The nature of phoneme representation in spoken word recognition

M. Gareth Gaskell, Philip T. Quinlan, Jakke Tamminen and Alexandra A. Cleland
University of York
York, United Kingdom

Four experiments used the Psychological Refractory Period logic to examine whether integration of multiple sources of phonemic information has a decisional locus. All experiments made use of a dual-task paradigm in which participants made forced-choice color categorization (Task 1), and phoneme categorization (Task 2) decisions at varying stimulus onset asynchronies. In Experiment 1, Task 2 difficulty was manipulated using words containing matching or mismatching coarticulatory cues to the final consonant. The results showed that difficulty and onset asynchrony combined in an underadditive way, suggesting that the phonemic mismatch was resolved prior to a central decisional bottleneck. Similar results were found in Experiment 2 using non-words. In Experiment 3, the manipulation of task difficulty involved lexical status, which once again revealed an underadditive pattern of response times. Finally, Experiment 4 compared this pre-bottleneck variable with a decisional variable: response key bias. The latter showed an additive pattern of responses. The experiments show that resolution of phonemic ambiguity can take advantage of cognitive "slack" time at short asynchronies, indicating that phonemic integration takes place at a relatively early stage of spoken-word recognition.

Keywords: phoneme, spoken word recognition, subcategorical mismatch, psychological refractory period, lexical access

To appear in *Journal of Experimental Psychology: General*.

M. Gareth Gaskell, Philip, T. Quinlan, Jakke Tamminen and Alexandra A. Cleland, Department of Psychology, University of York, UK.
Alexandra A. Cleland is now at the School of Psychology, University of Aberdeen, UK.
This study was supported by a Biotechnology and Biological Sciences Research Council research grant awarded to G.G. and P.Q. We thank Jim Magnuson and Ted Strauss for their advice on the use of the jTRACE software.
Correspondence concerning this article should be addressed to Gareth Gaskell, Department of Psychology, University of York, York YO10 5DD, UK. E-mail: g.gaskell@psych.york.ac.uk.

The nature of phoneme representation in spoken word recognition

The question of how listeners categorize the speech signal is of central importance to our understanding of language processing, both for normal (e.g., Marslen-Wilson & Warren, 1994; Morais & Kolinsky, 1994) and disordered populations (e.g., Serniclaes, Sprenger-Charolles, Carré & Demonet, 2001; Serniclaes, Ventura, Morais, & Kolinsky, 2005). A standard assumption in models of spoken word recognition (e.g., McClelland & Elman, 1986; Norris, 1994) is that auditory representations are mapped onto a phonemic level, and that word recognition operates as a comparison between this level of analysis and that of whole word forms, as specified in the mental lexicon.

Increasingly, however, the assumption of a mediating prelexical phonemic level of representation has been questioned. One alternative hypothesis is that phonemic awareness emerges largely as a consequence of learning to read an alphabetic language (Bertelson & De Gelder, 1991; Read, Zhang, Nie & Ding, 1986), and that phonemes do not have a role to play in the recognition of spoken words. In an influential study, Marslen-Wilson and Warren (1994) examined the perceptual consequences of cross-spliced syllables containing conflicting cues to the place of articulation of the final consonant (i.e., *subcategorical mismatch*; Streeter & Nigro, 1979; Whalen, 1984). For example, the initial consonant and vowel of *jog* might be spliced onto the final consonantal release burst of *job*. Cross-splicing in this case creates a token identifiable as *job*, but the relatively subtle pre-release cues are consistent with a /g/. If spoken word recognition requires identification of phonemes prior to lexical access then these kinds of mismatches should cause an identifiable delay in processing while the conflicting cues are resolved. Marslen-Wilson and Warren (1994) manipulated the lexical status of both pre-splice and post-splice component syllables across conditions in a lexical decision task. It was found that there were inhibitory effects of mismatching cues in all cases where pre- or post-splice stimuli were extracted from a word. However, mismatching cues had no delaying effect when both pre- and post-splice cues came from different non-word stimuli (e.g., *smod* and *smob*), as compared to a non-word control condition with no mismatching cues (e.g., two different tokens of *smob* cross-spliced). The absence of a mismatch effect in this condition is problematic for models of speech perception in which acoustic input is analyzed into categorical phonemic units prior to lexical access, because such models should always predict some delaying effect of subcategorical mismatch (i.e., the system would need to spend extra time working out the identity of the final phoneme before it could decide whether the stimulus was a word or not). Instead, Marslen-Wilson and Warren (1994) proposed that more fine-grained perceptual information is mapped directly onto the level at which word representations are stored and lexical status is evaluated. By this account, the degree to which the cues to the final phoneme are compatible with each other is irrelevant. The only factor that matters is the degree to which they match or mismatch word representations.

