Spoken word processing creates a lexical bottleneck

Alexandra A. Cleland¹, Jakke Tamminen², Philip T. Quinlan², and M. Gareth Gaskell²

¹School of Psychology, University of Aberdeen, Aberdeen, UK
²Department of Psychology, University of York, York, UK

We report 3 experiments that examined whether presentation of a spoken word creates an attentional bottleneck associated with lexical processing in the absence of a response to that word. A spoken word and a visual stimulus were presented in quick succession, but only the visual stimulus demanded a response. Response times to the visual stimulus increased as the lag between it and the spoken word decreased, suggesting a bottleneck in processing. This effect was modulated by the uniqueness point of the spoken word; bottleneck effects were strongest when the spoken word had a late uniqueness point (Experiment 1). The effect was also modulated by the nature of the second task, with the effect stronger when the visual stimulus was a word rather than a shape (Experiment 2) or face (Experiment 3). Word processing appears to create a transient lexical bottleneck that is driven by the magnitude of lexical activity.

Keywords: Spoken word processing; Attention; Uniqueness point.

Outside the laboratory setting, spoken word recognition under adverse conditions is the norm rather than the exception. Much of every day speech processing occurs under conditions of cognitive load; we listen to the radio or hold conversations as we walk, prepare food, drive, and work. Given that we must have limited attentional resources to deal with these tasks, cognitive
load is highly likely to adversely influence our ability to process language. However, despite several decades of research, key questions about the relationship between speech perception and attentional resources remain unanswered. Early research using selective listening tasks reported that participants repeating one of two dichotic speech streams remained unaware of the content of the unattended stream (e.g., Broadbent, 1954, 1958; Cherry, 1953; Moray, 1959; Treisman, 1964a, 1964b). Due to the capacity demanding nature of processing semantic content, Broadbent (1958) argued that irrelevant speech is filtered out at an early stage, before further processing (see Lachter, Forster, & Ruthruff, 2004, for a recent reaffirmation of Broadbent’s selective filter theory). Such an account would suggest that presentation of a spoken word should not be enough, in and of itself, to place demands upon central attentional resources.

Of some relevance is the evidence to suggest that attentional factors can influence lexical effects for tasks such as phoneme monitoring (e.g., Eimas, Marcovitz Hornstein, & Payton, 1990; Mirman, McClelland, Holt, & Magnusson, 2008). For instance, Mattys, Brooks and Cooke (2009) reported that cognitive load impacted upon listeners’ use of cues for speech segmentation. Listeners had to report whether they had heard one of two words (e.g., mile or mild) within a sequence of speech that could lead to a lexically acceptable (e.g., mild option) or unacceptable (e.g., mile doption) result. When attention was divided by introducing a secondary task, participants relied more heavily on lexical-semantic rather than on acoustic detail to make their decision (e.g., becoming more likely to report mild option); this suggests that cognitive load must have some effect on the processes underlying speech segmentation.

Recent evidence from the psychological refractory period (PRP) paradigm suggests that, while some aspects of word processing can proceed when attention is engaged with another task, others must be attentionally demanding (e.g., Cleland, Gaskell, Quinlan, & Tamminen, 2006; Gaskell, Quinlan, Tamminen, & Cleland, 2008; for similar research on visual word recognition, see e.g., McCann, Remington, & Van Selst, 2000; Reynolds & Besner, 2006; Ruthruff, Allen, Lien, & Grabbe, 2008). The PRP arises when participants are asked to make sequential speeded responses to two different tasks (task 1 and task 2) in quick succession (e.g., Telford, 1931), with performance on both tasks monitored as a function of stimulus onset asynchrony (SOA). Response times associated with the first task tend to be unaffected by the manipulation of SOA; however, response times to the second task increase as the SOA decreases. This pattern is interpreted as a consequence of a bottleneck or resource limitation that prevents one or more stages of processing from being carried out simultaneously for task 1 and task 2. Task 2 processing can proceed through its early stages; however, at
some point it will require access to limited central attentional resources. At this point, processing is halted until the bottleneck associated with task 1 processing clears.

A large part of the interest in the PRP in the language domain is driven by the fact that it can be used to determine the stage of processing influenced by difficulty manipulations of task 2. In short, if a difficulty manipulation reflects stages of processing that can proceed unhindered by the fact that central resources are engaged with another task, then the effect of the difficulty manipulation will be attenuated at short SOAs. This is because the extra processing associated with the difficult version of the task is completed while later stages of task 2 wait for the bottleneck to clear. PRP studies pairing lexical decision or naming with an unrelated task have demonstrated this “underadditive” PRP pattern for a number of manipulations of lexical difficulty; processing associated with phonemic ambiguity (Gaskell et al., 2008), repetition priming (Reynolds & Besner, 2006) and frequency (e.g., Cleland et al., 2006; Ruthruff et al., 2008; although cf. McCann et al., 2000) is unaffected by the bottleneck associated with an unrelated task. In contrast, aspects of word processing that are subject to the bottleneck include those affected by grapheme-phoneme complexity (Reynolds & Besner, 2006) and some manipulations of semantic difficulty (e.g., Lien, Ruthruff, Cornett, Goodin, & Allen, 2008; Tamminen, Cleland, Quinlan, & Gaskell, 2006). Regardless of more fine-grained manipulations of processing difficulty, however, these studies are striking because the first task never involves language processing and yet both tasks are subject to the central bottleneck. In all of these studies, responses to lexical stimuli are delayed when there is only a short stimulus asynchrony between the two tasks.

In other words, although some aspects of word recognition can proceed without access to central resources, there is still a cost associated with the presence of the first, unrelated, task. This must imply that some stages of word processing require access to central resources if they are to be completed, at least when a response is associated with that word. Of course, it may be that it is solely the selection of a response that requires access to central resources (and this would be entirely consistent with established PRP logic, which associates the bottleneck with response selection). It is less clear whether word recognition itself demands limited, central resources independent of response selection.

