Extrinsic Cognitive Load Impairs Low-Level Speech Perception

Sven L. Mattysa , Katharine Bardenb, and Arthur G. Samuelc, d, e

1. Department of Psychology, University of York, UK
2. School of Psychology, University of Bristol, UK
3. Psychology Department at Stony Brook University, United States
4. Basque Center on Cognition Brain and Language, Donostia, Spain
5. Ikerbasque, Basque Foundation for Science, Bilbao, Spain

Running title: Cognitive Load and Speech Perception

Word count for main text: 3936

Corresponding author:

Sven Mattys

Department of Psychology

University of York, UK

sven.mattys@york.ac.uk

Abstract

Recent research suggests that the extrinsic cognitive load generated by performing a non-linguistic visual task while perceiving speech increases listeners' reliance on lexical knowledge and decreases their capacity to perceive phonetic detail. In the present study, we asked whether this effect is better accounted for at a lexical or a sub-lexical level. The former implies that cognitive load directly affects lexical activation but not perceptual sensitivity. The latter implies that increased lexical reliance under cognitive load is only a secondary consequence of imprecise or incomplete phonetic encoding. Using the phoneme-restoration paradigm, we showed that perceptual sensitivity decreases (i.e., phoneme restoration increases) almost linearly with the effort involved in the concurrent visual task. However, cognitive load has only a minimal effect on the contribution of lexical information to phoneme restoration. We conclude that the locus of extrinsic cognitive load on the speech system is perceptual rather than lexical. Mechanisms by which cognitive load increases tolerance to acoustic imprecision and broadens phonemic categories are discussed.

Key words: speech perception, divided attention, cognitive load, phoneme restoration, lexical access

Attempts to understand the effects of adverse (i.e., everyday) listening conditions on speech perception have focused primarily on factors leading to variation or degradation of the speech signal (see Mattys, Davis, Bradlow, & Scott, 2012, for a review). Adverse conditions that do not affect the signal itself have been comparatively under-studied. In the studies that have investigated the effect of extrinsic cognitive load (e.g., an independent memory or attentional task) on speech perception, the cost is often measured as the degree to which listeners are impaired by interference from irrelevant perceptual information in the auditory signal while performing the secondary task. For example, Brungart et al. (2013) found that the interference of a competing talker was greater when listeners had to hold materials in working memory while performing the speech task than when they did the task without the memory load. The cost was less pronounced when the competing talker was replaced with meaningless speech-shaped noise, suggesting that cognitive load is particularly detrimental to speech-listening conditions requiring effortful stream segregation. Similarly, Francis (2010) found that selectively attending to one talker while ignoring a competing talker was harder when listeners were simultaneously asked to hold six digits in working memory rather than only one.

While these studies and others indicate that cognitive load might impair speech recognition by depleting processing resources needed for attentionally-demanding tasks (e.g., segregating signal from noise), the locus of this effect and its underlying mechanism are largely unknown. To address this question, Mattys and Wiget (2011) investigated the effect of cognitive load on phoneme categorization. In their experiments, listeners categorized the initial phoneme of a syllable as /g/ or /k/. The voice onset time of the phoneme was manipulated along a continuum such that the identity of the phoneme was either unambiguous (a clear /g/ or /k/) or ambiguous (a blend between /g/ and /k/). The /g-k/ continuum led to a word-nonword contrast (*gift-kift*) or a nonword-word contrast (*giss-kiss*). Listeners, who were asked to ignore the lexical status of the stimulus, performed the task in the presence or absence of a concurrent visual task. The visual task consisted of detecting a pre-specified target in an array of colored shapes displayed during the playback of the spoken syllable. In the absence of the visual task, listeners reported more /g/ responses along the *gift-kift* continuum and more /k/ responses along the *giss-kiss* continuum, suggesting that lexical knowledge biased phoneme categorization, as shown by Ganong (1980). Critically, this lexical effect was amplified when listeners performed the visual task. We will refer to this pattern as “lexical drift”: A stronger lexical influence on observed behavior under cognitive load (for a replication in Dutch, see Mattys & Scharenborg, in press). Importantly, in a separate experiment, Mattys and Wiget found that the concurrent visual task also impaired the listeners' capacity to discriminate pairs of syllables on the continuum, which suggests that cognitive load also has an effect on phoneme perception.