Marslen-Wilson and Warren (1994) argued, based on simulation analyses, that their data were incompatible with the TRACE model of spoken word recognition (McClelland & Elman, 1986). However, later data and more extensive simulations (Dahan, Magnuson, Tanenhaus, & Hogan, 2001) suggested that this is not so. TRACE is a connectionist interactive-activation model containing three levels of representation: a featural level, which provides input to the model, a phonemic level and a word level. Compatible units are connected between levels with positive weights, whereas connections between incompatible units within levels use negative weights. Word recognition is simulated in this model by activating the appropriate input units and allowing the network to settle into a stable activation state at the phoneme and word levels. Importantly, feedback is allowed, meaning that activation at the word level can propagate down to influence phoneme-level activations. Although TRACE contains a mediating prelexical phonemic representation, this level of representation operates in a cascaded way, meaning that categorical decisions do not need to be made at this level prior to the word recognition process. Thus, the Marslen-Wilson and Warren (1994) data do not rule out a prelexical phonemic level, but they are consistent with the idea that categorical *resolution* of any subcategorical conflict does not take place prior to the initiation of the word recognition process (cf. Andruski, Blumstein, & Burton, 1994). If a phoneme level does operate, then the system must be cascaded, in order for conflict at the phonemic level to percolate immediately through to the lexical competition level.

A second challenge to the claim that there is a pre-lexical phonemic level comes from evidence suggesting that a phoneme representation is inadequate for capturing effects of fine-grained information on word recognition. McQueen, Dahan, and Cutler (2003) discussed studies demonstrating that spoken word recognition is affected by aspects of speech that cannot be captured phonemically. For example, Davis, Marslen-Wilson, and Gaskell (2002; cf. Salverda, Dahan, & McQueen, 2003) examined perception of monosyllabic auditory words (e.g., *cap*), that matched with the onset of longer words (e.g., *captain*). Phonemically, these overlapping sequences are indistinguishable prior to the offset of the first syllable, but Davis et al. (2002) found that participants could distinguish between the two much earlier than this (i.e., during the vowel of the overlapping syllable). Acoustic analyses suggested that the key syllable in the case of short words had a longer duration and a higher mean fundamental frequency than the corresponding syllable in long words. Salverda and Tanenhaus (2006) recently demonstrated that syllable duration measured relative to overall speech rate has a potent effect on the word recognition process in cases like this. These studies suggest that some more detailed non-phonemic information must be preserved in the representation used for discriminating between words.

A third potential weakness of the prelexical phonemic hypothesis is revealed by a computational analysis of spoken word recognition. Scharenborg, Norris, ten Bosch, and McQueen (2005) reported data from simulations of a hybrid automatic speech recognition/psycholinguistic model. Two versions of this model were compared on ability to recognize words in real continuous speech. Both used phonemic representations of speech, but in one case the activation of this kind of representation was probabilistic (phoneme activations were graded), whereas in the other it was *categorical* (i.e., the recognition system was forced to select the best-fitting phoneme string before word recognition was initiated). The effect of enforcing a categorical decision at the phoneme level was to hamper recognition performance: the categorical model had an accuracy rate of 36%, whereas the comparable hit rate was 72% for the "soft" decision model. The detail employed by a probabilistic representation of phonemes appears greatly beneficial for the perceptual system.

Much of the above evidence does not deny the existence of a phoneme level (although data like the Davis et al., 2002, study are certainly problematic), yet it is becoming clear that categorical phonemic resolution does not occur as a necessary prerequisite to word recognition. So if phonemic resolution does not necessarily occur at this point, does the perceptual system produce categorical phonemic representations at any time during normal speech perception? A key goal of the current study was to address this question.

A parallel issue that has occupied researchers over several decades has been the involvement of lexical knowledge in phoneme perception. Numerous studies have demonstrated that listeners' judgments about the identity of phonemes are influenced by lexical information (e.g., Rubin, Turvey & van Gelder, 1976; Samuel, 1981; Warren, 1970). A classic example is the Ganong effect (Ganong, 1980), in which the categorization of an ambiguous phoneme depends to some extent on the lexical status of the embedding stimulus. For example, an ambiguous speech sound, [d/t], similar to both /d/ and /t/ might be categorized as a /d/ in the context of [d/t]ash (because *dash* is a word and *tash* is not), but categorized as /t/ in the context of [d/t]ask (because *task* is a word but *dask* is not). A simple explanation of this effect in the TRACE model would be that activated nodes at the lexical level feed activation back to the phonemic level. However, such effects can also be explained without any feedback links by assuming that a decisional system has access to information from both the lexical level and the phonemic pre-lexical level.

The latter bottom-up approach has been championed by Norris, McQueen and Cutler (2000) in their Merge model, which was set up as "a model of phonemic decision-making rather than of word recognition" (McQueen, Cutler & Norris, 2003, p. 260). As in TRACE, Merge has pre-lexical phoneme units, which feed into a separate word level. The normal operation of word recognition relies on just these two sets of units. However, a third decisional layer comes into play when listeners are asked to make judgments about speech sounds (such as is the case in a psycholinguistic experiment). Each decisional unit encodes a particular phoneme but at this level activations from the phoneme level and the lexical level are integrated (or merged). Decisions about which phonemes were present in the input are now based on activation at the decisional unit level in the network.