To address the question of whether lexical activity can create a bottleneck in processing, the current studies involved two stimuli presented in close succession. Stimulus one (S1) was a spoken word, which participants were not required to respond to. Stimulus 2 (S2) was presented visually and required a go/no-go response from participants. The go/no-go task typically produces faster response times than the lexical decision task (e.g., Perea,
Rosa, & Gómez, 2002), and as such should be more likely to show effects of S1 processing on S2 responses, assuming that these effects are transient. The crucial measure was whether S2 responses were influenced by the properties and timing of S1. If spoken word recognition does induce a processing bottleneck, even in the absence of response selection, then there should be a slowing of S2 responses when there is a short stimulus asynchrony between S1 and S2.

Of course, any such slowing could just be some kind of distraction effect, with the word presented as S1 slowing responses to S2 only because the acoustic stimulus momentarily directs attention away from the task in hand. In order to investigate whether spoken word processing itself creates a bottleneck, the uniqueness point (the point at which a word can be uniquely identified from phonologically overlapping competitors) of the word presented as S1 was manipulated. If any delay in S2 processing associated with the presentation of S1 at short asynchronies is attributable specifically to lexical activity, we would expect the effect of the uniqueness point manipulation to interact with any effect of stimulus asynchrony. This is based on the assumption that late uniqueness words will result in a greater degree of lexical activity than early uniqueness words (see e.g., Mattys & Clark, 2002; Mattys, Pleydell-Pearce, Melhorn, & Whitecross, 2005). If S2 is presented after S1 lexical access is completed (as would be the case where the uniqueness point is located towards the beginning of S1), we would not expect to find a delay in S2 processing at the short SOA relative to the long SOA. However, if S1 lexical processing is still ongoing (as would be the case with a late uniqueness of S1) then we would expect to see a bottleneck effect.

In Experiment 1, S1 was a spoken word and S2 was a visually presented letter string. The onset latency of S2 was also manipulated. According to the PRP logic, if a processing bottleneck has occurred, then the delay in processing S2 should be greatest at the shortest asynchrony, decreasing as the stimulus asynchrony increases. This is because, when S1 and S2 are presented with a relatively long stimulus asynchrony, the transient processing bottleneck associated with S1 will have cleared by the time S2 processing requires central resources. Assuming that any bottleneck observed should be associated with lexical activity, we predicted that the bottleneck effect should be greatest for late uniqueness items. As a further manipulation of lexical activity, the completeness of the words presented as S1 varied, with the final consonant excised in half of the trials.

The intention of this truncation manipulation was to examine whether PRP effects are stronger when the word recognition process is left suspended by the absence of the remaining speech. For late uniqueness point items this truncated condition should leave the system with a high level of competition,
EXPERIMENT 1

Method

Participants

Forty-eight undergraduate students (39 female and 9 male) from the University of York were either paid £2 or received course credit for participation. All were native English speakers and reported no speech, hearing or visual disorders.

Materials

S1 was a spoken word, either complete or truncated. Eighty items were taken from Gaskell and Marslen-Wilson (2002) (see Appendix 1), including the original sound files. For each of these words the truncation point used in Gaskell and Marslen-Wilson (described there as “Cut2”) was selected. This was generally three pitch periods into the second syllable vowel, allowing the vowel to be identified. When there were strong cues to the identity of the following consonant, the cut-off point was earlier. Forty items had an early uniqueness point (i.e. had no competitors fully matching the speech presented prior to the cut-off point) and forty had a late uniqueness point (i.e. had at least one competitor that matched the truncated speech). Items were matched on factors such as duration (638 ms for early uniqueness and 632 ms for late uniqueness), frequency (18.3 for early uniqueness and 16.5 for late uniqueness) and concreteness (mean 541 for both early and late uniqueness items). An additional 80 spoken words were also selected from the Gaskell and Marslen-Wilson set to be used as S1 words preceding nonword fillers.

S2 was a visually presented letter string. Eighty high-frequency words were selected from the Cleland et al. (2006) set of items (see Appendix 1). High-frequency words were chosen to elicit high degrees of accuracy from the participants; all had a frequency count of more than 100 per million based on the CELEX database (Baayen, Piepenbrock, & van Rijin, 1993). A further 80 nonwords were also selected from the Cleland et al. set. All words were monosyllabic.

Design

Visual S2 words were paired with the 80 spoken words such that there was no phonological, orthographic or obvious semantic link between the paired
words. The completeness of the spoken word was manipulated; S1 spoken words could either be whole or truncated. Stimulus asynchrony between S1 and S2 also varied; S2 could either be presented at the cut off point for each word (irrespective of whether the word was complete or truncated) (0 ms condition) or 100 ms later (100 ms condition). The selection of such short asynchronies was based on the assumption that any bottleneck associated with S1 would be short-lived in nature, and so as to allow comparison of any bottleneck effect with the shortest stimulus asynchronies in other PRP studies. In addition, the uniqueness point (early vs. late) of the S1 spoken word was manipulated. The experimental conditions were rotated through items and participants. Hence, each visual S2 word appeared once to each participant, in one of the eight conditions defined by the two levels of truncation (whole vs. truncated), the two levels of uniqueness point (early vs. late) and the two levels of stimulus asynchrony (0 ms vs. 100 ms). Additionally, each S2 word appeared equally often in each of these conditions across participants.

