LANGUAGE, COGNITION AND NEUROSCIENCE
https://doi.org/10.1080/23273798.2018.1562084

Routledge
Taylor & Francis Group

REGULAR ARTICLE

Check for updates

Noise increases listening effort in normal-hearing young adults, regardless of working memory capacity

Violet A. Brown and Julia F. Strand

Department of Psychology, Carleton College, Northfield, MN, USA

ABSTRACT
As listening conditions worsen (e.g. background noise increases), additional cognitive effort is required to process speech. The existing literature is mixed on whether and how cognitive traits like working memory capacity moderate the amount of effort that listeners must expend to successfully understand speech. Here, we validate a dual-task measure of listening effort (Experiment 1) and demonstrate that for normal-hearing young adults, effort increases as steady-state masking noise increases, but working memory capacity is unrelated to the amount of effort expended (Experiment 2). We propose that previous research may have overestimated the relationship between listening effort and working memory capacity by measuring listening effort using recall-based tasks. The present results suggest caution in making the general assumption that working memory capacity is related to the amount of effort expended during a listening task.

ARTICLE HISTORY
Received 21 June 2018
Accepted 14 December 2018

KEYWORDS
Listening effort; working memory; speech; dual-task; individual differences

For many listeners, understanding spoken language appears to occur effortlessly. However, this apparent ease belies the challenges of the task. Recognising spoken words requires extracting meaningful phonetic information from a rapidly-changing, variable signal and attempting to match that information to lexical representations stored in memory. Given the complexity of this process, there has been considerable research on whether and how individual differences in cognitive abilities relate to speech recognition. For example, in some populations, large working memory capacity (Arehart, Souza, Baca, & Kates, 2013; Desjardins & Doherty, 2013; Gordon-Salant & Cole, 2016; Koelewijn, Zekveld, Festen, Rönnberg, & Kramer, 2012; Lunner, 2003; Souza, Arehart, Shen, Anderson, & Kates, 2015), fast processing speed (Desjardins & Doherty, 2013), good inhibitory ability (Koelewijn et al., 2012), fast rhyme-judgements (Lunner, 2003), and good text reception thresholds (a visual analogue to speech reception thresholds; George et al., 2007; Koelewijn et al., 2012; Zekveld, George, Kramer, Goverts, & Houtgast, 2007) have been associated with good performance on speech recognition tasks. The relationship between cognitive ability and speech recognition appears to be relatively robust for older adults with hearing impairment, but much weaker when the sample consists of only younger adults with normal hearing (Besser, Koelewijn, Zekveld, Kramer, & Festen, 2013; Koelewijn et al., 2012; Surprenant & Watson, 2001; van Rooij, Plomp, & Orlebeke, 1989; Zekveld et al., 2011; Zekveld, Rudner, Johnsrude, Heslenfeld, & Rönnberg, 2012). Indeed, a recent meta-analysis by Füllgrabe and Rosen (2016) found that across 24 different datasets, working memory (WM) capacity accounted for an average of less than 2% of the variability in speech-in-noise performance for young listeners with normal hearing.

Individual differences in cognitive ability may not predict recognition accuracy in young adults with normal hearing due to a restriction of range in speech recognition abilities (see, for instance, van Rooij et al., 1989). However, when the listening conditions are difficult (resulting in an extended range of speech recognition performance), the recognition benefits associated with superior cognitive ability can reach a detectable level. For example, Zekveld et al. (2011) found a relationship between speech perception ability and both WM capacity (as measured by the RSpan test) and text reception threshold in young adults only under adverse conditions (i.e. conditions in which semantic interference was introduced). These results suggest that individual differences in cognitive ability are likely to have a more pronounced effect on speech recognition when the listening conditions are difficult. Note that although adverse conditions are most frequently introduced by increasing the level of the background noise, difficult listening conditions may also arise when speech is produced conversationally relative to clearly (Van Engen, Chandrasekaran, & Smiljanic, 2012), or by an accented

CONTACT Violet A. Brown violet.brown@wustl.edu; Julia F. Strand jstrand@carleton.edu

relative to a native talker (Adank, Evans, Stuart-Smith, & Scott, 2009). Further, adverse conditions can be introduced by properties intrinsic to the listener, including hearing loss (McCoy et al., 2005; Rabbitt, 1991; Wingfield, Tun, & McCoy, 2005) and cognitive decline due to aging (Desjardins & Doherty, 2013; Gosselin & Gagné, 2011b), among others (see Mattys, Davis, Bradlow, & Scott, 2012 for an excellent review of the challenges associated with speech perception). Given that older adults tend to have poorer hearing and have a smaller pool of cognitive resources at their disposal (Sommers, 1996), their listening conditions are intrinsically more adverse, which may explain the more robust relationship between performance on cognitive tasks and speech perception abilities in older adults.

Although certain cognitive advantages like larger WM capacity (Füllgrabe & Rosen, 2016; Zekveld et al., 2012), better SAT math and verbal scores, higher GPA (Surprenant & Watson, 2001), and better text reception thresholds (Zekveld et al., 2012) do not typically benefit speech recognition in young adults, they may reduce *listening effort* (LE). The concept of LE has been defined in terms of the cognitive resources necessary to comprehend speech (Bourland-Hicks & Tharpe, 2002; Fraser, Gagné, Alepins, & Dubois, 2010; Picou & Ricketts, 2014; Picou, Gordon, & Ricketts, 2016) or, more generally: "the deliberate allocation of mental resources to overcome obstacles in goal pursuit when carrying out a [listening] task" (Pichora-Fuller et al., 2016).

Given the common assumption that humans possess a limited pool of cognitive resources (Kahneman, 1973), as the listening task becomes more difficult, fewer resources remain available to complete any secondary cognitive tasks (Rabbitt, 1968). If a young adult's pool of cognitive resources is sufficiently large that the speech recognition task does not exhaust their resources, or if the range of abilities is restricted in young adults (see van Rooij et al., 1989), differences in speech recognition accuracy may not be observed. However, differences in LE may still be observed – although Besser et al. (2013) found no relationship between cognitive ability and recognition accuracy, they found that both WM capacity and text reception threshold were significantly correlated with listening span (note that although the authors used the listening span task as a measure of WM capacity, this task is also frequently used as a measure of LE; Johnson, Xu, Cox, & Pendergraft, 2015; Ng, Rudner, Lunner, Pedersen, & Rönnberg, 2013; Pichora-Fuller, Schneider, & Daneman, 1995; Sarampalis, Kalluri, Edwards, & Hafter, 2009; Strand, Brown, Merchant, Brown, & Smith, 2018). Further, effects that fail to emerge in speech recognition tasks sometimes emerge in LE tasks with the same speech materials. For example, although a visually presented dynamic circle that modulates with the acoustic amplitude envelope of the speech does not improve speech recognition in noise relative to an audio-only condition, it substantially reduces the effort listeners expend in order to complete the task (Strand, Brown, & Barbour, 2018). Similarly, some studies have shown that even when noise-reduction algorithms in hearing aids have little effect on the user's ability to recognise speech in noise, they can still reduce LE (Desjardins & Doherty, 2014; Sarampalis et al., 2009). Thus, although cognitive ability does not predict recognition accuracy in young adults, it may predict LE.