Norris et al. (2000) emphasized the fact that there is no feedback in this model, by which they meant that there is no backwards activation of the phoneme input units from either the lexical or decisional units. Nonetheless, findings such as the Ganong effect were explained quite straightforwardly in terms of a biasing influence of lexical information on activation at the decisional level. For example, in the case of [d/t]ask, the /d/ and /t/ decisional nodes both receive roughly equal levels of activation from the phoneme level /d/ and /t/ nodes, but the /t/ node would receive extra activation from the *task* word node, tipping the balance in favor of /t/.Attempts to discriminate between these two accounts of lexical effects have made use of increasingly intricate and ingenious methods (e.g., Magnuson, McMurray, Tanenhaus & Aslin, 2003; McQueen, 2003; Samuel & Pitt, 2003; see McClelland, Mirman & Holt, 2006, for a recent review), but it is fair to say that a consensus on this matter has not yet been reached.

For the current purposes, two aspects of the Merge model are particularly relevant. With respect to the issue of phonemic resolution, Norris et al. (2000) made use of a sublexical phoneme layer that preserved the relative activations of the various phoneme units even in cases of ambiguity. This aspect of the model allows it to avoid problems of prelexical categorical decisions discussed earlier, which may reduce the efficiency of recognition (cf. Scharenborg et al., 2005). Thus, subcategorical mismatch would initially produce graded pre-lexical phoneme activations just as in TRACE, but in TRACE these graded activations might gradually be eliminated through a settling process, leading to a categorical phonemic representation. In Merge these graded representations would remain frozen due to the lack of a settling network at this level. Instead, a settling network at the decisional level in Merge would then resolve the subphonemic ambiguity in the signal leading to a categorical representation only at this decisional level.

A second crucial aspect of the Merge account is that the settling network is explicitly a decisional one rather than a perceptual one. For instance, Norris et al. (2000; p. 316) stated that "connections from the lexical nodes to the phoneme decision nodes must be built on the fly, when the listener is required to make phonemic decisions." In the case of a subcategorical mismatch experiment, for example, a decision network with two response alternatives might be set up, either when the participant receives their instructions, or during the first part of the experiment (see Norris et al., 2000, p. 323). In fact, the decision network in Merge is viewed by its authors as an implementation of the flexible and configurable decision mechanism that could, in principle, be employed to "cope with all manner of bizarre tasks that experimenters, and the world in general, can throw at it." (Norris et al., 2000, p. 355). This very clear dependence on general purpose decisional mechanisms is a distinctive property of Merge, and provides it with a flexibility to deal with different response requirements in different situations. It also suggests that techniques developed in the study of processing bottlenecks and central resources may provide a fruitful and novel means of addressing the issues and problems summarized above. In the following section we review the Psychological Refractory Period (PRP) paradigm, and describe how it may apply to the issue of phonemic integration.

The PRP, PRP logic and PRP methods

In a PRP experiment participants perform two tasks in rapid succession (Pashler, 1994). Response times (RTs) to both tasks are measured as a function of the stimulus onset asynchrony (SOA) between the two tasks, defined as the time between the onset of the stimulus (known as S1) for Task 1 and the onset of the stimulus (known as S2) for Task 2. Typically, as SOA decreases RTs to Task 2 increase, whereas Task 1 RTs remain largely unchanged. The increase in Task 2 RTs is taken to reflect a processing limitation in being able to cope with two tasks that follow in quick succession.

There is an ongoing debate as to whether the limitation is best described as a "bottleneck" (Pashler, 1994) in which only one task is possible or as a general resource limitation (Navon & Miller, 2002; Tombu & Jolicoeur, 2003), in which limited parallel processing is possible. For the current purposes the predictions of these two accounts are equivalent (see Tombu & Jolicoeur, 2003), and so we will describe the theory in its more straightforward bottleneck terms. By this account, it is possible to decompose each task into three components – a pre-bottleneck stage of processing, a bottleneck stage of processing and a post-bottleneck stage of processing (see Figure 1). Pre- and post-bottleneck processes in both tasks (unshaded boxes in Figure 1) are assumed to operate without any resource limitation and so can occur in parallel. However, bottleneck processes (shaded boxes in Figure 1) must be carried out in serial fashion, which can lead to delays. Given that Task 1 starts prior to Task 2 the assumption is that Task 1 will engage the bottleneck prior to Task 2. As a consequence, Task 2 processes that make use of the bottleneck can only begin once Task 1 bottleneck processing is complete (i.e., the shaded boxes for Tasks 1 and 2 in Figure 1 can never overlap on the timeline). Such Task 2 delays are most likely to be evident at a short SOA (see panel A of Figure 1). In this case, pre-bottleneck processes associated with Task 1 and Task 2 proceed in parallel as there are no processing constraints at this stage. Task 2 may complete this stage shortly after Task 1, but because Task 1 has entered the bottleneck stage Task 2 is delayed until the bottleneck is cleared. This results in a period of "slack" time in the processing of Task 2. Such delays are the signature of the PRP, and result in a negative slope when RT is plotted against SOA. At very short SOAs this slope should be -1, because a reduction in SOA by some small amount of time simply increases the slack period by that same amount of time.

