**Divided attention disrupts perceptual encoding during speech recognition**

Sven L. Mattys

and

Shekeila D. Palmer

University of York

Department of Psychology, Heslington, York, YO10 5DD, United Kingdom

Corresponding author:

Sven Mattys

Department of Psychology

University of York

York, YO10 5DD

United Kingdom

sven.mattys@york.ac.uk

Abstract

Performing a secondary task while listening to speech has a detrimental effect on speech processing, but the locus of the disruption within the speech system is poorly understood. Recent research has shown that cognitive load imposed by a concurrent visual task increases dependency on lexical knowledge during speech processing, but it does not affect lexical activation per se. This suggests that ‘lexical drift’ under cognitive load occurs either as a post-lexical bias at the decisional level or as a secondary consequence of reduced perceptual sensitivity. This study aimed to adjudicate between these alternatives using a forced-choice task that required listeners to identify noise-degraded spoken words with or without the addition of a concurrent visual task. Adding cognitive load increased the likelihood that listeners would select a word acoustically similar to the target even though its frequency was lower than that of the target. Thus, there was no evidence that cognitive load led to a high-frequency response bias. Rather, cognitive load seems to disrupt sublexical encoding, possibly by impairing perceptual acuity at the auditory periphery.

**INTRODUCTION**

For models of human speech recognition, an account of how the system operates under suboptimal conditions is essential. Outside of the laboratory, it is seldom the case that listeners receive clear, uninterrupted speech signals in the absence of unwanted or distracting information. Instead, the speech recognizer is continually challenged by factors such as speaker variability, disfluencies, signal degradation, and the presence of distracting input. Although the factors affecting the quality of the speech input (i.e., noise, speaker variability, competing talkers) have received considerable attention (e.g., Luce & Pisoni, 1998; Samuel & Kraljic, 2009), far less is understood about the impact of attentional distraction on speech recognition.

In everyday life, speech processing commonly occurs under conditions of cognitive load. Here, we define cognitive load as a concurrent attentional or mnemonic task which places demand on central processing resources, thereby limiting the resources available for speech recognition (Mattys & Wiget, 2011). Within that definition, increased cognitive load can be elicited by contrasting focused attention (one stimulus) with selective attention (ignoring a concurrent stimulus), or focused attention with divided attention (processing both the target and the concurrent stimuli), or levels of divided attention in which the difficulty of a competing task is manipulated. Although speech processing might appear to proceed 'automatically’, a number of studies have shown that cognitive load has a detrimental effect on tone threshold perception (MacDonald & Lavie, 2011), phoneme perception (Casini, Burle, & Nguyen, 2009; Gordon, Eberhardt, & Rueckl, 1993), word segmentation (Fernandes, Kolinsky, & Ventura, 2010), and the ability to selectively attend to a single talker in a multi-talker environment (Francis, 2010). Findings such as these indicate that some aspects of speech processing are resource demanding, and hence vulnerable to attentional disruption, but the precise locus of the disruption within the speech recognition system is unclear.

As well as ‘bottom-up’ signal-driven processing, speech recognition is known to involve ‘top-down’ knowledge-driven processes, and the interplay between these components is likely to differ according to the listening conditions. For example, lexical top-down repair mechanisms can help compensate for degradation in the speech signal, although this process is dependent on the state of the signal itself (Baskent, 2012; Saija, Akyurek, Andringa, & Baskent, 2014). Increased reliance on lexical knowledge is also found under cognitive load (Mattys et al., 2009; Mattys & Wiget, 2011). In one such study, Mattys and Wiget (2011) examined the impact of a concurrent visual search task on the ‘Ganong’ effect (Ganong, 1980). The Ganong effect provides a measure of lexical influence on phoneme perception. For example, an ambiguous phoneme between /g/ and /k/ is more likely to be categorized as a /g/ if it is appended by ‘ift’ (i.e., gift) as opposed to ‘iss’ (i.e., giss), and as a /k/ if it is appended by ‘iss’ (i.e., kiss) as opposed to ‘ift’ (i.e., kift). The standard explanation is that ‘gift’ and ‘kiss’ activate lexical representations believed to support phoneme perception, whereas 'giss' and 'kift' are not represented in the lexicon, and therefore do not provide top-down facilitation. The Ganong effect emerges despite the fact that listeners are instructed to ignore the lexical status of the stimulus. Mattys and Wiget (2011) showed that performing a visual search task simultaneously with the phoneme categorization task inflated the Ganong effect, indicating that the relative contribution of lexical knowledge to phoneme categorisation increases when attentional resources are limited. They referred to the stronger lexical influence on tasks performed under cognitive load as ‘lexical drift’. The authors outlined three possible loci for the lexical drift effect. The first possibility is that lexical drift has a lexical locus, resulting from either increased lexical activation or a reduction in the recognition threshold for lexical representations. The second possibility is that the locus is sublexical, in that cognitive load impairs perceptual encoding of sublexical units during speech recognition, a stage often referred to as acoustic-to-phonetic mapping. In this case, increased reliance on lexical knowledge would occur only as a secondary consequence of impoverished acoustic encoding. A third alternative is that the impact of cognitive load is post-lexical, resulting from a lexical bias at the decisional stage. Here, lexical drift would be the outcome of a strategic attempt to make lexical-semantic sense of the signal under cognitive load, with no actual changes at the sublexical level or in lexical activation.

