How does cognitive load influence speech perception?
An encoding hypothesis

Holger Mitterer

Department of Cognitive Science

University of Malta

Sven Mattys

Department of Psychology

University of York

Two experiments investigated the conditions under which cognitive load exerts an effect on speech perception. These experiments extend earlier research by using a different speech perception task (four-interval oddity task) and by implementing cognitive load through a task often thought to be modular, namely, face processing. In the cognitive-load conditions, participants were required to remember two faces presented before the speech stimuli. In Experiment 1, performance in the speech-perception task under cognitive load was not impaired in comparison to a no-load baseline condition. In Experiment 2, we modified the load condition minimally such that it required encoding of the two faces simultaneously with the speech stimuli. As a reference condition, we also used a visual search task that in earlier experiments had led to poorer speech perception. Both concurrent tasks led to decrements in the speech task. The results suggest that speech perception is affected even by loads thought to be processed modularly, and that, critically, encoding in working memory might be the locus of interference.

One of the basic ideas in cognitive science is modularity, that is, the idea that certain specialized processes can operate independently of any other parallel processes. Language processing has often been at the heart of this debate, given the generative paradigm that assumes that language processing is based on special evolutionary adaptations (Hauser, Chomsky, & Fitch, 2002; Liberman, 1996). Modularity gives rise to the expectation that speech processing, given its auditory and linguistic nature, is an encapsulated process that is unlikely to be influenced by concurrent visual processing (and vice versa).

However, recent evidence suggests that this may not be the case. For example, Mattys and Wiget (2011) tested the acuity of speech perception under two conditions: A no-load baseline in which participants only performed a speech-perception task and a load condition in which participants performed a concurrent visual search task—finding a red square among an array of black squares and red triangles. The results showed that speech perception suffered in the dual-task condition.

Specifically, discrimination of short syllables along a voicing continuum (/gi/-/ki/) was worse under cognitive load than no cognitive load. Additionally, participants showed a larger lexical-bias effect under cognitive load. That is, under load, they relied more strongly on prior word knowledge than on the bottom-up signal to perform the task. This indicated that early speech processing—by which we mean the coding of the speech signal into a form compatible with lexical access—may not be modular and encapsulated. To further elucidate whether these patterns reflected a decrease in the fidelity of early speech perception processes or stronger strategic reliance on lexical knowledge, Mattys, Barden, and Samuel (2014) tested the effect of cognitive load on how well participants discriminated a word in which a segment had been *replaced* by noise from a word in which the same segment had been *overlaid* with noise. The rationale stemmed from research by Samuel (1996), who used it to test the effect of lexical knowledge on phonetic processing. With this task, it is possible to measure both perceptual sensitivity (i.e., how well noise-overlaid versus noise-replaced stimuli are discriminated) and lexical effects (i.e., whether sensitivity is affected by the lexical status of the stimuli). This allowed Mattys et al. to establish whether cognitive load leads to increased reliance on lexical knowledge or interferes with early speech processing. The results were in line with the latter: While the effect of lexicality was relatively stable over different levels of cognitive load, there was a clear and near-linear relation between the level of cognitive load and perceptual sensitivity. The higher the load, the lower the discrimination between overlaid and replaced stimuli. This was taken as an indication that cognitive load interferes with early speech perception processes.

Converging evidence for this assumption stems from studies that show that attention to or away from acoustic detail is under better control in individuals with high compared to low working-memory capacity (Rönnberg et al., 2013). Thus, since cognitive load would likely tax working memory, it would disrupt listeners' ability to pay attention to the acoustic details necessary to perform the speech discrimination task. Therefore, competition for attentional resources provides a possible mechanism in support of Mattys et al.'s claim that cognitive load impairs early speech perception.

The above results should be interpreted in light of several methodological considerations, however. In their speech discrimination task, Mattys and Wiget (2011, Exp. 6) used within-category and between-category pairs of phonemes. A within-category pair is one where both members of the pair are pre-dominantly perceived as either /gi/ or /ki/ (as revealed in a separate categorization task), whereas a between-category pair is one where one member is perceived as /gi/ and the other as /ki/. Cognitive load was found to predominantly impair the between-category pairs. While categorical perception was once considered a property of early speech processing (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967), more recent evidence suggests that early perceptual processing of speech is actually rather continuous, with categorical perception a marker of later

processes (Schouten, Gerrits, & van Hessen, 2003; Toscano, McMurray, Dennhardt, & Luck, 2010). The fact that a load effect was only found for the between-category pairs could therefore indicate that cognitive load affected predominantly post-categorization processes. A caveat here is that the overall level of discrimination on within-category trials in Mattys and Wiget (2011) was near chance in the no-load condition, which makes it impossible to distinguish a locus account from a floor effect.

