion of word response

time in the current study as a measure of listening effort is exploratory. Similarly, the response time for digit recall was used both as a check for a potential speed-accuracy trade-off in the recall task and as an additional, exploratory, measure of listening effort. A final motivation for including response time measures in the current study was to support closer comparison with prior simultaneous dual-task studies that did not observe secondary task benefits for predictable sentences and relied on reaction time measures (Desjardins & Doherty, 2013, 2014; Feuerstein, 1992). Note, however, that in the current study, the response time measures were on a longer time scale than these prior studies. Here, all dependent measures were obtained after the exertion of listening effort in a sequential dual-task design, rather than during listening effort in a simultaneous dual-task.

Third, an attempt was made was to determine whether the effect of predictable context on cognitive spare capacity could be decoupled from its effect on word recognition accuracy. In dual-task designs, increased difficulty of the speech recognition task is often accompanied by declines in both speech perception accuracy and secondary task performance. As described earlier, two prior studies using a dual-task memory load design have observed downstream effects of sentence predictability on recall accuracy in the secondary task (Hunter, 2020; Hunter & Pisoni, 2018). However, in both of these studies, the analysis of digit recall included trials on which word identification was both correct and incorrect. Further, when Hunter (2020) excluded trials of the secondary memory task on which the word had been incorrectly identified, the effect of predictability on (digit) recall was no longer significant. This could mean that the predictability effect observed when trials with incorrect word responses were included was caused by some cognitive process specific to trials on which words were not correctly recognized, and hence not separable from speech intelligibility. In the interest of greater ecological validity and to understand if the downstream effect of predictability on digit recall reflects allocation of working memory resources independently of speech intelligibility, the current study examined secondary task responses for trials on which words were correctly identified. A further motivation for analyzing responses from correct word identification trials separately was that, from a theoretical perspective, trials on which words were not correctly identified would reflect processes of unsuccessful word recognition, which could differ from the processes involved in successful word recognition. The aim here was to understand whether memory task accuracy, response times, or both would index effects of sentence predictability specifically on trials in which words were recognized correctly.

The final goal of the current study was to examine the (converse) impact of cognitive demands on speech recognition. Prior studies have found that the difficulty level of a concurrent memory task affects performance on speech perception tasks, consistent with models of effortful listening in which central cognitive resources are needed to perceive speech under adverse listening conditions. For example, Hunter and Pisoni (2018) found that on trials in which memory load was greater (more digits to remember), word recognition accuracy decreased. A series of studies by Mattys and colleagues have also found that speech perception accuracy decreases under high cognitive load (Mattys et al., 2009, 2014; Mattys & Palmer, 2015; Mattys & Wiget, 2011). In addition, recent studies using eye-tracking have observed a slowed time course of fixation to spoken words under cognitive load (Hadar et al., 2016; Nitsan et al., 2019). In the current study, the cognitive load of the memory task in a dual-task experiment was varied, and performance in the word recognition task was examined as a function of cognitive load to assess the impacts on speech perception.

In sum, the current study examined effects of sentence predictability and cognitive load on word identification and digit recall performance (accuracy and response time) in a dual-task memory load design. The following questions were investigated:

1. Does predictable sentence context increase or decrease listening effort, as measured by sequential dual-task digit recall performance, when sentences are presented in a noisy background?
2. Are participants' working memory and/or vocabulary correlated with the context benefit in either word identification or digit recall performance?
3. Do the effects (if any) of sentence predictability on digit recall persist when analysis is restricted to trials on which the speech was correctly identified?
4. Will prior findings of decreased word identification performance under high cognitive load be replicated with this design?

Materials and Methods

Participants

Twenty-three young adults recruited from the Indiana University campus (17 females, age range 18 to 30 years) participated in the study. All participants were native English speakers who reported no history of hearing or speech disorders and had normal hearing, defined as pure tone air conduction thresholds less than 15 dB HL (American National Standards Institute, 2010) at octave frequencies from 250 to 8000 Hz as well as 3150 Hz and 6300 Hz, and a normal tympanogram in the test ear. Participants all gave written informed consent and were paid $10 for each hour of

participation, in accordance with procedures approved by the Institutional Review Board at Indiana University at Bloomington.

Measures

The experiment took place over two sessions. In Session I, measures included the Words-in-Noise (WIN) test, a computerized reading span measure of working memory, and a vocabulary test. Session II included a dual-task experiment in which participants identified sentence-final words from the Speech Perception in Noise Test-Revised (SPIN-R) and also recalled sequences of digits that were presented visually before each spoken sentence. Session I scores on the WIN were used to set initial values for individualized SNRs for Session II.

WIN Test. A computerized version of the WIN test was implemented in MATLAB using prerecorded stimuli and noise samples (Disc 4.0 of Speech Recognition and Identification Materials, issued by the Department of Veterans Affairs). A set of 35 words was presented in a carrier sentence ("Please say the word ___") spoken by a female talker in a multitalker babble masker. The SNR decreased from 24 to 0 dB in 4 dB decrements, with five words presented at each SNR. Further details of the WIN test may be found in Wilson (2003) and Wilson and McArdle (2007). The 50% point of the psychometric function was calculated using the Spearman–Karber equation. This estimated SNR50 was used to set the initial SNR for the SPIN test in Session II. The individualized SNR was set to the estimated SNR50 minus 5 dB. Based on a prior comparison of psychometric functions for the SPIN and WIN tests (Wilson et al., 2012) and pilot data, this individualized SNR was expected to produce an accuracy level of approximately 50% correct on low-predictability SPIN sentences for each participant. This test took approximately 10 min to complete.

