



Published in final edited form as:

*Ear Hear.* 2018 ; 39(2): 378–389. doi:10.1097/AUD.0000000000000493.

## Extrinsic Cognitive Load Impairs Spoken Word Recognition in High- and Low-Predictability Sentences

Cynthia R. Hunter<sup>1</sup> and David B. Pisoni<sup>1</sup>

<sup>1</sup>Speech Research Laboratory, Department of Psychological and Brain Sciences, Indiana University, Bloomington, Indiana

### Abstract

**Objectives**—Listening effort (LE) induced by speech degradation reduces performance on concurrent cognitive tasks. However, a converse effect of extrinsic cognitive load on recognition of spoken words in sentences has not been shown. The aims of the current study were to (a) examine the impact of extrinsic cognitive load on spoken word recognition in a sentence recognition task, and (b) determine whether cognitive load and/or listening effort needed to understand spectrally degraded speech would differentially affect word recognition in high- and low-predictability sentences. Downstream effects of speech degradation and sentence predictability on the cognitive load task were also examined.

**Design**—One hundred and twenty young adults identified sentence-final spoken words in high- and low-predictability SPIN sentences. Cognitive load consisted of a pre-load of short (low-load) or long (high-load) sequences of digits, presented visually prior to each spoken sentence and reported either before or after identification of the sentence-final word. Listening effort was varied by spectrally degrading sentences with 4-, 6-, or 8-channel noise vocoding. Level of spectral degradation and order of report (digits first or words first) were between-participants variables. Effects of cognitive load, sentence predictability, and speech degradation on accuracy of sentence-final word identification as well as recall of pre-load digit sequences were examined.

**Results**—In addition to anticipated main effects of sentence predictability and spectral degradation on word recognition, we found an effect of cognitive load, such that words were identified more accurately under low load than high load. However, load differentially affected word identification in high- and low-predictability sentences depending on the level of sentence degradation. Under severe spectral degradation (4-channel vocoding), the effect of cognitive load on word identification was present for high-predictability sentences but not for low-predictability sentences. Under mild spectral degradation (8-channel vocoding), the effect of load was present for low-predictability sentences but not for high-predictability sentences. There were also reliable downstream effects of speech degradation and sentence predictability on recall of the pre-load digit sequences. Long digit sequences were more easily recalled following spoken sentences that were less spectrally degraded. When digits were reported after identification of sentence-final words, short digit sequences were recalled more accurately when the spoken sentences were predictable.

**Conclusions**—Extrinsic cognitive load can impair recognition of spectrally degraded spoken words in a sentence recognition task. Cognitive load affected word identification in both high- and low-predictability sentences, suggesting that load may impact both context use and lower-level perceptual processes. Consistent with prior work, listening effort also had downstream effects on memory for visual digit sequences. Results support the proposal that extrinsic cognitive load and listening effort induced by signal degradation both draw on a central, limited pool of cognitive resources that is used in order to recognize spoken words in sentences under adverse listening conditions.

---

## Introduction

Understanding speech often seems to be an effortless process for listeners with normal hearing. However, when there is degradation of the speech signal, whether originating in the signal itself, from a noisy environment, or as a consequence of hearing impairment, speech recognition becomes more effortful and demands additional information processing resources to maintain comprehension (Rönnberg et al., 2008; Mattys et al., 2012; Lemke & Besser, 2016). Listening effort (LE) refers to the use of central, re-allocable cognitive resources in order to recognize and understand degraded speech (cf. Pichora-Fuller et al., 2016). Performance on secondary information-processing tasks is reduced by LE, indicating that both tasks draw from a shared, limited pool of cognitive resources (Broadbent, 1958; Kahneman, 1973; Baddeley & Hitch, 1974). For example, words that were correctly perceived in background noise are later remembered less well than words that were correctly perceived in quiet (Rabbitt, 1968, 1991; Pichora-Fuller et al., 1995; Surprenant, 1999; 2007). In dual-task experiments, participants typically perform a speech task simultaneously with a secondary task in which decisions are made about stimuli presented in another modality, such as visual shapes, numbers, or letters. Declines in performance on the secondary task, most typically slowed reaction times, occur with parametric degradations of the speech signals (Rönnberg et al., 2010; Pals et al., 2013). In a “pre-load” experiment, cognitive load is sequential rather than simultaneous: listeners are required to encode and remember sequences of items that are presented prior to each speech stimulus (Baddeley and Hitch, 1974). Recall of pre-load sequences is reduced when the subsequent speech is degraded, indicating that the listening effort expended to recognize degraded speech has downstream effects on the rehearsal of items in working memory (Luce et al., 1983; see also Sarampalis et al., 2009). Convergent measures of pupil size also indicate that listening effort increases with greater degrees of speech degradation (Winn et al., 2015).

The central pool of cognitive resources involved in effortful listening has been identified with working memory, a limited-capacity cognitive system that is responsible for the storage and active manipulation of information in memory (Baddeley & Hitch, 1974; Daneman & Carpenter, 1980, 1983; Just & Carpenter, 1992; Wingfield, 2016). According to the Framework for Understanding Effortful Listening (FUEL), success in recognizing and understanding degraded speech may depend on deployment of cognitive resources (Pichora-Fuller et al., 2016). The Ease of Language Understanding (ELU) model describes how the understanding of degraded speech may recruit central cognitive resources (Rönnberg, 2003; Rönnberg et al., 2008, 2010). In the ELU model, the matching of acoustic-phonetic information to lexical representations in long-term memory occurs implicitly under optimal

conditions, but engagement of central cognitive resources may be triggered if the automatic matching process fails due to mismatch of a degraded speech signal with stored representations. An association between working memory capacity and degraded speech recognition has been established in the literature, at least for older adult listeners with hearing loss (Gatehouse et al, 2003; Lunner, 2003; Lunner, & Sundewall-Thorén, 2007; Akeroyd, 2008; Arlinger et al., 2009; Rudner et al., 2011). Fewer studies have examined this association in young adults, and whether this association is present in young adult listeners with normal hearing has been questioned (see Fullgrabe & Rosen, 2016). In studies that have observed a relation between working memory capacity and degraded speech recognition in both young and older adults with normal hearing, the relation has been weaker in the young adult group (Gordon-Salant & Cole, 2016; Ward, Shen, Souza, et al., 2017).

In the majority of studies that have manipulated cognitive processing load, investigation has focused on listening effort induced by the intrinsic cognitive load created by a degraded speech signal. In these studies, speech degradation is manipulated and the effect of the intrinsic cognitive load created by signal degradation is measured from performance on secondary tasks, such as the speed of responding to simultaneously presented visual stimuli or the accuracy of memory for correctly perceived words (e.g., Surprenant, 2007; Rönnerberg et al., 2010; cf. Mattys et al., 2012). Mattys and colleagues have pointed out a lack of research on the converse impact on speech perception of an extrinsic cognitive load, defined as the cognitive demand posed by a secondary attentional or memory task (Mattys et al., 2009; Mattys & Wiget, 2011; Mattys et al., 2012, 2014). Mattys and colleagues have examined the effects of extrinsic cognitive load on speech perception, often using dual-task paradigms. Results of these experiments indicate that the ability to perceive acoustic-phonetic detail is impaired by extrinsic cognitive load. For example, Mattys et al. (2014) found that greater levels of cognitive load led to increases in phonemic restoration, which refers to the illusory perception of phonemes within a word or nonword that have been removed and replaced by noise. The increase in phonemic restoration implies that reliance on acoustic-phonetic detail decreased under cognitive load. A further analysis indicated that the effect of cognitive load was not modulated by whether stimuli were words or nonwords, providing converging evidence that the locus of the impact of cognitive load was at the acoustic-phonetic level rather than at the lexical level (see also Francis, 2010). Thus, the line of work developed by Mattys and colleagues provides not only a clear demonstration of the reciprocal relation between cognitive demand and speech perception, but also indicates that the locus of the impact of cognitive capacity on speech perception may be at a low level, acoustic-phonetic process.

However, given the nature of the tasks involved in research on the impacts of extrinsic cognitive load on speech perception to date, it is not clear whether extrinsic cognitive load affects spoken word recognition in everyday listening situations, or if so, where the locus of such effects would be in the information-processing system. Specifically, the tasks used in the prior studies measured the ability to discriminate subtle acoustic-phonetic differences using a closed set of response choices (Francis & Nusbaum, 2009; Francis, 2010; Mattys & Wiget, 2011; Mattys et al., 2012, 2014). In contrast, success in everyday speech recognition typically involves recognizing words in sentence contexts from a large, open set of alternatives. Current theories of spoken word recognition propose that word recognition

occurs by mapping portions of the speech signal to the best matching representation in the mental lexicon. In this process, there are typically many possible mappings between the signal and word candidates (for reviews, see Cleary & Pisoni, 2001; Luce & McLennan, 2005; Dahan & Magnuson, 2006; McQueen, 2007). In fact, closed-set tasks in which participants choose among a limited set of response alternatives are often insensitive to lexical variables that have a well-established role in spoken word recognition (Sommers et al., 1997; Clopper et al., 2006). Thus, there is a need to replicate the effects of extrinsic cognitive load using other tasks that more closely approximate the task demands of everyday listening. The first objective of the current study was to determine whether extrinsic cognitive load has an impact on the recognition of spoken words in a sentence recognition task.

Additionally, most work to date on the effect of extrinsic cognitive load on speech recognition has used dual-task paradigms in which the speech is presented simultaneously with visual stimuli and concurrent tasks are performed for the speech and visual stimuli. Although this design has been useful in characterizing the impact of listening effort on the allocation of cognitive resources, the effects observed in dual-task experiments can be attributed to shifts of attention between auditory and visual stimuli (cf. Mattys et al., 2014). Indeed, based on results from dual-task paradigms, it has been hypothesized that the mechanism by which extrinsic cognitive load affects speech perception is competition between visual and speech processing for limited attentional resources that are needed for stimulus encoding (Mitterer & Mattys, 2017). In the current study, extrinsic cognitive load was manipulated using a pre-load design, in which visual stimuli for the load task were presented prior to the speech stimuli and load was operationalized as the number of pre-load items that were to be held in memory during presentation of the speech. Given the present experimental design, effects of cognitive load on spoken word recognition cannot be attributed either to a closed-set task in which attention is focused on acoustic-phonetic detail or to attentional shifting between concurrently presented stimuli. Rather, any effect of load should reflect the allocation of central cognitive resources for spoken word recognition.