In addition to testing discrimination, Mattys and Wiget (2011) tested speech categorization under load and no load. They did this using a two-alternative forced-choice (2AFC) task. They did not find an effect of cognitive load on the steepness of the categorization function. The steepness of a categorization function is often seen as reflecting the acuity of speech perception in native speakers as well as second-language learners or children with dyslexia (Díaz, Mitterer, Broersma, & Sebastián-Gallés, 2012; Godfrey, Sydral-Lasky, Millay, & Knox, 1981; MacKain, Best, & Strange, 1981). The fact that cognitive load did not modify the steepness of the identification function is at odds with the assumption that cognitive load affects early, acoustic-phonetic processes in speech perception. However, because the steepness of an average categorization function does not necessarily reflect the behavior of individual perceivers, we re-analyzed the data of Mattys and Wiget (2011), fitting logistic regressions for each participant. Yet, the results confirmed the average pattern: The slopes were not shallower under cognitive load than under no cognitive load[1].

An additional issue is the nature of the cognitive-load task used in the above studies. While listening to speech, the participants had to search for a red triangle in an array of (otherwise) red squares and black triangles. With these highly abstract and easily nameable objects, it cannot be ruled out that participants used verbal rehearsal during the visual-search task. Finally, the scanning-and-encoding nature of the visual search task makes it difficult to establish whether the reported effects were driven by cognitive load or by a mere perceptual load.

These considerations indicate that there is some ambiguity regarding the stage at which cognitive load affects speech perception. Therefore, the current experiments tested the effect of a cognitive load with a different speech-perception task as well as with a different concurrent task. As mentioned earlier, the fact that the AX task used by Mattys and Wiget (2011) gave rise to a categorical-perception effect (i.e., better discrimination for between-category pairs) indicates that this task might not truly reflect early processes in speech perception. If it did, it should have led to continuous rather

---

[1] For this analysis, the eight-step speech continuum variable was scaled from -3.5 to 3.5, with steps of 1 and used to predict the log-odds of /ki/-responses. As each individual responded to each stimulus only three times, the logistic regression often failed to converge when participants had a step-like identification function. For these cases, we assumed a regression weight (or slope) of 4, which gives rise to a nearly step-like function given the scaling of the predictor variable. These regression weights were used in a paired t-test, using each participant's regression weights with and without load. This led to non-significant differences for all three experiments (Exp 1: $t(110) = -1.72$, $p = 0.08$, Exp 2: $|t(33)| < 1$, Exp 3: $|t(33)| < 1$), overall: $t(179) = -1.21$, $p > 0.2$; note that the trend in Exp1 was in the opposite direction, with steeper slopes under load).

than categorical effects. We therefore substituted this task with a 4-interval oddity task. In this task, participants hear four consecutive stimuli, three of which are identical, with the odd one out in the second or third position. Participants indicate if the odd one out is the second or third stimulus, leading to chance performance of 50% correct. Discrimination tasks with four stimuli have generally provided a smaller advantage for between-category over within-category pairs than tasks with only two (AX) or three (ABX) stimuli (Gerrits & Schouten, 2004; Pisoni, 1975; Schouten et al., 2003) .

With regard to the cognitive load task, we made use of the fact that face processing seems to be quite modular and specialized (Kanwisher, 2000). As such, it provides a strong test of whether any concurrent task is able to lead to load effects on speech perception. We asked participants to remember two faces while performing the speech discrimination task. Afterwards, they were presented with a third face and had to judge whether the third face was amongst the two initially presented. With this design, we can potentially clarify whether load effects on speech perception generalize beyond Mattys and Wiget's (2011) search task.

Additionally, this design allows us to test whether a memory load is sufficient to impair speech perception or whether it is necessary to also encode new information. This was implemented by presenting the faces well before the speech stimuli (only a memory) or with the first speech stimulus (simultaneous auditory and visual encoding).

Experiment 1

Method

Participants

Twenty-one students from the University of Malta participated in the experiment. All of them used English regularly in their daily lives[2]. They were aged 18 to 29. Thirteen of them were female. They were paid for their participation in three sessions (see Materials and Procedure for details).

Materials and Procedure

Three syllables from Mattys and Wiget's (2011) auditory /gi/-/ki/ continuum were used in the current experiment (15, 33, and 48 ms VOT), with 15 ms versus 33 ms within category and 33 ms versus 48 ms between categories. For the 4I-oddity task, we generated sequences of four syllables, three of which were identical (the standard), and the odd-one-out appearing in second or third position. For each pair, four possible sequences were generated, counterbalancing the standard and

---

[2] The language situation in Malta is as such that both Maltese and English are official languages. All university students are (at least) bilingual, since English is the teaching language, already in secondary school. However, most Maltese prefer to speak Maltese in social settings. Importantly, like English, Maltese distinguishes stop consonants by their voicing duration (except, of course, for the glottal stop).

deviant, as well as their relative position. The stimulus onset asynchrony was 600 ms. This led to an inter-stimulus interval of approximately 480 ms (syllables varied in duration from 100 to 133 ms).