Working Memory. The sentence-span subtest from a MATLAB-based working memory test battery developed by Lewandowsky et al. (2010) was administered. The "easy" version of the sentence-span task was used for this study. In this task, subjects were presented with an alternating sequence of simple sentences (3–6 words in length) and single letters on the computer screen. Subjects judged whether the sentence was true or false on each presentation. Following the true–false response, a letter was presented. After between four and eight sentence/letter presentations, subjects were asked to recall the letters in the order they were presented. The test consisted of 15 trials (after three practice trials) with 3 trials for each number of sentence/letter presentations. No feedback was provided. The working memory score

was calculated as the proportion of items recalled correctly (Conway et al., 2005). This test took approximately 30 min to complete.

Vocabulary. The Shipley Institute of Living Scale is a vocabulary test including 40 progressively more difficult test words (Shipley, 1940; Zachary & Shipley, 1986). For each test word, participants choose one out of four possible synonyms. A computerized version of the test was administered. On each trial, the test word was presented in large font with the four alternatives presented below the test word as response buttons in a graphical user interface. Participants touched the appropriate button on the computer touch screen to make their response. The vocabulary score was calculated as the percent correct trials. This test took approximately 10 min to complete.

Dual-Task Memory Load Experiment. On each trial, a set of digits was presented on a computer screen and followed by a spoken sentence. Participants were required to hold in memory the visually presented digits, listen to the spoken sentence, and then report the sentence-final word and the digits. The sentence stimuli were taken from the SPIN-R, in which the sentence-final word of each sentence is either predictable (e.g., "Jane swept the floor with a broom") or unpredictable (e.g., "Jane did not discuss the broom"). The standard (original) recordings of the SPIN-R sentences by a male talker and multitalker babble were used (Bilger et al., 1984; Kalikow et al., 1977). The digit strings were randomly selected on each trial from a set of digit strings. The set contained all possible combinations of the digits 1 to 9 with a set size of three (low load) or six (high load), with no repetitions and no forward consecutive sequences (e.g, "1 2"). Each trial proceeded as follows. A string of three digits (low load) or six digits (high load) was presented on a computer screen in large font, remained on the screen for one second, and was followed by a one second interstimulus interval, after which a spoken sentence was presented. Following each sentence was another one second interstimulus interval, after which participants were prompted to type the sentence-final word. Immediately following the sentence-final word response, participants were prompted to type into a response box the digits that were presented at the beginning of the trial. Word responses were scored as correct if the typed response was an exact match to the target word. For example, if "broom" was the correct word, "brum" or "broon" would be scored as incorrect. Digit responses were scored as correct only if all digits were reported in the correct order. Response time to words was measured from the appearance of the word response text box to when the participant pressed ENTER to report their response. Response time to digits was

recorded from the appearance of the digit response text box to when the participant pressed ENTER to report their response. Participants were not informed that response times would be measured to avoid encouraging participants to speed-type, which could compromise response accuracy and induce speed-accuracy trade-offs.

Prior to beginning the test trials, 20 practice trials were presented. Practice trials were used both to familiarize participants with the task and to optimize the initial SNR to yield approximately 50% correct on low-predictability SPIN sentences. All sentences presented during the practice were low-predictability SPIN sentences from a different set than that used in the test trials. The initial SNR for each participant was adjusted for the test trials based on average performance on the practice trials. Specifically, if the mean word identification score on the practice trials was greater than 65%, the SNR was reduced by 3 dB, and if the mean score was less than 35%, the SNR was increased by 3 dB. This procedure resulted in a median performance on the low-predictability sentences in the experimental trials of 62.22% (range 30.00–78.89).

A total of 180 test sentences were presented, with 45 items per combination of sentence predictability (predictable, unpredictable) and memory load (high, low). A set of four counterbalanced lists were used such that across participants, each word was presented in a high- and a low-predictability sentence, and within each level of predictability, each word was presented with both a high- and low-digit preload. As such, across participants, the conditions were equivalent in the identity of the words presented and in the number of characters to be typed for each condition. Order of presentation of items within a list was randomized. The experiment lasted on average 60 min.

Equipment. Both sessions were conducted in a sound-treated booth. All presentation parameters including SNR, sound levels, and randomization were controlled through custom MATLAB (MathWorks) programs. Auditory stimuli were presented at an overall level of 68 dB sound pressure level using a computer interfaced to Tucker Davis Technologies System 3 hardware (RP2

16-bit D/A converter, HB7 headphone buffer) and routed through ER-3A insert earphones (E.A.R. Corporation). Both earphones were inserted during testing, with stimuli presented to the right ear only.

Results

Mean accuracy and response time for word and digit responses are shown in Table 1 and Figure 1. Accuracy and response time were analyzed using mixed-effects models with the lme4 package (version 1.1.23; Bates et al., 2014) in R version 4.0.3 (R Development Core Team, 2013). A mixed-effects analysis was chosen in part because such models are appropriate whether or not data are unbalanced, that is, having unequal numbers of items across conditions, and would thus be appropriate for excluding trials on which words were responded to incorrectly (Baayen et al., 2008). Analysis of word and digit response accuracy used generalized linear mixed models with a binomial distribution and logit link function. Analysis of word and digit response times used linear mixed models with a log transform of the data to approximate a Gaussian distribution and the lmerTest package (Kuznetsova et al., 2017) to provide p values. Fixed factors included the within-subjects factors of predictability and load. All contrasts for the fixed factors with two levels assessed the difference between the two levels of each factor (coded as –0.5, 0.5). For both factors, the level that was expected to impair performance (i.e., low predictability, high load) was coded as –0.5.

Random factors justified by the design included within-subjects and within-items factors of predictability and load (here, "items" refers to the sentence-final words). All models were initially fit with the maximal random-effects structure (Barr et al., 2013). However, the maximal models failed to converge, suggesting that the models were overparameterized. Thus, simpler models were run by removing any random slope that had a proportion of variance equal or close to zero as revealed by a principal components analysis of the random-effects variance-covariance estimates from a

Table 1. Dual-Task Accuracy and Response Time.