Procedure

Participants were seated in a quiet booth, a comfortable distance from the computer screen, with a button box on the table in front of them. They were instructed that they had to respond to a letter string presented on the screen. If the letter string was a real word, they should press the “word” button as quickly as they could. However, if they thought it was not a word, they were instructed to do nothing. Participants were told that they would hear a word over the headphones before they saw the word on the screen, but that it was the letter string on the screen which they should respond to. They were also told that they would be presented with a list of words at the end of the experiment, and that they would have to tick the ones that they thought they had heard during the session. They were told that they should therefore listen to the words over the headphones, but not to try and memorise them.

On each trial, a fixation cross (i.e., +) appeared on the screen (1000 ms). This was followed by the presentation of the spoken word over the headphones. Either at the cut off point (regardless of whether the word was actually truncated or not), or 100 ms after, the S2 word was presented on the screen. This remained on the screen until either the participant made a response, or 2000 ms had elapsed. The screen then went blank for 1000 ms, followed by the fixation point for the next trial. At the end of the session, participants were given a list of words and asked to tick any they thought they had heard during the experiment. The experimental session was preceded by a practice session of eight trials, with items not included in the experimental session.

RTs were measured from the onset of presentation of S2. The experimental stimuli were presented, and the participants’ response times recorded.
using E-Prime software on a Holly Systems PC running Windows XP. The stimuli were presented on an Iiyama 19" CRT monitor. Button presses were recorded from a Cedrus response box.

Results

Trials in which participants responded post-2000 ms or did not respond at all were excluded from the analysis (<0.5% of trials). Table 1 shows the RTs to S2 in all experimental conditions. Separate analyses treated participants ($F_1$) and items ($F_2$) as random effects. Analyses of variance (ANOVAs) on the RT data with uniqueness point (early vs. late), truncation (complete vs. truncated), and stimulus asynchrony (0 ms vs. 100 ms) revealed a statistically significant main effect of stimulus asynchrony, $F_1(1, 47) = 12.5, p < .001, \eta^2 = .21$; $F_2(1, 79) = 4.25, p = .043, \eta^2 = .051$. RTs were 12 ms slower at the 0 ms stimulus asynchrony than at the 100 ms stimulus asynchrony.

There was also a statistically significant interaction between stimulus asynchrony and uniqueness point, although this was marginal in the items analysis, $F_1(1, 47) = 4.25, p = .045, \eta^2 = .08$; $F_2(1, 79) = 4.00, p = .061, \eta^2 = .044$. For early uniqueness point words, there was a stimulus asynchrony difference of 5 ms; for late uniqueness words this rose to 20 ms. Simple planned comparisons revealed that the effect of stimulus asynchrony was only significant for words that had a late uniqueness point and so still had

<table>
<thead>
<tr>
<th>Stimulus asynchrony</th>
<th>0 ms</th>
<th>100 ms</th>
<th>Stimulus asynchrony effect (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Stimulus asynchrony</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0 ms</td>
<td>100 ms</td>
<td></td>
</tr>
<tr>
<td><strong>S1 condition</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Early uniqueness</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Whole</td>
<td>514 (13)</td>
<td>505 (14)</td>
<td>9</td>
</tr>
<tr>
<td>Truncated</td>
<td>509 (12)</td>
<td>508 (14)</td>
<td>1</td>
</tr>
<tr>
<td>Mean</td>
<td>511 (12)</td>
<td>506 (12)</td>
<td>5</td>
</tr>
<tr>
<td>Late uniqueness</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Whole</td>
<td>526 (13)</td>
<td>501 (12)</td>
<td>25*</td>
</tr>
<tr>
<td>Truncated</td>
<td>511 (13)</td>
<td>496 (13)</td>
<td>15*</td>
</tr>
<tr>
<td>Mean</td>
<td>518 (12)</td>
<td>498 (12)</td>
<td>20**</td>
</tr>
</tbody>
</table>

*Note: Values in brackets represent standard error.
*Effect of stimulus asynchrony was significant in the participants but not the items analysis.
*Effect of stimulus asynchrony was significant in pairwise comparisons by both participants and items, $p < .05$; **$p < .01$. 
matching competitors at the alignment point, $F_1(1, 47) = 16.99, p < .001; \eta^2 = .267$, $F_2(1, 79) = 8.61, p = .004, \eta^2 = .098$. There were no other statistically reliable effects: notably, the pattern of response times was not affected by the truncation manipulation.

**Discussion**

Experiment 1 showed a statistically significant slowing of responses to S2 at the short stimulus asynchrony following S1 relative to the longer stimulus asynchrony. This suggests that the processing of S2 was delayed until some aspect of processing S1 was complete. However, the bottleneck was selective for late uniqueness point stimuli; when S1 was an early uniqueness point word there was no effect of stimulus asynchrony on S2 processing. This effect was in the absence of any response requirement for S1 and fits the profile of a bottleneck or resource limitation that is transient in nature and occurs as a consequence of the normal word recognition process in speech perception. For early uniqueness point words, this bottleneck is cleared by the time S2 is processed, meaning that it has no effect on processing speed, whereas for late uniqueness point words the bottleneck exerts a cost on the processing of a closely following stimulus.

The fact that the size of the stimulus asynchrony effect depended on uniqueness point strengthens the case that the processing cost involves some aspect of lexical processing rather than being due merely to the presence of a distracting stimulus. We examined bottleneck effects for both complete words and words truncated by removing their final consonants. For the late uniqueness point items only, this changes the nature of the competition/recognition process, leaving it in an incomplete state. As this factor did not significantly interact with the stimulus asynchrony factor, it is difficult to draw strong conclusions about its effect. Given that the effect of stimulus asynchrony was numerically weaker for the late uniqueness point items when stimuli were truncated (14 ms vs. 25 ms), there is some indication that performance costs reflect delays that are contingent upon completion of the recognition process.