For the face-perception task, we used the freely available ORL face data base (AT & T Laboratories Cambridge, 2002), which contains 10 grayscale photographs of forty different persons' faces, all unknown to the participants. The concurrent task required participants to remember two faces (belonging to two different people) while listening to the speech stimuli and then decide whether a face displayed at the end of the trial matched one the two faces. In the same-face trials, the photographs always showed a different facial expression or orientation.

We also manipulated the relative timing of the visual and auditory stimuli. In an "early load" condition, participants were given time to encode the two faces before the speech stimuli were played. In this case, the faces were presented for 1.5 s prior to the trial and disappeared when the speech stimuli started. In a "late load" condition, the faces appeared synchronously with the first of the four speech stimuli and also disappeared after 1.5 s. As a baseline, we used a no-load condition that was identical to the late-load condition, except that participants were told to ignore the faces. In the baseline condition, there was no face probe after the decision about the speech stimuli.

Stimulus presentation was controlled by the ExperimentBuilder software from SR Research. Participants performed three sessions on three separate days, one session with no load, one with the early-load condition, and one with the late-load condition. Order of sessions was counterbalanced across participants using a Latin-square design. The order of presentation of different stimuli within sessions was randomized for each participant individually.

Results and Discussion

Figure 1 shows the mean proportion of correct responses obtained in Experiment 1. The data revealed an effect of speech pair, with better performance on the between- than within-category pair. Crucially, however, there was no effect of load. We analyzed the data with a generalized linear mixed effect model using a logistic linking function. The three-level Load factor was entered as two linearly independent contrasts. The first contrast was "load (early and late) versus no load" and the second contrast was "early load versus late load". For the binary variable Pair, the within-category condition was coded as -0.5 and the between-category condition was coded as 0.5. With this coding, the regression weights can be interpreted as d' effects, that is, how much separation there is between the two speech stimuli under different conditions, with zero showing that the stimuli cannot be discriminated, a value of 1 showing a good separation, and values above 3 near-perfect discriminability. A categorical-perception effect would then be reflected by a positive regression weight. Session number was entered as an additional control variable. Participant was used as a random effect and all possible random slopes were specified (Barr, Levy, Scheepers, & Tily, 2013), but

entered as uncorrelated random effects[3]. Table 1 shows the results, with no effect of Load but a significant effect of Pair and an effect of Session Number (better performance in later sessions).
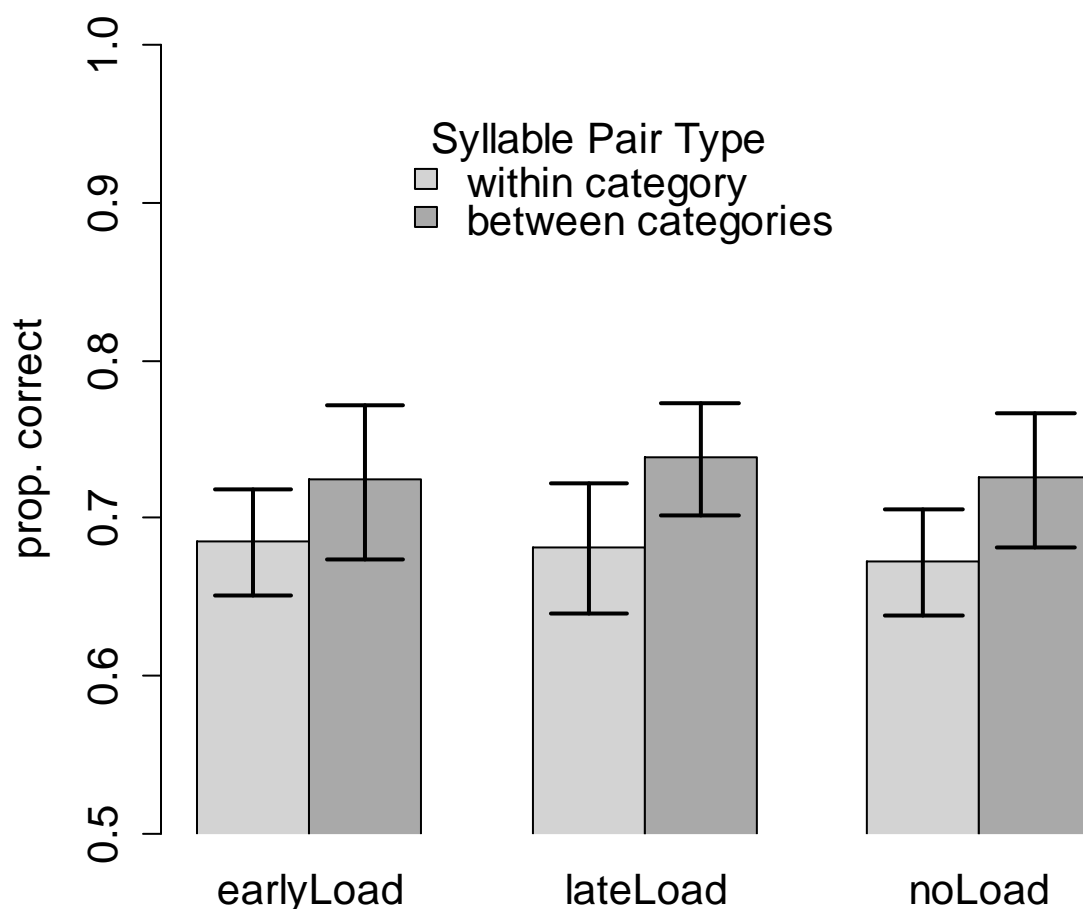