In a recent follow-up study, Mattys, Barden, and Samuel (2014) attempted to distinguish between the lexical and sublexical alternatives using a variant of the phoneme restoration paradigm (Warren, 1970). The phoneme restoration effect refers to a tendency for listeners to perceptually restore missing segments of speech which have been replaced by noise. Samuel (1981) developed the paradigm to assess the influence of lexical knowledge on phoneme perception. In Samuel's experiments, both words and nonwords were used as stimuli and participants were required to discriminate items that had a phoneme replaced by noise from items that simply had noise superimposed onto the critical phoneme. Participants’ discrimination performance was thought to reflect their ability to access acoustic detail in the speech signal, with low discrimination (a high restoration rate) being indicative of poor perceptual sensitivity. Samuel observed that discrimination performance was worse in words than nonwords, possibly because lexical knowledge took over (or obviated the need for) fine phonetic analysis. This result was taken to indicate that phoneme perception is influenced by top-down feedback. Using a similar restoration technique, Mattys et al. (2014) examined the effect of cognitive load on the interplay between phoneme discrimination and lexical knowledge. They reasoned that if cognitive load increases lexical activation, the standard lexical effect on phoneme restoration should be amplified under cognitive load (i.e., increased lexical activation under load should amplify lexically-driven restoration). However, they observed that although cognitive load caused a general reduction in perceptual sensitivity relative to a no load control condition, the performance decrement was comparable in words and non-words. This result was taken as evidence that cognitive load disrupts acoustic processing at the sublexical level, but it does not affect lexical activation per se.

Although these findings suggest that the impact of cognitive load on speech recognition is manifest in the early analysis of the signal, they leave several questions unanswered. The aim of this study is to address two of them. First, while the restoration data did not reveal an increase in lexical activation under cognitive load, post-lexical reporting biases were not considered. It is plausible that, in addition to impairing sublexical processing, cognitive load induces a bias to report lexically viable responses. In one of their experiments (Experiment 4), Mattys and Wiget (2011) attempted to address this issue by asking participants to decide whether each of the gift-to-kift and giss-to-kiss stimuli from the Ganong experiment was an acceptable token of the word gift and kiss, respectively. The task, which was meant to measure overt lexical biases, was performed under cognitive load and no cognitive load. The results showed no evidence of elevated acceptability judgement under cognitive load, thus theoretically undermining a post-lexical explanation. However, it is possible that the task, which had never been used before, lacked the sensitivity needed to capture a change in acceptability criterion. The lack-of-sensitivity issue could have been further amplified by the fact that cognitive load led to less pronounced acceptability judgements for both the word and nonword end-points of the continuum. Thus, a general decrease in confidence could have masked any post-lexical bias. Therefore, the first question is whether it is possible to identify a post-lexical bias in response to cognitive load when potentially interfering factors are minimized. Second, while the effects shown in previous studies were interpreted as being the result of a change in perceptual processes, they could also be the result of increased randomness in the listener's responses, with no fundamental change in perceptual processes. Indeed, in all the studies reported above, a reduction in perceptual precision coincided with poorer performance on the task. It was therefore impossible to decouple genuinely weakened perceptual processes from load-induced disengagement from the task. Thus, the second question is how much of a listener’s behavior under cognitive load is due to random responding as opposed to a change in perceptual sensitivity. To answer this question, we designed an experiment that dissociated reduced perceptual sensitivity from accuracy.

In the following experiment, participants heard a series of words played in masking noise. Noise was used to prevent a ceiling effect in word identification. With no a priori hypothesis about the level of noise required for cognitive load effects to emerge, we elected to play the stimuli in two noise levels: -6 dB (lower noise) and -12 dB (higher noise) signal to noise ratio (SNR). Following each noise-masked word (e.g., *peak*), listeners made a three-alternative forced choice to indicate what they heard. On critical trials, the possible response options included two foils: (1) a word that had a lower lexical frequency than the target but was acoustically close to it (acoustic response option, e.g., *teak*), and (2) a word that had a higher lexical frequency than the target but was comparatively less acoustically similar to it (high lexical frequency response option, e.g., *feet*). The third response option was ‘neither’, which on critical trials was the correct answer. The cognitive load was a visual search task previously used by Mattys and colleagues (Mattys & Wiget, 2011; Mattys et al., 2014), in which participants had to detect a red square in an array of red triangles and black squares. The array was displayed concurrently with the auditory stimulus and for the same duration.

It was hypothesized that if cognitive load leads to a decisional lexical bias, this should be reflected in an increased tendency for participants to select the high frequency (lexical) response option under conditions of cognitive load. On the other hand, if cognitive load leads to a decrease in auditory acuity, then tolerance to acoustic imprecision should be elevated under cognitive load, and hence, participants should be more likely to select the acoustically similar option, even though it has a lower lexical frequency. Note that, by pitting a frequency bias against an acoustic bias, this design forces the listeners to reveal which of the two is more prominent under cognitive load. With respect to the possibility that cognitive load might simply increase response randomness, this should manifest in a flatter distribution of responses across the three options, and an attenuation rather than an enhancement of any bias observed under no load.

**II METHOD**

**A. Participants**

Sixty-one students (46 female) participated in the experiment in return for either course credit or payment. All were British native English speakers with no known hearing difficulties and normal or corrected vision. Participant ages ranged from 18 to 27 years (mean age 20 years). In accordance with the University of York ethics regulations, informed consent was obtained from all participants before they took part in the experiment.