		Predictable		Unpredictable	
		High load	Low load	High load	Low load
Word identification	Accuracy	87.73 (35.86)	89.47 (34.48)	59.32 (55.13)	62.13 (54.82)
	RT (sec)	2.72 (1.31)	2.42 (1.14)	2.87 (1.41)	2.66 (1.31)
Digit recall	Accuracy	37.71 (50.01)	82.40 (43.00)	35.18 (48.54)	77.66 (46.93)
	RT (sec)	7.85 (2.97)	5.15 (1.96)	7.92 (2.82)	5.47 (2.35)

Note Mean and standard deviation for word identification and digit recall accuracy and response time for each level of predictability and memory load are shown. Italicized are the standard deviations.
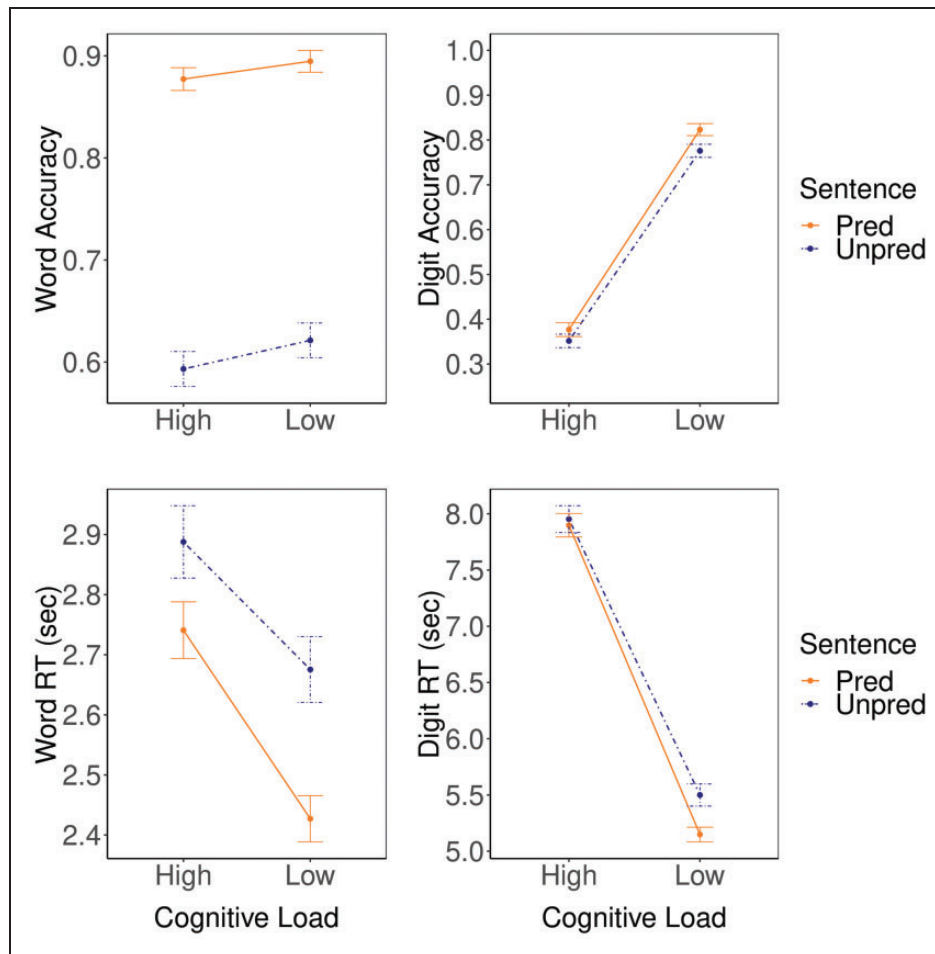RT = response time; sec = seconds.

Figure 1. Dual-Task Accuracy and Response Time. Mean word identification and digit recall accuracy and response times for each level of predictability and memory load are shown. Error bars show ± 1 SE, where SE is scaled to represent within-subjects variance for the repeated-measures design (Cousineau, 2005).
Pred = predictable; Unpred = unpredictable; RT = response time.

fitted mixed-effects model until convergence was achieved and the resulting model was not singular (Bates et al., 2015).

Word Identification Accuracy

For word accuracy, the random-effects structure included by-subjects and by-item random intercepts. A significant main effect of predictability (beta = −2.02, $SE = 0.10$, $z = −20.82$, $p < .001$) confirmed that words were identified more accurately in predictable than unpredictable sentences. The main effect of load was not significant (beta = 0.17, $SE = 0.09$, $z = 1.92$, $p = .055$), and the interaction of predictability and load was not significant ($z = 0.29$, $p = .77$).

Word Identification Response Time

The pattern of significant effects for word response time was the same whether trials on which words were

recognized incorrectly were included or excluded. Therefore, only the model that included only correct word responses is reported later. Response times for incorrect word responses ($n = 1,049$) were excluded from the model, leaving a total of 3,091 observations, and another 44 observations ($<2\%$) that were more than 3 standard deviations from the mean were replaced with the cutoff response time. The random-effects structure included by-subjects and by-item random intercepts, and by-subjects random slopes for predictability. Words were responded to more quickly in predictable than unpredictable sentences (beta = −0.07, $SE = 0.01$, $t = −5.05$, $p < .001$). Words were also responded to more quickly on low-load trials than high-load trials (beta = −0.09, $SE = 0.01$, $t = −7.09$, $p < .001$), indicating that reduced cognitive spare capacity when cognitive load was low had an impact on word processing. The interaction of predictability and load was not significant ($t = −1.33$, $p = .18$).