The second aim of this study was to investigate whether cognitive load and/or listening effort needed to understand speech differentially affect word recognition in high- and low-predictability sentences. Contextual facilitation refers to the more accurate recognition of words in predictable contexts than in unpredictable contexts or in isolation (Miller, Heise, & Lichten, 1951). Although it has been suggested that contextual facilitation may require central cognitive resources because it involves the maintenance of early parts of a sentence in memory and their integration with later parts, experimental evidence supporting this hypothesis has not been conclusive (Pichora-Fuller et al., 1995). However, several recent studies have reported correlations between working memory capacity and contextual facilitation. For example, Zekveld and colleagues found that larger working memory span was associated with the benefit obtained from a semantically-related text cue for sentences masked by interfering speech, although this association was not found for stationary or fluctuating noise maskers (Zekveld, Rudner, Johnstrude, et al., 2013). The authors suggested that the association between working memory and use of semantic cues emerged in the condition with the greatest informational, as opposed to energetic, masking because informational masking and the use of semantic cues both draw on attentional resources (see

also Koelewijn, Zekveld, Festen, et al. 2012). In another study, which used stationary noise as the masker, working memory was not associated with any benefit from a related cue but was related to reduced influence of misleading cues (Zekveld et al., 2011; see also Besser, Koelewijn, Zekveld, et al., 2013). Using a different design, Janse and Jesse (2014) found that working memory capacity, measured by a backwards digit span task, was associated with older adults' ability to benefit from context in a speeded phoneme-monitoring task. Specifically, digit span was related to facilitation of reaction time for detecting a word-initial target phoneme when the target phonemes were presented in high-predictability contexts. More recently, Gordon-Salant and Cole (2016) found that among older adults with normal hearing, those older adults who had high working memory capacity were better able than those with low working memory capacity to recognize degraded spoken words, both in isolation and in sentence contexts. Taken together, earlier studies of the relation between working memory and contextual facilitation suggest that individual differences in cognitive capacity may contribute to the ability to use context to recognize degraded speech signals. Yet, some studies have observed associations of word recognition with working memory in low-context sentences but not in high-context sentences (Meister, Schreitmüller, Ortmann, et al., 2016; Wayne, Hamilton, Huyck, et al., 2016), or for both lower- and higher-context speech stimuli (Benichov, Cox, Tun, et al., 2012; Gordon-Salant & Cole, 2016). Further research is therefore needed in order to identify the locus of effects of cognitive capacity on the recognition of words in spoken sentences.

In the current study, we investigated the effect of extrinsic cognitive load on the identification of degraded sentence-final words of high- and low-predictability sentences. If extrinsic cognitive load impacts word identification, then the effect of load on contextual facilitation can also be assessed. For high-predictability sentences, participants can identify sentence-final words by attending to the preceding context. Thus, an effect of extrinsic cognitive load on the identification of words in high-predictability sentences could be due to an impact of load on the use of an informative sentence context to “fill in the blanks” of sentence-final words, or on lower levels of perceptual processing such as the resolution of fine-grained acoustic-phonetic details of speech. In contrast, because the final words of low-predictability sentences can be identified by attending to acoustic-phonetic or lexical information but not by attending to context, effects of cognitive load on word recognition in low-predictability sentences must be due to an impact of load on information processing at a lower level than context use. Thus, the effect of cognitive load on the difference in performance for high- and low-predictability sentences can be used to index the impact of load on contextual facilitation. However, given that prior work has observed effects of cognitive load on acoustic-phonetic processing of speech (Mattys et al., 2009; Mattys & Wiget, 2011; Mattys et al., 2012, 2014), it will be important to assess the effect of cognitive load in low- and high-predictability conditions separately, rather than focusing only on the difference score of performance between low- and high-predictability conditions. For example, if cognitive load decreases accuracy in low-predictability sentences but has no effect in high-predictability sentences, the difference score of contextual facilitation would be greater under cognitive load, suggesting an effect of cognitive load on contextual facilitation, yet the effect would be entirely due to a change in the baseline, low-predictability sentences.

Contextual facilitation generally increases when signal quality is low (Boothroyd, & Nittrouer, 1988; Miller et al., 1951; Morton, 1969; Kalikow, Stevens, & Elliott, 1977; Obleser, Wise, Dresner, et al., 2007; Rönnerberg, 2003; Zekveld et al. 2011). For example, speech that is presented in background noise or spectrally degraded to approximate the signal from a cochlear implant is understood with far greater accuracy when a sentence context is present than when words are presented in isolation (Obleser et al. 2007; Sheldon, Pichora-Fuller, & Schneider, 2008). Similarly, word recognition accuracy is typically lower in older adults and individuals with hearing impairment than young adults or individuals with normal hearing, particularly when words are presented without supporting context. However, these age and hearing status differences are reduced or eliminated when words are presented in highly predictive sentence contexts (Gordon-Salant & Fitzgibbons, 1997; Pichora-Fuller, Schneider, & Daneman, 1995; Sommers & Danielson, 1999; Wingfield, 1996). In the current study, sentences were mildly, moderately, or highly spectrally degraded in order to determine whether the level of signal degradation would modulate effects of cognitive load in high- and low-predictability sentences.

Performance on the secondary task of memory for visually presented digits was also examined. As discussed above, in earlier studies of listening effort the downstream effects of speech degradation on secondary tasks have been examined more frequently than the converse effects of cognitive load on speech recognition (Luce et al., 1983; Pals et al., 2013; Rönnerberg, et al., 2010; Sarampalis et al., 2009). In the current study, both listening effort created by degradation of the speech stimuli and sentence predictability could potentially affect performance in the secondary task of digit memory. With respect to sentence predictability, prior studies have shown that a semantically-related context improves memory for words that are correctly perceived, suggesting that the greater ease of recognition provided by a semantic context has the downstream benefit of leaving more cognitive resources available for encoding the perceived words in memory (McCoy et al., 2005; Sarampalis et al., 2009). Results of the current study using a digit pre-load design to manipulate extrinsic cognitive load will be informative as to whether this downstream benefit of sentence context is present for visually presented stimuli that are unrelated to the sentence context itself.

Extrinsic cognitive load and sentence predictability were the major variables of interest in the current study, and effects of these variables both depend on a baseline level of performance. Effects of cognitive load reflect the difference between performance under high and low load, and effects of sentence predictability reflect the difference in performance between high- and low-predictability sentences. Thus, these variables were selected to be within-subjects factors. Level of spectral degradation of the spoken sentences was a between-subjects factor in the current study. Degradation was expected to have a large effect on baseline accuracy and was of interest to the extent that it would modulate the effects of load and predictability on sentence-final word identification or affect digit recall accuracy. Given that degradation was a between-subjects factor, effects of degradation may reflect the use of different listening strategies developed over the course of exposure to sentences at a given level of degradation. An additional between-subjects factor was order of report of the sentence-final word and digit sequence. The purpose of the variable of order of

report was to determine if effects of the other variables would depend on holding the to-be-reported word or digit sequence in memory during report of the first item.

## Materials and Methods

### Participants

A total of 120 young adults participated for payment or course completion credit in an introductory psychology course at Indiana University. Participants gave written informed consent to the protocol approved by the local Institutional Review Board. Participants reported having no history of a hearing or speech disorder. This study used six groups of participants, with 20 participants in each group. Sentence degradation (mild, moderate, and severe) and order of report (sentence-final word reported before or after the digits) were the between-subjects variables. Thus, there were 20 participants for each combination of order of report (word first, digit first) and sentence degradation (mild, 8-channel vocoded; moderate, 6-channel vocoded; and high, 4-channel vocoded).

### Stimuli

The speech materials were a subset of the sentences from the revised version of the Speech Perception in Noise (SPIN-R) test (Bilger, Nuetzel, Rabinowitz, & Rzeczkowski, 1984). The SPIN test was designed as a clinically useful assessment of both the ability to recognize speech based on bottom-up processing of the acoustic signal, and the ability to recognize speech based on top-down processing using linguistic knowledge (Kalikow et al., 1977). The target stimuli for the SPIN test are the final words of sentences, which are either predictable from the preceding context (e.g., “The girl swept the floor with a broom”) or are not easily predicted from the preceding context (e.g., “Ruth’s grandmother discussed the broom.”). For each word there is a high-predictability and a low-predictability sentence context. A total of 148 sentence-final words were selected. A male talker with a Midwestern accent recorded high- and low-predictability sentences for each sentence-final word. Stimuli were recorded in a sound-attenuating booth using a free-field microphone. Each sentence was digitally spliced into a separate .wav file and normalized to a root-mean-square amplitude of 68 dB SPL.

Spectral degradation was accomplished by noiseband vocoding using Tiger CIS (<http://www.tigerspeech.com>). Noise vocoding involved an analysis phase, which divides the signal into frequency bands and derives the amplitude envelope from each band, and a synthesis phase, which replaces the frequency content of each band with noise that is modulated with the appropriate amplitude envelope. Stimuli were bandpass filtered into 4, 6, or 8 spectral channels between 200 and 7000 Hz using Greenwood’s filter function (24 dB/octave slope). The number of spectral bands used for each level of degradation (mild, 8-channels; moderate, 6-channels; and severe, 4-channels) was chosen based on prior, unpublished experiments from our laboratory. The temporal envelope of each channel was then derived using a low pass filter with an upper cutoff at 160 Hz with a 24 dB/octave slope. In the synthesis phase, the spectral information in each channel was replaced with band-pass noise that was modulated by the corresponding temporal envelope.

The visual stimuli for the digit pre-load task were strings of either three (low-load) or seven (high-load) digits in 36-point font. Digit strings were randomly selected on each trial from a set of digit strings. The set contained all possible combinations of the digits 1 through 9 with a set size of three (low-load) or seven (high-load), with no repetitions and no forward consecutive sequences (e.g., “1 2” did not occur in any of the digit sequences, although “1 3” and “2 1” did occur).