Figure 1: Mean proportion of correct responses on the 4I-oddity task in Experiment 1. Error bars represent the standard error of the mean in logOdds space, projected back on the proportion scale.

Parameter	B	SE	z	p
Intercept	0.77	0.11	6.82	< 0.001
speechPair	0.32	0.13	2.46	0.01
Load				
Load vs no Load	-0.04	0.05	-0.25	0.41
early vs late Load	0.02	0.13	0.19	0.85
SessionNumber	0.42	0.12	3.46	< 0.001
speechPair x Load				
speechPair x (Load vs no Load)	-0.01	0.06	-0.13	0.90
speechPair x (early vs late Load)	-0.03	0.19	-0.17	0.86

[3] With logistic regression models, correlated random effects often lead to convergence problems, hence we used uncorrelated random effects throughout this project. Uncorrelated effects for the two-factor design (Load and Pair) were specified in R as "(1|subject) + (0 + speechPair|subject) + (0 + LoadContrast1| subject) + (0+LoadContrast2| subject) + (0 + speechPair:LoadContrast1|subject) + (0 + speechPair:LoadContrast2|subject) +(0+session|subject)".

Table 1: Results from the mixed-effect model analysis of the data from Experiment 1.

The absence of an effect of load is unexpected given earlier results (Mattys, Barden, & Samuel, 2014; Mattys & Wiget, 2011). One possibility is that participants paid less attention to the face task when the faces were presented simultaneously with the first speech stimulus (late load) than when they were presented prior to it (early load), hence decreasing the chances of finding an effect in what can be thought of as the more challenging condition. However, performance on the face task was numerically better in the late-load than early-load condition (79 % vs 74 %). Even though this difference was not significant (b = 0.29, SE = 0.26, z = 1.1, *p* = 0.25), it shows that participants did not neglect the face task in the presumably more challenging load condition.

Experiment 2 investigated three possible explanations for the contrast between these results and those of Mattys et al. (2011, 2014). The first one is that face-processing is indeed a much more modular task than the visual-search task they used, and therefore might not interfere with speech perception. A second possibility is that that the 4I-oddity task is targeting an early stage of speech processing that is not susceptible to load effects, independent of the type of the secondary task used. Finally, an effect of cognitive load may only occur if visual and auditory information is encoded continuously and concurrently. It might be argued that this would predict an effect of the late-load condition in Experiment 1, in which the faces appeared simultaneously with the first speech stimulus. Note, however, that the first speech stimulus is not crucial for the success on the speech task, since the odd-one-out is always the second or the third, and this might weaken an effect of encoding two faces on the 4I-oddity task. Experiment 2 was designed to tease apart these possibilities.

Experiment 2

Experiment 2 used three different task settings: A no-load baseline and two different types of cognitive load (see Figure 2). The speech task was the same as in Experiment 1 (4I-oddity discrimination task). One of the cognitive-load conditions relied on the visual-search task that previously led to decrements in speech perception (Mattys et al., 2014; Mattys & Wiget, 2011). The other cognitive-load condition was a modified version of the face-recognition task used in Experiment 1. In Experiment 2, participants were presented with the two faces sequentially while listening to the speech stimuli (see Figure 2), as opposed to simultaneously, as in Experiment 1. The rationale is that both this modified face task and the visual search task require continuous scanning and encoding of the visual array during the auditory stimuli.

How do these conditions address the three possible explanations for the results of Experiment 1? If Mattys et al.'s results were specific to the AX discrimination task they used, we should not observe an effect of either type of cognitive load here. On the assumption that the 4I-oddity task taps early speech perception processes, this would indicate that early stages of processing are immune to

cognitive load. A second possibility is that face processing is modular and hence can proceed in parallel with speech perception without interference. If this is the case, we should find a load effect with the visual search task, but not with the modified face perception task. Finally, if concurrent encoding is the crucial component that gives rise to load effects, both types of load should lead to poorer performance in the speech perception task.

Participants

Twenty-four students from the University of Malta participated in the experiment. All of them used English regularly in their daily lives. They were aged 18 to 32. Fifteen of them were female. They were paid for their participation in three sessions (see Materials and Procedure for details).