Further predictions can be made when task difficulty is manipulated, using the "locus-of-slack" methodology (see Pashler, 1994, for a thorough exposition). In particular, if Task 2 stimuli vary along some dimension affecting processing speed, then inferences about the locus of this variable can be made based on the pattern of responses to Task 2. Manipulation of such a variable should lead to slower RTs to the difficult condition at long SOAs regardless of the locus of the variable with respect to the bottleneck (see Figure 1, panel B). Such cases are assumed to be essentially the same as when Task 2 is carried out in isolation from Task 1, because there comes a point where the delay between the tasks is so long that neither task affects the other.

The pattern of performance is more revealing at short SOAs (see Figure 1, upper panel). If Task 2 difficulty affects processes at or after the bottleneck (i.e., BN or Post BN boxes in Figure 1), then the effect of the difficulty variable should be just the same as it is in the long SOA condition (i.e., the effects of SOA and Task 2 difficulty should be *additive*). However, if the manipulation affects early, pre-bottleneck processes (marked Pre BN in Figure 1), then the difficult condition can exploit the slack time created when the tasks are processed closely together, and the RT difference between easy and difficult conditions should be reduced or even completely absorbed at short SOAs. Such an *underadditive* combination of SOA and Task 2 difficulty is a hallmark of a pre-bottleneck task manipulation.

Identification of pre- or post-bottleneck tasks is only likely to be useful if the locus of the bottleneck itself is known. It is standardly argued that the central bottleneck occurs because of an inability to organize more than one response at once (e.g., Pashler, 1998). Part of this response organization involves deciding which response to output. In this respect it is understood that the bottleneck reflects a limitation at a decisional stage of processing—it is difficult if not impossible to make more that one response selection at the same time (e.g., Van Galen & ten Hoopen, 1976).

For example, Fagot and Pashler (1992) used Stroop interference as their Task 2 difficulty variable in a word naming task. Task 1 involved responding to a tone (high vs. low) using a button press, whereas Task 2 involved naming the color of ink of a visually presented word. Most critically, naming responses to congruent cases were quicker than those to incongruent cases, and the size of this Stroop effect was additive with SOA. In applying standard PRP logic, this means that the color of a word could not influence its processing prior to the bottleneck stage. Fagot and Pashler (1992) took the additive pattern to reflect the fact that the Stroop manipulation "affects a conceptual decision stage" of processing (p. 1075).

The precise nature of the processing bottleneck itself remains controversial (Hazeltine, Ruthruff & Remington, 2006; Meyer & Kieras, 1997; Navon & Miller, 2002; Tombu & Jolicoeur, 2003; Sigman & Dehaene, 2005), but the various different explanations are generally agreed in assigning a central or decisional locus to the bottleneck (Tombu & Jolicoeur, 2005). For example, Sigman and Dehaene (2005) characterize task processing in terms of (i) perception, (ii) decision based on noisy integration of evidence, and (iii) response, with only the decision process subject to a bottleneck. Thus, in the context of speech perception, variables that affect the early stages of word recognition should show a pattern of underadditivity (pre-bottleneck), whereas variables influencing response-based or decisional components should, like the Stroop study (Fagot & Pashler, 1992), show additivity (at or post-bottleneck).

Recent studies have demonstrated that the PRP bottleneck can provide a valuable means of separating early and late components of aspects of language processing (Cleland, Gaskell, Quinlan & Tamminen, 2006; Ferreira & Pashler, 2002; McCann, Remington, and Van Selst, 2000; Reynolds & Besner, 2006). McCann et al. (2000) used the PRP paradigm to identify the locus of word frequency effects in visual word recognition. Following up on this study, Cleland et al. (2006) used the PRP to examine the same effects in both spoken and visual word recognition. The overall conclusion from these studies is that frequency-sensitive processes are located throughout the word recognition system, both prior to, and either during or after the bottleneck. For instance, McCann et al. (2000) reported additive effects of word frequency with SOA and concluded that the effects of word frequency exerted its influence only relatively late, and subsequent to any pre-bottleneck stage of processing. In contrast though, Cleland et al. (2006) reported underadditive effects of word frequency with SOA, but with frequency effects remaining even at short SOAs. Considering the overall pattern of performance they concluded that the data were more in keeping with a system in which word frequency can exert its influence at several stages of processing. An early frequency component was associated with pre-bottleneck stage of stimulus encoding, but the frequency effect was not eliminated at short SOAs, suggesting some frequency effects at or subsequent to the central decisional bottleneck. These studies demonstrate that the PRP methodology provides a discriminating tool in the examination of human word recognition, and that linguistic variables can be found that operate on either side of the bottleneck boundary.

The main goal of the experiments reported below was to use the PRP paradigm as a new means of addressing the issue of phoneme processing in speech perception. Specifically we addressed whether integration of phonemic information to form categorical (i.e., non-graded) representations is subject to central/decisional bottleneck constraints. As we have mentioned, the Merge model was developed specifically to model phonemic decision-making, on the understanding that listeners build up a decision network during the course of an experiment in response to task demands. This flexible decision process is not part of normal speech recognition (Norris et al., 2000, p. 323), and therefore it should consume processing resources just as any other decision requiring the ad hoc integration of perceptual information should. Other models however (e.g., McClelland & Elman, 1986; Gaskell & Marslen-Wilson, 1997) do not make such explicit usage of decisional units, but of course they all rely on decisional mechanisms at some level. We will return to this point in the General Discussion.