## Procedure

Data collection used a custom script written for AppleScript, and implemented on four Apple PowerMac G4 computers. Audio signals were presented over calibrated Beyer Dynamic DT-100 headphones. Prior to the experimental trials, participants received a pre-practice familiarization with vocoded sentences as well as a block of practice trials. In the pre-practice block, ten SPIN-R sentences that were degraded to the same number of vocoded channels as the stimuli in the experimental block were presented. Following each spoken sentence, its written version was displayed on the computer screen for one second and was immediately followed by an on-screen response box in which participants were asked to type the final word of each sentence. The practice block had the same structure as the experimental block and used a set of ten additional SPIN sentences. None of the items for the experimental trials were used in the practice or pre-practice. On-screen instructions preceded each block to orient participants to the materials and requirements of the upcoming task.

On each trial, a digit string (either high- or low-load) was displayed, centered on a computer monitor for one second, and was immediately followed by a spoken sentence. After an inter-stimulus-interval (ISI) that was randomly drawn from the set of 1.0, 1.25, or 1.5 seconds, the first response box appeared on the screen. Immediately after the participant entered their response, the second response box appeared. Participants either typed the sentence-final word in the first response box and the digits in the second response box, or vice versa, depending on the order-of-report condition. Each trial was followed by an inter-trial-interval (ITI) that was randomly drawn from the set of 1.75, 2.0, or 2.25 seconds. The ISI and ITI were randomly jittered in order to match with the design of a parallel study of event-related potentials that is currently in progress. Prior to beginning the practice and experimental blocks, participants were given verbal and written instructions on the order of report of sentence-final words and digits. A total of 148 trials were presented to each participant, consisting of 37 trials for each combination of sentence type (high- or low-predictability) and level of digit load (high- or low-load). 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 pre-load. Order of presentation of items within a list was randomized. The experiment lasted on average 45 minutes.

## Results

Accuracy for sentence-final word recognition and digit recall was analyzed using mixed-model ANOVAs with between-subjects factors of degradation (mild, moderate, high) and



order of report (words first, digits first), and within-subjects factors of digit sequence (high load, low load) and sentence predictability (high, low).

### Word Accuracy

Accuracy for sentence-final word identification is shown in Figure 1A. In the figure, accuracy is collapsed across order of report, because order of report did not produce any significant main effects or interactions for word identification accuracy. Figure 1A indicates large, expected effects of spectral degradation and sentence predictability on accuracy. Sentence-final words were identified more accurately in sentences that were less degraded and in sentences that were high-predictability rather than low-predictability. Statistical analysis confirmed that sentence-final words were recognized more accurately with less spectral degradation (main effect of degradation on word accuracy,  $F(2,114) = 431.64$ ,  $MSE = .02$ ,  $p < .001$ ,  $\eta^2_p = .88$ ) and when sentence contexts were more predictable (main effect of predictability,  $F(1,114) = 891.27$ ,  $MSE = .01$ ,  $p < .001$ ,  $\eta^2_p = .89$ ). Figure 1B illustrates contextual facilitation, the increase in accuracy for high-predictability compared to low-predictability sentences. Contextual facilitation was similar across levels of degradation and load. Contextual facilitation was approximately 23 percent on average across conditions, but was below this average value in two conditions: under high cognitive load and severe spectral degradation (4 Ch) and under low cognitive load and mild spectral degradation (8 Ch). Figures 1A and 1C show the effects of extrinsic cognitive load, wherein sentence-final words were identified more accurately when cognitive load was low. Figure 1C shows the cognitive load effect, the difference in accuracy across high and low load conditions, as a function of predictability and spectral degradation. In Figure 1C, negative difference scores indicate lower accuracy under high than low cognitive load. The effect of cognitive load appears to depend on sentence predictability as well as level of spectral degradation: from Figure 1C, under severe degradation (4 Ch) the effect of load appears differentially in the high-predictability sentences, whereas under mild degradation (8 Ch), the effect of load appears differentially in the low-predictability sentences. This pattern can also be seen in Figure 1A, in which the effects of cognitive load appear in the high-predictability condition for the severe degradation (4 Ch) condition, but in the low-predictability condition when sentences were mildly degraded (8 Ch), and appear similar across high- and low-predictability conditions for moderate degradation (6 Ch). These observations were supported statistically by a main effect of load,  $F(1,114) = 16.05$ ,  $MSE = .004$ ,  $p < .001$ ,  $\eta^2_p = .12$ , indicating that sentence-final words were identified more accurately under low than high extrinsic cognitive load, and a three-way interaction of predictability, load, and degradation,  $F(2,114) = 5.27$ ,  $MSE = .004$ ,  $p < .01$ ,  $\eta^2_p = .09$ , confirming that the effect of cognitive load varied across levels of predictability and spectral degradation. The three-way interaction was assessed with separate ANOVAs for each level of spectral degradation.

In the severe (4 Ch) degradation condition, sentence-final words were identified more accurately in high-predictability sentences (main effect of predictability ( $F(1,38) = 169.95$ ,  $MSE = .01$ ,  $p < .001$ ,  $\eta^2_p = .82$ ), and under low than high extrinsic cognitive load (main effect of load, ( $F(1,38) = 7.15$ ,  $MSE = .005$ ,  $p < .05$ ,  $\eta^2_p = .16$ ). However, an interaction between predictability and load,  $F(1,38) = 9.87$ ,  $MSE = .004$ ,  $p < .01$ ,  $\eta^2_p = .21$ , indicated that the effects of cognitive load varied as function of sentence predictability. Specifically,

words were recognized more accurately under low than high cognitive load for the high-predictability sentences (main effect of load,  $F(1,38) = 14.30$ ,  $MSE = .005$ ,  $p < .001$ ,  $\eta^2_p = .27$ ), but not for the low-predictability sentences ( $F_s < 1$ ). The effect of load in high-predictability sentences is illustrated in Figure 1C, and is also reflected in Figure 1B, where it can be seen that less contextual facilitation in the severe degradation (4 Ch) condition under high than low cognitive load reflects the selective effect of load in the high-predictability condition (see also Figure 1A).

In the moderate (6 Ch) degradation conditions, sentence-final words were identified more accurately in high-predictability sentences (main effect of predictability ( $F(1,38) = 392.14$ ,  $MSE = .006$ ,  $p < .001$ ,  $\eta^2_p = .91$ ). The effect of load was not significant ( $F(1,38) = 2.93$ ,  $MSE = .004$ ,  $p < .10$ ,  $\eta^2_p = .07$ ).

In the mild (8 Ch) degradation conditions, sentence-final words were identified more accurately in high-predictability sentences (main effect of predictability ( $F(1,38) = 502.79$ ,  $MSE = .004$ ,  $p < .001$ ,  $\eta^2_p = .93$ ), and under low than high extrinsic cognitive load (main effect of load, ( $F(1,38) = 6.62$ ,  $MSE = .003$ ,  $p < .05$ ,  $\eta^2_p = .15$ ). The interaction between predictability and load was not significant,  $F(1,38) = 2.72$ ,  $MSE = .003$ ,  $p = .11$ ,  $\eta^2_p = .07$ . However, separate planned comparisons were conducted for the high- and low-predictability conditions given the a priori research question of whether the locus of the effects of load would appear in high- or low-predictability sentences. There were no significant effects for the high-predictability condition ( $F_s < 1$ ). For the low-predictability condition, there was a main effect of load,  $F(1,38) = 6.26$ ,  $MSE = .004$ ,  $p < .05$ ,  $\eta^2_p = .14$ , indicating that word recognition was more accurate under low than high cognitive load. The effect of load in low-predictability sentences is illustrated in Figure 1C, and is also reflected in Figure 1B, where in the mild degradation (8 Ch) condition a decrease in contextual facilitation under low cognitive load compared to high cognitive load is the result of the selective effect of load in the low-predictability condition (see also Figure 1A).

In sum, extrinsic cognitive load reduced accuracy for identifying spoken words in sentences. This result indicates that effects of cognitive load impact spoken word recognition in a task that has relatively high face validity in terms of the level(s) of information processing to which listeners direct attention during everyday listening. In addition, whether the effect of load was observed in low- or high-predictability sentences depended on the level of spectral degradation. Under severe degradation, cognitive load impaired word recognition in high-predictability sentences, whereas with milder degradation, cognitive load impaired word recognition in low-predictability sentences. One account of this asymmetry is that the different levels of intelligibility may have led to ceiling and floor effects that selectively impacted high- and low-predictability sentences as a function of signal degradation. As can be seen in Figures 1A and 1C, effects of load were not observed in the condition with the highest overall accuracy, the high-predictability sentences under mild degradation (8-Channels). Under mild degradation, ceiling effects may have prevented effects of cognitive load from emerging in the high-predictability sentences ( $M = 93.74$ ,  $SD = 4.77$ , range = 77 - 100) but not in the low-predictability sentences ( $M = 71.44$ ,  $SD = 8.41$ , range = 53 - 86). Effects of load were also not observed in the condition with the lowest overall accuracy, the low-predictability sentences under severe degradation (4-Channels). Here, floor effects may

have prevented effects of load from emerging in the low-predictability sentences ( $M = 30.30$ ,  $SD = 7.33$ , range 18 – 45) but not in the high-predictability sentences ( $M = 51.55$ ,  $SD = 10.18$ , range = 35 – 77).