Several studies have demonstrated associations between various cognitive abilities and the amount of LE expended. In general, larger WM capacity (Mishra, Lunner, Stenfelt, Rönnberg, & Rudner, 2013a, 2013b; Ng et al., 2013; Strand, Brown, Merchant, et al., 2018), lower text reception thresholds (Mishra et al., 2013a, 2013b; Strand, Brown, Merchant, et al., 2018), and better inhibitory control (Mishra, Lunner, Stenfelt, Rönnberg, & Rudner, 2013b) are associated with less LE expenditure. Although some studies have failed to find a relationship between certain cognitive abilities and LE (e.g. Desjardins & Doherty, 2014; Zekveld, Festen, & Kramer, 2013), given the wide variety of LE measures that are frequently employed in the literature (e.g. subjective, behavioural, and physiological measures), and the large number of cognitive traits that are often correlated with these various measures of LE (e.g. reading span, letter memory score, text reception threshold, processing speed), it is not surprising that the literature is mixed on the extent to which certain cognitive abilities are related to LE (see Strand, Brown, Merchant, et al., 2018).

Researchers have proposed four possible frameworks for how cognitive abilities may affect LE (see Ahern & Beatty, 1979 for the three original hypotheses on which the more recent frameworks are based; Strand, Brown, Merchant, et al., 2018 ; van der Meer et al., 2010 ; Zekveld et al., 2011). The *effort hypothesis* predicts that individuals with greater cognitive capacity expend *more* LE because they invest more resources in a task, regardless of the listening conditions. Similarly, the *resource hypothesis* predicts that individuals with greater cognitive capacity expend more LE, but only in difficult listening conditions. The *cognitive efficiency hypothesis* predicts that individuals with greater cognitive capacity can use their resources more efficiently, and therefore expend less LE overall. Finally, the *Ease-of-Language Understanding* (ELU; Rönnberg, 2003; Rönnberg et al., 2013; Rönnberg, Rudner, Foo, & Lunner, 2008) model also predicts that those with greater cognitive capacity expend less LE, but this is especially true in difficult listening conditions.

In order to reconcile some of the inconsistent findings in the LE literature and distinguish between these hypotheses, Strand, Brown, Merchant, et al. (2018) conducted a large-scale validation study in which 111 young adults with normal hearing completed five cognitive tasks and seven measures of LE in both an easy and a difficult signal-to-noise ratio (SNR). One goal of this study was to determine the extent to which measures of LE depend on various cognitive abilities. Results showed that larger WM capacities, faster processing speeds, and better text reception thresholds tended to be associated with less LE expenditure (consistent with the general claims of the cognitive efficiency and ELU hypotheses). However, the correlations among LE measures were relatively weak and the extent to which each measure of LE related to cognitive abilities varied markedly. For example, some measures of LE were robustly correlated with WM and others were not at all. These results may explain why some studies have failed to find correlations between some measures of LE and certain cognitive traits – the relationship between cognitive ability and LE may depend on the specific tasks used. Though six of the seven LE tasks indicated that participants expended more effort in the difficult SNR than the easy one, there was mixed evidence for the ELU and cognitive efficiency hypotheses. That is, some of the relationships between cognitive tasks and LE measures showed that greater cognitive ability reduces LE specifically in difficult listening conditions, but others showed that greater cognitive ability reduces LE regardless of listening difficulty (and yet others showed that greater cognitive ability reduces LE more in an easy listening condition than a hard one, a finding that is not consistent with any of the four hypotheses).

One possible explanation for the varying support for the ELU and cognitive efficiency hypotheses is that the two SNRs employed by Strand, Brown, Merchant, et al. (2018) were not sufficiently different for an interaction between SNR and cognitive ability to become apparent. Figure 1 depicts a hypothetical distribution of the relationship between acoustic challenge (i.e. SNR) and LE for two different levels of cognitive ability. For all listeners, LE is highest at an intermediate level of acoustic challenge; when the listening situation becomes too challenging, the listener may give up on the speech task, which manifests as a reduction in LE. Thus, the relationship between LE and acoustic challenge may be inverted U-shaped rather than linear, and this relationship may only be observable across a sufficiently large range of SNRs. Indeed, Zekveld and Kramer (2014; Experiment 2) demonstrated exactly this trend – they found an inverted U-shaped relationship such that participants expended the most LE at intermediate levels of acoustic challenge, but this relationship only emerged when they used a range of nine (rather than three) SNRs. Similarly, Wu, Stangl, Zhang, Perkins, and Eilers (2016) used two different dual-task paradigms across a range of eleven SNRs and found that more LE was expended at intermediate compared to easy and difficult SNRs (see also Ohlenforst et al., 2018). Thus, certain effects may only emerge when the range of listening conditions is adequately large.

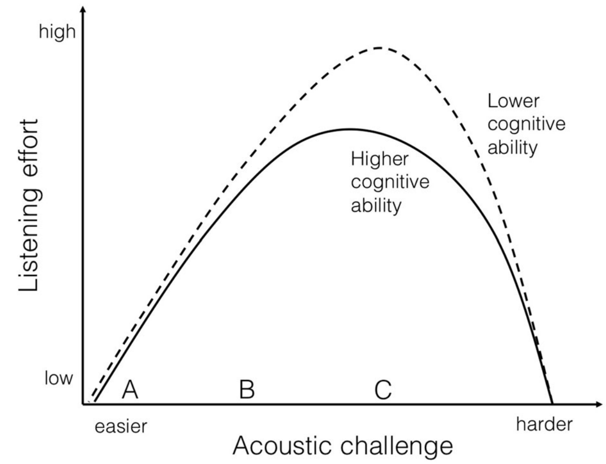

Figure 1. Hypothetical distribution depicting the relationship between acoustic challenge (i.e. SNR) and LE for two levels of cognitive ability.

In Figure 1, the reduction in LE for individuals with superior cognitive capacity is most pronounced at intermediate SNRs, but is not discernable at every SNR. If the two SNRs selected by Strand, Brown, Merchant, et al. (2018) happened to fall at points A and B in Figure 1, the interaction between SNR and cognitive ability would likely fail to reach significance, thereby supporting the cognitive efficiency hypothesis. If, however, the SNRs fell at points A and C, a significant interaction may have emerged as a result of a stronger relationship between LE and cognitive ability at the more difficult SNR, providing support for the ELU model. Note that these hypothetical distributions depict only one of numerous ways in which cognitive ability and SNR could interact; indeed, another way in which the interaction could manifest is a horizontal shift towards more difficult listening conditions for individuals with higher cognitive ability compared to those with lower ability. That is, individuals with larger WM capacities may not expend less LE overall, but the point at which they expend their maximum amount of LE may occur at a more difficult SNR than those with smaller WM capacities (see Ohlenforst et al., 2018 for a demonstration of a horizontal shift in pupil dilation with and without a noise-reduction algorithm). Regardless of the nature of the interaction, a larger range of SNRs in Strand, Brown, Merchant, et al. (2018) may have enabled more conclusive support for either the ELU or cognitive efficiency hypothesis to emerge.