To address the involvement of decisional resources we ran four experiments in the PRP paradigm using a non-linguistic Task 1 (a simple color discrimination task) and a linguistic Task 2. Task 2 was a speeded phoneme categorization task, and in all experiments there was some manipulation of the difficulty of this task (via either subcategorical mismatch or variation in lexical status), allowing us to see whether the difficulty variable had a pre-bottleneck locus or not. If this variable affects processing early on in word recognition then difficulty should interact in an underadditive fashion with the SOA between Task 1 and Task 2, implying a pre-bottleneck manipulation. If however, such a manipulation influences a later decisional stage of processing (i.e., at or after the bottleneck stage) then an additive pattern of responding is predicted. Experiments 1 and 2 used the presence or absence of subcategorical mismatch as the key Task 2 variable (cf. Marslen-Wilson & Warren, 1994), whereas Experiment 3 examined the lexical advantage in categorizing final phonemes. As a comparison, Experiment 4 included a further task difficulty variable that we expected would affect decisional processing: response key bias.

Experiment 1

Method

Participants. Thirty-six native English speakers (29 female and 7 male) from the University of York were paid or received course credit for participation. The age of participants ranged from 18-26 years (mean age = 20), and participants reported no speech, hearing, or visual disorders.

Materials. In Task 1, participants were presented with a sequential pair of geometric figures, one of which was colored. Participants made a speeded two-alternative forced choice (2AFC) response to the color of this item by pressing either a designated Blue or a designated Green response key. The figures comprised a circle and a square, and for each shape blue, green, and unfilled versions were created. In Task 2, participants made a speeded 2AFC response to the final phoneme of a spoken word. Henceforth S1 refers to the imperative stimulus in Task 1, namely the colored shape; and S2 refers to the Task 2 stimulus (i.e., the spoken word). The use of shape pairs rather than single shapes extended the duration of the visual sequence, and ensured that the onset of the visual sequence always preceded the onset of the auditory stimulus, in congruence with the expected ordering of responses (cf. Cleland et al., 2006). For the test materials, the colored shape was always the second member of the visual sequence, whereas filler trials were used to ensure that an equal number of trials contained a colored shape as the first member of the sequence.

For Task 2 seventy-eight monosyllabic word pairs ending in stop consonants were chosen as the experimental stimuli (see Appendix A). The members of each pair (e.g., *job – jog*) were matched in terms of the initial consonant(s) and vowel, but differed on the place of articulation of the final consonant (henceforth, the target phoneme). In generating the critical pairs of words contrasting final consonants were /k/-/t/, /p/-/t/, /p/-/k/, /d/-/g/, /b/-/g/, and /b/-/d/. All words had a CELEX frequency of at least four per million (Baayen, Piepenbrock, & Van Rijn, 1993).

Multiple recordings of each pair were made by a native English speaker (MGG). Following Marslen-Wilson and Warren (1994), the audio files were cross-spliced by taking, for each pair, the word with the higher frequency (e.g. *job*), and excising the final consonant release burst (/b/). This was replaced with the equivalent burst from the second member of the pair (the /g/ from *jog*) and so gave rise to a mismatching item. The /g/ burst was also spliced onto the final section of a second recording of *jog* and this gave rise to a matching item. For the matching item the coarticulatory information in the initial portion matched with the target phoneme and for the mismatching item the coarticulatory information in the initial portion contrasted with target phoneme. By using this method of cross-splicing with all the items there was no confound between cross-splicing and subcategorical mismatch. Spliced filler items were also constructed from 111 word pairs. The final consonants of these items included stops, nasals, and fricatives. Ten more stimuli were used as practice items.

Design. The experiment contained six conditions, defined by combinations of three levels of SOA (100, 200, or 1000 ms) and two levels of coarticulatory cue correspondence (henceforth cue correspondence; matching vs. mismatching). The use of two short SOAs is common in PRP experiments (e.g., McCann et al., 2000), and provides an opportunity to test whether the slope of the SOA by RT function is roughly -1, as predicted by PRP theory. In typical PRP experiments Task 1 and Task 2 stimuli are temporally-aligned with respect to the onsets of S1 and S2. However, a special concern arises here because of the use of spoken words. With such stimuli information unfolds over time and the key phoneme occurs several hundred milliseconds after the onset of the word. Cleland et al. (2006) found that if SOAs were defined relative to word onset then the PRP slope was relatively shallow, suggesting that Task 2 processing often reached the bottleneck after the termination of Task 1 bottleneck processing even for short SOAs. Our solution was to align the onset of S1 relative to the onset of the final phoneme burst of S2 (which in the current case is the splice point). This method revealed the typical PRP -1 slope, and so was used in the current experiments. The RTs to Task 2 were also measured from this splice point.

Experimental items were counterbalanced across six lists so that each word appeared at each SOA condition, in both mismatching and matching conditions across all participants, but each participant heard only one version of each item.