**B. Stimuli**

The critical stimuli consisted of 68 Consonant-Vowel-Consonant (CVC) triplets. Each triplet contained a target word, an acoustically similar low frequency word (LF or ‘acoustic’ foil) and a less acoustically similar high frequency word (HF or 'lexical' foil). Target words had a mean frequency of 412.13, LF words had a mean frequency of 59.12, and HF words had a mean frequency of 6727.56 (frequency ratings obtained from CELEX lexical database, 1995). In instances where the target word was a homophone, the highest frequency value was used. Acoustic similarity between the target and the LH and HF words was established based on Luce’s (1986) phoneme confusion matrices. Luce’s confusion matrices were established by tabulating listeners’ transcriptions of CV and VC nonsense syllables played in various noise levels. The LF words, selected on the basis that they were highly confusable with the target word, differed from the target word by one phoneme, and the replacement phoneme was highly confusable with the target phoneme. Based on Luce's matrices, a target phoneme was deemed 'highly confusable' with a replacement phoneme when more than 10% of tokens of the target phoneme were identified as the replacement phoneme in Luce’s -5 dB SNR matrix. The HF words, chosen on the basis that they were less confusable with the target words, differed from the target by two phonemes. One replacement phoneme was moderately confusable with the target phoneme (3% to 10% confusion), whereas the other could be any phoneme. Within the 68 triplets, the confusable phoneme could occur in the onset position (27 cases), nucleus position (18 cases), or in the coda position (23 cases). In addition to the critical stimuli, there were 40 filler stimuli. These consisted of CVC word pairs containing a target word (20 high frequency and 20 low frequency) and a distractor word. Distractor words had one phoneme in the same syllable position as the target. The confusability of the phonemes in the distractor words was not controlled. All stimuli are presented in the Appendix.

The spoken stimuli were recorded by a female speaker of standard Southern British English who was instructed to read the words one at a time in a citation form. Each word was recorded twice and the clearer of the two tokens was selected. The average word duration was 595 ms. Recording took place in a sound attenuated booth at a sampling rate of 44.1 kHz with 16-bit resolution. Noise was added at -6 dB SNR and -12 dB SNR to each of the stimuli. Noise consisted of white noise low-pass filtered at 5 kHz. The 5 kHz filter was meant to efficiently mask the average frequency spectrum of speech without the discomfort of high-frequency white noise. Noise intensity was set at 76 dB SPL. The average intensity of the stimuli was normalized to either 70 dB SPL or 64 dB SPL, creating the -6 dB SNR and -12 dB SNR conditions, respectively. Noise started 50 ms prior to the onset of the word and continued for 50 ms after its offset. Noise intensity was linearly ramped up during the lead time and gradually ramped down during the lag time.

The visual arrays in the cognitive load condition were the same as those used by Mattys and Wiget (2011) and consisted of 6 x 6 grids made up of black squares and red triangles appearing in a random configuration on a white background. Half of the grids contained a red square that the participant was asked to detect. The red square could appear anywhere within the grid. The size of the grid on the computer monitor was approximately 9x9 cm, and the size of each square or triangle was approximately 5x5 mm (for an example, see Mattys & Wiget, 2011). The distance between the participants' eyes and the monitor was approximately 50 cm. Stimuli were presented and responses collected using DMDX display software (Forster & Forster, 2003) and testing took place in a sound-attenuated booth. The auditory stimuli were played over Sony MDR V700 headphones.

**C. Design and Procedure**

The experiment followed a 2 x 2 design with Cognitive Load (Cognitive Load vs. No Load) and Noise (-6 dB SNR vs. -12 dB SNR) as within-subjects factors. The Cognitive Load factor was blocked, so that half of the participants completed the Cognitive Load condition first, whereas the other half completed the No Load condition first. The two blocks were separated by a break interval paced by the participant. Within each block, Noise was randomized from trial to trial, with half of the target stimuli presented at -6 dB SNR and the other half at -12 dB SNR. The stimuli appearing in each of the four conditions were counterbalanced between participants, such that participants never heard the same word more than once. Each participant was tested in a single session which lasted approximately 20 minutes.

Each trial started with the presentation of a spoken word in noise. Following a 50 ms interval, three response options appeared on the computer monitor. On critical trials, these consisted of the LF and HF words from the triplet appearing on the left and right hand side of the screen (position counterbalanced across trials), and a ‘Neither’ option appearing top center. On filler trials, the target word and the distractor word appeared in place of the LF and HF options. The ‘Neither’ option appeared in the same position as in the critical trials. Participants were required to indicate their response by pressing one of three keys on a computer keyboard mimicking the position of the three options on the monitor (C, left option; M, right option; Y, ‘Neither’ option). Participants had up to 20 s to give an answer before the program moved on to the next trial. A 2-s interval separated trials.

During the cognitive load block, a visual array appeared on the screen 50 ms after the onset of the sound file, to coincide with the onset of the target word after the 50 ms noise lead. The array remained on the screen for 692 ms which was the duration of the longest word. Participants were instructed to search for the red square in the array while simultaneously listening to the target. After indicating which word they heard, a second screen appeared asking the participant to indicate whether or not they had seen a red square in the visual array. Participants had 20 s to respond ‘no’ or ‘yes’, using the left and right shift keys, respectively. The computer screen remained blank (white background) throughout the experiment, except when the visual arrays or response screens were presented. No feedback was provided and participants did not have the opportunity to correct their responses.