The results of Experiment 1 suggest that there is a cost directly associated with lexical processing. We see two possible ways in which such a cost might come about. One is that the cost applies only to subsequent lexical (i.e., word recognition) processes. By this account, the recognition process uses up modular lexical resources that are required for the recognition of the word presented as S2. The second possibility is that the bottleneck observed in Experiment 1 has more general consequences; in other words, word recognition generates a processing cost that limits the capacity of decision-making mechanisms in general.
To examine these issues in more detail, Experiment 2 was designed to directly compare lexical and non-lexical tasks associated with S2. Participants performed different tasks in the two blocks of the experiment. In one block (in the lexical condition), they pressed a button if the S2 letter string was a word (making no response if it was a nonword), and in the other block they pressed a button if a coloured rectangle was blue (making no response if it was green). Based on the interaction of stimulus asynchrony by uniqueness point in Experiment 1, we made separate predictions for early and late uniqueness items. For early uniqueness items, we predicted that the effect of stimulus asynchrony would be small, and similar for both tasks since Experiment 1 showed that for early uniqueness items the bottleneck clears before impacting on S2 processing. For late uniqueness items we predicted that the effect of stimulus asynchrony should be stronger. Furthermore, if the bottleneck caused by S1 was a general resource limitation rather than being limited to lexical processing then the asynchrony effect should be similar in magnitude across the two tasks. Alternatively, if the bottleneck includes a component specific to lexical processing then the stimulus asynchrony effect should be stronger for the visual word task than the colour task.

One issue that arises from the methodology of Experiment 1 relates to the list of words presented at the end of the study. Although participants did not have to respond to the spoken words, they were aware that they would be presented with a list of words at the end of the experiment, and that they would be asked to indicate which of the words they remembered hearing during the course of the study. The motivation for this was to ensure participants remained engaged with the task. However, it raises the problem that, although participants did not have to respond to the words at the time, they did have to attend to them. One possibility is that the lexical bottleneck is driven by intentional encoding; that is, it only occurs when participants have to identify the word. In order to examine this possibility, in Experiment 2 there was no task associated with the spoken words whatsoever. Participants were told that they would hear words presented over headphones but that it was only the stimuli on the screen that they had to respond to. If the effect of stimulus asynchrony disappears in Experiment 2, then this would suggest that the lexical bottleneck arises due to intentional encoding of S1 rather than arising from the lexical activity associated with S1.

EXPERIMENT 2

Participants

Sixty-four undergraduate students (24 male and 40 female) from the University of Aberdeen received course credit for participation. Participants
were native English speakers and did not report any speech, hearing or visual disorders.

Materials
S1 was a spoken word. The same set of 80 words was used in Experiment 2 as in Experiment 1; however, only complete words were used. S2 was either a visually presented word (see Appendix 1) or nonword, or a blue or green rectangle. As before, a further 80 spoken words were used as S1 words preceding nonword fillers.

Design
As before, the S2 items were paired with the 80 spoken words. The stimulus asynchrony between S1 and S2 was manipulated such that despite S1 always being presented whole, S2 was either presented at the same cut-off alignment point as in Experiment 1 (0 ms condition) or 100 ms later (100 ms condition). This was rotated through items and participants as before. In addition, the nature of S2 was rotated through the uniqueness point and SOA conditions. Hence, each item could occur in one of the eight conditions defined by the two levels of uniqueness point (early vs. late), the two levels of stimulus asynchrony (0 ms vs. 100 ms), and the two levels of S2 type (lexical vs. colour). Each participant took part in two blocks; in one of these blocks they were presented with the lexical decision task, and in the other they were presented with the colour decision task. The order of these two blocks was counterbalanced across participants.

Procedure
Participants were instructed that they would complete two blocks of the study. They were told that in both blocks, they would hear a spoken word over headphones, but that their responses were only to the stimulus presented on the screen. They were instructed that in one block they would be required to make a response as quickly as they could if they saw a word on the screen and not to respond if they saw a nonword; in the other block they should make the response as quickly as they could if the rectangle they saw was blue and make no response if it was green.

The experimental stimuli were presented, and the participants’ response times recorded using E-Prime software on a Dell PC running Windows XP. The stimuli were presented on a Dell 19" flat panel monitor. Button presses were recorded from an E-Prime SR Box. In all other aspects, the procedure was similar to Experiment 1.
Results

Trials in which participants responded post-2000 ms or did not respond at all were excluded from the analysis (<0.5% of all trials). Table 2 shows the RTs to S2 in all experimental conditions. ANOVAs on the RT data with uniqueness point (early vs. late), stimulus asynchrony (0 ms vs. 100 ms), and task type (lexical vs. colour) revealed a statistically significant main effect of stimulus asynchrony, $F_1(1, 63) = 47.18, p < .001, \eta^2 = .428$; $F_2(1, 78) = 25.12, p < .001, \eta^2 = .244$, with RTs 15 ms slower at the 0 ms stimulus asynchrony than the 100 ms stimulus asynchrony. There was also a statistically significant main effect of task type, $F_1(1, 63) = 580.17, p < .001, \eta^2 = .902$; $F_2(1, 78) = 948.01, p < .001, \eta^2 = .924$, with responses to the colour task 153 ms faster than the lexical task.

There was also an interaction of task by stimulus asynchrony, although this did not reach significance in the items analysis, $F_1(1, 63) = 4.61, p = .036, \eta^2 = .068$; $F_2(1, 78) = 2.37, p = .13, \eta^2 = .029$, with a 19 ms effect of stimulus asynchrony for the lexical task and a 10 ms effect for the colour task. No other effects were significant. In particular, the three-way interaction of task by stimulus asynchrony by uniqueness did not reach significance, $F_1(1, 63) = 2.70, p = .11, \eta^2 = .11$; $F_2(1, 78) = 2.81, p = .098, \eta^2 = .035$.