The goal of the present study was to assess LE across a wide range of SNRs in young adults with normal hearing to determine whether superior cognitive ability mitigates the detrimental effects of difficult listening conditions. According to the ELU model, noisy conditions elicit a mismatch between the acoustic-phonetic input and representations stored in the listener's mental lexicon, so listeners must recruit storage and processing components of WM to retain the input in memory and resolve the mismatch (Rönnberg et al., 2013). If this is indeed the case, then WM capacity may be expected to be a particularly salient cognitive predictor of LE (see also Mishra et al., 2013a; Rönnberg, 2003; Rönnberg et al., 2008; Strand, Brown, Merchant, et al., 2018). Further, given that greater reliance on WM capacity is expected in more difficult listening conditions, the large range of SNRs we employ should provide ample opportunity for any possible relationship between WM capacity and LE to emerge. Thus, we opted to use a WM capacity task as our measure of cognitive ability given its direct implication in the ELU model, regular use in the LE literature, and the fact that it tends to be associated with other cognitive abilities (Daneman & Carpenter, 1980; Fry & Hale, 1996; Strand, Brown, Merchant, et al., 2018).

Researchers have used numerous behavioural measures of LE (see Strand, Brown, Merchant, et al., 2018) that tend to fall into one of two general categories: dual-task paradigms (in which greater LE is revealed as impaired performance on a secondary task) or recall paradigms (in which greater LE results in poorer encoding and therefore recall of to-be-remembered stimuli). We opted to use a dual-task paradigm rather than a recall task because a relationship between a recall measure of LE and a WM task could be attributable to general WM capacity rather than LE specifically. Further, dual-task paradigms are commonly used in the LE literature and there is consensus that these tasks are reliable and robust measures of LE (see Gagné, Besser, & Lemke, 2017). Finally, the dual-task paradigm we employed allows for a clear dissociation between performance on the primary speech recognition task and effort on an unrelated visual task, which will allow us to detect effects of cognitive overload if they emerge (i.e. listeners abandon the speech recognition task when it becomes too difficult, resulting in improved performance on the secondary task; see Zekveld & Kramer, 2014). We opted to use the complex dual-task (see Picou & Ricketts, 2014; Sarampalis et al., 2009; Strand, Brown, Merchant, et al., 2018), which involves listening and responding to speech while simultaneously completing a visual response time task.

Prior work from our lab using similar cognitive and LE measures (Strand, Brown, Merchant, et al., 2018) found some evidence for both the cognitive efficiency and ELU hypotheses. Both frameworks predict that higher cognitive ability is associated with reductions in LE, but the ELU hypothesis predicts that this relationship will be more pronounced in difficult listening situations, because those conditions are more likely to elicit a mismatch between the acoustic-phonetic input and representations of words stored in memory. Given that we are using a large range of SNRs in the current study, we hypothesised (in line with the predictions of the ELU model) that WM capacity and SNR would interact such that WM capacity is more beneficial at difficult SNRs, indicating that individuals with large WM capacities are less adversely affected by difficult SNRs than those with small WM capacities.

Research on LE tends to manipulate SNR by holding the level of the speech constant and varying the level of the background noise (Desjardins & Doherty, 2014; Fraser et al., 2010; Gosselin & Gagné, 2011a, 2011b; Picou & Ricketts, 2014). However, it is possible that increasing the level of the background noise may adversely affect performance on the secondary tasks used in dual-task paradigms, independent of the challenges associated with attempting to understand the speech (see Szalma & Hancock, 2011 for a review and meta-analysis of the effects of noise on cognitive and perceptual tasks). If this were the case, then effects of LE would be confounded with effects of noise on secondary tasks. Therefore, to assess whether the level of the background noise affects performance on the visual task used in Experiment 2, we first conducted a preliminary validation experiment in which participants performed the visual task without the associated speech recognition task at two extreme levels of background noise.

Experiment 1

Method

All raw data, stimuli, and code for analysis are available via the Open Science Framework at https://osf.io/yzp79/.

Participants

30 participants between the ages of 18 and 23 participated in Experiment 1. This preliminary experiment

took approximately 10 min, and participants were compensated \$2 for their time. Based on the recent recommendation that response time studies have at least 1,600 observations per condition (Brysbaert & Stevens, 2018), and given that there were 80 items per condition (see "Stimuli" section below), 20 participants should result in a sufficiently powered study. We analysed data from 30 participants to increase the power of the study. To reach this sample size, we collected data from 32 participants, but excluded two participants' data from analysis (see "Exclusion criteria" section below). All research methods were approved by the Carleton College Institutional Review Board. Participants signed a consent form prior to participation.

Stimuli and procedure

Participants were seated a comfortable distance away from a 21.5-inch iMac computer running SuperLab 5 (Cedrus) for stimulus presentation and data collection. Masking noise was presented via Seinheisser HD 280 Pro headphones. In this experiment, participants completed the visual component of the complex dual-task (Picou & Ricketts, 2014; Sarampalis et al., 2009; Strand, Brown, Merchant, et al., 2018) in two different levels of masking noise without speech. In this task, two square boxes (measuring approximately 5 cm across) appeared on the screen – one on the left side and one on the right side – and in random intervals between 500 ms and 2,000 ms (in 250 ms intervals), a digit between one and eight appeared in one of the boxes. Participants responded to the numbers by pressing buttons on a button box (Cedrus RB-740) that depicted either a left-facing or a right-facing arrow. If the number was even, participants were instructed to press the button with the arrow that pointed toward the box with that number (e.g. if the number 4 appeared in the left box, they pushed the left-facing arrow), and if the number was odd, they pressed the button with the arrow that pointed away from the box with that number. The number remained on the screen until the participant responded or 2,500 ms elapsed. Participants were instructed to respond as quickly and accurately as possible, and the dependent variable of interest was response time to correct responses, recorded from the onset of the number appearing in one of the boxes. Participants were presented 80 trials in each of the two conditions (160 trials total), but due to experimental error, one participant was presented only 64 trials per condition (128 trials total). The noise level was blocked and counterbalanced across participants. Prior to beginning the task, participants completed approximately 20 s of practice trials.

The background noise consisted of speech-shaped noise that matched the long-term average spectrum of the speech files used in Experiment 2. The two levels of background noise we selected corresponded to the easiest and most difficult SNRs in Experiment 2 because we wanted to determine whether response times to the cognitive task differed between the most extreme conditions when speech was not present. Thus, the background noise was set to approximately 46 dB SPL in the quiet condition and 70 dB SPL in the loud condition. If we do not observe any response time differences between these two levels of background noise, then it is unlikely that any differences exist when the noise levels are less extreme.

Exclusion criteria

Participants ($N = 1$) were excluded from analyses if their accuracy at performing the odd/even judgement task was worse than three SDs below the mean for either noise condition. Participants ($N = 1$) were also excluded if their mean response time was more than three SDs above or below the mean for either condition. At the trial level, we planned to exclude individual trials if the response time was more than three median absolute deviations above or below the participant's median response time for that condition (Leys, Ley, Klein, Bernard, & Licata, 2013), but no individual trials met this criterion.