Digit Recall Accuracy

There were two analyses for digit recall responses, either including or excluding trials on which the word response was incorrect. In the following, results are reported for the model that included trials with incorrect word responses, and where the models differed, results are also reported for the model that excluded these trials. From both models, responses that were not digits (e.g., "?") were excluded, leaving a total of 3,089 observations in the model without trials with incorrect word responses, and a total of 4,137 observations in the model with incorrect word responses. For both models, the random-effects structure included by-subjects and by-item random intercepts and by-subject random slopes for load. Digits were recalled more accurately on trials in which the spoken sentence was predictable (beta = 0.24, $SE = 0.08$, $z = 3.01$, $p < .01$), indicating that cognitive spare capacity was greater on trials in which the spoken sentence was predictable. However, in the model of digit accuracy that included only trials on which the word was responded to correctly, the effect of predictability was not significant ($z = -1.13$, $p = .26$), indicating that the effect of predictability on digit recall accuracy largely reflected trials on which participants made an error in responding to the word. Unsurprisingly, digit sequences were recalled more accurately under low load, that is, on trials for which there were fewer digits to remember (beta = 2.29, $SE = 0.17$, $z = 13.85$, $p < .001$).

Digit Recall Response Time

Results are reported for the model that included trials with incorrect word responses, and where the models differed, results are also reported for the model that excluded these trials. . In the model that included trials with incorrect word responses, 73 observations ($<2\%$) that were more than 3 standard deviations from the mean were replaced with the cutoff response time. In the model that excluded these trials, another 59 observations ($<2\%$) that were more than 3 standard deviations from the mean were replaced with the cutoff response time. For both models, the random-effects structure included by-subjects and by-item random intercepts, and by-subjects random slopes for load. Digits were recalled more quickly on trials in which the spoken sentence was predictable (beta = -0.05, $SE = 0.01$, $t = -5.81$, $p < .001$), indicating that cognitive spare capacity was greater on trials in which the spoken sentence was predictable. Unsurprisingly, digits were also recalled more quickly under low load, that is, on trials for which there were fewer digits to remember (beta = -0.39, $SE = 0.02$, $t = -18.21$, $p < .001$), presumably reflecting a longer time needed to both recall and type longer digit sequences. There was also an interaction of load and predictability (beta = -0.05, $SE = 0.02$, $t = -2.71$, $p < .01$), which appears to reflect an effect of predictability on digit response time on low-load trials that was not present on high-load trials (see Figure 1). This interaction was not significant, however, in the model that did not include trials with incorrect word responses (beta = -0.04, $SE = 0.02$, $t = -1.92$, $p = .055$).

Correlations of Context Benefit With Working Memory and Vocabulary

Pearson pairwise correlations were analyzed for working memory span with the context benefit in both accuracy and response time for both word and digit responses. For each dependent variable, context benefit was calculated for each participant as the mean difference in performance across high- and low-predictability conditions. Partial pairwise correlations controlling for baseline performance in the low-predictability condition were also analyzed. None of these correlations was significant (all $r < .26$; all $p > .24$). Pairwise correlations and partial correlations of vocabulary with context benefit in word and digit response accuracy and response time were also not significant (all $r < .4$; all $p > .11$).

Discussion

The current study examined the effects of varying both sentence predictability and cognitive load on word identification and digit recall performance in a dual-task memory load design. With respect to effects of sentence predictability on listening effort, goals of the current study were to understand if the downstream effects of sentence predictability on digit recall that have been shown in prior studies extend to scenarios in which speech is degraded by background noise and digit recall is analyzed only for trials with correct word responses. Unlike these prior studies, both accuracy and reaction time in the word recognition and digit recall tasks were examined.

Words were identified more quickly and accurately in predictable than unpredictable sentences. Both effects were in the direction of better performance with predictable sentences, indicating that the predictability benefits do not owe to a speed-accuracy trade-off. Further, the finding that words from predictable sentences were identified more quickly when only correct trials were analyzed suggests that listening effort was reduced for predictable sentences even when intelligibility was matched at ceiling. However, note that word identification response time was an exploratory measure of listening effort in the present study.

Digit recall was both more quick and accurate on trials in which sentences were predictable. Thus, there

was no evidence for a speed-accuracy trade-off in the digit responses. More accurate recall of digits on trials in which an unrelated spoken sentence was predictable indicates that cognitive spare capacity was greater on predictable-sentence trials, leaving more cognitive resources available for rehearsing and remembering the digits. However, when analysis of the digit recall task was restricted to only trials in which the word was correctly recognized, sentence predictability effects remained for digit recall response time but were no longer significant for digit recall accuracy. Thus, it is possible that cognitive processes that are specifically elicited on trials with incorrect word responses were responsible for the effects of sentence predictability on recall accuracy. Of course, this finding does not establish that a distinct cognitive process that is elicited on incorrect word trials but not on correct word trials did underlie the effect of sentence predictability on digit recall accuracy, but it does fail to rule out the possibility. Alternatively, excluding incorrect word trials may simply have removed the most challenging and effortful trials from analysis, reducing the impact on digit recall accuracy of the same set of cognitive processes that are involved on trials with both correct and incorrect word responses.

What cognitive process might be specifically elicited on incorrect word trials and potentially responsible for the effect of sentence predictability on digit recall accuracy? A candidate process is error monitoring of the word response. According to dual-process theories, control of attention is accomplished by two main processes: the maintenance of task goals, subserved by prefrontal cortex, and the detection of errors that conflict with those goals, subserved by anterior cingulate cortex (Braver et al., 2007). The construct of listening effort includes many mental processes under the umbrella of those processes that demand central and capacity-limited cognitive resources (McGarrigle et al., 2014; Pichora-Fuller et al., 2016). It would be consistent with current theories of attentional control and working memory to include error monitoring as one of these processes (Kane & Engle, 2002; Miller et al., 2012). Nevertheless, the extent to which error monitoring is involved in listening effort during everyday spoken word recognition may be less than in a laboratory experiment. Thus, the fact that the effect of predictability on digit recall accuracy was not observed when incorrect word trials were excluded indicates the potential for limited ecological validity of this finding.