To evaluate the possibility that extrinsic cognitive load has an effect on performance only when performance is in the middle range, we examined “load cost” using a difference score of accuracy in low-load minus accuracy in high-load conditions. Figure 2 shows scatterplots of load cost as a function of overall accuracy for each level of degradation and predictability. The main effect of spectral degradation on overall accuracy is shown in the stratification of data points by level of degradation. The main effect of sentence predictability is evident in comparing the high- and low-predictability sentences across the top and bottom panels of the figure. A positive value for “load cost” indicates the expected direction of the difference between low- and high-load conditions, with higher accuracy of word identification under low than high load. Figure 2 also shows that there was substantial individual variability in both overall accuracy and load cost at each level of sentence predictability and degradation. For example, for the severely degraded low-predictability sentences, overall accuracy ranged from 18 to 45 percent, overlapping with accuracy in the high-predictability condition (range = 35 – 77). Thus, if a floor effect determined whether load impacted word identification in this condition, one might expect a correlation between accuracy and load cost. However, there does not appear to be a relation between participants' overall accuracy and load cost in any condition. Accuracy and load cost were not correlated either in the severely degraded four-channel conditions (high-predictability,  $r(38) = .15$ ; low-predictability,  $r(38) = .10$ ), the moderately degraded six-channel conditions (high-predictability,  $r(38) = -.21$ ; low-predictability,  $r(38) = .17$ ), or the mildly degraded eight-channel conditions (high-predictability,  $r(36) = .20$ ; low-predictability,  $r(36) = .14$ ). Thus, although ceiling and floor effects should be considered as potential causes of the asymmetry in the interaction of cognitive load and sentence predictability across levels of sentence degradation, there is limited evidence to support this explanation in the current data. Nevertheless, the lack of any correlation between load cost and accuracy is a weaker indication of the absence of a ceiling effect than the absence of a floor effect. The mean accuracy in the condition with the highest overall accuracy, of approximately 94 percent, would be reached by missing approximately two out of 37 items. This high level of accuracy reduces the variability that could be accounted for by an effect of load, which should decrease the likelihood of observing a correlation between load cost and accuracy. By contrast, the mean condition accuracy in the condition with the lowest overall accuracy, of approximately 30 percent, would be reached by missing approximately 26 out of 37 items. Thus, a ceiling effect is more likely than a floor effect to have attenuated the effects of load in the current data.

Another potential explanation for the finding of load effects in high-predictability sentences under severe degradation but in low-predictability sentences under mild degradation is that cognitive load drew limited processing resources from distinct subtypes of controlled information processing when the stimuli were mildly and severely degraded. In the mildly degraded listening conditions in which acoustic-phonetic information was generally available, listeners' attention may have been directed to acoustic-phonetic or lexical features, such that cognitive load drew resources selectively from these lower-level perceptual processes. Successful word recognition in low-predictability sentences can only be based on

acoustic-phonetic and lexical-level information (e.g., word frequency). In contrast, word recognition in high-predictability sentences can also be based on contextual facilitation. Thus, given that listening effort was directed to lower-level processing under mild spectral degradation, cognitive load should have selectively decreased identification accuracy for sentence-final words of low-predictability sentences. Conversely, under severe spectral degradation, listeners' attention may have been directed to context-level information in order to facilitate top-down completion of the highly degraded speech signal, such that cognitive load selectively drew resources from an effortful process involved in contextual facilitation. Only in high-predictability sentences can performance be based on contextual facilitation. Thus, given that under severe spectral degradation listening effort was directed to top-down lexical access based on contextual facilitation, cognitive load should have selectively decreased word identification accuracy for high-predictability sentences. On this account of the interaction with cognitive load, wherein load affects performance in high-predictability sentences under severe degradation but in low-predictability sentences under mild degradation, the level of degradation of the spoken sentences influenced whether the listening effort of each group of participants was directed towards lower-level decoding of acoustic-phonetic information or instead towards using sentence context to facilitate word identification. The level of processing to which attentional resources were deployed then determined whether extrinsic cognitive load selectively impacted high-predictability sentences, suggesting an impact of cognitive load on contextual facilitation, or low-predictability sentences, suggesting an impact of cognitive load on lower-level perceptual processes.

### Digit Accuracy

Accuracy for digit recall is shown in Figure 3. For clarity, the variable of cognitive load is referred to as digit sequence length in the analysis of digit recall accuracy. Digit accuracy was scored using Levenshtein edit distance in order to avoid floor effects for digit accuracy in the longer sequences of seven digits. Percent correct was calculated as 100 percent minus the Levenshtein edit distance between the reported digit sequence and the correct digit sequence. As can be seen in Figure 3, there were two large, expected effects that were not under investigation: the longer sequences of seven digits were recalled less accurately than the shorter sequences of three digits (main effect of digit sequence length,  $F(1,114) = 694.36$ ,  $MSE = .01$ ,  $p < .001$ ,  $\eta^2_p = .86$ ), and digit sequences were recalled more accurately when the order of report was digits first and words second than vice versa (main effect of order of report,  $F(1,114) = 53.18$ ,  $MSE = 0.04$ ,  $p < .001$ ,  $\eta^2_p = .32$ ). There was also an interaction of digit sequence length with order of report,  $F(1,114) = 7.82$ ,  $MSE = 0.01$ ,  $p < .01$ ,  $\eta^2_p = .06$ , indicating an exaggerated difference in accuracy between short and long digit sequences when digits were reported after reporting the sentence-final word rather than immediately following sentence presentation.

The pattern of results shown in Figure 3 also suggests two, smaller effects that are relevant to the current research questions. First, for short digit sequences reported after reporting the sentence-final word (“W/D”), accuracy was affected by sentence predictability. Specifically, recall of short digit sequences was better following high- than low-predictability sentences. Although the three way interaction of order of report, digit sequence length, and sentence-

predictability was not significant,  $F(1,114) = 2.05$ ,  $MSE = .001$ ,  $p = .16$ ,  $\eta^2_p = .02$ , there was a significant two-way interaction of sentence predictability and digit sequence length,  $F(1,114) = 8.57$ ,  $MSE = .001$ ,  $p < .01$ ,  $\eta^2_p = .07$ , reflecting a downstream effect of sentence predictability that was present primarily for the short digit sequences.

Second, for the long digit sequences, digit recall was more accurate when sentences were less degraded. Statistically, this was indicated by a marginal main effect of degradation,  $F(2,114) = 3.06$ ,  $MSE = 0.05$ ,  $p = .051$ ,  $\eta^2_p = .05$ , reflecting a trend for higher accuracy for remembering digits that preceded less degraded sentences, and an interaction of digit sequence length with degradation,  $F(2,114) = 6.55$ ,  $MSE = .01$ ,  $p < .01$ ,  $\eta^2_p = .10$ , suggesting that a downstream effect of speech degradation on digit recall was present primarily for the long digit sequences.

Given that digit sequence length interacted with each of the other three factors, separate ANOVAs were conducted for each level of digit sequence length. For the short digit sequences, this analysis confirmed downstream effects of sentence predictability on digit recall. Short digit sequences were reported more accurately when the preceding sentences were predictable (main effect of predictability,  $F(1,114) = 8.38$ ,  $MSE = .001$ ,  $p < .01$ ,  $\eta^2_p = .07$ ), and this effect was present primarily when digits were reported after the sentence-final word (interaction of predictability with order of report,  $F(1,114) = 5.17$ ,  $MSE = .001$ ,  $p < .05$ ,  $\eta^2_p = .04$ ). Report of short digit sequences was also more accurate when digits were reported before rather than after the sentence-final words (main effect of order of report,  $F(1,114) = 48.91$ ,  $MSE = .01$ ,  $p < .001$ ,  $\eta^2_p = .30$ ). For the long digit sequences, the analysis confirmed downstream effects of sentence degradation on recall. Long digit sequences were reported more accurately when preceding sentences were less degraded (main effect of degradation,  $F(2,114) = 5.19$ ,  $MSE = .04$ ,  $p < .01$ ,  $\eta^2_p = .08$ ), and when digits were reported before rather than after the sentence-final word (main effect of order of report,  $F(1,114) = 38.36$ ,  $MSE = .04$ ,  $p < .001$ ,  $\eta^2_p = .25$ ).

In summary, two effects of interest to the aims of the current study were observed for accuracy in the secondary, digit recall task. Effects of sentence degradation and sentence predictability on digit report showed that the greater listening effort exerted when sentences were more degraded and less predictable had downstream impacts on the cognitive processes used to recall the digit sequences. First, sentence predictability affected the short sequences of three digits when the digits were reported after sentence-final words. Second, sentence degradation affected the long sequences of seven digits and did not depend on whether digits were reported before or after words.

## Discussion

The first aim of the current study was to determine whether extrinsic cognitive load would impair spoken word recognition in a sentence recognition task. Results showed that cognitive load did impact accuracy for identifying the final words of spoken sentences. In the current study participants identified spoken words in an open-set task in which there were unlimited response choices. Similar to a typical everyday listening situation, the spoken words occurred in sentence contexts. As described in the Introduction, the impact of

extrinsic cognitive load on early acoustic-phonetic registration and encoding has been shown in prior experimental work (Francis, 2010; Mattys & Wiget, 2011; Mattys et al., 2009; 2014). The present result indicates that the impact of extrinsic cognitive load on speech perception is not limited to closed-set tasks in which participants direct attention to subtle acoustic-phonetic distinctions. Rather, the effects of cognitive load in the current study were observed for the identification of words in sentences, a task that has relatively high face validity as an approximation of the level(s) of spoken word processing to which effortful listening is directed during everyday listening.

Following the ELU model, the present results indicate that central cognitive resources were used to match degraded acoustic-phonetic information in the spoken sentences to representations of words in long-term memory, and that the common pool of these processing resources was depleted by the load task, such that fewer resources were available to support effortful listening. This finding is consistent with a large body of correlational research in which working memory capacity is related to recognition of degraded speech (Akeroyd, 2008; Arlinger et al., 2009; Foo, Rudner, Rönnberg, et al. 2007; Rudner et al., 2011; Rönnberg, et al., 2008; 2010). In addition, substantial evidence using secondary memory or dual tasks indicates that processing degraded speech requires listening effort (LE), which draws on a limited pool of central processing resource, resulting in declines in secondary task performance (see Mattys et al., 2012; Rönnberg et al., 2008). However, relatively few studies have examined the converse influence of an external cognitive demand on the recognition of degraded speech (Francis, 2010; Mattys & Wiget, 2011; Mattys et al., 2009; 2014). The present finding of an impact of extrinsic cognitive load on word identification indicates such a converse influence, wherein as a non-speech task demands more processing resources, those resources become less available to support effortful listening. These results suggest that effortful listening to speech does not have an exclusive priority over other cognitive tasks in the allocation of domain-general cognitive resources. Further, the effects of extrinsic cognitive load in the sentence-final word identification task suggest that cognitive demand influences the effectiveness of effortful listening in everyday listening situations.