Results and discussion

Data were analysed using linear mixed effects models via the *lme4* package in R (version 3.3.3; Bates et al., 2014). In all analyses, participants and items were entered as random effects, and we utilised the maximal random effects structure justified by the design (see the R script at https://osf.io/yzp79/ for more details on implementation, including random effects structures and issues of non-convergence). We built two nested models predicting response times to correct responses (95.6% of trials; note that accuracy did not differ between the quieter (95.7%) and louder (95.6%) noise levels) – a full model with noise level as a fixed effect and participants and items as random effects, and a reduced model with only random effects. A likelihood ratio test indicated that the reduced model provided a better fit for the data than the model that included noise level ($\chi^2_1 = 0.46$; $p = 0.50$), suggesting that noise did not have an effect on response times to the visual task. Indeed, response times in the quieter noise level (mean = 700 ms; SD = 96 ms) were nearly identical to those in the louder noise level (mean = 693 ms; SD = 85 ms), Cohen's $d = -0.03$. These results demonstrate that changing the level of the background noise did not affect performance on the visual task. In Experiment 2, participants completed the same visual task in noise, but were also asked to simultaneously listen to and repeat a stream of isolated words. The results of Experiment 1 suggest that any differences in response times to the visual task at different noise levels in Experiment 2 are attributable to LE, rather than simply being a reflection of increased processing demands for the visual portion of the dual-task in noisier conditions.

Experiment 2

Method

Sample size, experimental method, exclusion criteria, and analysis plan were pre-registered via the Open Science Framework (https://osf.io/5n9my). All raw data, stimuli, and code for analysis are available at https://osf.io/yzp79/.

Participants

Participants were young adults (ages 18–28) from the Carleton College community. All participants had self-reported normal hearing, normal or corrected-to-normal vision, and had not participated in Experiment 1. The number of responses to the visual task varied by participant depending on response speed, but based on the response rates from another study using the complex dual-task (Strand, Brown, Merchant, et al., 2018), we expected that participants would respond to approximately 45–50 visual stimuli in the time it took to complete the speech task at each SNR. We planned to include a total of 75 participants; given the recommendation for a minimum of 1,600 observations per condition for experiments measuring response time (Brysbaert & Stevens, 2018), 75 participants each responding to approximately 45 items per condition (3,375 observations per condition) is a conservative estimate for a well-powered study. To achieve the final sample of 75, we ran a total of 105 participants. Three participants' data were lost due to experimenter error, and three participants failed to complete the reading span task correctly. Strict pre-registered exclusion criteria (i.e. a participant was excluded from all analyses if they met any exclusion criterion in just one of the nine SNR levels; see results for more details), eliminated an additional 21 participants. One additional participant was excluded due to unanticipated poor performance on one task (see results section). We therefore had usable data from 77 participants, but the analyses included data from only the first 75, given that the sample size was pre-registered. Carleton College's Institutional Review Board approved all procedures.

Complex dual-task

Unless otherwise specified, the equipment and procedures were the same as in Experiment 1. Speech stimuli were recorded at 16-bit, 44100 Hz using a Shure KSM-32 microphone with a plosive screen by a female talker without a discernible regional accent. All auditory stimuli were equalised on root-mean-square amplitude using Adobe Audition (version 9.2.0).

Each participant heard words at nine different SNRs (−10 dB, −7 dB, −4 dB, −1 dB, 2 dB, 5 dB, 8 dB, 11 dB, 14 dB). These noise levels were intended to cover a range of word intelligibility from floor to ceiling. Speech files were consistently presented at approximately 60 dB SPL and the SNR was manipulated by increasing or decreasing the level of the background noise. Speech stimuli consisted of nine lists of 30 consonant-vowel-consonant words (e.g. "bear", "rain", "call"), and lists were counterbalanced across conditions using a Latin Squares design such that each list appeared in each of the nine conditions. Counterbalancing ensured that any observed effects could not be attributed to the order of presentation of the noise conditions or the particular words in each list. The ISIs between speech stimuli were 2,000 ms, 2,500 ms, and 3,000 ms, presented in a randomised order.

Participants listened to and verbally repeated the words while simultaneously completing the visual task described in Experiment 1. They were told that they should complete both tasks to the best of their ability, but the speech recognition task was more important, so they should focus their attention on that task (Bourland-Hicks & Tharpe, 2002; Desjardins & Doherty, 2013; Downs, 1982; Fraser et al., 2010). Response times to correct responses to the visual task were taken as a measure of LE. Prior to starting the task, participants completed 10 s of practice on the visual task alone, followed by approximately 15 s of practice (five words) on the word recognition and visual tasks simultaneously, at an SNR of 0 dB. Verbal responses to the word identification task were recorded via Audacity (version 2.1.2), and were later scored off-line by research assistants. Responses were scored as correct only if they matched the target word exactly (i.e. inflected forms of the target word were scored as incorrect). Following the procedures of Sarampalis et al. (2009), speech stimuli and visual complex dual-task stimuli were presented at random intervals so did not precisely coincide. This ensured that participants could not use the appearance of one stimulus as a cue to when the other would appear (see discussion for more on this issue). After completing the complex dual-task, participants completed a WM task.

Reading span test (Rspan; working memory capacity)

We included the Rspan task (Conway et al., 2005; Daneman & Carpenter, 1980; Foo, Rudner, Rönnberg, & Lunner, 2007; Lunner, 2003; Rudner, Rönnberg, & Lunner, 2011) as a measure of WM capacity given its prevalence in the LE literature (see Desjardins & Doherty, 2014; Mishra et al., 2013a, 2013b; Ng et al., 2013; Rudner, Lunner, Behrens, Thorén, & Rönnberg, 2012; Strand, Brown, Merchant, et al., 2018; Zekveld et al., 2013). Stimuli for the Rspan task were obtained from the Engle lab (Redick et al., 2012). Participants were presented with declarative sentences that were either sensical (e.g. "During winter you can get a room at the beach for a very low rate") or nonsensical (e.g. "During the week of final spaghetti, I felt like I was losing my mind"). Participants silently read each sentence and committed the last word to memory. In order to ensure that they processed the entire sentence and did not simply read the last word (Mishra et al., 2013a, 2013b), participants were required to judge whether the sentence was sensical or nonsensical by pressing a button on a button box. The next sentence was presented after a 1,750 ms inter-sentence interval (Mishra et al., 2013a, 2013b). After a series of 3, 4, 5, or 6 sentences, participants were prompted to verbally recall the final words of all the sentences in the set in any order, and responses were recorded via Audacity (version 2.1.2). Following the procedures of Pichora-Fuller et al. (1995), three trials of each set size were presented in ascending order. Verbal responses were scored off-line by research assistants, and Rspan score was calculated by dividing the number of words correctly recalled by 54 (the maximum possible score) to determine the percent correct. Participants completed three practice trials (lengths 3, 4, and 5) prior to beginning the task.