Although the three-way interaction did not reach significance, we nevertheless considered the task by asynchrony interaction separately for early and late uniqueness point items, in line with our prediction that the stimulus asynchrony effect should be carried by the late uniqueness point condition. For late uniqueness items, the stimulus asynchrony effect was significantly greater for the lexical task (26 ms) than the colour task (9 ms), $F_1(1, 63) = 5.28, p = .025, \eta^2 = .077$; $F_2(1, 78) = 5.56, p = .021, \eta^2 = .067$. By

### Table 2

Response times to S2 across conditions for Experiment 2

<table>
<thead>
<tr>
<th>Stimulus asynchrony</th>
<th>0 ms</th>
<th>100 ms</th>
<th>Stimulus asynchrony effect (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Early uniqueness</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Colour task</td>
<td>353 (6)</td>
<td>342 (7)</td>
<td>11*</td>
</tr>
<tr>
<td>Lexical task</td>
<td>508 (9)</td>
<td>496 (9)</td>
<td>12</td>
</tr>
<tr>
<td>Late uniqueness</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Colour task</td>
<td>357 (8)</td>
<td>348 (8)</td>
<td>9*</td>
</tr>
<tr>
<td>Lexical task</td>
<td>516 (9)</td>
<td>490 (8)</td>
<td>26**</td>
</tr>
</tbody>
</table>

Note: Values in brackets represent standard error.

*Effect of stimulus asynchrony was significant in pairwise comparisons by both participants and items, $p < .05$; **$p < .01$. 

582 CLELAND ET AL.
contrast, when only early uniqueness items were considered, the stimulus asynchrony effect was equivalent for the lexical task (12 ms) and the colour task (11 ms), both $F$s < 1.

Discussion

Experiment 2 replicated the effect of stimulus asynchrony observed in Experiment 1; however, this was modulated by the nature of task 2. When S2 was a visual word and the task involved lexical judgement, the stimulus asynchrony was stronger (19 ms) than when the task was a colour judgement to rectangles (10 ms). This effect was carried by the late uniqueness stimuli alone; when only early uniqueness items were considered, there was no difference in the effect of stimulus asynchrony for the colour and lexical task.

Experiment 2 additionally provided a stronger test of whether a bottleneck was observed in the complete absence of a response to S1. As the stimulus asynchrony effect remained for late UP items in the lexical task, there was no evidence that the bottleneck effect observed in Experiment 1 relied on intentional or explicit encoding of S1; passive listening had the same effect. To further address this question, we carried out a direct comparison of the two studies omitting the truncated trials from Experiment 1 and the colour trials from Experiment 2. ANOVAs with uniqueness point (early vs. late) and stimulus asynchrony (0 ms vs. 100 ms) as within participants factors and memory task (word list vs. no word list) as a between participants factor revealed no interactions with the manipulation of word list (both $F$s < 1.2). When participants were instructed that they would be asked to indicate which words they remembered at the end of the experiment, there was a 17 ms effect of stimulus asynchrony; when there was no task associated with the spoken word whatsoever, there was a 19 ms effect of stimulus asynchrony. The stimulus asynchrony effect demonstrated in Experiment 1 cannot therefore be attributed to task-related encoding of S1.

The difference in the size of the stimulus asynchrony effect for the lexical and colour task is particularly striking given the fact that responses to the coloured rectangle were 153 ms faster than the visual word; as processing of the coloured rectangle was much faster than the visual word, we might have expected to see a larger bottleneck effect for the colour task under the assumption that the colour task requires earlier access to central resources than the lexical task and so would be more likely to be halted by a transient bottleneck and, more importantly, halted for a longer duration. However, there are also two potential problems with the colour task used in Experiment 2.

The first relates to whether the reduced stimulus asynchrony effects are due to the difficulty of the colour task. Unlike the lexical task, the stimuli
were the same on every “blue” trial; there was no variation between trials as there would be for the different word stimuli. Response times to the colour task were 153 ms faster than to the word task, suggesting that participants found it easier than the lexical task. In other words, it is possible that the interaction between task and stimulus asynchrony was due to the difficulty of the tasks rather than the nature of the tasks.

Secondly, we could not rule out the possibility that participants carried out some limited lexical processing in performing the colour task; it is possible that they used verbal labels to help them make their decision on how to respond to the two colours. Indeed, recent evidence suggests that performance on some colour discrimination tasks is improved when those colours correspond to language-specific category labels (e.g., Winawer et al., 2007). The effect of stimulus asynchrony, although significantly reduced, was still significant for the colour task in the pairwise comparisons (see Table 2). This could reflect the fact that the spoken word stimulus creates a general resource limitation, but that the colour task was just not difficult enough to be affected comparably to the lexical task (as suggested above). Alternatively, it may be that participants were engaging in some linguistic strategy to complete the task, and so the small but significant effect of stimulus asynchrony in the colour task was actually due to a lexical bottleneck rather than a more general resource limitation.

Experiment 3 was designed to address these issues. As in Experiment 2, participants performed different tasks in the two blocks of the experiment. In one block (the lexical condition), they pressed a button if the S2 visual string was a word (making no response if it was a nonword), and in the other block they pressed a button if a face presented on the screen showed a fearful expression (making no response if it was neutral). Experiment 3 required a non-lexical task comparable in difficulty to lexical decision, but for which participants were unlikely to use a verbal labelling strategy. The face task was selected after pilot work demonstrated that RTs to the faces were broadly comparable to those for the lexical decision task. Unlike the colour task, a different stimulus was encountered on each trial of the face block, making the task more similar in nature to the lexical task. In addition, responses to emotional content were deemed less likely to engage a verbal labelling strategy than the colour task.