Procedure. Participants were seated in a quiet booth, a comfortable distance from the computer screen. They were instructed that they had two tasks to perform. The first was to respond to a colored shape that would appear within a sequence of two shapes presented rapidly on the screen, and the second was to identify the final sound in each of the words presented over the headphones. For Task 1, they had to press one button if the shape was blue and a different button if it was green. For Task 2, they were told that they would see two letters on the screen, and that they should press the button corresponding to the letter that they heard at the end of the word (the upper button if the letter appeared above, and the lower button if the letter appeared below). The button box was laid out so that participants could rest the index finger and the thumb of their left and right hands on the corresponding buttons, making responses to the color stimulus with their left hand and responses to the spoken stimulus with their right hand. The positioning of the response buttons was kept constant across participants regardless of which was their dominant hand. Participants were told that the letters corresponding to the sounds would appear first, but also that the string of shapes would be presented before the word. They were told that they should respond to the color of the shape before making their phoneme response, and that although they should try and respond quickly and accurately to both the colors and the words, they should give special emphasis to responding quickly to the color.

Each trial began with the presentation of the two response options for Task 2 (the phonemic decision task) as a pair of lowercase letters, one above and one below the centre of the screen. These remained on screen until the trial was complete in order that participants did not have to remember what the phoneme alternatives were. After a pause of 2000 ms a central fixation cross was presented for 1000 ms which, in turn, was replaced by the first shape of the shape pair sequence. Each shape was presented for 375 ms and only one of the shapes (i.e., the target shape, S1) was colored (either blue or green). In half of the trials S1 occurred as the first member of the shape pair and on the remaining trials it occurred as the second member of the pair. S1 only occurred as the first of the pair on filler trials, and responses on these trials were not examined further.

The auditory stimulus was presented through headphones, with stimulus alignment manipulated with respect to the onset of the colored shape and the splice point in the spoken word. In this context the SOA between S1 and S2 was defined as the time between the onset of the colored shape in S1 and the onset of the final phoneme burst (i.e. splice point) in the spoken S2.

Task 1 responses to the colored shape were made using upper and lower keys of a Cedrus button box with the left hand, and Task 2 responses to the target phoneme were made using separate upper and lower keys with the right hand. The response mapping for a given phoneme pair was kept constant across trials, so if the choice was between /b/ and /p/, /b/ was always assigned to the upper key and /p/ to the lower key. The experiment began with two practice blocks, the first containing single-task trials with Task 1 only (30 trials), and the second containing dual-task trials (50 trials).

Task 1 RTs were measured from the onset of the colored shape, and Task 2 RTs from the splice point of the spoken word. We presented the experimental stimuli and recorded the participants' RTs using E-Prime software on a PC running Windows XP. The stimuli were presented on a Sharp 17-in TFT monitor. Timing tests of the experimental set-up, conducted via a Black Box Toolkit (Plant, Hammond, & Whitehouse, 2002), verified the consistency of the timings requested by the experimental script. The experimental session took around 30 minutes to complete and included one rest break.

Results

Trials where an error was made in either task were excluded (8% of all trials). Outlying responses above 3000 ms or below 100 ms were excluded (none in response to T1, 0.07% of trials in response to T2), and the remaining data are summarized in Table 1. Repeated-measures Analyses of Variance (ANOVAs) involving SOA (100, 200, 1000 ms) and cue correspondence (matching vs. mismatching) were conducted. Separate analyses treated participants (F_1) and items (F_2) as random effects.

Task 1 RTs. Neither the main effect of cue correspondence, nor the interaction between cue correspondence and SOA reached statistical significance (all ps > .05). Only the main effect of SOA was statistically reliable, $F_1(2, 70) = 5.21$, $p < .05$, $F_2(2, 308) = 6.50$, $p < .05$. Pairwise comparisons revealed that this was due to RTs being longer (47 ms) at the 200 ms than the 1000 ms asynchrony (see Figure 2).

Task 2 RTs. Analyses revealed significant main effects of cue correspondence, $F_1(1, 35) = 10.00$, $p < .01$, $F_2(1, 154) = 6.35$, $p < .05$, with responses to words with matching co-articulatory cues being on average 37 ms faster than to words with mismatching cues, and SOA $F_1(2, 70) = 224.63$, $p < .001$, $F_2(2, 308) = 444.67$, $p < .001$, with an 85 ms decrease in RTs between the 100 and 200 ms SOAs, and a 418 ms decrease between the 100 and 1000 ms SOAs.

In addition, the cue correspondence by SOA interaction was statistically significant, $F_1(2, 70) = 4.73$, $p < .05$, $F_2(2, 308) = 4.07$, $p < .05$. Planned comparisons showed that the effect of cue correspondence was statistically significant only at the longest asynchrony, $F_1(1, 35) = 60.48$, $p < .001$, $F_2(1, 135) = 24.35$, $p < .001$. Planned interaction contrasts confirmed this, showing that the effect was larger at the 1000 ms delay (85 ms) than both the 200 ms (11 ms), $F_1(1, 35) = 8.88$, $p < .01$, $F_2(1, 154) = 8.40$, $p < .01$, and the 100 ms delay (18 ms), $F_1(1, 35) = 7.88$, $p <.01$, $F_2(1, 154) = 4.81$, $p <.05$. There was no effect of cue correspondence at the two shortest delays (both ps > .05).