In contrast, the effect of predictability on digit recall response time was robust to exclusion of incorrect word trials. Responses were faster on trials in which the spoken sentence was predictable, particularly on low-load trials, even when only trials on which the sentence-final word was responded to correctly were

analyzed. These results indicate that cognitive spare capacity was greater on trials in which sentences were predictable and that cognitive processes elicited by incorrect word responses, such as error monitoring, are not fully responsible for the effect. Interestingly, the effect of predictability on digit recall response time was diminished on high (memory) load trials in the model that included both correct and incorrect word identification trials. A similar interaction of memory load and predictability, in which the downstream effects of sentence predictability on digit recall performance were greater for short compared to long sequences of to-be-recalled digits, was also observed by Hunter and Pisoni (2018), albeit in digit accuracy rather than response time. It may be that on high-load trials, the amount of cognitive resources needed to produce a measurable difference in digit recall performance exceeds the magnitude of the release of cognitive spare capacity by sentence predictability. An implication may be that sentence predictability is protective of cognitive spare capacity when other mental demands create low-to-moderate cognitive load, but less protective when cognitive load is very high. Overall, the finding that digit recall response times were faster on predictable trials, whether or not incorrect word trials were included, indicates that listening effort was reduced on trials in which sentences were predictable, independent of speech intelligibility. However, given the exploratory status of this dependent variable as a measure of listening effort, this result should be considered preliminary.

Taken together, the current results for digit recall accuracy and word and digit response times indicate that listening effort is reduced for predictable compared to unpredictable sentences presented in background noise. As such, the findings replicate and extend prior work with the dual-task memory load design that had used spectral vocoding to degrade spoken sentences, providing evidence that decreased listening effort for processing predictable compared to unpredictable sentences presented in background noise also results in a release of cognitive resources that can then be used to support memory processes. In terms of the cognitive processes underlying context benefit in speech perception, reduced listening effort when sentences are predictable supports the idea that high-context sentences improve word recognition accuracy through a relatively automatic, effortless process. As such, the results are consistent with long-standing proposals that associative spreading activation to semantically related words underlies at least some forms of linguistic prediction (for a review, see Huettig, 2015). More specifically, the current results suggest that a mechanism for the context benefit in spoken word recognition in adverse listening conditions is a relatively automatic process such as spreading activation in lexical networks. In the

framework of the Ease of Language Understanding (ELU) model, which aims to specify the role of working memory in language understanding, working memory resources are recruited when rapid and automatic speech recognition processes fail to match an incoming speech signal with phonetic representations in long-term memory (Rönnberg et al., 2013). In the ELU framework, rapid and automatic speech recognition processes can be preset by factors such as regional accent or semantic context. The current results appear to be in line with the ELU model, in that by preactivating lexical representations, sentence context may serve to preset the rapid automatic matching process to recognize phonetic representations of words that fit with the context. However, it should be noted that the current results demonstrate only that cognitive spare capacity was relatively greater in predictable sentences than unpredictable sentences. It remains possible that an effortful, controlled process was involved in context benefit but to a lesser extent than was required to recognize low-predictability sentences. Together with prior findings from the dual-task memory load design, in which spoken sentences were spectrally degraded (Hunter & Pisoni, 2018), the current findings with noise-degraded sentences support the idea that a relatively effortless route to prediction underlies context benefit in adverse listening conditions.

The view that sentence context reduces listening effort is further supported by the finding that working memory capacity was not correlated with the context benefit in either word identification accuracy or any of the measures of listening effort. Given that previous studies with young adults with normal hearing have observed a relation of working memory with context benefit in intelligibility (Zekveld et al., 2012, 2013) as well as with delayed sentence recognition accuracy (Zekveld et al., 2013), what can account for the difference in findings across studies? The current and prior studies are comparable in using background noise rather than vocoding to degrade the spoken stimuli. It should be noted that the current study may be underpowered to detect correlations with approximately 20 participants, although this sample size is comparable with the prior studies. One potentially important difference is the type of noise used to degrade the stimuli. The current study used multitalker babble, whereas prior studies used either stationary noise or a single-talker masker. Although Zekveld et al. (2012) observed a correlation of working memory and context benefit in intelligibility with a stationary noise masker, Zekveld et al. (2013) found that the correlation was absent for noise maskers and present only for a single-talker masker, suggesting that informational masking may mediate the relation with working memory. Given that the amount of informational masking in multitalker babble is greater than in stationary noise but less than for a single-talker masker, it may be that a relation of context benefit with working memory would have been observed in the current experiment if a single-talker masker had been used instead of multitalker babble.

Another factor is that the context cue provided in the studies by Zekveld et al. was not a sentence context as in the current study, but rather a semantically related text cue or prime presented prior to each sentence. Importantly, storing a text cue in memory while processing a spoken sentence may require a different use of working memory resources than processing a spoken sentence context as it unfolds. For example, participants might accomplish this task by storing the text cue in short-term memory while processing the incoming spoken sentence, which would necessarily involve working memory. As such, context benefit in these studies may have involved working memory to a greater extent than would sentential context.

In addition, the current study examined recognition only for sentence-final words, whereas prior studies that found associations of context benefit and working memory in young adults with normal hearing (Zekveld et al., 2012, 2013), older adults (Janse & Jesse, 2014), and postlingually deaf adult cochlear implant users (Dingemanse & Goedegebure, 2019) examined recognition for words throughout the sentence. This is potentially a key difference in terms of the association of working memory with context benefit, because context that follows a keyword has been argued to require more working memory resources than context that precedes a keyword (Wingfield et al., 1994). Thus, the current focus on sentence-final words may have sampled a type of context benefit that is less taxing on working memory than the benefit for keywords that are followed by (rather than preceded by) disambiguating context. To understand how predictability affects cognitive spare capacity throughout a sentence, further work will be needed to compare dual-task performance and correlations of working memory with context benefit for words preceding and following sentential context.