Here, we ask whether lexical drift is better accounted for at a lexical or sub-lexical level. A lexical locus implies that cognitive load directly affects lexical activation (which could be implemented by, e.g., decreasing the recognition threshold of lexical representations) or post-lexical decisions, without fundamentally altering early perceptual stages. Modulation of lexical activation and/or inhibition has been proposed as a potential mechanism for changes in attentional focus between lexical and sub-lexical levels (Mirman, McClelland, Holt, & Magnuson, 2008). In contrast, a sub-lexical locus implies that cognitive load impairs early perceptual stages, with the greater lexical contribution being only a secondary consequence of imprecise or incomplete phonetic encoding. Effects of cognitive load on phoneme perception have indeed been shown by Casini, Burle, and Nguyen (2009) and Gordon, Eberhardt, and Rueckl (1993). In this conceptualization, cognitive load reduces attention to phonetic detail, which in turn leads to greater reliance on lexical plausibility as a compensatory response.

 To adjudicate between these two possibilities, the present experiment assesses the effect of cognitive load using a task based on the phoneme-restoration effect (Warren, 1970). Warren found that when he replaced a speech segment with noise, listeners consistently reported the speech as being intact – they perceptually restored the replaced speech. Samuel (1981) developed a paradigm to separate lexical influences on the effect from basic acoustic-phonetic perception. In this paradigm, a phoneme is either replaced with noise or has noise added to it. Listeners' ability to discriminate between the added and replaced conditions reflects their capacity to perceive the fine details of the speech signal. If a listener perceptually restores the missing speech in a replaced item, the result should be similar to an added item, reducing discriminability. Thus, high restoration indicates poor perceptual sensitivity and low restoration reflects good perceptual sensitivity. Samuel tested whether lexical activation affects acoustic-phonetic perception by comparing discrimination of added/replaced segments in real words (with lexical support) versus matched nonwords. The words and nonwords present comparable acoustic challenges, but the words potentially add a lexical basis to restore the missing segments. Discrimination between the added and replaced stimuli was poorer in the word than the nonword condition, suggesting that lexical knowledge constrains phoneme perception.

We used this paradigm because it simultaneously provides a measure of basic acoustic-phonetic performance and of lexical influences. A comparable change in performance for words and nonwords is an effect on basic acoustic-phonetic processing, while a change that selectively affects words is due to lexical activation. If the lexical drift seen in Mattys and Wiget (2011) was due to impaired acoustic-phonetic processing under cognitive load rather than increased lexical activation per se, then discrimination performance on the phoneme-restoration task under a cognitive load should be impaired compared to no cognitive load, but this decrease should be of comparable magnitude in words and nonwords. In contrast, if cognitive load directly increases lexical activation, then the lexical effect on phoneme restoration should be more pronounced under load: Discrimination between the added and replaced stimuli should be more impaired by cognitive load in words than nonwords.

For the cognitive load, we chose a non-linguistic visual search task to prevent any effect from resulting from simple modality-/domain-specific interference. Cognitive load was implemented using arrays of colored shapes similar to those in Mattys and Wiget (2011). We used five levels of load: No load, Very Low load, Low load, High load, Very High load. In the four proper load conditions (Very Low, Low, High, Very High), listeners saw an array of colored shapes, differing in size and complexity, during the playback of a speech stimulus and had to detect the presence or absence of a red square.