Errors. Participants made an error in either Task 1 or Task 2 in a total of 221 trials (8% of all trials). Of these, 63% were errors in response to the phoneme decision task (e.g. a /t/ response when /k/ was correct), and 37% were errors in responding to the color. Table 1 shows the proportions of errors produced in all conditions. There were no statistically reliable effects in the accuracy analyses.

Discussion

The critical result here is the underadditive interaction between the manipulation of cue correspondence and SOA (see Figure 2). At the long SOA, Task 2 response times showed the standard pattern of delayed responses for subcategorical mismatch (e.g., Marslen-Wilson & Warren, 1994, McQueen, Norris & Cutler, 1999; Whalen, 1984). In our case, this basic mismatch cost was particularly strong (85 ms). Nonetheless, at the shorter SOAs Task 2 RTs were unaffected by whether or not there was agreement between the initial co-articulation cues and the target phoneme included in the spoken word. Applying standard PRP logic (Pashler, 1994), this underadditive pattern indicates that resolution of subcategorical mismatch occurs during the slack period at short SOAs and is therefore a pre-bottleneck process. This suggests that the integration of such cues in this case does not depend on a central or decisional mechanism.

Before a more thorough theoretical exposition of the underadditive pattern of performance in Experiment 1 is attempted, it is important to consider the degree to which the findings reflect the influence of the lexicon. In Experiment 1 all of the items used in Task 2 were spoken words. Even in the mismatching cases each spoken utterance corresponded to a real word. Given this, the phoneme decisions could have been based not on listeners' categorization of the speech stream, but upon the retrieval of lexical representations of phonological form (cf. Cutler & Norris, 1979). One simple way to assess this theoretically important possibility is to examine performance with non-words. For these stimuli a simple lexical locus of the effects is ruled out because non-words do not have associated lexical representations. Therefore in Experiment 2 we replaced the spoken words with spoken non-words. If resolution of subcategorical mismatch does not require access to particular lexical representations then the switch from words to non-words should not affect the underadditive pattern of responding in Task 2. However, if such an underadditive pattern is a signature of lexical access then it ought not recur in cases where non-words are used.

Experiment 2

Method

Participants. Thirty-six native English speakers (26 female and 10 male) from the University of York were paid or received course credit for participation. The age of the participants ranged from 18 to 41 years (mean age = 21), and participants reported no speech, hearing or visual disorders. None had participated in Experiment 1.

Materials and Design. Following on from Experiment 1 our intention was to generate a Task 2 stimulus set comprising spoken non-words. Non-words were based on the words from Experiment 1 but had markedly different initial consonants (e.g. *bug-nug*; see Appendix A). In a small proportion of items the vowel was changed instead of the final phoneme, as a change of consonant would have resulted in a real word (e.g. the non-word equivalent of *hit-hip* was *het-hep*). Filler and practice non-words were derived in the same way. Non-words were recorded in the same session as the real words, and were edited and cross-spliced as described in Experiment 1.

In all other respects the experiment was the same as Experiment 1.

Results

Trials with an error in either task were excluded (10%). Responses with an RT above 3000 ms or below 100 ms were again treated as outliers, although in this case none were present. Table 2 summarizes the data in both tasks.

Task 1 RTs. There were no statistically significant effects of cue correspondence or SOA (all *p*s > .05).

Task 2 RTs. Whereas the main effect of cue correspondence (10 ms) was not statistically reliable, $F_1(1, 35) = 0.65$, $p > .05$, $F_2(1, 153) = 0.77$, $p > .05$, both the main effect of SOA, $F_1(2, 70) = 216.91$, $p < .001$, $F_2(2, 306) = 486.92$, $p < .001$, and the cue correspondence by SOA interaction, $F_1(2, 70) = 4.76$, $p < .05$, $F_2(2, 306) = 4.39$, $p < .05$, were statistically significant (see Figure 3). RTs decreased by 90 ms between the 100 and 200 SOAs, and by 398 ms between the 100 and 1000 SOAs. Planned comparisons showed that the effect of cue correspondence was statistically significant at the longest asynchrony, $F_1(1, 35) = 16.74$, $p < .001$, $F_2(1, 153) = 6.79$, $p < .05$, but not at the two short asynchronies (all *p*s > .05). Planned interaction contrasts showed the effect to be significantly larger at the 1000 ms asynchrony (46 ms) than the 100 ms asynchrony (-31 ms), $F_1(1, 35) = 11.68$, $p < .01$, $F_2(1, 153) = 8.55$, $p < .01$. No other contrasts were statistically reliable, although the effect at 200 ms SOA (16 ms) was marginally larger than at the 100 ms SOA, $F_1(1, 35) = 3.07$, $p = .09$, $F_2(1, 153) = 4.03$, $p < .05$.

Errors. Participants made an error in either Task 1 or Task 2 in a total of 280 trials (10% of all trials). Of these, 66% were errors in response to the phoneme decision task (e.g. a /t/ response when /k/ was correct), and 34% were errors in responding to the color. Table 2 shows the proportions of errors produced in all conditions. There were no statistically reliable effects in the accuracy analyses.