Finally, given that prior research has shown a weak or absent relation of working memory capacity to overall speech intelligibility and listening effort for young adults with normal hearing (Brown & Strand, 2018; Füllgrabe & Rosen, 2016), context benefit in this population may be less likely to correlate with working memory capacity than in other groups in which working memory capacity has been associated with overall speech intelligibility, such as older adults with hearing loss (Akeroyd, 2008). Further work will be needed to understand the relation of working memory capacity to context benefit in both intelligibility and listening effort in such populations.

The final aim of the current study was to examine the effects of cognitive load on sentence-final word recognition. Prior work has found that sentence-final words in spectrally degraded spoken sentences are recognized less accurately under high cognitive load (Hunter & Pisoni, 2018). In addition, several studies have shown that performance on a variety of acoustic-phonetic discrimination tasks decreases when cognitive load is high (Mattys et al., 2009, 2014; Mattys & Palmer, 2015; Mattys & Wiget, 2011). Such experimental results are in line with evidence that overall speech perception accuracy in adverse listening conditions is correlated with working memory capacity (for a review, see Akeroyd, 2008; but see Füllgrabe & Rosen, 2016 for evidence that the correlation is not present in young adult listeners). That is, both prior experimental findings of effects of cognitive load on speech perception and correlations of working memory with speech perception indicate that cognitive capacity is required to perceive speech in adverse listening conditions. Here, cognitive load did not have a significant effect on word identification accuracy, although there was a numerical difference of the means in the expected direction. However, word response times were significantly slower under high cognitive load. This effect of cognitive (memory) load on word response time suggests that reduced cognitive spare capacity under high load impacted perceiving and/or reporting the sentence-final word, in accord with prior experimental findings of decreased speech perception performance under cognitive load (Hunter & Pisoni, 2018; Mattys et al., 2009, 2014; Mattys & Palmer, 2015; Mattys & Wiget, 2011) and slower eye movements to spoken words under cognitive load (Hadar et al., 2016; Nitsan et al., 2019) in young adults.

In conclusion, the current study provides evidence that the benefit of a predictive sentence context for listeners is not only a boost in word identification accuracy but also an increase in available cognitive capacity. Digit recall accuracy was higher, and response time for both words and digits was faster on trials on which the spoken sentence was predictable, indicating that sentence context reduces listening effort. The effect of sentence predictability on word identification and digit recall response times, but not digit recall accuracy, was robust to exclusion of trials on which word identification was incorrect. As such, the effect of predictability on secondary task response time was independent of speech intelligibility. However, given that the response time measures have not been independently validated by prior studies as measures of listening effort, these findings should be considered exploratory. In accord with the experimental dual-task findings, working memory was not correlated with context benefit in either word identification or digit recall, consistent with the idea that processing predictable sentences does not require more

working memory resources than processing unpredictable sentences. Finally, although word identification accuracy was not affected by cognitive load, word identification response times were slower under high cognitive load, in line with prior findings that speech perception in adverse listening conditions is impacted by cognitive load.

Data Accessibility Statement

Declaration of Conflicting Interests

Funding

ORCID iD

Cynthia R. Hunter 🔟 https://orcid.org/0000-0003-2695-9896

References

Akeroyd, M. A. (2008). Are individual differences in speech reception related to individual differences in cognitive ability? A survey of twenty experimental studies with normal and hearing-impaired adults. *International journal of audiology, 47*(sup2), S53–S71.

Alhanbali, S., Dawes, P., Lloyd, S., & Munro, K. J. (2017). Self-reported listening-related effort and fatigue in hearing-impaired adults. *Ear and Hearing, 38*(1), e39–e48. https://doi.org/10.1097/AUD.0000000000000361

Alhanbali, S., Dawes, P., Lloyd, S., & Munro, K. J. (2018). Hearing handicap and speech recognition correlate with self-reported listening effort and fatigue. *Ear and Hearing, 39*(3), 470–474. doi: 10.1097/AUD.0000000000000515

American National Standards Institute Standards. (2010). *American national standard specification for audiometers.* American National Standards Institute.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*(4), 390–412.

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.) Psychology of learning and motivation (Vol. 8, pp. 47–89). Elsevier.

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), 255–278.

Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). *Parsimonious mixed models*. ArXiv Preprint ArXiv:1506.04967.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). *Fitting linear mixed-effects models using lme4*. ArXiv Preprint ArXiv:1406.5823.

Bilger, R. C., Nuetzel, J. M., Rabinowitz, W. M., & Rzeczkowski, C. (1984). Standardization of a test of speech perception in noise. *Journal of Speech, Language, and Hearing Research*, *27*(1), 32–48.

Bogacz, R., Wagenmakers, E.-J., Forstmann, B. U., & Nieuwenhuis, S. (2010). The neural basis of the speed–accuracy tradeoff. *Trends in Neurosciences*, *33*(1), 10–16. https://doi.org/10.1016/j.tins.2009.09.002

Braver, T. S., Gray, J. R., & Burgess, G. C. (2007). Explaining the many varieties of working memory variation: Dual mechanisms of cognitive control. *Variation in Working Memory*, *75*, 106.

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*(5), 628–640.

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, *82*(6), 407.

Conway, A. R., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review*, *12*(5), 769–786.

Cousineau, D. (2005). Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson's method. *Tutorials in Quantitative Methods for Psychology*, *1*(1), 42–45.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Memory and Language*, *19*(4), 450.

Desjardins, J. L., & Doherty, K. A. (2013). Age-related changes in listening effort for various types of masker noises. *Ear and Hearing*, *34*(3), 261–272. doi: 10.1097/AUD.0b013e31826d0ba4

Desjardins, J. L., & Doherty, K. A. (2014). The effect of hearing aid noise reduction on listening effort in hearing-impaired adults. *Ear and Hearing*, *35*(6), 600–610. doi: 10.1097/AUD.0000000000000028

Dingemanse, J. G., & Goedegebure, A. (2019). The important role of contextual information in speech perception in cochlear implant users and its consequences in speech tests. *Trends in Hearing*, *23*, 2331216519838672. https://doi.org/10.1177/2331216519838672

Doherty, J. M., Belletier, C., Rhodes, S., Jaroslawska, A., Barrouillet, P., Camos, V., Cowan, N., Naveh-Benjamin, M., & Logie, R. H. (2019). Dual-task costs in working memory: An adversarial collaboration. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*(9), 1529.

Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science*, *11*(1), 19–23.

Feuerstein, J. F. (1992). Monaural versus binaural hearing: Ease of listening, word recognition, and attentional effort. *Ear and Hearing*, *13*(2), 80–86.

Francis, A. L., & Nusbaum, H. C. (2009). Effects of intelligibility on working memory demand for speech perception. *Attention, Perception, & Psychophysics*, *71*(6), 1360–1374.

Füllgrabe, C., & Rosen, S. (2016). On the (un) importance of working memory in speech-in-noise processing for listeners with normal hearing thresholds. *Frontiers in Psychology*, *7*, 1268. https://doi.org/10.3389/fpsyg.2016.01268

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

Hadar, B., Skrzypek, J. E., Wingfield, A., & Ben-David, B. M. (2016). Working memory load affects processing time in spoken word recognition: Evidence from eye-movements. *Frontiers in Neuroscience*, *10*, 221. https://doi.org/10.3389/fnins.2016.00221

Hornsby, B. W. (2013). The effects of hearing aid use on listening effort and mental fatigue associated with sustained speech processing demands. *Ear and Hearing*, *34*(5), 523–534.

Hornsby, B. W., & Kipp, A. M. (2016). Subjective ratings of fatigue and vigor in adults with hearing loss are driven by perceived hearing difficulties not degree of hearing loss. *Ear and Hearing*, *37*(1), e1–e10. doi: 10.1097/AUD.0000000000000203

Hornsby, B. W., Naylor, G., & Bess, F. H. (2016). A taxonomy of fatigue concepts and their relation to hearing loss. *Ear and Hearing*, *37*(Suppl 1), 136S. doi: 10.1097/AUD.0000000000000289

Houben, R., van Doorn-Bierman, M., & Dreschler, W. A. (2013). Using response time to speech as a measure for listening effort. *International Journal of Audiology*, *52*(11), 753–761.

Huettig, F. (2015). Four central questions about prediction in language processing. *Brain Research*, *1626*, 118–135.

Hunter, C. R. (2020). Tracking cognitive spare capacity during speech perception with EEG/ERP: Effects of cognitive load and sentence predictability. *Ear and Hearing*, *41*(5), 1144–1157. doi: 10.1097/AUD.0000000000000856

Hunter, C. R., & Pisoni, D. B. (2018). Extrinsic cognitive load impairs spoken word recognition in high-and low-predictability sentences. *Ear and Hearing*, *39*(2), 378–389. doi: 10.1097/AUD.0000000000000493

Janse, E., & Jesse, A. (2014). Working memory affects older adults' use of context in spoken-word recognition. *The Quarterly Journal of Experimental Psychology*, *67*(9), 1842–1862.

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.

Kahneman, D. (1973). *Attention and effort* (Vol. 1063). Citeseer.

Kalikow, D. N., Stevens, K. N., & Elliott, L. L. (1977). Development of a test of speech intelligibility in noise

using sentence materials with controlled word predictability. *The Journal of the Acoustical Society of America*, 61(5), 1337–1351.

Kane, M. J., & Engle, R. W. (2002). The role of prefrontal cortex in working-memory capacity, executive attention, and general fluid intelligence: An individual-differences perspective. *Psychonomic Bulletin & Review*, 9(4), 637–671.

Kidd, G., Mason, C. R., Richards, V. M., Gallun, F. J., & Durlach, N. I. (2008). Informational masking. In W. A. Yost, A. N. Popper, & R. R. Fay (Eds.). *Auditory perception of sound sources* (pp. 143–189). Springer.

Kuperberg, G. R. (2007). Neural mechanisms of language comprehension: Challenges to syntax. *Brain Research*, 1146, 23–49.

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

Lewandowsky, S., Oberauer, K., Yang, L.-X., & Ecker, U. K. H. (2010). A working memory test battery for MATLAB. *Behavior Research Methods*, 42(2), 571–585. https://doi.org/10.3758/BRM.42.2.571

Luce, P. A., Feustel, T. C., & Pisoni, D. B. (1983). Capacity demands in short-term memory for synthetic and natural speech. *Human Factors*, 25(1), 17–32.

Mattys, S. L., Barden, K., & Samuel, A. G. (2014). Extrinsic cognitive load impairs low-level speech perception. *Psychonomic Bulletin & Review*, 21(3), 748–754.

Mattys, S. L., Brooks, J., & Cooke, M. (2009). Recognizing speech under a processing load: Dissociating energetic from informational factors. *Cognitive Psychology*, 59(3), 203–243.

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.

Mattys, S. L., & Palmer, S. D. (2015). Divided attention disrupts perceptual encoding during speech recognition. *The Journal of the Acoustical Society of America*, 137(3), 1464–1472.

Mattys, S. L., & Wiget, L. (2011). Effects of cognitive load on speech recognition. *Journal of Memory and Language*, 65(2), 145–160.

McGarrigle, R., Munro, K. J., Dawes, P., Stewart, A. J., Moore, D. R., Barry, J. G., & Amitay, S. (2014). Listening effort and fatigue: What exactly are we measuring? A British Society of Audiology Cognition in Hearing Special Interest Group 'white paper'. *International Journal of Audiology*, 53, 433–445.

Meister, H., Rählmann, S., Lemke, U., & Besser, J. (2018). Verbal response times as a potential indicator of cognitive load during conventional speech audiometry with matrix sentences. *Trends in Hearing*, 22, 2331216518793255. https://doi.org/10.1177/2331216518793255

Miller, A. E., Watson, J. M., & Strayer, D. L. (2012). Individual differences in working memory capacity predict action monitoring and the error-related negativity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 38(3), 757.