**Method**

**Participants**

144 monolingual native English speakers participated (114 female; mean age 20 yrs, range 18 - 30 yrs). None reported hearing or speech impairments. To keep testing time within reasonable limits, two of the load conditions (Low and High) were administered to 72 participants and the other two load conditions (Very Low and Very High) to the other 72 participants. Both groups were also run on the No load condition. In each of the two groups, 36 participants were assigned to the word condition and the other 36 to the nonword condition. They received course credit or payment.

**Materials**

Sixty test word-nonword pairsand 30 filler word-nonword pairs were selected. All stimuli were four or five syllables long. In each word-nonword pair, the critical phoneme was the first phoneme of the final syllable (or ambisyllabic between the penultimate and final syllables). The critical phoneme was either a nasal (30 items – 17 /n/ and 13 /m/) or a liquid (30 items – 15 /l/ and 15 /r/). Previous restoration work (e.g., Samuel, 1981, 1996) had suggested that liquids and nasals were in a good "medium" range in terms of restorability, away from ceiling (fricatives) or floor (vowels). The word and nonword within a test pair had the same stress pattern and the same final two syllables. Wherever possible, the phoneme preceding the penultimate syllable was also matched between the word and the nonword of a pair to facilitate subsequent splicing, as described below (e.g., word: *discrimi****n****ate*; nonword: *notromi****n****ate*, with the bold letter indicating the critical phoneme). Filler stimuli were included to ensure that participants listened to the stimuli as a whole, rather than just the final syllable. Therefore, the critical phoneme in the filler stimuli was within the first syllable or the onset of thesecond syllable. Filler words and nonwords were matched for stress pattern. The first syllable and the onset of the second syllable were phonemically matched across filler words and nonwords (e.g., ***a****cknowledgement,* ***a****cknallutstump*).

The cognitive load was a visual search task in which participants had to judge whether a red square was present in an array of colored shapes, as described in Mattys and Wiget (2011). In addition to a baseline No-load condition, there were four levels of Load: Very Low, Low, High, Very High (see Figure 1). The Very Low load condition consisted of 2 x 2 odd-one-out arrays. In the target-present arrays, the target red square was accompanied by three black squares or three red triangles. The red square could be anywhere in the array. The target-absent arrays consisted of four black squares or four red triangles. The Low load condition also consisted of 2 x 2 arrays, but, in the target-present condition, the red square target was accompanied by a combination of red triangles, black squares and black triangles. Thus, the target did not pop out from the display as easily as it did in the Very Low load condition. The High load condition was similar to the Low load condition except that the arrays contained 6 x 6 elements. In the Very High load condition, the arrays contained 10 x 10 elements. The size of the array on the monitor was approximately 4x4cm (2x2 arrays), 8x8cm (6x6), and 19x19cm (10x10).

**Procedure**

**Recording and sound file processing.** The stimuli were recorded by a female phonetician who spoke Standard Southern British English. Recording took place in a sound-treated room using a Shure WH20 cardioid dynamic headset microphone at a sampling rate of 44.1 kHz with 16-bit resolution. Several recordings were made of each word and nonword. Word-nonword pairs that were judged most similar in pitch and prosody were selected. Insufficiently well-matched pairs were re-recorded. The onsets and offsets of the added/replaced segments were identified auditorily such that there was no easily audible coarticulation that would cue the critical segment outside the noise window.

In the Added condition, signal-correlated noise was added at 0 dB SNR to the critical segment. In the Replaced condition, the same critical segment was replaced by signal-correlated noise of the same amplitude. The average duration of a noise segment (excluding fillers) was 111 ms (range 74 – 213 ms). These durations were identical in the words and matched nonwords due to the splicing procedure, as described below.

The word-nonword test pairs were cross-spliced such that the portion containing the critical phoneme was identical in the word and the nonword of the pair. The splice point was the onset of the penultimate syllable. Thus, the last two syllables of a word and its matched nonword (which included the critical segment) were acoustically identical. For counterbalancing purposes, half of the test pairs used the word ending as the common slice, and the other half used the nonword ending as the common slice. The average duration of the test stimuli was 766 ms, ranging from 502 ms to 1069 ms. Fillers were not cross-spliced.