The current study is one of only a few studies of listening effort to use the pre-load design (see also Luce et al., 1983; Mattys et al., 2009; Sarampalis et al., 2009). In this design, the extrinsic load stimulus has undergone sensory and perceptual processing prior to any processing of the sentence stimuli. This is in contrast with the more frequently used dual-task paradigm in which speech signals are presented simultaneously with visual stimuli. Effects of extrinsic cognitive load on speech perception in the dual-task design may be due to attention switching between the encoding of auditory and visual stimuli (Mitterer & Mattys, 2017). In contrast, in the current study the processing of the digits during presentation of the sentences was memory-based, given that the load stimuli were presented prior to presentation of the spoken sentence. Specifically, the spoken sentences began immediately after the digits disappeared from the computer screen. Given that the duration of visual sensory memory is no more than half a second, any processing of the digits after the first few syllables of the spoken sentence must have been based on representations in short-term or working memory (). The effects of load in the present experiment therefore reflect the use of central cognitive resources to further process or to rehearse the digits in

short-term/working memory. A possible mechanism for these effects is attentional switching between the digit representations in memory and the encoding and processing of the spoken sentence. Alternatively, the effects of load could be due to a flexible division of central processing resources rather than to switching between the full allocation of processing resources to the sentence recognition and digit memory tasks (Ma, Husain, & Bays, 2014; Morey, Cowan, Morey, & Rouder, 2011). In general, the present results provide important converging evidence that the recognition of degraded speech is impacted by the availability of central cognitive resources.

The second aim of the current study was to investigate differential effects of extrinsic cognitive load in high- and low-predictability sentences in order to probe the level(s) of spoken word processing affected by load. Prior studies have reported correlations between working memory capacity and contextual facilitation, the increase in recognition accuracy for words in high-predictability compared to low-predictability sentences. These correlations suggest that contextual facilitation may not be an automatic process, but rather a process that demands cognitive resources (Gordan-Salant & Cole, 2016; Janse & Jesse, 2014; Zekveld et al., 2011; Zekveld, Rudner, Johnstrude, et al., 2013). However, other studies have observed larger correlations of working memory capacity with word recognition in low-predictability sentences than high-predictability sentences, a result that suggests that lower-level processing operations demand greater cognitive resources than contextual facilitation (Benichov, Cox, Tun, et al., 2012; Meister, Schreitmüller, Ortmann, et al., 2016; Wayne, Hamilton, Huyck, et al., 2016). In the current study, extrinsic cognitive load impacted word recognition accuracy in both high- and low-predictability sentences, but whether the effects of load were observed in high- or low-predictability sentences depended on the level of spectral degradation of the sentences. For the group of participants who listened to spoken sentences that were mildly degraded, cognitive load impacted word recognition for low-predictability sentences and not for high-predictability sentences. In contrast, for the group of participants who listened to spoken sentences that were highly degraded, cognitive load impacted word recognition for high-predictability sentences and not for low-predictability sentences. This crossover interaction may reflect the use of fundamentally different listening strategies for sentences that are mildly versus highly degraded. That is, when spoken sentences are spectrally degraded but still fairly intelligible, attentional resources may be allocated to acoustic-phonetic and/or lexical processing; however, when acoustic-phonetic information is highly degraded, attentional resources may instead be allocated to top-down recognition of sentence-final words based on contextual facilitation. If this account is correct, the effect of cognitive load observed in the highly degraded condition suggests that at least when listening conditions are quite adverse, using sentence contexts to recognize spoken words becomes an effortful, capacity-demanding process that makes use of central cognitive resources. Indeed, in the earlier correlational studies in which contextual facilitation of accurate speech recognition was significantly related to working memory capacity, signal-to-noise ratios were quite adverse (Gordon-Salant & Cole, 2016; Zekveld et al., 2012, 2013). In contrast, when speech is less degraded, effortful listening may be directed to attending to acoustic-phonetic information in order to recognize words through bottom-up processing. It should be noted that the level of spectral degradation in the current

study was a between-subjects variable, so that participants had ample time to adjust their listening strategy.

Current models of listening effort postulate that working memory is the central cognitive resource used for effortful listening to degraded speech (Pichora-Fuller et al., 2016; Rönnerberg et al., 2008). Contemporary models of working memory include mechanisms for the flexible direction of attention (Baddeley & Hitch, 1974; Ma et al., 2014; Morey et al., 2011). Thus, the idea that the crossover interaction in the current results may have been caused by the flexible allocation of listening effort either to top-down contextual facilitation or to bottom-up processing, depending on signal degradation, is in line with current models of listening effort. On this account, the level of processing at which cognitive capacity influences the recognition of words in spoken sentences may depend on signal degradation, with an influence on top-down processes such as contextual facilitation being more likely as the speech signal becomes more degraded. In broad terms, this account is in line with the well-known principle that top-down processing becomes more important for word recognition when speech signals are more highly degraded (Rönnerberg et al., 2008). Yet, it should also be noted that predictability had a relatively stable effect on accuracy across levels of degradation in the current study – sentence-final words in high-predictability sentences were recognized approximately 20 percent more accurately than low-predictability sentences at each level of degradation (see Figure 1B). We suggest that this effect of contextual facilitation may have occurred automatically under mild degradation, but was maintained through effortful, controlled listening to sentence context under severe degradation. That is, the sentence predictability effect may have been accomplished with relatively automatic processing when sentences were less degraded, whereas cognitive resources were needed to maintain the predictability benefit when sentences were highly degraded. Finally, the SPIN sentence stimuli and procedure that were used in the current study are often used to measure contextual facilitation in a test of sentence recognition. Nevertheless, given that the words to be reported were always sentence-final, participants may have been encouraged to strategically focus listening effort on using the sentence contexts more than would be typical of everyday listening, particularly for the participants who listened to the most degraded sentences.

An alternative account of the crossover interaction is that the interaction of predictability, spectral degradation, and cognitive load reflects an absence of extrinsic load effects when recognition accuracy was near ceiling or floor levels. This explanation is viable given that cognitive load did not have an effect on performance in the conditions in which overall accuracy was lowest (low-predictability sentences under severe spectral degradation) or in which overall accuracy was highest (high-predictability sentences under mild spectral degradation), but was instead observed when performance was in the middle range. If this account of the current results were correct, one might expect correlations between the load effect and overall accuracy. Given that such correlations were not observed for any condition, support is limited for the idea that the observed pattern of load effects reflects ceiling and floor effects. In addition, relations between working memory capacity and use of semantic cues for the identification of words in sentences have been observed at similar accuracy levels to those that were potentially at floor in the current study (Zekveld et al., 2012). However, it is more likely that load effects were prevented by a ceiling effect than a



floor effect given the levels of overall accuracy in the current data. Further work will be needed to systematically investigate whether overall accuracy or the level of processing to which attention is directed determines whether effects of extrinsic cognitive load on word identification appear in low- or high-context sentences.

Earlier experimental work on listening effort (LE) has generally focused on the effects of signal degradation on secondary task performance, showing that declines in performance on secondary tasks occur as the speech signal in the primary task of speech recognition becomes more degraded. Downstream effects of LE were observed on performance in the cognitive load task of digit recall in the current study. First, the longer digit sequences were reported more accurately when the preceding sentences were less spectrally degraded. This result is in line with earlier studies, indicating that common, central cognitive resources were recruited for sentence recognition as signal degradation increased, and that these resources were unavailable for a concurrent cognitive task (e.g., Luce et al., 1983; Pals et al., 2013; Rönnberg et al., 2010). Most prior studies of LE have used dual-task methodologies, in which stimuli for the primary and secondary tasks are presented simultaneously and effects of LE are observed on secondary task performance. Given that the digit sequences in the current study were presented and removed from the computer screen before the speech signals began, the downstream effect of LE observed on digit recall cannot be attributed to attentional switching between the auditory and visual stimuli. Instead, interference in the digit recall task must have occurred during the processing or rehearsal of the representations of the digits in short-term/working memory. Thus, the current results from the memory load paradigm provide additional, converging support for the proposal that downstream effects of LE reflect the allocation of a central, limited pool of cognitive resources.

Another downstream effect of LE on digit recall was that the shorter digit sequences were reported more accurately after high- than low-predictability sentences when participants reported the digit sequences after reporting the sentence-final word. This effect indicates that the high-predictability sentences may have required fewer central cognitive resources to recognize than the low-predictability sentences, creating a trade-off with the processing resources available for the active rehearsal of the short digit sequences in memory. Very few studies have investigated the downstream effects of sentence predictability on listening effort. In the study of McCoy et al. (2005), a series of spoken words were presented to older adult listeners either with or without hearing loss. The series of words, which varied in the degree to which they approximated meaningful sentences, were paused intermittently, at which point the most recent three items were to be reported. Accuracy was quite high for the most recent item, but for those sequences that least approximated meaningful sentences, accuracy for the first two words of the recall set was lower than accuracy for the third, most recent item. This result suggests that a meaningful context supported maintenance of words in working memory. In the McCoy et al. study, these effects of sentence context were assessed using the same words for which recognition was facilitated by the sentence context (see also Sarampalis et al., 2009). In the current study, the downstream effects of sentence predictability were observed on recall of short digit sequences that were presented prior to the spoken sentences. Therefore, the current results cannot be attributed to facilitation of the processing of digit sequences by a sentence context with a related meaning. Here, the facilitation of recall of short digit sequences by sentence predictability suggests that fewer

central cognitive resources were needed to process high- versus low-predictability sentences, such that greater central cognitive resources were available for processing and rehearsal of the short digit sequences. An alternative explanation is that sentence-final words were reported more quickly after high- versus low-predictability sentences, such that there was less time for the low-load digit sequences to decay from memory after high- than low-predictability sentences.