**D. Data analysis**

The data were analyzed using mixed effects logistic regression models with Cognitive Load and Noise as fixed factors. Although the level of noise (-6 vs -12 dB SNR) was only an exploratory factor, we decided to include it in the analyses to document a potentially relevant variable for further research. For both fixed factors, the less challenging condition was coded -1 (No Load and Low Noise) and the more challenging condition was coded 1 (Load and High Noise). For all models, we included by-subject and by-item intercepts, as well as by-subject random slopes for Cognitive Load and Noise. By-item random slopes for Cognitive Load and Noise were initially considered, but the models failed to converge, so these two terms were removed. The fixed factors, as well as their interaction, were considered incrementally, starting from a base model that included only the random terms. Improved fit when a fixed-effect or an interaction term was added was assessed using the likelihood ratio test.

In the Cognitive Load condition, all trials were entered in the analyses, whether or not the response on the visual task was correct. We chose to include all trials rather than limiting the analyses to the trials in which the visual response was correct because incorrect visual responses do not necessarily imply a lack of engagement in the visual task. Note, however, that when analyses restricted to correct responses were run, they did not differ from the results reported below.

**III RESULTS**

The distribution of responses across cognitive load and noise conditions is shown in Figure 1a (Critical trials) and Figure 2a (Filler trials). These figures show that adding cognitive load did not lead to a flatter distribution of responses across categories, as would have been expected if cognitive load had simply increased randomness in responding. Instead Figure 1b suggests that the low frequency response bias present in the no load condition was enhanced under cognitive load. This pattern was explored relative to one of our main questions, namely, whether cognitive load leads to a decisional lexical bias or to a reduction in auditory acuity.

1. **Critical trials**

In relation to the hypothesis that cognitive load may induce a high-frequency bias in responding, the first contrast examined the overall proportion of HF responses relative to all other responses. This analysis showed that Cognitive Load did not increase the proportion of HF responses, *β* = -.010, *SE* = .051, *χ*2(1) = .03, *p =* .85 (Figure 1b). However, Noise did, *β* = .130, *SE* = .047, *χ2*(1) = 7.37, *p* = .007, with participants giving more HF responses under -12 dB SNR than under -6 dB SNR (Figure 1c). The interaction between Cognitive Load and Noise was not significant, *β* = -.031, *SE* = .047, *χ2*(1) = .43, *p* = .51. Thus, cognitive load did not increase the listeners’ tendency to report a high-frequency word, but the higher noise level did (-12 dB vs. -6 dB SNR).

The alternative possibility, that cognitive load may increase acoustic imprecision, was tested in a second set of analyses which examined the proportion of LF responses relative to all other responses. In these analyses, there were significant main effects of both Cognitive Load, *β* = .124, *SE* = .040, *χ2*(1) = 8.58, *p =* .003, and Noise, *β* = -.260, *SE* = .036, *χ2*(1) = 39.52, *p* < .001, indicating that cognitive load increased participants’ tendency to report a response acoustically similar to the target, even though it had lower lexical frequency (Figure 1b). Increasing noise had the opposite effect (Figure 1c). Including the interaction term between Cognitive Load and Noise did not significantly improve the fit of the model, *β* = -.054, *SE* = .036, *χ2*(1) = 2.21, *p* = .14.

Given that cognitive load did not affect the proportion of HF responses, the increase in proportion of LF responses under cognitive load was most likely associated with a corresponding decrease in the proportion of ‘neither’ responses. To test this, a third set of analyses was conducted examining the overall proportion of ‘neither’ responses. A significant main effect of Cognitive Load was observed, *β* = -.154, *SE* = .044, *χ2*(1) = 11.11, *p <* .001, confirming that participants made fewer ‘neither’ responses under cognitive load, relative to the no load condition (Figure 1b). There was also a significant main effect of Noise, *β* = .237, *SE* = .046, *χ2*(1) = 20.75, *p* < .001, with participants making more ‘neither’ responses under -12 dB SNR relative to -6 dB SNR (Figure 1c). Again, the interaction between Cognitive Load and Noise was not significant, *β* = -.038, *SE* = .041, *χ2*(1) = .83, *p* = .33. This analysis suggests that the effect of cognitive load consists of a change in the balance between LF (acoustic) responses and 'neither' responses, but not in a modulation of HF responses.

1. **Filler trials**

Unlike the critical trials, for which the correct answer was "neither," the filler trials contained the correct answer in one of the two response options. The proportion of correct responses on the filler trials was affected by both Cognitive Load, *β* = .206, *SE* = .054, *χ2*(1) = 13.27, *p* < .001 (Figure 2b), and Noise, *β* = -.521, *SE* = .057, *χ2*(1) = 56.53, *p* < .001 (Figure 2c), with participants making more correct responses in the Cognitive Load condition relative to the No Load condition, and under -6 dB than -12 dB SNR. The interaction between Cognitive Load and Noise was not significant, *β* = .084, *SE* = .057, *χ2*(1) = 2.14, *p* = .14. While it might seem counterintuitive that cognitive load should increase accuracy, it is noteworthy that cognitive load simultaneously decreased the proportion of ‘neither’ responses, *β* = -.623, *SE* = .061, *χ2*(1) = 16.91, *p* < .001 (Figure 2b). This pattern is similar to the one observed for the critical trials, where responses shifted from the ‘neither’ category to the LF acoustic foil under cognitive load. This indicates that, in both the test and filler trials, cognitive load caused a shift in the acoustic acceptance criterion, such that listeners were more likely to choose a response that sounded like the target (the acoustic foil in the critical trials and the actual target in the filler trials) under cognitive load. Cognitive Load did not impact the likelihood that the participant chose the distractor response, *β* = -0.021, *SE* = .106, *χ2*(1) = .03, *p* = .84 (Figure 2b).