**Experiment and trial structures.** Participants were assigned to either the word or the nonword condition. Within each group, participants were further assigned to one of two sub-groups. One received the No load, Low load, and High load conditions. The other received the No load, Very Low load, and Very High load conditions. Level of Load was blocked and the order of the three levels counterbalanced across participants. We blocked Load levels to minimize the added effort that would be involved in switching from one visual search strategy to another from trial to trial. The test pairs were broken down into three sub-sets (each containing an equal number of nasal and liquid critical phonemes) and each sub-set was assigned to one of the three Load blocks. The assignment between sub-set and Load block was counterbalanced between participants; a pair of stimuli was heard in only one of the Load conditions. Participants heard both the added and the replaced versions of a given word or nonword. Trials within each block were randomized, with a different random order for each block and each participant. Each block started with eight practice trials representative of the upcoming type of load; these items were not presented in the test block. Within each Load condition and stimulus set, visual arrays were randomly assigned to auditory stimuli, with the constraint that half of the auditory stimuli in each Load condition for each stimulus set were paired with target-present arrays. Pairings of auditory stimuli with visual arrays were identical across word and nonword conditions. For example, the word *discriminate* under High Load was paired with the same visual array as the nonword *notrominate* under High Load.

Participants were tested individually in a sound-treated room. We used a version of the phoneme restoration task developed by Samuel (1996). Participants were informed that, for each trial, they would hear a stimulus with some white noise added, and then the same stimulus with no noise. Their task was to rate the similarity of the two stimuli using an 8-point scale with the labels ‘Not similar’ (1) and ‘Very similar’ (8) at either end. This rating procedure is more sensitive to the difference between the added and replaced conditions than Samuel’s (1981) approach of explicitly asking participants to decide if the critical phoneme has been replaced or superimposed with noise. In particular, the rating task provides empirical distributions of intactness ratings for both truly intact (Added) and disrupted (Replaced) stimuli, whereas the original procedure followed traditional signal detection norms and simply assumed normal distributions.

Participants were instructed to try to ignore the noise. The interval between the onset of the Added/Replaced stimulus and the onset of the intact stimulus was 2220 ms, giving an average silent interval of 1454 ms. 1270 ms after the onset of the intact stimulus, the confidence scale appeared on screen. Participants had up to 20 s to respond. In the No Load condition, the next trial started 2 s after a response was made or at the end of the 20-s window.

In the four Load conditions, the timing of events was identical to the No-load condition, except that the visual array was displayed during the Added/Replaced stimulus. A prompt reading "No……red square? ……Yes" appeared on the monitor after participants had registered their similarity ratings. Participants were asked to press the left shift key for No and the right shift key for Yes. There was a 2-s interval between the response and the onset of the next trial. If there was no response, the program moved on to the next stimulus after 20 s. The experiment lasted approximately 20 minutes.

**Results**

Table 1 shows average ratings broken down by Noise (added, replaced), Lexicality (nonwords, words), and Load (No load, Very Low, Low, High, Very High). Discrimination scores, calculated as the difference between the ratings in the Added and the Replaced conditions are displayed in Figure 2. The main panel (2a) shows these scores for the words and the nonwords, broken down by level of Load. Recall that the restoration paradigm provides indices for (1) how load affects acoustic-phonetic perception, and (2) any lexical effect on such perception. Accordingly, Figure 2b shows how acoustic-phonetic perception was affected by Load (averaged across words and nonwords). Figure 2c shows how Load moderated the lexical effect (the difference between words and nonwords).