In the current study all of the participants were normal-hearing young adults. In contrast, the majority of research on listening effort has been conducted with older adults who have hearing loss. Young adults may differ from older adults in how much they rely on sentence context for word recognition and in the mechanisms of context use. In particular, long-term hearing loss in older adults may lead to habitual reliance on compensatory strategies such as context use. Thus, it is possible that context use is more automatic, and makes fewer demands on limited central cognitive resources for hearing-impaired older adults than normal-hearing younger adults (Pichora-Fuller et al., 1995). For example, a recent study has found that older adults rely on sentence context even when the context is misleading, which suggests a bias for reliance on top-down processing strategies (Rogers, Jacoby, & Sommers, 2012). Yet, there is also some evidence that when listening conditions are adverse, older adults may focus limited cognitive resources on word recognition at the expense of engaging in the higher-level processing that is needed to comprehend extended monologues and conversations (Avivi-Reich, Jackupczyk, Daneman, & Schneider, 2016; Schneider, Avivi-Reich, & Daneman, 2015). Interestingly, studies of another population with long-term experience with hearing a degraded auditory signal, cochlear implant users implanted in childhood, have shown decreased ability to take advantage of semantic context compared to normal-hearing controls (Conway, Deocampo, Walk, et al., 2014; Smiljanic, & Sladen, 2013). Further research using multiple methodologies with participants from a variety of clinical populations with chronic hearing impairment will be needed in order to determine the generalizability of the effects of extrinsic cognitive load on spoken word recognition in high- and low-predictability sentence contexts, and, in the larger picture, to determine what factors lead individuals to direct their listening effort to particular cognitive processes during speech understanding.

In summary, the present findings indicate that extrinsic cognitive load from a visual digit pre-load procedure can impair spoken word recognition in a sentence recognition task. In line with the ELU model and the Framework for Understanding Effortful Listening (FUEL), these findings suggest that word recognition in everyday listening situations is likely to rely on central cognitive resources (Pichora-Fuller et al., 2016; Rönnberg et al., 2008). Notably, the present findings were obtained with young adult college students, a population with relatively high cognitive and hearing functioning. Extrinsic cognitive load affected spoken word recognition in both high- and low-predictability sentences, indicating that central cognitive resources may contribute to word recognition both when a predictive context is used to compensate for a degraded speech signal and also when word recognition relies on lower-level perceptual processes in less adverse listening situations. The present findings provide converging evidence, using an experimental manipulation of cognitive capacity, for the correlations that have been observed between working memory and word recognition in both high- and low-predictability sentence contexts. The complex interaction between

extrinsic cognitive load, sentence predictability, and spectral degradation observed in the current study suggests that varying levels of speech intelligibility may contribute to differences across studies in whether cognitive capacity is associated with word recognition in low-predictability or high-predictability contexts.

## Acknowledgments

Financial Disclosures / Conflicts of Interest: This research was supported by grants from the National Institutes of Health, National Institute on Deafness and Other Communication Disorders: R01 DC009581 and National Center for Advancing Translational Sciences, Clinical and Translational Sciences Award: TL1TR001107

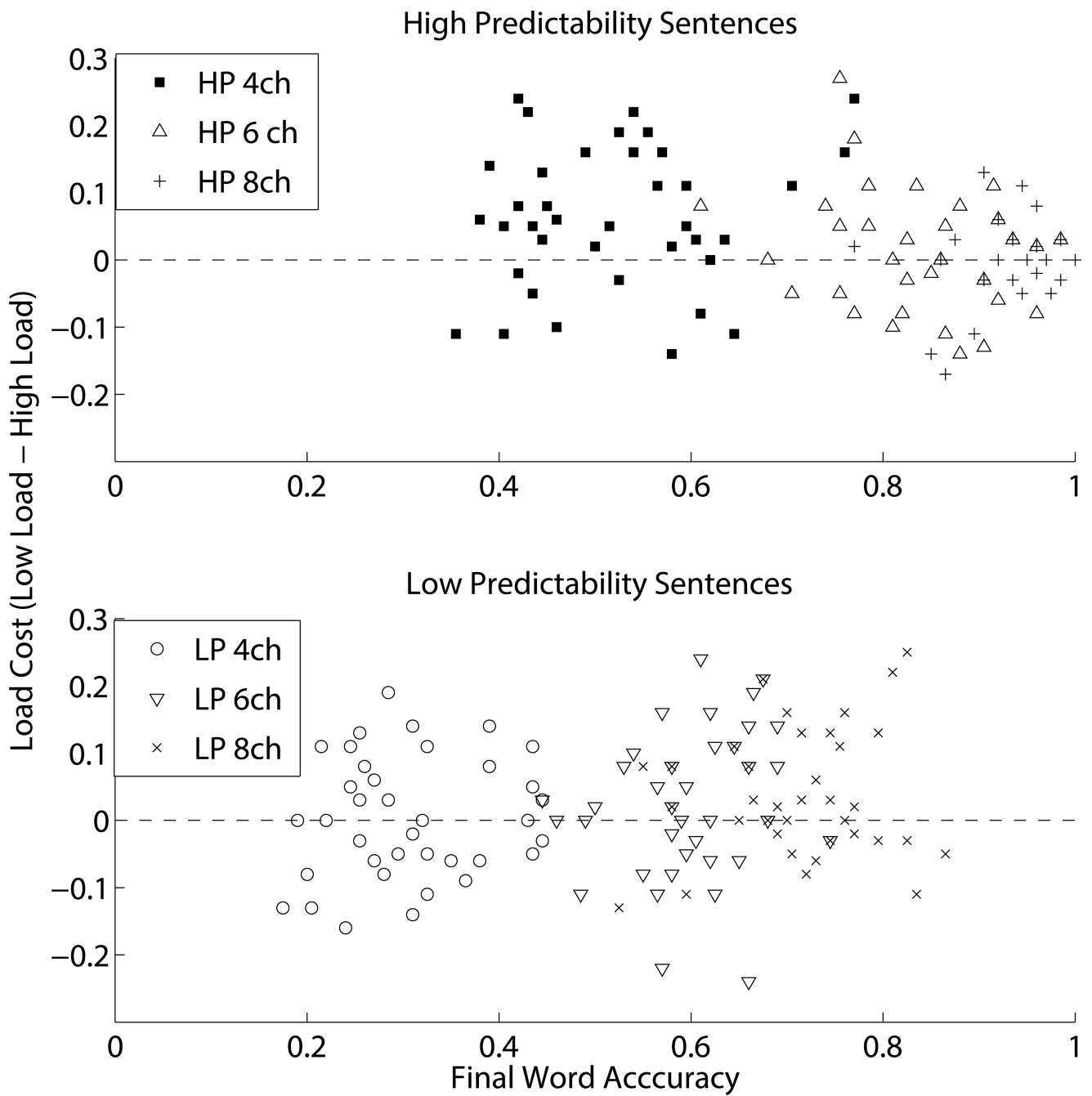
## References

- Akeroyd MA. 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*. 2008; 47(S2):S53–S71. [PubMed: 19012113]
- Arlinger S, Lunner T, Lyxell B, et al. The emergence of cognitive hearing science. *Scandinavian journal of psychology*. 2009; 50(5):371–384. [PubMed: 19778385]
- Avivi-Reich M, Jakubczyk A, Daneman M, Schneider BA. How age, linguistic status, and the nature of the auditory scene alter the manner in which listening comprehension is achieved in multitalker conversations. *Journal of Speech Language and Hearing Research*. 2015; 58:1570–1591.
- Baddeley AD, Hitch G. Working memory. *Psychology of learning and motivation*. 1974; 8:47–89.
- Baltes PB, Lindenberger U. Emergence of a powerful connection between sensory and cognitive functions across the adult life span: a new window to the study of cognitive aging? *Psychology and aging*. 1997; 12(1):12–21. [PubMed: 9100264]
- Benichov J, Cox LC, Tun PA, et al. Word recognition within a linguistic context: Effects of age, hearing acuity, verbal ability and cognitive function. *Ear and Hearing*. 2012; 32(2):250.
- Besser J, Koelewijn T, Zekveld AA, et al. How linguistic closure and verbal working memory relate to speech recognition in noise—a review. *Trends in amplification*. 2013; 17(2):75–93. [PubMed: 23945955]
- Bilger RC, Nuetzel JM, Rabinowitz WM, et al. Standardization of a test of speech perception in noise. *Journal of Speech, Language, and Hearing Research*. 1984; 27(1):32–48.
- Boothroyd A, Nittrouer S. Mathematical treatment of context effects in phoneme and word recognition. *The Journal of the Acoustical Society of America*. 1988; 84(1):101–114. [PubMed: 3411038]
- Broadbent, DE. In *perception and communication*. Elmsford, NY, US: Pergamon Press; 1958. The effects of noise on behaviour; p. 81-107.
- Cleary, M., Pisoni, DB. *Speech Perception and Spoken Word Recognition: Research and Theory*. In: Goldstein, EB., editor. *Blackwell Handbook of Sensation and Perception*. Malden, MA: Blackwell; 2001. p. 499-534.
- Clopper CG, Pisoni DB, Tierney AT. Effects of open-set and closed-set task demands on spoken word recognition. *Journal of the American Academy of Audiology*. 2006; 17(5):331–349. [PubMed: 16796300]
- Conway CM, Deocampo JA, Walk AM, Anaya EM, et al. Deaf children with cochlear implants do not appear to use sentence context to help recognize spoken words. *Journal of Speech, Language, and Hearing Research*. 2014; 57(6):2174–2190.
- Dahan D, Magnuson JS. Spoken word recognition. *Handbook of psycholinguistics*. 2006; 2:249–284.
- Daneman M, Carpenter PA. Individual differences in working memory and reading. *Journal of verbal learning and verbal behavior*. 1980; 19(4):450–466.
- Daneman M, Carpenter PA. Individual differences in integrating information between and within sentences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 1983; 9(4): 561–583.