Noise also influenced the distribution of responses on filler trials: Listeners were more likely to choose the distractor,  *β* = 0.569, *SE* = .111, *χ2*(1) = 22.36, *p* < .001, or ‘Neither’, *β* = .434, *SE* = .067, *χ2*(1) = 32.15, *p* < .001, under -12dB than -6dB SNR (Figure 2c). This pattern probably reflects increased response randomness under noisier conditions. Including an interaction term did not improve the fit of the model in any of the filler analyses (all *p* > .1).

On the visual task, participants obtained an average proportion correct of .76. Model comparisons similar to those reported above showed no effect of Noise, *β* = .029, *SE* = .073, *χ2*(1) = .15, *p =* .70, Trial Type (Critical vs Filler, coded as -1 and 1 , respectively), *β* = -.012, *SE* = .042, *χ2*(1) = .08, *p* = .77, or interaction, *β* = .058, *SE* = .043, *χ2*(1) = 1.75, *p* = .19.

**IV DISCUSSION**

Previous research has shown that limiting central processing resources by adding cognitive load during speech recognition leads to a ‘lexical drift’ (Mattys & Wiget, 2011). The purpose of the current study was to determine whether this increased reliance on lexical knowledge is the result of a decisional response bias, or whether it is a secondary consequence of impoverished acoustic encoding under cognitive load. During the experiment, participants heard a series of words presented at two different noise levels: -6 dB SNR and -12 dB SNR.

Following each word, listeners made a three-alternative forced choice to indicate what they heard. On critical trials, the response options were: (1) an *acoustic foil,* which was a word with lower lexical frequency than the target, differing from it by only one phoneme, (2) a *lexical foil,* which was a word with higher lexical frequency than the target, differing from it by more than one phoneme, and (3) a *neither* option, which was the correct response. Participants performed the task under cognitive load (a concurrent visual search task) or no cognitive load. It was hypothesized that if cognitive load induces a lexical bias at the decisional level, then this would be reflected in an increased tendency for participants under cognitive load to select responses that have a relatively high lexical frequency (the lexical foils). On the other hand, if cognitive load decreases auditory acuity, then participants should opt for similar-sounding responses, even if they have low lexical frequency (the acoustic foils).

Analyses showed that cognitive load did not affect the likelihood that participants would select a high frequency response, indicating that cognitive load does not create a lexical bias in responding. That is, cognitive load does not seem to encourage listeners to select a communicatively more likely response in a strategic attempt to alleviate the strain on processing resources. Instead, participants under cognitive load were more likely to select a similar-sounding, yet just as incorrect, response, suggesting increased acceptability for an acoustic competitor (rather than for the neither answer, which is both the correct response and a non-committal decision). This finding is consistent with Mattys et al.'s (2014) claim that cognitive load increases tolerance to acoustic imprecision. Moreover, the fact that cognitive load increased the already strong tendency to pick a similar-sounding response also allows us to reject the hypothesis that cognitive load simply leads to greater randomness in the participants’ responses. Had this been the case, responses would have been distributed more evenly, rather than less evenly, across response categories in the cognitive load condition.

Interestingly, increasing noise in this experiment had the opposite effect to adding cognitive load. In the noisier condition, participants made fewer acoustic responses whereas the proportion of high frequency and ‘neither’ responses increased. The bias for high-frequency responses under higher noise is generally consistent with previous research showing that top-down compensatory mechanisms are deployed when sensory information is severely compromised (Pichora-Fuller, 2008; Riecke, Esposito, Bonte, & Formisano, 2009; Sivonen, Maess, & Friederici, 2006; Zekveld, Heslenfeld, Festen, & Schoonhoven, 2006). Computationally (e.g., TRACE, McClelland & Elman, 1986; Shortlist, Norris, 1994) too, a highly degraded sensory input is shown to lead to more widespread lexical activation than a less degraded sensory input, with the relative frequency of the activated candidates becoming a more prominent decisional factor when the sensory evidence is incomplete. On the other hand, it is possible that louder noise simply increased randomness in the participants’ responses, without fundamentally altering the input-to-lexicon mapping process. Since the design of the experiment induced a general bias for responses that were acoustically similar to the target (the acoustic foils), the attenuation of this bias under high noise could indeed reflect an increased tendency to ‘guess’ when the target stimulus was ambiguous.

The contrasting effect of cognitive load and noise was also evident on filler trials, where the correct response was always one of the two words presented. As would be expected, increasing the noise level decreased the proportion of correct responses, and increased the proportion of distractor and ‘neither’ responses. Adding cognitive load had the opposite effect: Cognitive load increased the proportion of correct responses and decreased the proportion of ‘neither’ responses. As proposed earlier, this pattern is consistent with the idea that cognitive load broadens phonemic categories (or blurs their boundaries), and hence, increases the likelihood of accepting percepts that fall within a broader range of acoustic similarity. On a computational level, this can be achieved if cognitive load acts as a reducer of lateral inhibition between phoneme nodes (see Mirman, McClelland, Holt, and Magnuson, 2008, for a similar approach to attentional modulation of lexical knowledge). An impaired ability for a phoneme node to inhibit neighbouring nodes has indeed been shown to be one of the consequences of a general reduction of activation across the phoneme layer (McClelland & Elman, 1986).