Data were analysed using mixed-effects regression models, with the reflect log-transformed rating values as the dependent variable. The ratings were transformed in order to attenuate the negative skew of the rating distribution.1 The fixed factors were Noise, Lexicality, and Load. The two levels of the binary factors were coded as -1 and +1, which was the case for Noise and Lexicality. The five levels of Load were coded as -1, -.5, 0, +.5, and +1 for No load, Very Low Load, Low Load, High Load, and Very High Load, respectively. The fixed factors, as well as their interactions, were added incrementally and improved fit was assessed using the likelihood ratio test. For all models, we included by-subject and by-item intercepts, and by-subject random slopes for Noise and by-item random slopes for Noise and Lexicality.2

The base model included only the random terms. The main effects of Noise, Lexicality, and Load were assessed by testing the increase in model fit when each one of these factors, considered individually, was added to the base model.3 A main effect of Noise, *β* = .042, SE = .005, *χ2*(1) = 57.30, *p* < .001, showed that stimuli containing a phoneme with added noise were judged to be more similar to their intact counterpart than stimuli containing a phoneme replaced with noise. Thus, listeners were able to discriminate between presence and absence of the test phoneme. Lexicality also significantly improved the base model, *β* = .027, SE = .012, *χ2*(1) = 5.36, *p* < .05: Overall, similarity was rated as higher in the nonwords than the words, suggesting that because participants knew what words should sound like, any disruption (added or replaced) was less acceptable than in the nonword counterparts. Likewise, a Load effect, *β* = -.005, SE = .002, *χ2*(1) = 5.03, *p* < .05, showed that ratings were influenced by the presence of the visual task and its complexity. A series of models restricted to pairwise comparisons indicated that ratings were lower in the No load condition than in the Very Low Load and very High Load conditions (*p*s < .05). The other load conditions did not significantly differ from each other. Thus, when listeners could completely focus on the rating task, they were slightly less likely to give high similarity ratings but this pattern was not consistently observed across the Load levels.

The critical analyses focused on (1) the effect of Load on the listeners' capacity to discriminate between the added and replaced conditions, that is, the interaction between Load and Noise (Figure 2b), and (2) the effect of Lexicality on discrimination, that is, the interaction between Lexicality and Noise (Figure 2c).

The first analysis compared a model including Load and Noise as fixed factors to a model including Load, Noise, and their interaction. The model containing the interaction term provided a better fit, *β* = .008, SE = .002, *χ2*(1) = 16.83, *p* < .001, with discrimination (i.e., the Noise effect) decreasing as a function of level of Load. This pattern is clearly visible in Figure 2b.

The second analysis compared a model including Lexicality and Noise with a model including Lexicality, Noise, and their interaction. The model with the interaction term provided a better fit, *β* = .010, SE = .002, *χ2*(1) = 15.76, *p* < .001: Discrimination was significantly worse in the Word than Nonword conditions, replicating Samuel's (1981, 1996) finding that lexical knowledge can be used to restore missing phonetic segments, reducing the perceptual difference between Added and Replaced stimuli. This effect is illustrated in the consistently positive difference scores in Figure 2c. Critically, the Lexical effect on discrimination was not modulated by Load. That is, the fit by a Noise-by-Lexicality-by-Load model was not affected by the presence or absence of the three-way interaction term, *χ2*(1) = 1.24, *p* = .26: The lexically-driven reduction in discrimination was not modulated by the level of Load. As a further check of this point, we also tested this 3-way interaction for each combination of two levels of Load (No Load vs Very Low Load, Very Low Load vs Low Load, etc.). None of the 10 3-way interactions reached *p* < .14.

Performance on the visual search task in the four Load conditions was analyzed in a similar fashion (Noise x Lexicality x Load). The only factor that significantly contributed to the variance in detection accuracy was Load, *β* = -.093, SE = .003, *χ2*(3) = 1132.6, *p* < .001 (Very Low: 96%; Low: 95%; High: 89%; Very High: 69%). All pairwise comparisons reached *p* < .001, except for Very Low vs. Low, *χ2*(1) = 2.44, *p* = .12. All other effects and interactions were non-significant.