### EXPERIMENT 3

**Participants**

Sixty-four undergraduate students (48 female and 16 male) from the University of Aberdeen received course credit for participation. Participants
were native English speakers and did not report any speech, hearing or visual disorders.

**Materials**

S1 was a spoken word, with the same set of 80 words used in Experiment 2 used in Experiment 3. S2 was either a visually presented word (see Appendix 1) or nonword, or a face with either a fearful or neutral expression. Facial (full face) images were taken from The Karolinska Directed Emotional Faces database (Lundqvist, Flykt, & Öhman, 1998).

As before, a further 80 spoken words were used as S1 words preceding nonword and neutral face fillers.

**Design**

As before, the S2 items were paired with the 80 spoken words. As in Experiment 2, the stimulus asynchrony between S1 and S2 was manipulated such that S2 was either presented at the same cut-off alignment point as in Experiment 1 (0 ms condition) or 100 ms later (100 ms condition). This was rotated through participants and items as before. In addition, the nature of S2 was rotated through the uniqueness point and SOA conditions. Hence, each item could occur in one of the eight conditions defined by the two levels of uniqueness point (early vs. late), the two levels of stimulus asynchrony (0 ms vs. 100 ms), and the two levels of S2 type (lexical vs. face). Each participant took part in two blocks; on one of these blocks they were presented with the lexical decision task, and in the other they were presented with the face task. The order of these two blocks was counterbalanced across participants.

**Procedure**

Participants were instructed that they would complete two blocks of the study. They were told that in both blocks, they would hear a spoken word over headphones, but that their responses were only to the stimulus presented on the screen. They were instructed that in one block they would be required to make a response as quickly as they could if they saw a word on the screen and not to respond if they saw a nonword; in the other block they should make the response as quickly as they could if the face they saw on the screen was fearful and make no response if it had a neutral expression. In all other aspects, the procedure was similar to Experiment 2.

**Results**

Trials in which participants responded post-2000 ms or did not respond at all were excluded from the analysis (<4% of trials). The increase in error rate
compared to Experiments 1 and 2 was driven by the face task, with 77% of post-2000 ms or no-response trials occurring for the face task. Table 3 shows the RTs to S2 in all experimental conditions. ANOVAs on the RT data with uniqueness point (early vs. late), stimulus asynchrony (0 ms vs. 100 ms), and task type (lexical vs. emotion) revealed a statistically significant main effect of stimulus asynchrony, $F_1(1, 63) = 15.62, p < .001, \eta^2 = .199$; $F_2(1, 78) = 8.13, p = .006, \eta^2 = .094$ with RTs 13 ms slower at the 0 ms stimulus asynchrony than the 100 ms stimulus asynchrony. There was also a statistically significant main effect of task type, $F_1(1, 63) = 7.83, p = .007, \eta^2 = .11; F_2(1, 78) = 12.07, p = .001, \eta^2 = .13$, with RTs to the face task 31 ms slower than the lexical task. There was a marginal interaction of task by stimulus asynchrony, $F_1(1, 63) = 2.88, p = .09, \eta^2 = .044; F_2(1, 78) = 2.93, p = .09, \eta^2 = .036$, with an 8 ms effect of stimulus asynchrony for the face task, but an 18 ms for the lexical task. Planned comparisons revealed that the effect of stimulus asynchrony was in fact only significant for the lexical task, $F_1(1, 63) = 24.96, p < .001, \eta^2 = .28; F_2(1, 78) = 10.14, p = .002, \eta^2 = .12$, with no significant effect of stimulus asynchrony for the face task, $F_1(1, 63) = 2.62, F_2(1, 78) < 1$. No other effects were significant. In particular, the three-way interaction of task by stimulus asynchrony by uniqueness did not reach significance, $F_1(1, 63) = 2.11, p = .15, \eta^2 = .032; F_2(1, 78) = 0.45, p = .50, \eta^2 = .006$.

As with Experiment 2, we considered the task by asynchrony interaction separately for early and late uniqueness point items, in line with the prediction that the stimulus asynchrony effect should be carried by the late uniqueness

TABLE 3
Response times to S2 across conditions for Experiment 3

<table>
<thead>
<tr>
<th>Stimulus asynchrony</th>
<th>0 ms</th>
<th>100 ms</th>
<th>Stimulus asynchrony effect (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Early uniqueness</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Face task</td>
<td>537 (15)</td>
<td>525 (16)</td>
<td>12*</td>
</tr>
<tr>
<td>Lexical task</td>
<td>507 (9)</td>
<td>492 (10)</td>
<td>15**</td>
</tr>
<tr>
<td><strong>Late uniqueness</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Face task</td>
<td>539 (15)</td>
<td>534 (18)</td>
<td>5</td>
</tr>
<tr>
<td>Lexical task</td>
<td>516 (10)</td>
<td>495 (10)</td>
<td>21**</td>
</tr>
</tbody>
</table>

Note: Values in brackets represent standard error.

*Effect of stimulus asynchrony was significant in the participants but not the items analysis.

*Effect of stimulus asynchrony was significant in pairwise comparisons by both participants and items ($p < .05$); **$p < .01$. 
point condition. For late uniqueness items, the stimulus asynchrony effect was greater for the lexical task (21 ms) than the face task (5 ms), $F_1(1, 63) = 4.24$, $p = .044$, $\eta^2 = .063$; $F_2(1, 78) = 3.03$, $p = .086$, $\eta^2 = .037$, although this interaction was marginal in the items analysis. Pairwise comparisons revealed that the effect of stimulus asynchrony was only significant for the late uniqueness point items in the lexical task (see Table 3). By contrast, when early uniqueness items were considered, the stimulus asynchrony effect was equivalent for the lexical task (15 ms) and the face task (12 ms), both $Fs < 1$.