- Foo C, Rudner M, Rönnberg J, et al. Recognition of speech in noise with new hearing instrument compression release settings requires explicit cognitive storage and processing capacity. *Journal of the American Academy of Audiology*. 2007; 18(7):618–631. [PubMed: 18236648]
- Francis AL. Improved segregation of simultaneous talkers differentially affects perceptual and cognitive capacity demands for recognizing speech in competing speech. *Attention, Perception, & Psychophysics*. 2010; 72(2):501–516.
- Francis AL, Nusbaum HC. Effects of intelligibility on working memory demand for speech perception. *Attention, Perception, & Psychophysics*. 2009; 71(6):1360–1374.
- Fullgrabe C, Rosen S. On the (un)importance of working memory in speech-in-noise processing for listeners with normal hearing thresholds. *Frontiers in Psychology*. 2016; 7:1268. [PubMed: 27625615]
- Gatehouse S, Naylor G, Elberling C. Benefits from hearing aids in relation to the interaction between the user and the environment. *International Journal of Audiology*. 2003; 42(sup1):77–85.
- Gordon-Salant S, Cole SS. Effects of Age and Working Memory Capacity on Speech Recognition Performance in Noise Among Listeners With Normal Hearing. *Ear and hearing*. 2016; 37(5):593–602. [PubMed: 27232071]
- Gordon-Salant S, Fitzgibbons PJ. Selected cognitive factors and speech recognition performance among young and elderly listeners. *Journal of Speech, Language, and Hearing Research*. 1997; 40(2):423–431.
- Janse E, Jesse A. Working memory affects older adults' use of context in spoken-word recognition. *The Quarterly Journal of Experimental Psychology*. 2014; 67(9):1842–1862. [PubMed: 24443921]
- Just MA, Carpenter PA. A capacity theory of comprehension: individual differences in working memory. *Psychological review*. 1992; 99(1):122–149. [PubMed: 1546114]
- Kahneman, D. *Attention and effort*. Englewood Cliffs, NJ: Prentice-Hall; 1973. p. 246
- Kalikow DN, Stevens KN, Elliott LL. Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. *The Journal of the Acoustical Society of America*. 1977; 61(5):1337–1351. [PubMed: 881487]
- Koelewijn T, Zekveld AA, Festen JM, et al. Pupil dilation uncovers extra listening effort in the presence of a single-talker masker. *Ear and Hearing*. 2012; 33(2):291–300. [PubMed: 21921797]
- Lemke U, Besser J. Cognitive Load and Listening Effort: Concepts and Age-Related Considerations. *Ear and Hearing*. 2016; 37:77S–84S. [PubMed: 27355774]
- Lindenberger U, Baltes PB. Sensory functioning and intelligence in old age: a strong connection. *Psychology and aging*. 1994; 9(3):339–355. [PubMed: 7999320]
- Luce PA, Feustel TC, Pisoni DB. Capacity demands in short-term memory for synthetic and natural speech. *Human Factors: The Journal of the Human Factors and Ergonomics Society*. 1983; 25(1):17–32.
- Luce, PA., McLennan, CT. Spoken word recognition: The challenge of variation. In: Pisoni, DB., Remez, RE., editors. *Handbook of Speech Perception*. Malden, MA: Blackwell; 2005. p. 591–609.
- Lunner T. Cognitive function in relation to hearing aid use. *International journal of audiology*. 2003; 42:S49–S58. [PubMed: 12918610]
- Lunner T, Sundewall-Thorén E. Interactions between cognition, compression, and listening conditions: Effects on speech-in-noise performance in a two-channel hearing aid. *Journal of the American Academy of Audiology*. 2007; 18(7):604–617. [PubMed: 18236647]
- Ma WJ, Husain M, Bays PM. Changing concepts of working memory. *Nature neuroscience*. 2014; 17(3):347–356. [PubMed: 24569831]
- Mattys SL, Barden K, Samuel AG. Extrinsic cognitive load impairs low-level speech perception. *Psychonomic bulletin & review*. 2014; 21(3):748–754. [PubMed: 24189992]
- Mattys SL, Davis MH, Bradlow AR, et al. Speech recognition in adverse conditions: A review. *Language and Cognitive Processes*. 2012; 27(7-8):953–978.
- Mattys SL, Wiget L. Effects of cognitive load on speech recognition. *Journal of Memory and Language*. 2011; 65(2):145–160.
- Mattys SL, Brooks J, Cooke M. Recognizing speech under a processing load: Dissociating energetic from informational factors. *Cognitive Psychology*. 2009; 59(3):203–243. [PubMed: 19423089]

- McCoy SL, Tun PA, Cox LC, et al. Hearing loss and perceptual effort: Downstream effects on older adults' memory for speech. *The Quarterly Journal of Experimental Psychology Section A*. 2005; 58(1):22–33.
- McQueen JM. Eight questions about spoken-word recognition. *The Oxford handbook of psycholinguistics*. 2007:37–53.
- Meister H, Schreitmüller S, Ortmann M, et al. Effects of hearing loss and cognitive load on speech recognition with competing talkers. *Frontiers in psychology*. 2016; 7:301. [PubMed: 26973585]
- Miller GA, Heise GA, Lichten W. The intelligibility of speech as a function of the context of the test materials. *Journal of experimental psychology*. 1951; 41(5):329–335. [PubMed: 14861384]
- Mitterer H, Mattys SL. How does cognitive load influence speech perception? An encoding hypothesis. *Attention, Perception, & Psychophysics*. 2017; 79(1):344–351.
- Morton J. Interaction of information in word recognition. *Psychological review*. 1969; 76(2):165–178.
- Morey CC, Cowan N, Morey RD, Rouders JN. Flexible attention allocation to visual and auditory working memory tasks: Manipulating reward induces a trade-off. *Attention, Perception, & Psychophysics*. 2011; 73(2):458–472.
- Obleser J, Wise RJ, Dresner MA, et al. Functional integration across brain regions improves speech perception under adverse listening conditions. *The Journal of Neuroscience*. 2007; 27(9):2283–2289. [PubMed: 17329425]
- Pals C, Sarampalis A, Ba kent D. Listening effort with cochlear implant simulations. *Journal of Speech, Language, and Hearing Research*. 2013; 56(4):1075–1084.
- Pichora-Fuller MK, Kramer SE, Eckert MA, et al. Hearing Impairment and Cognitive Energy: The Framework for Understanding Effortful Listening (FUEL). *Ear and Hearing*. 2016; 37:5S–27S. [PubMed: 27355771]
- Pichora-Fuller MK, Schneider BA, Daneman M. How young and old adults listen to and remember speech in noise. *The Journal of the Acoustical Society of America*. 1995; 97(1):593–608. [PubMed: 7860836]
- Rabbitt PM. Channel-capacity, intelligibility and immediate memory. *The Quarterly journal of experimental psychology*. 1968; 20(3):241–248. [PubMed: 5683763]
- Rabbitt P. Mild hearing loss can cause apparent memory failures which increase with age and reduce with IQ. *Acta otolaryngologica*. 1991; 111(sup476):167–176.
- Rogers CS, Jacoby LL, Sommers MS. Frequent false hearing by older adults: The role of age differences in metacognition. *Psychology and aging*. 2012; 27(1):33. [PubMed: 22149253]
- Rönnberg J. Cognition in the hearing impaired and deaf as a bridge between signal and dialogue: A framework and a model. *International Journal of Audiology*. 2003; 42:S68–S76. [PubMed: 12918612]
- Rönnberg J, Rudner M, Foo C, et al. Cognition counts: A working memory system for ease of language understanding (ELU). *International Journal of Audiology*. 2008; 47(sup2):S99–S105. [PubMed: 19012117]
- Rönnberg J, Rudner M, Lunner T, et al. When cognition kicks in: Working memory and speech understanding in noise. *Noise and Health*. 2010; 12(49):263–269. [PubMed: 20871181]
- Rudner M, Rönnberg J, Lunner T. Working memory supports listening in noise for persons with hearing impairment. *Journal of the American Academy of Audiology*. 2011; 22(3):156–167. [PubMed: 21545768]
- Sarampalis A, Kalluri S, Edwards B, et al. Objective measures of listening effort: Effects of background noise and noise reduction. *Journal of Speech, Language, and Hearing Research*. 2009; 52(5):1230–1240.
- Schneider BA, Avivi-Reich M, Daneman M. How spoken language comprehension is achieved by older adults in difficult listening situations. *Experimental Aging Research*. 2016; 42:31–49. [PubMed: 26683040]
- Sheldon S, Pichora-Fuller MK, Schneider BA. Priming and sentence context support listening to noise-vocoded speech by younger and older adults. *The Journal of the Acoustical Society of America*. 2008; 123(1):489–499. [PubMed: 18177176]
- Smiljanic R, Sladen D. Acoustic and semantic enhancements for children with cochlear implants. *Journal of Speech, Language, and Hearing Research*. 2013; 56(4):1085–1096.