Materials and Procedure

The same auditory materials and photographs of faces were used as in Experiment 1. For the concurrent visual-search task, we used the visual arrays from Mattys and Wiget (2011), in which a red square is present (or not) in an array of red triangles and black squares (see Figure 2). These stimuli were used to generate three conditions: Speech perception under no load, speech perception with a simultaneous visual-search task, and speech perception with a simultaneous face-recognition task. The speech perception task was the 4I-oddity discrimination task used in Experiment 1. For the visual-search task, participants had to find a red square amongst a 10x10 array of red triangles and black squares (approximately 24 cm x 24 cm). Half the arrays did not contain a red square. The array appeared at the onset of the first syllable and disappeared at the offset of the last syllable.

For the face-recognition task, the first face appeared between the first and second syllables, and the second face between the third and fourth syllables (see Figure 2). Each face stayed on the screen for 800ms. For the no-load condition, the faces also appeared on the screen, but participants were told to ignore them. At the end of the four syllables, participants were first asked whether the second or the third syllable was the odd one out. In the load conditions, they were then asked whether the array contained a red square or not (in the visual-search condition) or whether a probe picture displayed at the end of the trial matched one of the two faces (in the face-recognition condition).

As in Experiment 1, stimulus presentation was controlled by the ExperimentBuilder software from SR Research. Participants performed three sessions on separate days, one with no load, one with the face task as load, and one with the visual-search task as load. Order of sessions was counterbalanced across participants using a Latin square design. Trial presentation order within sessions was randomized offline for each participant individually. The selection of faces and visual arrays was random, with the constraint that half of the trials were "present" trials (there was a red square and the face at test was amongst the two faces presented with the auditory stimuli) and half of the trials were "absent" trials.
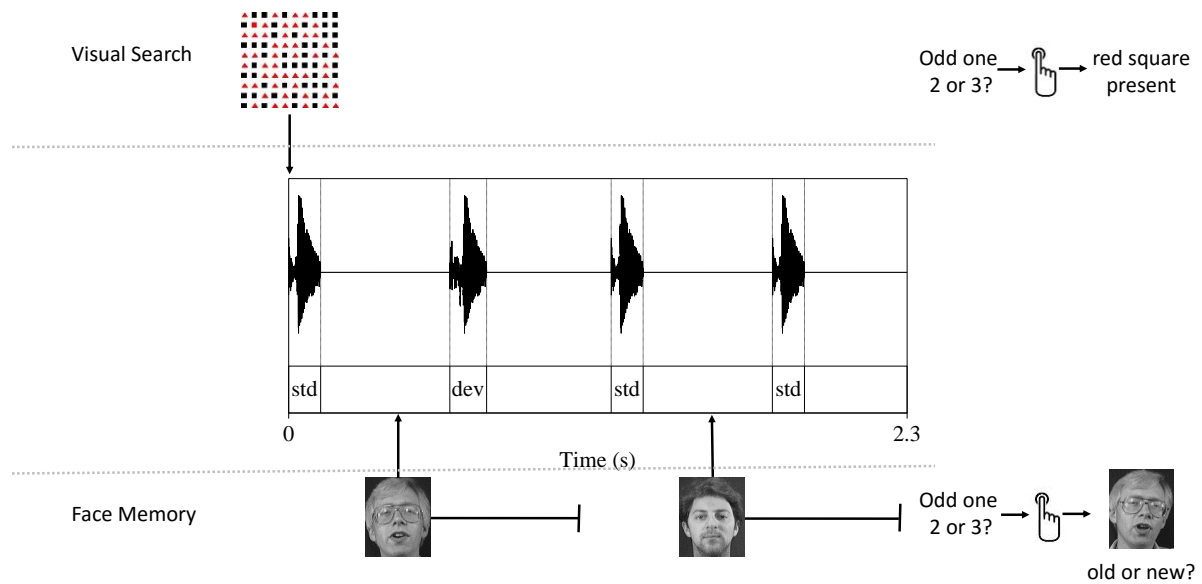
Figure 2: Layout of the stimulus presentation in Experiment 2. The speech stimuli were the same for both load tasks: Four short syllables were played one after the other, with three standards (std) and one deviant (dev) played in either the second or the third position (2nd position in the example). The upper row shows the timing of the concurrent visual-search task. The lower row shows the timing of the concurrent Face-Memory Task.