**Discussion**

Recent data by Mattys and Wiget (2011) suggest that cognitive load leads to both greater reliance on lexical knowledge and an impoverishment in listeners' ability to process acoustic detail. Whether this lexical drift under cognitive load reflects a direct change in activation within the lexicon or whether it is the cascaded effect of impoverished perceptual processes was the main question for this study. Our data showed that cognitive load impaired listeners' capacity to discriminate phonemes replaced by noise from phonemes that have noise added to them but, critically, this discrimination decrease was comparable in words and nonwords. Thus, cognitive load did not directly modulate lexical activation. The reduction in perceptual sensitivity did not lead to a cascaded increase in lexical activation, as it did in Mattys and Wiget's categorization experiment, probably because the nonwords in our study did not overlap with existing words enough to be confusable with them, even with reduced perceptual precision.

However, it would be premature to conclude that the locus of extrinsic cognitive load in the speech system is entirely perceptual. Although Load had no significant effect on the lexical influence, the lexical effect in the No-Load condition was numerically smaller than in the other load conditions. One interpretation of this trend is that full attention was needed to minimize the lexical influence – note that in this task, the lexical effect reduces the ability to make the required precise acoustic-phonetic judgment. Once attention was divided by imposing a second task, there was no extra cost for higher processing demands in the secondary task: The size of the lexical effect was virtually unchanged across the four degrees of load. Thus, suppressing the effects of lexical activation may require considerable attention, but this effect is clearly quite different than the very systematic decrease in acoustic-phonetic discrimination as load increased.

This graded effect could be due to the amount of time attention is engaged in the secondary task. If a visual target can be detected more quickly in a small than a large array, attention could be redirected to the speech signal earlier in the former than the latter, a difference that would be all the more critical here because the manipulated phonemes came relatively late in the test stimuli. There are, of course, alternative or additional mechanisms that could produce the cognitive-load-induced reduction in perceptual sensitivity. The effect could be due to a change in the depth/complexity of acoustic processing, or to a disruption in the sampling rate of the acoustic information. In the former, cognitive load could act as a regulator of signal-to-noise ratio leading to a re-prioritization of perceptual cues (e.g., Gordon, Eberhardt, & Rueckl, 1993). In the latter, cognitive load would disrupt the intake of speech samples, causing samples to be "skipped" while attention is directed to the secondary task (e.g., Casini, Burle, & Nguyen, 2009).

In addition to the issue of mechanism, other important questions remain. For example, how early in the speech-perception system is cognitive load having an effect? Does cognitive load change low-level hearing mechanisms (sub-cortically, e.g., Rinne et al., 2008) or does it only affect the output of such mechanisms (cortically, e.g., Rinne et al., 2007)? Although the results of the current study have shown that the lexical drift observed in previous studies is primarily driven by a decrement in acoustic-phonetic precision under cognitive load, much remains to be learned about how attentional mechanisms interact with language processes.

# References

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*, 255-278.

Brungart, D., Iyer, N., Thompson, E. R., Simpson, B. D., Gordon-Salant, S., Schurman, J., Vogel. C., & Grant, K. (2013). Interactions between listening effort and masker type on the energetic and informational masking of speech stimuli. In *Proceedings of Meetings on Acoustics* (Vol. 19, p. 060146).

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

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.

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., and Rueckl, J. G. (1993). Attentional modulation of the phonetic significance of acoustic cues. *Cognitive Psychology*, *25*, 1-42.

Mattys, S. L. & Scharenborg, O. (in press). Phoneme categorization and discrimination in younger and older adults: A comparative analysis of perceptual, lexical, and attentional factors. *Psychology and Aging.*

Mattys, S. L., Davis, M.H., Bradlow, A.R., & Scott, S.K. (2012). Speech recognition in adverse conditions: A review. *Language & Cognitive processes*, *27*, 953-978.