Results

Twenty-one participants were excluded due to pre-registered criteria: four participants due to poor word recognition accuracy in at least one condition, eleven participants for low response rates in at least one condition, three for poor accuracy on the secondary odd/even judgment task, and three for extreme response times (six total, but three of the six were already eliminated based on one of the other exclusion criteria). One additional participant was excluded due to poor performance on the Rspan sensicality judgment task (7% correct, compared to the overall participant average of 70%). We did not anticipate that any participant would perform poorly on this very simple task, so did not pre-register this as an exclusion criterion. The decision to eliminate that participant was made prior to conducting the main analysis.[1] No participants were excluded for low response rates to the secondary odd/even judgment task, and no individual trials were excluded for having response times more than three median absolute deviations above or below the participant's median for that condition. We analysed a total of 28,256 dual-task trials that participants had correctly classified as odd or even (94.4% of trials). The condition with the smallest number of correct responses to the odd/even judgement task (SNR = −10 dB) had 2,912 individual trials, well above the 1,600 benchmark (Brysbaert & Stevens, 2018).

The average by-participant response time on the complex dual-task (collapsing across SNR) was 799 ms (SD = 156 ms), and the average Rspan score was 70% (SD = 17%). Although the mean Rspan score is higher than has been reported in some experiments (49% in Desjardins & Doherty, 2014; 55% in Mishra et al., 2013a, 43% in 2013b), scores ranged from 33% to 100% correct, suggesting that there is substantial variability in Rspan scores, even in our population of college students, and these values and ranges are very similar to prior work in our lab that used the same tasks (Strand, Brown, Merchant, et al., 2018). Word recognition accuracy ranged from 5.0% in the hardest SNR (−10 dB) to 92.8% in the easiest SNR (14 dB), indicating that the SNRs elicited a very wide range of accuracies (see Figure 2, grey line).

Listening effort analysis

All data were analysed using linear mixed effects models via the *lme4* package in R (version 3.3.3; Bates et al., 2014). In all the models described below, participants and complex dual-task stimuli were entered as random effects, and the maximal random effects structure justified by the design was used (see the R script at the link above for more details). We first built two nested models, one that included SNR and one that did not, to determine whether performance on the complex dual-task suffered as the SNR became more difficult. A likelihood ratio test indicated that the model including SNR provided a better fit for the data ($\chi^2_1 = 53.69$; $p < 0.001$), and examination of the summary output of this model indicated that response times were faster in easier SNRs ($\beta = -4.96$, $SE = 0.57$, $t = -8.78$, $p < 0.001$; see Figure 2, black line).

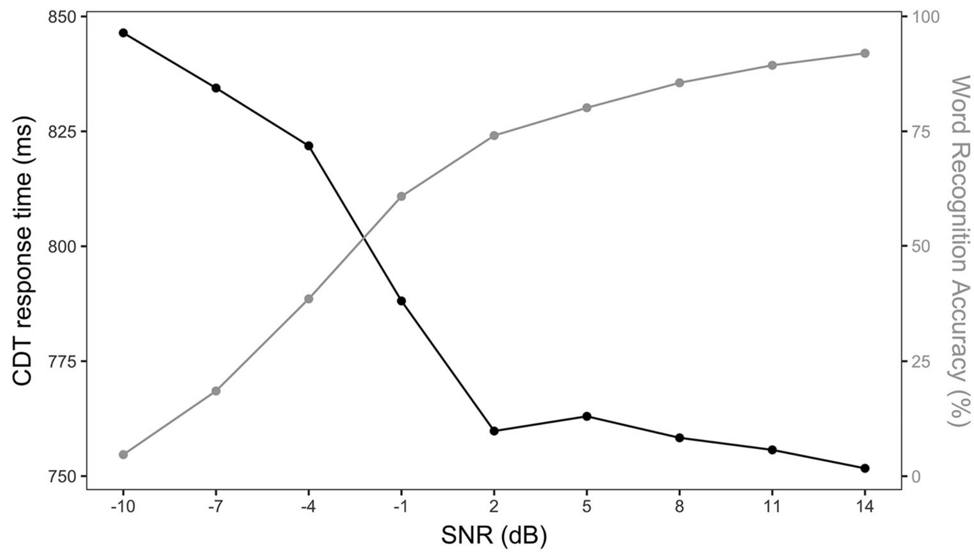

Figure 2. Scatterplot showing how complex dual-task (CDT) response times (black line) and word recognition accuracy (grey line) change as a function of SNR. Points represent average values for each SNR, collapsing across participants and items. Easier SNRs tended to be associated with faster CDT RTs and higher word recognition accuracy.

To examine the nature of the relationship between LE and SNR, we plotted response times to the complex dual-task against SNR, collapsed across participants and items so that the mean response time for each SNR is displayed in the plot (see Figure 2, black line). The fact that the slowest response times were found in the most difficult SNR suggests that participants did not abandon the primary task (i.e. there was no evidence of cognitive overload). The figure reveals that performance steadily decreased with the addition of more noise below an SNR of 2 dB, but response times were less affected by SNRs above 2 dB.

Next, we built a model that included both SNR and Rspan as fixed effects and compared it to the model that included only SNR to determine whether Rspan affects response times to the visual task. A likelihood ratio test indicated that the model without Rspan provided a better fit for the data ($\chi^2_1 = 0.48$; $p = 0.49$), indicating no effect of Rspan score on response time. We then built a model that included SNR, Rspan, and the interaction between SNR and Rspan as fixed effects and compared it to the model lacking the interaction term to determine whether WM has a different effect on LE at different levels of listening difficulty. A likelihood ratio test indicated that the model without the interaction term provided a better fit for the data ($\chi^2_1 = 0.65$; $p = 0.42$), indicating that the relationship between LE and Rspan did not differ by SNR (see Figure 3).

General discussion

The current study makes two contributions to the literature. First, these results validate the complex dual-task as a measure of LE by demonstrating that response times to the visual task differ robustly as a function of noise level, but only when the visual task is completed with a concurrent speech task. This finding supports the claim that the complex dual-task is in fact measuring LE, rather than simply reflecting increased arousal or stress from the presence of loud noise. Increased background noise levels can certainly impair cognitive function in some domains (Szalma & Hancock, 2011), but these results indicate that the changes in noise that significantly affect LE are not sufficient to lead to impairments in the visual task when completed in isolation.

The second notable finding is that although participants expended more LE at more challenging SNRs, superior cognitive capacity did not mitigate the detrimental effects of louder background noise as we hypothesised. These null effects are not likely attributable to poorly measured constructs – when we have implemented these specific versions of the complex dual-task and the Rspan task in our lab, they correlated with performance on other tasks as expected. For example, complex dual-task scores were significantly correlated with other response time measures, and Rspan scores were correlated with other recall-based tasks (Strand, Brown, Merchant, et al., 2018). The average values and ranges of the complex dual-task response times and Rspan scores shown here are nearly identical to those in our prior work, suggesting that the lack of relationship between them is not attributable to restricted range in the current dataset. Even if the relationship between LE and WM capacity differs as a function of SNR, the very large range of SNRs employed