Results and Discussion

Figure 3 shows the mean proportion of correct responses in Experiment 2. The data revealed no clear effect of speech pair, but there was an effect of load. For the statistical analyses, we used a linear mixed-effect model with a logistic linking function to account for the categorical nature of the dependent variable. As fixed effects, Pair (within versus between category) and Load were coded as numeric contrasts. For the binary variable Pair, within-category was coded as -0.5 and between-category was coded as 0.5. The three levels of Load were re-coded as two linearly independent contrasts, one comparing both load conditions to the no-load condition, and one comparing the two load conditions with each other. Participant was used as a random effect and all possible random slopes were entered as uncorrelated random effects. Table 2 shows that there was an overall effect of Load, but no significant difference between the two load conditions.
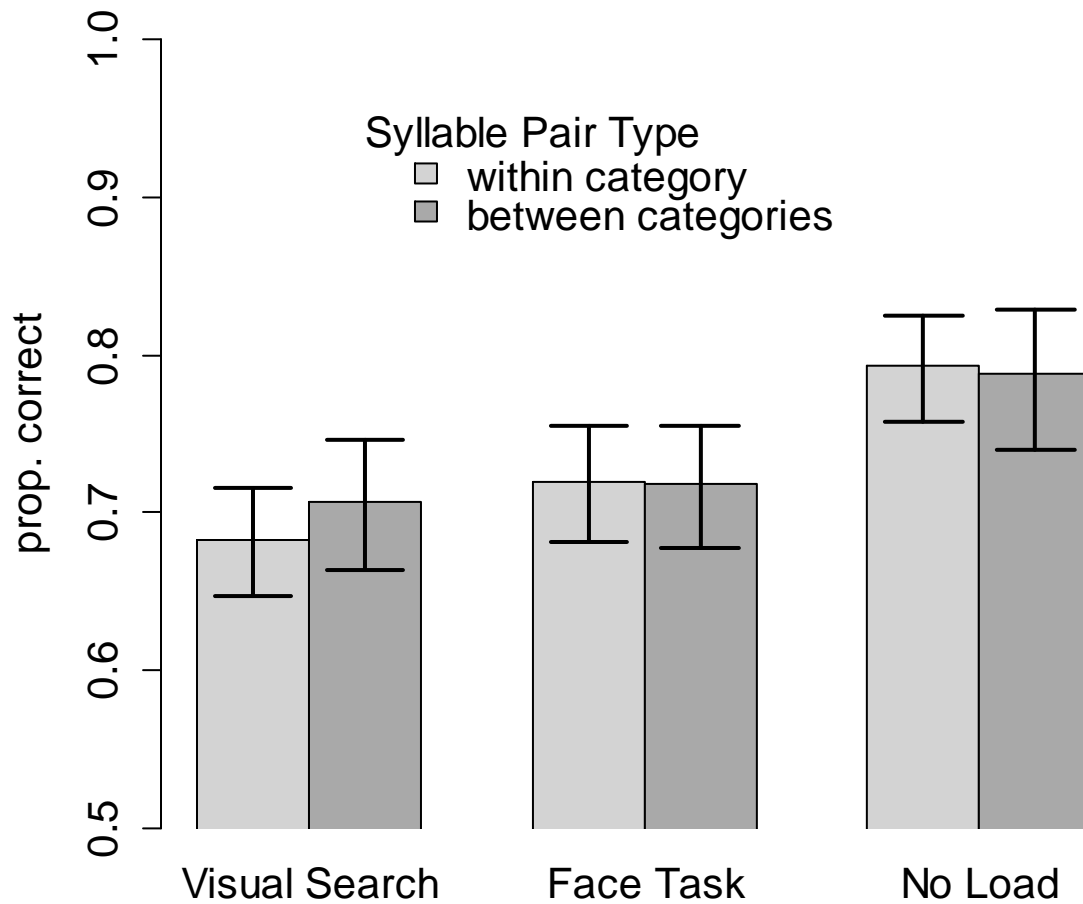
Figure 3: Mean proportion of correct responses on the 4I-oddity task in Experiment 2. Error bars represent the standard error of the mean in logOdds space, projected back on the proportion scale.

Parameter	B	SE	z	p
Intercept	1.16	0.11	10.19	< 0.001
speechPair	0.09	0.17	0.51	0.61
Load				
Load vs no Load	0.35	0.06	5.45	<0.001
Type of Load	0.12	0.08	1.50	0.13
SessionNumber	0.47	0.10	4.75	<0.001
speechPair x Load				
speechPair x (Load vs no Load)	-0.05	0.09	-0.55	0.59
speechPair x (Type of Load)	-0.15	0.11	-1.40	0.16

Table 2: Results from the mixed-effect model analysis of the data from Experiment 2.

Interestingly, we observed a dissociation between interference effects and task difficulty. The visual-search task had higher scores than the face-recognition task (81.2 % vs 72.8 %, respectively, *b* = 0.43, *SE* = 0.10, *z* = 4.28, *p* < .001), suggesting that the face-recognition task is the more difficult one of the two tasks. This is not mirrored in the amount of interference the two tasks caused in speech perception. There was no significant difference in the amount of interference these tasks caused on speech perception, and the descriptive tendency went in the opposite direction, with more interference being caused by the visual-search task. Thus, task difficulty (measured as percentage correct) seems to be independent of the amount of interference the task causes.