The underlying mechanism linking cognitive load to reduced activation across the phoneme layer is unclear at this stage, but possibilities can be sought both in higher-order temporal computation and in sound sensitivity at the auditory periphery. In relation to the temporal computation of the signal, Casini, Burle, and Nguyen (2009) have proposed that a cognitively controlled central timer is used to estimate segment duration during speech perception. When central processing resources are limited by a competing task, ‘temporal pulses’ in the auditory signal (equivalent to input samples) are missed, causing phoneme identification to be based on incomplete input. This hypothetical mechanism aligns with the notion that attention operates through a single channel that assigns its resources to one and only one object at a time, switching between the auditory object and the secondary object on a moment-by-moment basis (e.g., Moray, 1986). This would amount to processing the auditory object at a “lower resolution” under cognitive load than under focused attention, increasing the overlap between similar-sounding phonemes—i.e., blurring phoneme category boundaries.

At a more peripheral level, attention is thought to modulate cochlear activity via the medial olivocochlear (MOC) efferent pathway descending from the cortex (Froechlich, Collet, & Morgon, 1993; Maison, Micheyl, & Collet, 2001; Meric & Collet, 1992). Of particular relevance to our cross-modal data is evidence that selective attention to visual stimuli can transiently reduce cochlear sensitivity (e.g., Hernandez-Peon, Scherrer, & Jouvet, 1956; Lukas, 1980; Oatman, 1976; Puel, Bonfils, & Pujol, 1988) and that this reduction is commensurate to the demands of the visual task (Delano, Elgueda, Hamame, & Robles, 2007). It should be noted that this line of research typically examines focused attention, with the underlying assumption that the system suppresses task-irrelevant sensory information. In our study, participants were required to simultaneously attend to the visual and auditory channels. However, given that the two tasks did not overlap in terms of their modality or their goals, it is possible that while attention is oriented towards the visual task, a central inhibitory mechanism transiently dampens auditory sensitivity via the MOC pathway.

Overall, the data reported here align well with the results of previous studies aiming to determine the locus of cognitive load effects within the speech perception system. Mattys et al. (2014) demonstrated that cognitive load does not increase lexical activation, and therefore suggested that lexical drift effects occur as a cascaded consequence of impoverished sensory encoding. However, they could not rule out the possibility that cognitive load causes a lexical bias at the decisional level. The results of the present study discount this as a likely explanation, and further demonstrate that cognitive load does not simply increase randomness in responding. Instead, cognitive load seems to specifically impair acoustic acuity at the sublexical level.

**APPENDIX: EXPERIMENTAL STIMULI**

Critical Stimuli

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Target Word | Frequency |  | HF Word | Frequency |  | LF Word | Frequency |
| peak | 406 |  | feet | 4104 |  | teak | 3 |
| keys | 289 |  | deep | 2413 |  | tease | 33 |
| pool | 593 |  | soon | 5116 |  | tool | 285 |
| tin | 513 |  | hill | 1258 |  | kin | 60 |
| tip | 364 |  | fit | 689 |  | kip | 5 |
| dad | 560 |  | cap | 1032 |  | tad | 0 |
| pile | 407 |  | fine | 2563 |  | tile | 58 |
| cot | 380 |  | dog | 1246 |  | tot | 21 |
| beam | 168 |  | keep | 1566 |  | deem | 4 |
| curse | 148 |  | church | 2844 |  | terse | 22 |
| died | 451 |  | sight | 1685 |  | tied | 136 |
| sheep | 359 |  | seen | 6750 |  | jeep | 128 |
| code | 447 |  | hole | 1015 |  | toad | 56 |
| barn | 186 |  | part | 8628 |  | darn | 15 |
| bike | 150 |  | sign | 1621 |  | tyke | 3 |
| boss | 398 |  | god | 4668 |  | toss | 60 |
| gain | 370 |  | came | 3127 |  | dane | 36 |
| sit | 537 |  | thin | 1345 |  | kit | 152 |
| soil | 742 |  | choice | 1828 |  | coil | 49 |
| seed | 498 |  | chief | 1816 |  | deed | 92 |
| ship | 793 |  | sick | 1226 |  | chip | 101 |
| shade | 403 |  | date | 1007 |  | jade | 31 |
| nut | 123 |  | love | 4614 |  | mutt | 3 |
| moon | 951 |  | you've | 3573 |  | noon | 290 |
| male | 1546 |  | same | 12082 |  | nail | 189 |
| mat | 192 |  | sad | 827 |  | gnat | 11 |
| pub | 371 |  | sun | 2689 |  | tub | 141 |
| map | 541 |  | gas | 1269 |  | mac | 52 |
| tap | 325 |  | cash | 981 |  | tat | 9 |
| pipe | 396 |  | nice | 2528 |  | pike | 54 |
| hit | 324 |  | miss | 3721 |  | hitch | 40 |
| port | 438 |  | force | 2589 |  | porch | 225 |
| fate | 585 |  | race | 1340 |  | fake | 122 |
| lab | 338 |  | bad | 3755 |  | lag | 44 |
| teach | 226 |  | week | 4853 |  | teat | 82 |
| patch | 304 |  | mass | 2023 |  | pat | 40 |
| ridge | 219 |  | did | 5237 |  | rig | 72 |
| boss | 398 |  | top | 4174 |  | botch | 9 |
| juice | 365 |  | youth | 1159 |  | jute | 6 |
| kiss | 298 |  | wish | 777 |  | kitsch | 7 |
| fuss | 287 |  | cup | 1072 |  | phut | 3 |
| lose | 358 |  | food | 4552 |  | loon | 11 |
| bars | 482 |  | hard | 4744 |  | barn | 186 |
| laugh | 558 |  | sharp | 1033 |  | larch | 21 |
| tooth | 233 |  | use | 3828 |  | toot | 4 |
| myth | 356 |  | this | 84927 |  | mitt | 0 |
| pin | 175 |  | live | 1004 |  | ping | 15 |
| pan | 470 |  | have | 21383 |  | pang | 47 |
| dim | 277 |  | big | 5677 |  | din | 74 |
| firm | 1222 |  | word | 3950 |  | fern | 21 |
| leap | 137 |  | look | 4710 |  | loop | 99 |
| heap | 182 |  | ship | 793 |  | hoop | 27 |
| sheet | 616 |  | met | 2652 |  | chute | 30 |
| fan | 165 |  | fall | 1113 |  | fen | 25 |
| ham | 123 |  | house | 10001 |  | hem | 45 |
| jam | 227 |  | warm | 1517 |  | gem | 17 |
| pack | 329 |  | walk | 1126 |  | peck | 26 |
| dock | 188 |  | down | 21923 |  | dyke | 25 |
| loud | 652 |  | head | 4046 |  | lad | 186 |
| bite | 216 |  | back | 22071 |  | bout | 106 |
| hide | 179 |  | hat | 950 |  | hoard | 20 |
| mice | 184 |  | much | 21286 |  | morse | 11 |
| dull | 580 |  | doubt | 2467 |  | dell | 29 |
| folk | 307 |  | fed | 16339 |  | fake | 122 |
| bone | 378 |  | when | 46311 |  | bane | 17 |
| loose | 748 |  | bus | 1155 |  | lease | 71 |
| soup | 362 |  | some | 34232 |  | seep | 5 |
| mood | 932 |  | good | 16874 |  | mead | 31 |