**Discussion**

The results of Experiment 3 are highly consistent with those of Experiments 1 and 2. When S2 was a visual word and the task involved lexical judgement, the stimulus asynchrony was stronger (18 ms, compared to 19 ms in Experiment 2) than when the task was non-lexical (8 ms, compared to 10 ms in Experiment 2). As in Experiment 2, the interaction of stimulus asynchrony by task type appeared to be carried largely by the late uniqueness items; when only early uniqueness items were considered, there was no difference in the effect of stimulus asynchrony for the face and lexical task. We note that the three-way interaction of task by stimulus asynchrony by uniqueness did not reach significance in either Experiment 2 or 3; however, the replication in the separate analyses of early and late uniqueness items demonstrates a consistent pattern.

The face task was more comparable to the lexical task in terms of RTs and distribution than the colour task. Inspection of the raw RTs produced across all trials found that the overall mean RT for the face task was 530 ms ($SD = 172$) and the overall mean RT for the lexical task 502 ms ($SD = 115$). The difference between the two tasks was significant; however, unlike in Experiment 2 (where the overall mean RT for the colour task was 352 ms, $SD = 92$), the lexical task was responded to slightly faster than the non-lexical task. Hence, we can rule out the possibility that the pattern of results in Experiment 2 was due to the fact that the colour task was less difficult than the lexical task. In summary, the results of Experiment 3 support the conclusion that the cognitive load associated with S1 is lexical in nature, rather than being a more general resource limitation.

**GENERAL DISCUSSION**

The results of the current studies demonstrate that processing of a spoken word stimulus creates a brief resource limitation that slows processing of a subsequent word at short asynchronies. Crucially, this effect was modulated by the nature of the spoken word; slowing was only apparent for words that
could not be uniquely identified at the point in time at which the second stimulus was presented. This suggests that the bottleneck itself is created by lexical activity; in the case of early uniqueness words, where lexical access had already occurred, the bottleneck was eliminated (Experiment 1). Furthermore, components of this bottleneck appear to be unique to lexical processing. When late uniqueness point items were considered in Experiment 2, the effect of stimulus asynchrony was greater for the lexical task than the colour task. In Experiment 3, where the lexical task was compared with an emotion recognition task, the effect of stimulus asynchrony was only significant for the lexical task. This pattern of results suggests a bottleneck associated with lexical processing even when no response to the word is required.

Previous findings suggest that early stages of spoken word recognition can proceed unhindered while central resources are engaged upon another task. Cleland et al. (2006; cf. Reynolds & Besner, 2006; Ruthruff et al., 2008) found evidence that frequency-sensitive processing occurred prior to a central processing bottleneck. Similarly, Gaskell et al. (2008) found that lexical and coarticulatory cues as to phonemic identity are integrated prior to the bottleneck. Taken together, these studies show that many aspects of linguistic processing can proceed without placing any demands on central resources. However, the fact that the current studies found S2 task slowing at all must imply that some aspect of word processing nevertheless places demands on some form of limited resource; if this was not the case, then S2 processing would have immediate access to resources and no slowing would be observed. As participants did not have to make a response to the spoken word in the current studies, we can only speculate as to the cause of the resource limitation; however, given the evidence that early stages of word recognition are not attentionally demanding, we suggest that access to lexical content is a potential cause (an assumption consistent with the numerically smaller stimulus asynchrony effect observed for truncated words in Experiment 1).

An account that attributes the lexical bottleneck to lexical access and activation of word meaning would accord with data from Gaskell and Marslen-Wilson (2002), suggesting that the semantic activation of word meaning during the recognition of a spoken word occurs in a very selective time window. Their study used semantic priming to assess semantic activation during the perception of spoken words varying in uniqueness point and competition environment. When many lexical candidates matched a speech fragment (i.e., early on in the processing of a word) semantic priming was absent. Priming effects emerged only when competition was relatively weak (few lexical candidates remained) and tended to diminish again soon after the word became uniquely identifiable. For early uniqueness point words, there
was some evidence that semantic activation had peaked by the end of the word and diminished soon after (Gaskell & Marslen-Wilson, 2002, p. 255). The time-course of semantic activation observed by Gaskell and Marslen-Wilson corresponds to the degree to which bottleneck effects were found in the current studies, suggesting that retrieval of word meaning may underlie the lexical bottleneck effect.

Initially, such an account appears to be incompatible with the findings of Broadbent and others (e.g., Broadbent, 1954, 1958; Cherry, 1953; Treisman, 1964a, 1964b) who reported that participants can selectively filter streams of speech in dichotic listening shadowing tasks without accessing the meaning of the non-selected speech stream. However, we suggest that, once the to-be-shadowed speech-stream occupies the resources associated with the lexical bottleneck, this speech gains a monopoly on the limited resources available. By this view, the fact that participants have no awareness of the content of the second speech stream is not because they have filtered it out prior to deeper processing; rather, processing of this speech stream does not occur because there are no resources available to complete lexical access. Although there was an observable resource limitation in the current studies, this was only under the most extreme conditions of cognitive load (i.e., only when words had late diverging competitors and were followed in less than a tenth of a second by another stimulus that required lexical processing). In this respect, the current findings suggest that, in naturalistic settings, word recognition processes for speech proceed in an extremely efficient manner. More generally, the kind of adverse conditions created in the current studies are only likely to arise when two different sources of language input co-occur.