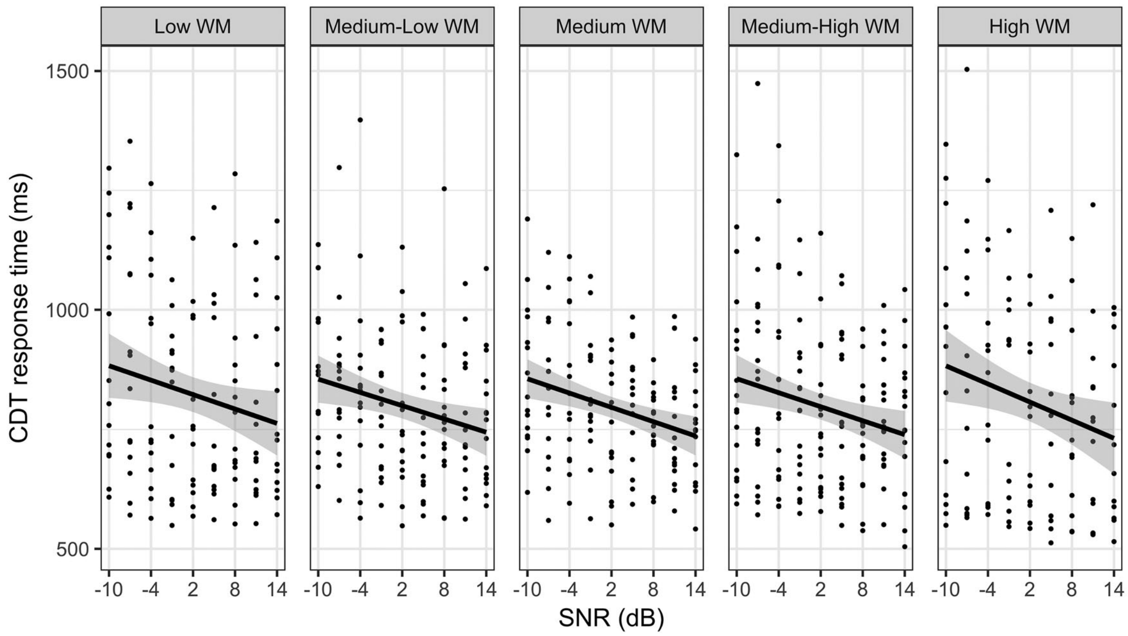

Figure 3. Scatterplots showing the relationship between SNR and complex dual-task (CDT) response time, faceted by WM scores. Note that WM scores were separated into bins for visualisation purposes, but measured continuously in the analyses. WM bins were chosen such that there were approximately equal numbers of participants in each of the five bins, resulting in breaks at Rspan scores of 54%, 62%, 75%, and 88%. The negative slopes of the lines indicate that participants tended to respond to the visual task more quickly (expended less LE) in easier listening situations. The fact that the slopes are similar in each panel indicates that the relationship between CDT response time and SNR is relatively consistent across a range of WM scores.

here, spanning floor- to ceiling-level performance in the word recognition task, should have provided ample opportunity for LE and WM to interact. Finally, the lack of relationship between Rspan scores and LE is not likely to be a function of low power, as the number of trials in each condition was well above the number recommended for response time data of this nature (Brysbaert & Stevens, 2018).

How might we reconcile the lack of relationship between WM and LE with prior research that has indicated the opposite? We posit that the relationship between WM and LE may be highly dependent on the measure of LE the researchers employ. Previous studies that have demonstrated a link between LE and Rspan have generally measured LE using recall-based tasks (e.g. listening span, cognitive spare capacity, sentence-final word recall; Besser et al., 2013; Daneman & Carpenter, 1980; Mishra et al., 2013a; Ng et al., 2013; Strand, Brown, Merchant, et al., 2018) rather than dual-task paradigms. Recall measures of LE require that listeners manipulate and store aurally-presented items briefly in memory. Although these tasks are assumed to measure LE, they also draw heavily on domain-general WM ability. Thus, the correlation between recall-based measures of LE and WM tasks may reflect the WM capacity that both rely on, rather than the true effect of WM capacity on the amount of effort listeners expend. In support of this idea, in Strand, Brown, Merchant, et al. (2018) we found that all three recall-based measures of LE correlated with performance on one or more of the WM tasks we included (Rspan or letter memory, which involves updating serially presented visual letters). In contrast, neither of the two dual-task measures of LE were correlated with either measure of WM capacity. Taken together, these results suggest that recall and dual-task measures of LE differ in the extent to which they rely on WM.

To our knowledge, only one study using a dual-task paradigm has found a relationship between WM capacity and LE, and this study included older adults in their sample (Desjardins & Doherty, 2013). It is conceivable that the relationship emerged in older adults because the difficulty of the task for this age group necessitated the recruitment of additional cognitive resources that were not needed for young adult listeners. That is, because older adults' listening conditions are intrinsically more adverse, the ELU model would posit that the greater mismatch between the acoustic-phonetic input and the representations stored in the listeners' memories would require the recruitment of more resources than would be required by young adults (just as processing speech in more difficult SNRs would require more resources; Rönnberg et al., 2013). Further, since older adults have a reduced pool of cognitive resources at their disposal (e.g. Sommers, 1996), any differences in WM capacity may have a larger effect on LE – or even speech recognition accuracy – compared to young adults. However, given that the current study only included young adults in the sample, our results cannot directly speak to the relationship between WM capacity and LE for older adults. Future research should assess the relationship between Rspan and LE across the lifespan, thereby determining the extent to which the effects observed by Desjardins and Doherty (2013) were driven by the older adults in the sample, as well as the extent to which our results may apply to other samples.

It is also possible that the lack of relationship between LE and WM capacity is attributable to the particular stimulus materials we employed. It may be that monosyllabic words are not sufficiently cognitively taxing to require recruiting WM, so a relationship between LE and WM capacity may have emerged if we had included longer words or sentences (e.g. see Signoret, Johnsrude, Classon, & Rudner, 2018). Given that sentences contain contextual cues – both semantic and grammatical – about upcoming words, the greater role of prediction inherent in sentential processing may necessitate the recruitment of additional cognitive resources, including WM, relative to lexical processing alone. However, sentences with low-predictability words have been shown to be more sensitive to changes in SNR than those with high-predictability words, suggesting that semantic context may reduce LE (Strand, Brown, Merchant, et al., 2018). Further, given that monosyllabic words tend to have more lexical competitors than longer words (see Vitevitch & Luce, 2016), monosyllabic word recognition in noise may be expected to be more cognitively taxing than recognising longer words. It therefore is not clear whether a task with sentences or longer words would be expected to put greater strain on WM. Thus, if a relationship between LE and WM capacity exists in healthy, young adults with high cognitive ability, a monosyllabic word recognition task in high levels of background noise should be sufficiently difficult for it to emerge.

Following the procedures of Sarampalis et al. (2009), the visual and speech stimuli were presented at random intervals and thus were not synchronised. The fact that RTs were still slower at more difficult SNRs suggests that the effects of LE are not so transient that they disappear during small delays between when LE is induced and when it is measured. Future research should evaluate whether effects of SNR on complex dual-task RTs are even stronger if the visual and speech stimuli are presented concurrently. Such research could shed light on the persistence of LE effects and the time course with which they decay.