It could be argued that the higher percentage of correct responses in the visual-search task indicates that participants paid more attention to this task and hence deployed resources differently for the speech task. However, previous results make this unlikely since Mattys et al. (2014), as noted above, found a clear relationship between difficulty of a search task and the amount of interference that task caused in the concurrent speech-perception task. Nevertheless, if such trading effects occurred in our experiment, they should have led to worse performance on the speech-discrimination task when performance was high on the visual task. To assess such trading effects, we tested whether a correct response on the visual tasks was associated with worse performance on the speech-discrimination task. To do so, we ran another statistical model that included success on the visual task as an additional predictor for the success on the speech-discrimination task. This analysis revealed no evidence of a trading relation; in fact, participants were more likely to answer correctly in the speech-discrimination task if they also answered correctly in the visual task ($b_{\text{visualSuccess}}$ = 0.30, *SE* = 0.10, *z* = 2.95, *p* = 0.003), and this did not interact with the type of visual task (*b* = -0.16, *SE* = 0.13, *z* = 2.95, *p* = 0.22). Thus, there is no evidence that participants were trading resources between the two tasks.

The effect of speech pair (better performance on the between-category pair), which was found in Experiment 1, was not found here. However, a cross-experiment analysis revealed no difference between the two experiments (*t*(44) = 0.95, p = 0.35, adjusted *df* = 41.49), with the main categorical-perception effect constituting only a trend (*t*(44) = 1.84, *p* = 0.07). This shows that, overall, categorical-perception effects were very small if not absent if tested with a 4I-oddity task. Thus, considered together, the results of Experiments 1 and 2 underscore the tendency for categorical-perception effects to be larger with two-interval tasks than with four-interval tasks (Gerrits & Schouten, 2004).

More importantly, this experiment allows us to reconcile Mattys et al.'s results (2011; 2014) with those of Experiment 1. First, we can rule out that the format of the speech-perception task matters. Indeed, both the 4I-oddity task used here and the AX task used in their experiments showed a load effect with the visual-search task. We can also rule out the possibility that face processing, as a

particularly modular domain, does not interfere with speech perception, since the face-recognition task in Experiment 2 gave rise to a load effect. This leaves the possibility that concurrent encoding of visual and auditory information is the crucial condition for load effects to occur.

General Discussion

In two experiments, we aimed to elucidate how speech perception is compromised by cognitive load. Earlier research (Mattys et al., 2014; Mattys & Wiget, 2011) indicated that cognitive load influences aspects of speech perception associated with the early encoding of speech (e.g., phoneme discrimination) but not others (e.g., steepness of phoneme identification function). To investigate this discrepancy, we used a different speech task (oddity discrimination) and a different concurrent task (face recognition). With this combination of tasks, Experiment 1 did not reveal an effect of cognitive load. However, the results showed an effect of speech pair, with better discrimination of syllables straddling a phoneme boundary (/ki/ vs /gi/) than syllables within the same category (e.g., /ki$_1$/ vs /ki$_2$/).

Experiment 2 tested three possible accounts of the lack of a cognitive load effect in Experiment 1, considering the nature of the speech perception task, the nature of the secondary task, and the relative encoding timecourse of the two tasks. Regarding the speech perception task, unlike the AX task used in previous studies, the 4I-oddity task in this study tends to probe pre-categorical, possibly auditory representations of the speech signal (Gerrits & Schouten, 2004; Schouten et al., 2003). Thus, the difference between the earlier research and Experiment 1 could be due to the locus of interference probed by the tasks. However, Experiment 2 showed an effect of cognitive load on the 4I-oddity task, ruling out the possibility that the nature of the speech-perception task is responsible for the null-effects of cognitive load in Experiment 1.

Another possibility was that face perception, which is often thought of as a fairly specialized process, is less likely to interfere with the speech perception task than the visual search task. However, the face-perception load impaired speech perception in Experiment 2, which indicates that it is not the content of the concurrent task but its format that is pivotal in explaining whether or not cognitive load produces interference. Specifically, a cognitive load effect was observed whenever it involved simultaneous encoding of both the auditory and the visual stimuli.

Taken together, the data are consistent with the assumption that continuous scanning and encoding of the visual environment is what hampers speech perception. In fact, this encoding account could also explain the observed dissociation between performance on the load task and the magnitude of the cognitive load effect on speech perception. Recall that, in Experiment 2, performance on the visual-search task was better than performance on the face-recognition task. Yet, the former was numerically more detrimental to speech perception than the latter. This pattern

cannot be explained by an account whereby the mere complexity of the load task determines the size of the detrimental effect on speech perception effect. Instead, under the assumption that processing a 10 x 10 array requires more encoding of visual information than the face-recognition task, the hypothesis that encoding new information is what interferes with speech perception is entirely consistent with our results.