Morey, C. C., & Cowan, N. (2004). When visual and verbal memories compete: Evidence of cross-domain limits in working memory. *Psychonomic Bulletin & Review*, 11(2), 296–301.

Neher, T., Grimm, G., & Hohmann, V. (2014). Perceptual consequences of different signal changes due to binaural noise reduction: Do hearing loss and working memory capacity play a role? *Ear and Hearing*, 35(5), e213–e227.

Nitsan, G., Wingfield, A., Lavie, L., & Ben-David, B. M. (2019). Differences in working memory capacity affect online spoken word recognition: Evidence from eye movements. *Trends in Hearing*, 23, 2331216519839624.

Obleser, J., & Kotz, S. A. (2009). Expectancy constraints in degraded speech modulate the language comprehension network. *Cerebral Cortex*, 20(3), 633–640.

Pals, C., Sarampalis, A., van Rijn, H., & Başkent, D. (2015). Validation of a simple response-time measure of listening effort. *The Journal of the Acoustical Society of America*, 138(3), EL187–EL192.

Pichora-Fuller, M. K. (2016). How social psychological factors may modulate auditory and cognitive functioning during listening. *Ear and Hearing*, 37, 92S–100S. https://doi.org/10.1097/AUD.0000000000000323

Pichora-Fuller, M. K., Kramer, S. E., Eckert, M. A., Edwards, B., Hornsby, B. W., Humes, L. E., Lemke, U., Lunner, T., Matthen, M., & Mackersie, C. L. (2016). Hearing impairment and cognitive energy: The framework for understanding effortful listening (FUEL). *Ear and Hearing*, 37, 5S–27S. doi: 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.

R Development Core Team. (2013). R: A language and environment for statistical computing.

Rakerd, B., Seitz, P., & Whearty, M. (1996). Assessing the cognitive demands of speech listening for people with hearing losses. *Ear and Hearing*, 17(2), 97–106.

Rönnberg, J., Lunner, T., Zekveld, A., Sörqvist, P., Danielsson, H., Lyxell, B., Dahlström, Ö., Signoret, C., Stenfelt, S., & Pichora-Fuller, M. K. (2013). The Ease of Language Understanding (ELU) model: Theoretical, empirical, and clinical advances. *Frontiers in Systems Neuroscience*, 7, 1–17. https://doi.org/10.3389/fnsys.2013.00031

Rosen, S., Souza, P., Ekelund, C., & Majeed, A. A. (2013). Listening to speech in a background of other talkers: Effects of talker number and noise vocoding. *The Journal of the Acoustical Society of America*, 133(4), 2431–2443.

Rudner, M. (2016). Cognitive spare capacity as an index of listening effort. *Ear and Hearing*, 37, 69S–76S.

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*, 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*, 58(6), 1781–1792.

Shipley, W. C. (1940). A self-administering scale for measuring intellectual impairment and deterioration. *The Journal of Psychology*, 9(2), 371–377.

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*, 61(6), 1463–1486.

Surprenant, A. M. (1999). The effect of noise on memory for spoken syllables. *International Journal of Psychology*, 34(5–6), 328–333.

Surprenant, A. M. (2007). Effects of noise on identification and serial recall of nonsense syllables in older and younger adults. *Aging, Neuropsychology, and Cognition*, 14(2), 126–143.

Wagenmakers, E.-J., Van Der Maas, H. L., & Grasman, R. P. (2007). An EZ-diffusion model for response time and accuracy. *Psychonomic Bulletin & Review*, 14(1), 3–22.

Wickelgren, W. A. (1977). Speed-accuracy tradeoff and information processing dynamics. *Acta Psychologica*, 41(1), 67–85.

Wilson, R. H. (2003). Development of a speech-in-multitalker-babble paradigm to assess word-recognition performance. *Journal of the American Academy of Audiology*, 14(9), 453–470.

Wilson, R. H., & McArdle, R. (2007). Intra-and inter-session test, retest reliability of the Words-in-Noise (WIN) test. *Journal of the American Academy of Audiology*, 18(10), 813–825.

Wilson, R. H., McArdle, R., Watts, K. L., & Smith, S. L. (2012). The Revised Speech Perception in Noise Test (R-SPIN) in a multiple signal-to-noise ratio paradigm. *Journal of the American Academy of Audiology,* 23(8), 590–605.

Wingfield, A., Alexander, A. H., & Cavigelli, S. (1994). Does memory constrain utilization of top-down information in spoken word recognition? Evidence from normal aging. *Language and Speech*, 37(3), 221–235.

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, 2331216516669723. https://doi.org/10.1177/2331216516669723

Wu, Y.-H., Stangl, E., Zhang, X., Perkins, J., & Eilers, E. (2016). Psychometric functions of dual-task paradigms for measuring listening effort. *Ear and Hearing*, 37(6), 660. doi: 10.1097/AUD.0000000000000335

Zachary, R. A., & Shipley, W. C. (1986). *Shipley institute of living scale: Revised manual*. WPS, Western Psychological Services.

Zekveld, A. A., Rudner, M., Johnsrude, I. S., Festen, J. M., Van Beek, J. H., & Rönnberg, J. (2011). The influence of semantically related and unrelated text cues on the intelligibility of sentences in noise. *Ear and Hearing*, 32(6), e16–e25.

Zekveld, A. A., Rudner, M., Johnsrude, I. S., Heslenfeld, D. J., & Rönnberg, J. (2012). Behavioral and fMRI evidence that cognitive ability modulates the effect of semantic context on speech intelligibility. *Brain and Language*, 122(2), 103–113.

Zekveld, A. A., Rudner, M., Johnsrude, I. S., & Rönnberg, J. (2013). The effects of working memory capacity and semantic cues on the intelligibility of speech in noise. *The Journal of the Acoustical Society of America*, 134(3), 2225–2234.