Finally, it is worth noting that the pattern of results reported here may be specific to the steady-state speech-shaped noise we employed. This type of noise results in energetic masking, which occurs when the target and masker contain simultaneous overlapping frequency bands, rendering the target speech inaudible due to low-level sensory interference at the auditory periphery (Brungart, Simpson, Ericson, & Scott, 2001). In contrast, informational masking impairs intelligibility of the target speech due to difficulty segregating the target from intelligible background noise (e.g. two-talker babble), despite near-perfect peripheral audibility (cf. Freyman, Helfer, McCall, & Clifton, 1999). Informational masking is typically attributed to higher-level cognitive processing, so may necessitate the recruitment of additional cognitive resources to resolve stimulus uncertainty beyond those recruited in the presence of energetic masking. A relationship between LE and WM capacity may therefore emerge in the more cognitively demanding listening conditions induced by informational masking. Future research should explore this possibility, and should additionally explore the effects of various masking types on LE (e.g. informational masking, steady-state energetic masking, modulating energetic masking; see Koelewijn et al., 2012; Mishra et al., 2013a).

The results of the current study suggest that the four hypotheses described above (effort, resource, cognitive efficiency, and ELU) may oversimplify the apparently tenuous relationship between cognitive abilities and LE. The lack of a main effect of Rspan indicates that we did not find support for the cognitive efficiency or effort hypotheses, and the lack of an interaction between Rspan and SNR indicates that we did not find support for the resource or ELU hypotheses. Of course, these results may only apply to the particular measures of cognitive ability and LE used here, so future research should employ different tasks across a range of SNRs to shed light on the interactive effects of cognitive ability and listening difficulty on LE. Further, these results may only apply to the specific population we had in the current study; young, motivated college students with normal hearing may not need to recruit WM capacity to complete the listening task, even in difficult SNRs.

Prior research has robustly indicated that some tasks assumed to measure LE are correlated with WM capacity (Ng et al., 2013; Strand, Brown, Merchant, et al., 2018). Here we demonstrate a situation in which this relationship is absent. The results support the use of the complex dual-task as a measure of LE, given that response times vary predictably with changes in the level of the background noise, but only in the presence of speech. Despite well-validated measures of LE and WM capacity, a well-powered design, and a wide range of SNRs, the relationship between WM capacity and LE failed to emerge in our sample of normal-hearing young adults. Thus, these results suggest caution in making the general assumption that cognitive ability (as measured by WM capacity) is related to the amount of effort expended during a listening task.

Note

1. An exploratory analysis that included the twenty-two eliminated participants was also conducted out of concern that removing those participants might mask any effects of cognitive overload. That analysis revealed the same conclusions as the analysis reported here.

Acknowledgements

Carleton College supported this work. We are grateful to Naseem Dillman-Hasso, Kate Finsteun-Magro, Alexander Frieden, Maryam Hedayati, Sasha Mayn, Lucia Ray, Julia Smith, Janna Wennberg, and Annie Zanger for assistance with data collection and helpful conversations, and to the Engle lab for providing Reading Span materials.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by Carleton College.

ORCID

Violet A. Brown http://orcid.org/0000-0001-5310-6499
Julia F. Strand http://orcid.org/0000-0001-5950-0139

References

Adank, P., Evans, B. G., Stuart-Smith, J., & Scott, S. K. (2009). Comprehension of familiar and unfamiliar native accents under adverse listening conditions. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(2), 520–529.

Ahern, S., & Beatty, J. (1979). Pupillary responses during information processing vary with scholastic aptitude test scores. *Science*, *205*(4412), 1289–1292.

Arehart, K. H., Souza, P., Baca, R., & Kates, J. M. (2013). Working memory, age and hearing loss: Susceptibility to hearing aid distortion. *Ear & Hearing*, *34*(3), 251–260.

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R., Singmann, H., … Green, P. (2014). *Package "lme4."* R foundation for statistical computing, Vienna, 12. Retrieved from https://github.com/lme4/lme4/

Besser, J., Koelewijn, T., Zekveld, A. A., Kramer, S. E., & Festen, J. M. (2013). How linguistic closure and verbal working memory relate to speech recognition in noise—a review. *Trends in Amplification*, *17*(2), 75–93.

Bourland-Hicks, C., & Tharpe, A. M. (2002). Listening effort and fatigue in school-age children with and without hearing loss. *Journal of Speech, Language, and Hearing Research: JSLHR*, *45*(3), 573–584.

Brungart, D. S., Simpson, B. D., Ericson, M. A., & Scott, K. R. (2001). Informational and energetic masking effects in the perception of multiple simultaneous talkers. *The Journal of the Acoustical Society of America*, *110*(5 Pt 1), 2527–2538.

Brysbaert, M., & Stevens, M. (2018). Power analysis and effect size in mixed effects models: A tutorial. *Journal of Cognition*, *1*(1), 9.

Conway, A. R. A., 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.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, *19*, 450–466.

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.

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.

Downs, D. W. (1982). Effects of hearing aid use on speech discrimination and listening effort. *The Journal of Speech and Hearing Disorders*, *47*(2), 189–193.

Foo, C., Rudner, M., Rönnberg, J., & Lunner, T. (2007). 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*, *18*(7), 618–631.

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

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

Fry, A. F., & Hale, S. (1996). Processing speed, working memory, and fluid intelligence: Evidence for a developmental cascade. *Psychological Science*, *7*(4), 237–241.

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.

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

George, E. L. J., Zekveld, A. A., Kramer, S. E., Goverts, S. T., Festen, J. M., & Houtgast, T. (2007). Auditory and nonauditory factors affecting speech reception in noise by older listeners. *The Journal of the Acoustical Society of America*, *121*(4), 2362–2375.

Gordon-Salant, S., & Cole, S. S. (2016). Effects of age and working memory capacity on speech recognition performance in