An interesting aspect of the encoding account is that it would also be able to deal with the lack of effect of cognitive load on the steepness of phoneme-categorization functions in Mattys and Wiget (2011). As mentioned in the Introduction, that finding was somewhat surprising because the steepness of a categorization function is usually associated with discrimination ability. However, a difference between speech categorization and speech discrimination is that the latter, but not the former, requires encoding of the stimuli in working memory. In a categorization task, participants simply have to categorize an incoming stimulus without the need to maintain it in memory. In contrast, a discrimination task necessarily entails that one stimulus is maintained in memory and compared with an incoming stimulus. Assuming that the concurrent encoding of visual stimuli creates a memory load, this would explain why speech discrimination, but not speech identification, is affected by cognitive load.

Finally, our data also touch upon theories of limits in working memory, for which a multitude of accounts exist (cf. Cowan, Rouder, Blume, & Scott, 2012). The focus of these accounts is on whether and how more than one stimulus can be held in memory at the same time, and limits are explained either by limited resources or interference between items. Since these models focus on storage and maintenance of information rather than on encoding, the link to our findings is only partial. Nevertheless, in Experiment 2, it is difficult to explain the decrement in dual-task performance purely by interference between items in working memory. Face processing and speech processing have both been described as relatively modular domains, and it is difficult to see how there could be shared features between these two different domains, which is a prerequisite for interference accounts to explain a decrement in performance under dual task conditions (Oberauer & Kliegl, 2006). Our data would in fact fit better with the classical multi-storage model of working memory (Baddeley, 2000), in which a central executive is necessary for encoding information in different domains. The limits of this central executive in encoding information rapidly would be prime candidate for the interference between the tasks observed in the second experiment (cf. Morris & Jones, 1990).

In summary, the current data indicate that the interfering effect of cognitive load on speech perception may not necessarily or solely be due to a reduction of fidelity of early perceptual processes. Instead, a crucial, or at least additional bottleneck may be competition for working memory resources at the encoding stage.

References

AT & T Laboratories Cambridge. (2002). ORL face database. Retrieved from http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html

Baddeley, A. (2000). The episodic buffer: a new component of working memory? *Trends in Cognitive Sciences*, *4*(11), 417–423. http://doi.org/10.1016/S1364-6613(00)01538-2

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. http://doi.org/10.1016/j.jml.2012.11.001

Cowan, N., Rouder, J. N., Blume, C. L., & Scott, J. (2012). Models of verbal working memory capacity: What does it take to make them work? *Psychological Review*, *119*(3), 480–499. http://doi.org/10.1037/a0027791

Díaz, B., Mitterer, H., Broersma, M., & Sebastián-Gallés, N. (2012). Individual differences in late bilinguals' L2 phonological processes: From acoustic-phonetic analysis to lexical access. *Learning and Individual Differences*, *22*(6), 680–689.

Gerrits, E., & Schouten, M. E. H. (2004). Categorical perception depends on the discrimination task. *Perception & Psychophysics*, *66*, 363–376.

Godfrey, J. J., Sydral-Lasky, A. K., Millay, K. K., & Knox, C. M. (1981). Performance of dyslectic children on speech perception tests. *Journal of Experimental Child Psychology*, *32*, 401–424.

Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). The faculty of language: What is it, who has it, and how did it evolve? *Science*, *298*(5598), 1569–1579. http://doi.org/10.1126/science.298.5598.1569

Liberman, A. M. (1996). *Speech: a special code*. Cambridge, Mass: MIT Press.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, *74*(6), 431.

MacKain, K. S., Best, C. T., & Strange, W. (1981). Categorical perception of English /r/ and /l/ by Japanese bilinguals. *Applied Psycholinguistics*, *2*(4), 369–390. http://doi.org/10.1017/S0142716400009796

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

Mattys, S. L., & Wiget, L. (2011). Effects of cognitive load on speech recognition. *Journal of Memory and Language*, *65*(2), 145–160. http://doi.org/10.1016/j.jml.2011.04.004

Morris, N., & Jones, D. M. (1990). Memory updating in working memory: The role of the central executive. *British Journal of Psychology*, *81*(2), 111–121. http://doi.org/10.1111/j.2044-8295.1990.tb02349.x

Oberauer, K., & Kliegl, R. (2006). A formal model of capacity limits in working memory. *Journal of Memory and Language*, *55*(4), 601–626. http://doi.org/10.1016/j.jml.2006.08.009

Pisoni, D. B. (1975). Auditory short-term memory and vowel perception. *Memory & Cognition*, *3*(1). http://doi.org/10.3758/BF03198202

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*. http://doi.org/10.3389/fnsys.2013.00031

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

Schouten, M. E. H., Gerrits, E., & van Hessen, A. (2003). The end of categorical perception as we know it. *Speech Communication*, *41*, 71–80.

Toscano, J. C., McMurray, B., Dennhardt, J., & Luck, S. J. (2010). Continuous perception and graded categorization electrophysiological evidence for a linear relationship between the acoustic signal and perceptual encoding of speech. *Psychological Science*, *21*(10), 1532–1540.