Filler Stimuli

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Target Word |  | HF Word | Frequency |  | LF Word | Frequency |
| bead |  | heat | 2005 |  | bead | 33 |
| dude |  | dark | 3331 |  | dude | 21 |
| mop |  | mean | 2339 |  | mop | 47 |
| rob |  | red | 3030 |  | rob | 26 |
| loom |  | team | 1514 |  | loom | 34 |
| coot |  | cause | 1563 |  | coot | 8 |
| pip |  | park | 1063 |  | pip | 5 |
| fib |  | rich | 2153 |  | fib | 3 |
| sag |  | said | 49759 |  | sag | 13 |
| moss |  | move | 1574 |  | moss | 114 |
| rum |  | home | 8831 |  | rum | 98 |
| huff |  | had | 22393 |  | huff | 13 |
| laze |  | loss | 1401 |  | laze | 2 |
| rife |  | safe | 1426 |  | rife | 26 |
| chive |  | wife | 3777 |  | chive | 2 |
| gel |  | will | 9411 |  | gel | 6 |
| chef |  | wet | 1099 |  | chef | 48 |
| tome |  | form | 4221 |  | tome | 8 |
| mope |  | note | 1257 |  | mope | 1 |
| vole |  | voice | 4159 |  | vole | 4 |
| rain |  | rain | 1261 |  | rag | 102 |
| bed |  | bed | 4736 |  | beige | 51 |
| page |  | page | 1074 |  | poise | 30 |
| full |  | full | 4818 |  | howl | 24 |
| large |  | large | 6638 |  | sage | 63 |
| van |  | van | 971 |  | vet | 158 |
| five |  | five | 5226 |  | lime | 144 |
| book |  | book | 4857 |  | bug | 42 |
| rock |  | rock | 1405 |  | jog | 41 |
| death |  | death | 4008 |  | moth | 51 |
| thick |  | thick | 1186 |  | rook | 5 |
| king |  | king | 1598 |  | fang | 8 |
| thought |  | thought | 4426 |  | rot | 78 |
| ten |  | ten | 50742 |  | bun | 67 |
| worth |  | worth | 1830 |  | hearth | 80 |
| phone |  | phone | 1159 |  | coin | 134 |
| march |  | march | 1043 |  | perch | 38 |
| took |  | took | 2127 |  | soak | 32 |
| noise |  | noise | 1142 |  | wheeze | 9 |
| shock |  | shock | 941 |  | rack | 145 |

**V Acknowledgements**

This research was supported by a grant from the ESRC to Sven Mattys (RES-062-23-2746). We thank Katharine Barden for help with recordings.

**VI REFERENCES**

Baskent, D. (2012). Effect of speech degradation on top-down repair: phonemic restoration with simulations of cochlear implants and combined electric–acoustic stimulation. *Journal of the Association for Research in Otolaryngology, 13,* 683–692.

Casini, L., Burle, B., & Nguyen, N. (2009). Speech perception engages a general timer: Evidence from a divided attention word identification task. *Cognition, 112,* 318–322.

Delano, P. H., Elgueda, D., Humame, C. M. & Robles, L. (2007). Selective attention to visual stimuli reduces cochlear sensitivity in chinchillas. *The Journal of Neuroscience, 27,* 4146-4153*.*

Fernandes, T., Kolinsky, R., & Ventura, P. (2010). Cognitive noise is also noise: The impact of attention load on the use of statistical information and coarticulation as speech segmentation cues. *Attention, Perception, & Psychophysics, 72,* 1522–1532.