The current findings are relevant to the debate over the validity of uniqueness point effects and their role in lexical processing. Radeau, Morais, Mousty and Bertelson (2000) argued that uniqueness point effects arise due to strategic control on the part of the participant. They demonstrated that, when words were presented to participants at a speed similar to everyday speech, uniqueness point effects disappeared, suggesting that they are a phenomenon restricted to experimental settings. As there was no response required to S1 in the current studies, it is difficult to see what kind of strategy would lead participants to respond differently to S2 depending on the uniqueness point of S1. Our results are more compatible with Mattys and Clark’s (2002) conclusion that late uniqueness point words generate greater lexical activity than early uniqueness point words, and that lexical representations are activated automatically upon the reception of speech. In fact, Mattys and Clark speculated that the reason late uniqueness points slowed pause detection RTs was that “entertaining multiple lexical candidates may use up some of the processing resources otherwise allocated to pause detection, and hence, slow down participants’
responses” (p. 355). The current results suggest that Mattys and Clark were correct in this proposal; in the current studies we only found consistent evidence of a bottleneck effect when S1 had a late uniqueness point. The current studies provide a stronger test of the processing resources hypothesis than the pause detection paradigm because participants did not have to make any response at all to the spoken words. These findings fit well with models of spoken word recognition such as the Distributed Cohort Model (Gaskell & Marslen-Wilson, 1997), Shortlist (Norris, 1994), and TRACE (McClelland & Elman, 1986), in which the uniqueness point is strongly correlated with the speed at which lexical competition is resolved and word recognition is completed.

The current results are also relevant to the literature on the PRP and response selection. Accounts of the PRP generally assume that the bottleneck stage of processing is a general resource limitation associated with response selection (Pashler, 1994). Jolicoeur and Dell’Acqua (1998) found slowing of task 2 responses when no response was required to task 1; however, note that it is likely that this bottleneck was associated with stimulus encoding, as participants had to recall the characters presented as the first stimulus at the end of some trials. Others have found bottlenecks in the absence of a response to task 1 (e.g., Bertelson & Tisseyre, 1969; De Jong, 1993; Van Selst & Johnston, 1997), but in these cases a response was sometimes required to S1, and sometimes not (i.e., the task was a go/no-go task). It is entirely plausible that deciding not to respond in such a task constitutes a form of response selection. There is therefore limited evidence as to whether more central resource bottlenecks exist without a response component required for S1, but the current data show that response selection is not required in the case of a more specific bottleneck associated with lexical processing.

The “lexical bottleneck” we observed also differs from more general resource limitations in terms of the magnitude of the slowing effect for the subsequent task. In a PRP study, Ferreira and Pashler (2002) found that responses to a tone discrimination task were 74 ms longer at the 50 ms SOA than at the 150 ms SOA when they were preceded by a picture naming task (Experiment 1, p. 1192). This suggests that processing of the second task was almost completely postponed by processing of the first. The current results show a numerically more modest effect of stimulus asynchrony; in Experiment 1 S2 RTs were 12 ms longer at the 0 ms stimulus asynchrony than at the 100 ms stimulus asynchrony, in Experiment 2 there was a 15 ms difference, and in Experiment 3 a 13 ms difference. Given this pattern of results, it is likely that the bottleneck in processing associated with S1 presentation is very short-lived; when presentation of S2 is delayed by 100 ms, the bottleneck has already cleared...
by the time S2 processing requires central resources. This is all in line with the idea that lexical access in word recognition is extremely swift and efficient. The system has to be stressed to an extreme and unnatural level (with overlapping presentation of words and spoken stimuli that become unambiguous late on) to see any processing limitation at all.

In conclusion, the current findings suggest that there is a processing bottleneck that is specific to the language system and has a different set of properties to the more central resource limitations associated with PRP effects. Our results are most compatible with the notion that recognising a spoken word places demands on these resources. This capacity-limited process seems to be obligatory, in that it is invoked without any task-specific requirements. The bottleneck in processing is cleared quickly and so is only observed in quite extreme circumstances (perhaps within 100 ms of the recognition point of the word), but it nonetheless represents a limiting factor in the speed of communication. Speculatively, the bottleneck might relate to the time it takes to retrieve stored lexical knowledge relating to a single word before the subsequent word can be processed (cf. Pulvermüller, Shtyrov, & Hauk, 2009; Starr & Rayner, 2001).

References


Experiments 1–3 experimental items

**Early uniqueness spoken words**

Abode, annex, athlete, biscuit, blanket, bracelet, breakfast, closet, comfort, cottage, deluge, disease, dungeon, garment, husband, invoice, lantern, lizard, mackerel, minstrel, mortgage, mural, onion, parade, platform, pocket, pulpit, rebel, reflex, report, rifle, segment, sergeant, squirrel, stomach, trumpet, tulip, union, vessel, volume.

**Late uniqueness spoken words**

Captain, compass, design, discourse, forest, goblin, pennant, refuse, revolt, satin, wicket, advert, album, cabbage, carbon, cement, comic, concern, furnace, lattice, parish, solid, surface, tunic, village, bandage, device, mallet, meringue, passage, relic, salad, saloon, cavern, convent, insect, locust, camel, college, fillet.

**Visually presented words (Experiment 1)**

Ball, break, bring, care, catch, cause, church, claim, club, cost, course, deal, death, door, drive, drop, fact, fall, fear, fight, film, floor, food, form, game, give, hall, hand, head, hide, hill, hold, join, kill, land, lead, life, lose, mean, mile, need, note, play, point, prove, pull, push, rate, road, rock, room, rule, save, school, sell, sense, share, side, size, skin, smile, sort, source, staff, star, start, stone, stop, team, throw, time, touch, trade, truth, type, view, wall, week, wind, work.

**Visually presented words (Experiment 2 and 3)**

Ball, bit, break, care, catch, claim, course, cut, door, drive, drop, end, fight, food, game, hall, hand, head, hill, hit, land, life, play, pull, rate, road, rock, rule, sell, sense, share, star, stone, stop, team, throw, time, truth, wall, week.