noise among listeners with normal hearing. *Ear and Hearing*, *37*(5), 593–602.
Gosselin, P. A., & Gagné, J.-P. (2011a). Older adults expend more listening effort than young adults recognizing audiovisual speech in noise. *International Journal of Audiology*, *50*(11), 786–792.
Gosselin, P. A., & Gagné, J.-P. (2011b). Older adults expend more listening effort than young adults recognizing speech in noise. *Journal of Speech Language and Hearing Research*, *54*(3), doi:10.1044/1092-4388(2010/10-0069.
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*. Englewood Cliffs, NJ: Prentice-Hall.
Koelewijn, T., Zekveld, A. A., Festen, J. M., Rönnberg, J., & Kramer, S. E. (2012). Processing load induced by informational masking is related to linguistic abilities. *International Journal of Otolaryngology*, *2012*, 865731.
Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology*, *49*(4), 764–766.
Lunner, T. (2003). Cognitive function in relation to hearing aid use. *International Journal of Audiology*, *42*(Suppl 1), S49–S58.
Mattys, S. L., Davis, M. H., Bradlow, A. R., & Scott, S. K. (2012). Speech recognition in adverse conditions: A review. *Language and Cognitive Processes*, *27*(7–8), 953–978.
McCoy, S. L., Tun, P. A., Cox, L. C., Colangelo, M., Stewart, R. A., & Wingfield, A. (2005). Hearing loss and perceptual effort: Downstream effects on older adults' memory for speech. *The Quarterly Journal of Experimental Psychology. A, Human Experimental Psychology*, *58*(1), 22–33.
Mishra, S., Lunner, T., Stenfelt, S., Rönnberg, J., & Rudner, M. (2013a). Seeing the talker's face supports executive processing of speech in steady state noise. *Frontiers in Systems Neuroscience*, *7*, 96.
Mishra, S., Lunner, T., Stenfelt, S., Rönnberg, J., & Rudner, M. (2013b). Visual information can hinder working memory processing of speech. *Journal of Speech, Language, and Hearing Research*, *56*, 1120–1132.
Ng, E. H. N., Rudner, M., Lunner, T., Pedersen, M. S., & Rönnberg, J. (2013). Effects of noise and working memory capacity on memory processing of speech for hearing-aid users. *International Journal of Audiology*, *52*(7), 433–441.
Ohlenforst, B., Wendt, D., Kramer, S. E., Naylor, G., Zekveld, A. A., & Lunner, T. (2018). Impact of SNR, masker type and noise reduction processing on sentence recognition performance and listening effort as indicated by the pupil dilation response. *Hearing Research*, *365*, 90–99.
Pichora-Fuller, M. K., Kramer, S. E., Eckert, M. A., Edwards, B., Hornsby, B. W. Y., Humes, L. E., … Wingfield, A. (2016). Hearing impairment and cognitive energy: The framework for understanding effortful listening (FUEL). *Ear and Hearing*, *37*(Suppl 1), 5S–27S.
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.
Picou, E. M., Gordon, J., & Ricketts, T. A. (2016). The effects of noise and reverberation on listening effort in adults with normal hearing. *Ear and Hearing*, *37*(1), 1–13.
Picou, E. M., & Ricketts, T. A. (2014). The effect of changing the secondary task in dual-task paradigms for measuring listening effort. *Ear and Hearing*, *35*(6), 611–622.
Rabbitt, P. M. (1968). Channel-capacity, intelligibility and immediate memory. *The Quarterly Journal of Experimental Psychology*, *20*(3), 241–248.
Rabbitt, P. M. (1991). Mild hearing loss can cause apparent memory failures which increase with age and reduce with IQ. *Acta Oto-Laryngologica. Supplementum*, *476*, 167–175. discussion 176.
Redick, T. S., Broadway, J. M., Meier, M. E., Kuriakose, P. S., Unsworth, N., Kane, M. J., & Engle, R. W. (2012). Measuring working memory capacity with automated complex span tasks. *European Journal of Psychological Assessment: Official Organ of the European Association of Psychological Assessment*, *28*(3), 164–171.
Rönnberg, J. (2003). Cognition in the hearing impaired and deaf as a bridge between signal and dialogue: A framework and a model. *International Journal of Audiology*, *42*, S68–S76.
Rönnberg, J., Lunner, T., Zekveld, A., Sörqvist, P., Danielsson, H., Lyxell, B., … Rudner, M. (2013). The ease of language understanding (ELU) model: Theoretical, empirical, and clinical advances. *Frontiers in Systems Neuroscience*, *7*, 31.
Rönnberg, J., Rudner, M., Foo, C., & Lunner, T. (2008). Cognition counts: A working memory system for ease of language understanding (ELU). *International Journal of Audiology*, *47*(sup2), S99–S105.
Rudner, M., Lunner, T., Behrens, T., Thorén, E. S., & Rönnberg, J. (2012). Working memory capacity may influence perceived effort during aided speech recognition in noise. *Journal of the American Academy of Audiology*, *23*(8), 577–589.
Rudner, M., Rönnberg, J., & Lunner, T. (2011). Working memory supports listening in noise for persons with hearing impairment. *Journal of the American Academy of Audiology*, *22*(3), 156–167.
Sarampalis, A., Kalluri, S., Edwards, B., & Hafter, E. (2009). Objective measures of listening effort: Effects of background noise and noise reduction. *Journal of Speech, Language, and Hearing Research: JSLHR*, *52*(5), 1230–1240.
Signoret, C., Johnsrude, I., Classon, E., & Rudner, M. (2018). Combined effects of form- and meaning-based predictability on perceived clarity of speech. *Journal of Experimental Psychology. Human Perception and Performance*, *44*(2), 277–285.
Sommers, M. S. (1996). The structural organization of the mental lexicon and its contribution to age-related declines in spoken-word recognition. *Psychology and Aging*, *11*(2), 333–341.
Souza, P. E., Arehart, K. H., Shen, J., Anderson, M., & Kates, J. M. (2015). Working memory and intelligibility of hearing-aid processed speech. *Frontiers in Psychology*, *6*, 526.
Strand, J. F., Brown, V. A., & Barbour, D. L. (2018). Talking points: A modulating circle reduces listening effort without improving speech recognition. *Psychonomic Bulletin & Review*, doi:10.3758/s13423-018-1489-7.
Strand, J. F., Brown, V. A., Merchant, M. M., Brown, H. E., & Smith, J. (2018). Measuring listening effort: Convergent validity, sensitivity, and links with cognitive and personality measures. *Journal of Speech, Language, and Hearing Research: JSLHR*, *61*, 1463–1486.
Surprenant, A. M., & Watson, C. S. (2001). Individual differences in the processing of speech and nonspeech sounds by normal-hearing listeners. *The Journal of the Acoustical Society of America*, *110*(4), 2085–2095.

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

van der Meer, E., Beyer, R., Horn, J., Foth, M., Bornemann, B., & Ries, J. (2010). Resource allocation and fluid intelligence: Insights from pupillometry. *Psychophysiology*, *47*, 158–169.

Van Engen, K. J., Chandrasekaran, B., & Smiljanic, R. (2012). Effects of speech clarity on recognition memory for spoken sentences. *PloS One*, *7*(9), e43753.

van Rooij, J. C. G. M., Plomp, R., & Orlebeke, J. F. (1989). Auditive and cognitive factors in speech perception by elderly listeners. I: Development of test battery. *The Journal of the Acoustical Society of America*, *86*, 1294–1309.

Vitevitch, M. S., & Luce, P. A. (2016). Phonological neighborhood effects in spoken word perception and production. *Annual Review of Applied Linguistics*, *2*(1), doi:10.1146/annurev-linguist-030514-124832.

Wingfield, A., Tun, P. A., & McCoy, S. L. (2005). Hearing loss in older adulthood: What it is and how it interacts with cognitive performance. *Current Directions in Psychological Science*, *14*(3), 144–148.

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), 660–670.

Zekveld, A. A., Festen, J. M., & Kramer, S. E. (2013). Task difficulty differentially affects two measures of processing load: The pupil response during sentence processing and delayed cued recall of the sentences. *Journal of Speech, Language, and Hearing Research: JSLHR*, *56*(4), 1156–1165.

Zekveld, A. A., George, E. L. J., Kramer, S. E., Goverts, S. T., & Houtgast, T. (2007). The development of the text reception threshold test: A visual analogue of the speech reception threshold test. *Journal of Speech, Language, and Hearing Research: JSLHR*, *50*(3), 576–584.

Zekveld, A. A., & Kramer, S. E. (2014). Cognitive processing load across a wide range of listening conditions: Insights from pupillometry. *Psychophysiology*, *51*(3), 277–284.

Zekveld, A. A., Rudner, M., Johnsrude, I. S., Festen, J. M., van Beek, J. H. M., & 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.