- Sommers MS, Danielson SM. Inhibitory processes and spoken word recognition in young and older adults: The interaction of lexical competition and semantic context. *Psychology and aging*. 1999; 14(3):458. [PubMed: 10509700]
- Sommers MS, Kirk KI, Pisoni DB. Some considerations in evaluating spoken word recognition by normal-hearing, noise-masked normal-hearing, and cochlear implant listeners. I: The effects of response format. *Ear and Hearing*. 1997; 18(2):89. [PubMed: 9099558]
- Surprenant AM. The effect of noise on memory for spoken syllables. *International Journal of Psychology*. 1999; 34(5-6):328–333.
- Surprenant AM. Effects of noise on identification and serial recall of nonsense syllables in older and younger adults. *Aging, Neuropsychology, and Cognition*. 2007; 14(2):126–143.
- Ward KM, Shen J, Souza PE, Grieco-Calub TM. Age-Related Differences in Listening Effort During Degraded Speech Recognition. *Ear and hearing*. 2017; 38(1):74–84. [PubMed: 27556526]
- Wayne RV, Hamilton C, Huyck JJ, et al. Working Memory Training and Speech in Noise Comprehension in Older Adults. *Frontiers in aging neuroscience*. 2016; 8(49)
- Wingfield A. Evolution of Models of Working Memory and Cognitive Resources. *Ear and Hearing*. 2016; 37:35S–43S. [PubMed: 27355768]
- Wingfield A. Cognitive factors in auditory performance: context, speed of processing, and constraints of memory. *Journal of the American Academy of Audiology*. 1996; 7(3):175–182. [PubMed: 8780990]
- Winn MB, Edwards JR, Litovsky RY. The impact of auditory spectral resolution on listening effort revealed by pupil dilation. *Ear and hearing*. 2015; 36(4):e153–e165. [PubMed: 25654299]
- Zekveld AA, Rudner M, Johnsrude IS, et al. The influence of semantically related and unrelated text cues on the intelligibility of sentences in noise. *Ear and hearing*. 2011; 32(6):e16–e25. [PubMed: 21826004]
- Zekveld AA, Rudner M, Johnsrude IS, et al. The effects of working memory capacity and semantic cues on the intelligibility of speech in noise. *The Journal of the Acoustical Society of America*. 2013; 134(3):2225–2234. [PubMed: 23967952]



**Figure 1. Accuracy of sentence-final word recognition**

*Note.* **A.** Accuracy of sentence-final word recognition as a function of sentence predictability and cognitive load. **B.** Contextual facilitation effect, i.e., the difference in word recognition accuracy between high- and low-predictability sentences, shown as a function of cognitive load. **C.** Cognitive load effect, i.e., the difference in word recognition accuracy under high versus low cognitive load, shown as a function of sentence predictability. Separate panels show the spectral degradation conditions: 4 Ch, highly degraded 4-channel vocoding; 6 Ch, moderately degraded 6-channel vocoding; 8 Ch, mildly degraded 8-channel

vocoding. Data are collapsed across order of report (words first or digits first). Error bars show plus or minus one standard error (SE) from the mean.

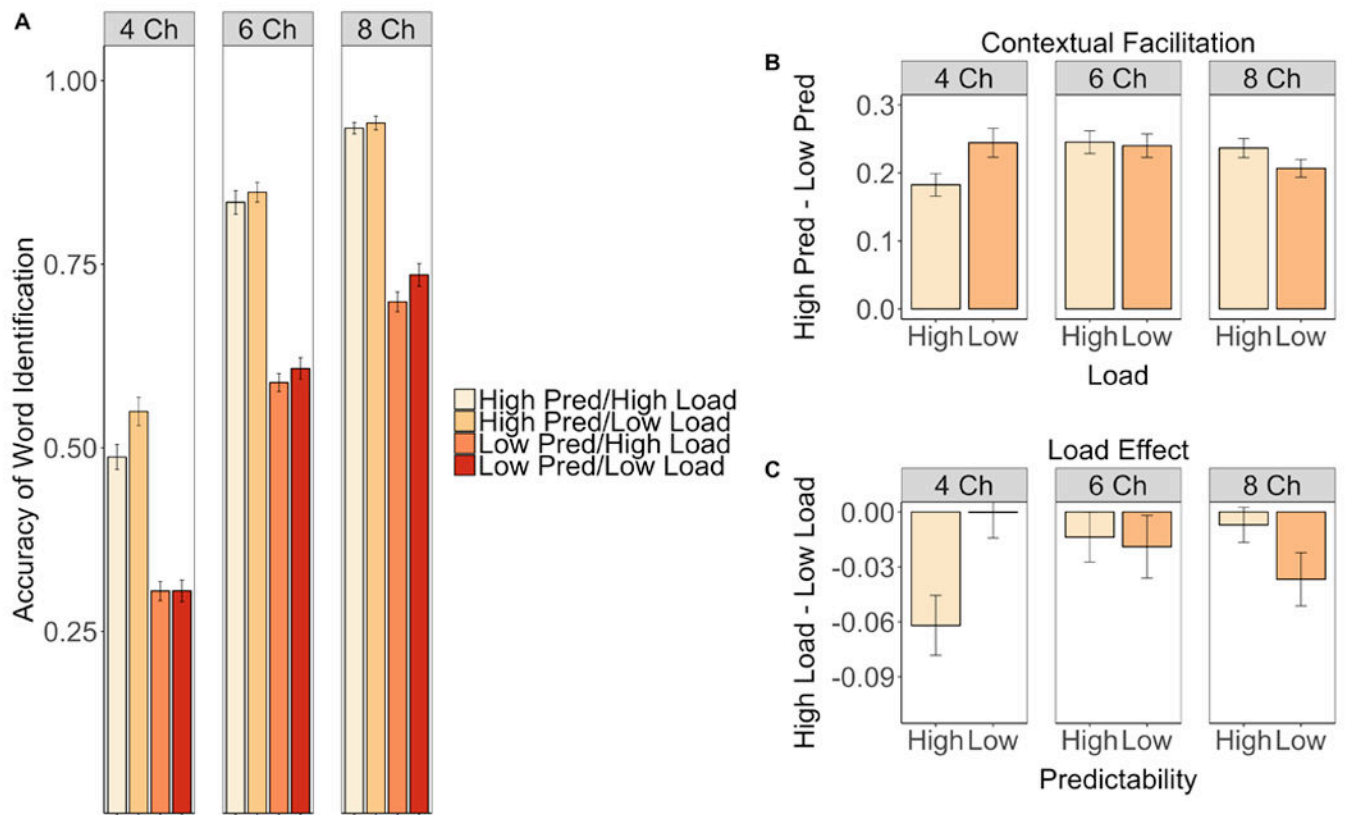
Author Manuscript

Author Manuscript

Author Manuscript

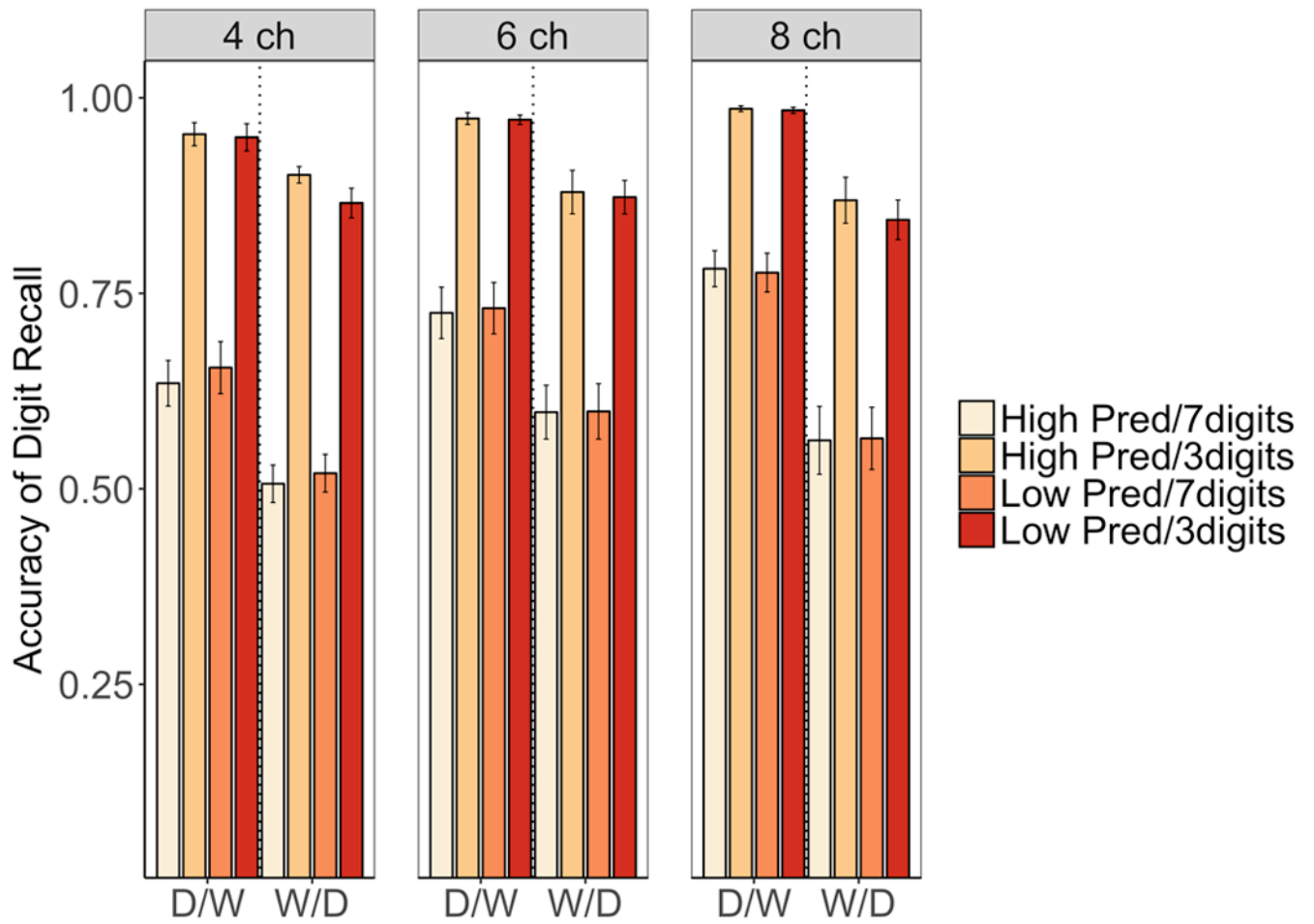
Author Manuscript





**Figure 2. Load cost as a function of accuracy for words**

*Note.* Scatterplots of Load Cost (low load accuracy minus high load accuracy) as a function of word identification accuracy per condition (mean accuracy for high and low load conditions). Top panel, high-predictability (HP) sentences; bottom panel, low-predictability (LP) sentences HP 4ch, highly degraded 4-channel vocoding; 6ch, moderately degraded 6-channel vocoding; 8ch, mildly degraded 8-channel vocoding. Data are collapsed across order of report (words first or digits first). Note that a positive value for “load cost” indicates the expected direction of the difference, with higher accuracy of word identification under low than high load.



**Figure 3. Accuracy of digit recall**

*Note.* D/W, digits reported first; W/D, words reported first. Separate panels show the spectral degradation conditions: 4 Ch, highly degraded 4-channel vocoding; 6 Ch, moderately degraded 6-channel vocoding; 8 Ch, mildly degraded 8-channel vocoding. Error bars show plus or minus one standard error (SE) from the mean.