Forster, K. I. & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavioral Research Methods, Instruments, & Computers, 35,* 116-124.

Francis, A. L. (2010). Improved segregation of simultaneous talkers differentially affects perceptual and cognitive capacity demands for recognizing speech in competing speech. Attention, Perception, & Psychophysics, *72*, 501–516.

Froehlich, P., Collet, L., & Morgon, A. (1993). Transiently evoked otoacoustic emission amplitude change with changes of directed attention. *Physiology and Behavior*, *53*, 679–682.

Ganong, W. F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance, 6,* 110–125.

Gordon, P. C., Eberhardt, J. L., & Rueckl, J. G. (1993).Attentional modulation of the phonetic significance of acoustic cues. *Cognitive Psychology, 25,* 1–42.

Hernandez-Peon R, Scherrer H, Jouvet M (1956). Modification of electric activity in cochlear nucleus during attention in unanesthetized cats. *Science, 123,* 331–332.

Luce, P. A. (1986). *Neighborhood of words in the mental lexicon.* Research on speech perception, Technical report No. 6. Bloomington, IN: Speech Research Laboratory, Department of Psychology, Indiana University.

Luce, P. A. & Pisoni, D. B. (1998).Recognizing spoken words: The neighbourhood activation model. *Ear & Hearing*, *19*, 1-36.

Lukas, J. H. (1980). Human auditory attention: the olivocochlear bundle may function as a peripheral filter. *Psychophysiology 17,* 444–452.

Macdonald, J. S., & Lavie, N. (2011). Visual perceptual load induces inattentional deafness. *Attention, Perception, & Psychophysics*, *73*, 1780-1789.

Maison, S., Micheyl, C., & Collet, L. (2001). Influence of focused auditory attention on cochlear activity in humans. *Psychophyiology, 38,* 35-40.

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

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

McClelland, J., & Elman, J. (1986). The TRACE model of speech perception. *Cognitive Psychology, 18,* 1-86.

Meric, C., & Collet, L. (1992) Visual attention and evoked otoacoustic emissions: a slight but real effect. *International Journal of Psychophysiology, 12,* 233–235.

Mirman, D., McClelland, J. L., Holt, L. L., & Magnuson, J. S. (2008). Effects of attention on the strength of lexical influences on speech perception: Behavioral experiments and computational mechanisms. *Cognitive Science*, *32*, 398-417.

Moray, N. (1986). *Monitoring behavior and supervisory control*. In K. R. Boff, L. Kaufman, and J.P. Thomas (Eds.), *Handbook of Perception and Human Performance. Chapter 45*. New York: Wiley.

Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition, 52,* 189-234.

Oatman, L. C. (1976). Effects of visual attention on the intensity of auditory evoked potentials. *Experimental Neurology, 51*, 41–53.

Puel, J. L., Bonfils, P., & Pujol, R. (1988). Selective attention modifies the active micromechanical properties of the cochlea. *Brain Research, 447*, 380 –383

Pichora-Fuller, M. K. (2008). Use of supportive context by younger and older adult listeners: balancing bottom-up and top-down information processing. *International Journal of Audiology, 47,* S72-S82.

Riecke, L., Esposito, F., Bonte, M., & Formisano, E. (2009). Hearing illusory sounds in noise: The timing of sensory-perceptual transformations in auditory cortex. *Neuron, 64,* 550–561.

Samuel, A. G. (1981). Phonemic restoration: Insights from a new methodology. *Journal of Experimental Psychology: General, 110,* 474–494.

Samuel, A. G. & Kraljic, T. (2009).Perceptual learning for speech. *Attention, Perception, & Psychophysics, 71,* 1207-1218

Saija, J. D., Akyurek, E. G., Andringa, T. C., & Baskent, D. (2014). Perceptual restoration of degraded speech is preserved with advancing age. *Journal of the Association for Research in Otolaryngology, 15,* 139–148.

Sivonen, P., Maess, B., & Friederici, A. D. (2006).Semantic retrieval of spoken words with an obliterated initial phoneme in a sentence context. *Neuroscience Letters, 408,* 220–225.

Warren, R. M. (1970). Perceptual restoration of missing speech sounds. *Science, 167,* 392–393.

Zekveld, A. A., Heslenfeld, D. J., Festen, J. M., & Schoonhoven, R. (2006). Top-down and bottom-up processes in speech comprehension. *NeuroImage, 32,* 1826-1836.

**Figure Captions**

FIGURE 1. (a) Distribution of responses in the Cognitive Load and No Load conditions under -6 dB SNR and -12 dB SNR on critical trials. (b) Distribution of responses in the Cognitive Load and No Load conditions on critical trials collapsed across SNR levels. (c) Distribution of responses under -6 dB SNR and -12 dB SNR on critical trials collapsed across load levels. Error bars represent the standard error of the mean.

FIGURE 2. (a) Distribution of responses in the Cognitive Load and No Load conditions under -6 dB SNR and -12 dB SNR on filler trials. (b) Distribution of responses in the Cognitive Load and No Load conditions on filler trials collapsed across SNR levels. (c) Distribution of responses under -6 dB SNR and -12 dB SNR on filler trials collapsed across load levels. Error bars represent the standard error of the mean.