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

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.

Rinne, T., Balk, M. H., Koistinen, S., Autti, T., Alho, K., & Sams, M. (2008). Auditory selective attention modulates activation of human inferior colliculus. Journal of Neurophysiology, *100*, 3323-3327.

Rinne, T., Stecker, G. C., Kang, X., Yund, E. W., Herron, T. J., & Woods, D. L. (2007). Attention modulates sound processing in human auditory cortex but not the inferior colliculus. *Neuroreport*, *18*, 1311-1314.

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

Samuel, A. G. (1996). Does lexical information influence the perceptual restoration of phonemes? *Journal of Experimental Psychology: General*, *125*, 28-51.

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

**Footnote**

1The reflect log transform consisted of subtracting each rating score from 9 (the highest possible rating plus one) and apply a logarithmic transformation to that new value. All the analyses were also run on the untransformed ratings. There was no substantial difference between the outcome of the analyses with and without transformation.

2To establish the adequate random effects structure for our analyses, we followed Barr, Levy, Scheepers, and Tily 's (2013) recommendation that one should use the maximal random effects structure justified by the experimental design. To do so, we initially included random intercepts for subjects and for items, as well as by-subjects random slopes for Noise and random by-items slopes for Noise, Lexicality, and Load. The random slopes were introduced to reflect the within-subjects/items nature of those factors. However, including random by-items slopes for Load did not significantly improve the model. Therefore, that term was omitted. The remaining structure is the one we used in the analyses. Note that we also initially included by-subjects random intercepts and slopes for Type of Segment (Nasals, Liquids) but these terms did not improve the model either. They were therefore left out.

3We initially included Type of Segment (Nasals, Liquids) as an additional fixed effect, but this term did not significantly improve the base model (untransformed ratings for Nasals *M*: 6.02, *SD*: 1.63 vs. Liquids *M*: 5.77, *SD*: 1.83),

**Authors’ Note**

Correspondence can be sent to Sven Mattys, Department of Psychology, University of York, York, UK, Sven.Mattys@york.ac.uk.This study was made possible thanks to a grant from the ESRC (RES-062-23-2746) to S. L. Mattys. A. G. Samuel received support from grant PSI2010-17781 from the Spanish Ministerio de Economia y Competividad.**.**

|  |  |
| --- | --- |
|

|  |
| --- |
| Table 1.Average similarity ratings (and standard deviations). Ratings are made on a scale from 1 (not similar) to 8 (very similar). |

 |
|   |  **No Load** |  **Very Low**  | **Low** | **High** |  **Very High** |
| **Nonwords** |  |  |  |  |  |
|  Added | 6.44 (1.52) | 6.34 (1.62) | 6.49 (1.38) | 6.47 (1.45) | 6.36 (1.47) |
|  Replaced | 5.59 (2.04) | 5.51 (2.02) | 5.74 (1.88) | 5.74 (1.89) | 5.72 (1.85) |
| **Words** |  |  |  |  |  |
|  Added | 5.96 (1.61) | 6.04 (1.56) | 6.05 (1.46) | 6.01 (1.49) | 5.96 (1.60) |
|  Replaced | 5.28 (1.88) | 5.56 (1.73) | 5.60 (1.72) | 5.64 (1.71) | 5.64 (1.65) |

**Figure captions**

Figure 1. Small-scale examples of target-present arrays in the Very Low (a), Low (b), High (c), and Very High (d) load conditions. The target to detect is a red square.

Figure 2. (a) Discrimination (rating in the added condition minus rating in the replaced condition) and standard error for the words and the nonwords, broken down by level of cognitive load; (b) Discrimination scores averaged across words and nonwords as a function of level of cognitive load; (c) Lexical effect, i.e., difference between discrimination in the word condition and discrimination in the nonword condition (dark grey bar minus light grey bar from Figure 2a), as a function of level of cognitive load.

Figure 1



Figure 2