Discussion

The key finding in Experiment 2 was an underadditive interaction between cue correspondence and SOA. As in Experiment 1, there was a clear effect of subcategorical mismatch at the long asynchrony, but there was no such effect at the short asynchronies. Following standard PRP logic, the implication is that the resolution of phonemic conflict takes place early on, that is, at a pre-bottleneck stage of processing.

More particularly, Experiment 2 demonstrated that phonemic resolution does not depend on access to stored representations of words because the effects of cue correspondence arose in the context of the perception of spoken non-words. Against this it could perhaps be argued that the non-words were sufficiently similar to the existing words for normal lexical access to take place. Hence, phonemic decisions may have been made on the basis of sufficiently activated lexical entries. Indeed, evidence from eye-tracking studies (e.g., Allopenna, Magnuson, & Tanenhaus, 1998) suggests that maximal mismatches (in longer words) can activate the lexical representations albeit rather weakly. In contrast though, for short monosyllabic items even minor mismatches between speech input and lexical representations are enough to block lexical access (Marslen-Wilson, Moss, & van Halen, 1996). On the whole our stimuli were constructed so that there was maximal deviation between the non-words and probable target words, with less than 1 in 5 deviations corresponding to a minimal pair (e.g., *bob-pob*). Hence it seems highly unlikely that the underadditive pattern with the non-word stimuli in the current experiment could be attributable to lexical access.

The results of Experiments 1 and 2 are of particular relevance to the Merge model (Norris et al., 2000), because of the emphasis it places on decisional mechanisms for the integration of phonemic information in an experiment like ours. The pre-lexical phoneme and lexical competition levels of Merge do not involve decisional resources and are part of the normal perception of speech. As such they are pre-bottleneck processes, and, according to standard PRP logic, may operate in parallel with any bottleneck processing that Task 1 demands (cf. Cleland et al., 2006). However, the flexible decision level of Merge is intended to model decision processes and so should operate at or after the dual-task bottleneck. As a consequence, and in order to perform a response, the integration of the pre-bottleneck activation can only operate once bottleneck resources are freed up (see Figure 1). Consider the state of the pre-lexical phoneme layer in Merge in a long SOA trial of the current experiment. As mentioned, this type of trial is most similar to the single task case studied previously (e.g., Marslen-Wilson & Warren, 1994). If the stimulus contains mismatching subcategorical cues to the final consonant (e.g., information compatible with /b/ in the initial CV, followed by the final release burst of a /g/), then this conflict should be reflected in the state of the pre-lexical units (e.g., an activation level of .15 for the /b/ node and an activation of .85 for the /g/ node; see Norris et al., p. 314). Given that there is no settling or clean-up at this level, these units are essentially "frozen" with the mismatch apparent. Looking now at the decisional units in Merge, we can assume that their only coherent input (in Experiment 2) comes from the pre-lexical units, since the stimulus does not match a lexical representation. Because there is no decisional bottleneck at the long SOA, the decision units receive the graded activation from the pre-lexical units immediately, and the settling network at this level allows the conflict to be resolved, leading to a /g/ response in this case. The extra settling time at the decisional level compared to a no-mismatch stimulus allows Merge to predict the observed mismatch cost.

Now consider the same situation but with a short SOA, leading to a delay in freeing up processing resources required to carry out the decision. There is no mechanism in Merge to "clean-up" activations at the prelexical level, and so this level remains frozen, with the subcategorical conflict still evident.[1] When the bottleneck does clear, the situation at the pre-lexical level is the same, leading to the prediction that mismatch effects should still be observed at the short SOAs. It is feasible that Merge might predict a somewhat smaller effect of subcategorical mismatch at this later timepoint because the temporal ordering of mismatching (vowel transition) prior to matching (release) information will not be so relevant, but as long as the mismatch is represented at the input level there should be some cost at short SOAs. Strikingly, though, this prediction is not borne out in the data (across the two experiments, mismatching items at the 100 ms SOA were responded to 7 ms *faster* than matching items; $F < 1$). Instead, it seems that the perceptual system can make use of the slack time at short SOAs to clean up the conflicting representation, transforming graded activations (e.g., .85 and .15) into more categorical ones (e.g., 1.0 and 0.0). One means of accomplishing this clean-up might be a settling network at the pre-lexical level.

The mechanistic account of Merge sketched out above is still a little simplistic, because it ignores the fact that the perceptual information at the pre-lexical level accumulates over time. This is an important point and we will return to it after the data for the next experiment have been presented. In Experiment 3 we wanted to extend the scope of the research to look at the integration of different kinds of

---

[1]. The situation here is similar to McQueen, Cutler & Norris (2003), in which phonetic categorization responses were examined for fast, medium and slow participants separately. The authors (p. 267) state that "the activation of the pre-lexical nodes does not die away immediately after stimulus offset" because they form a kind of memory buffer, meaning that a "slow decision in Merge can thus be based on an accurate description of the signal."

information in the judgment of speech forms. The decisional mechanism in Merge (Norris et al., 2000) was explicitly set up to handle the integration of lexical and perceptual information in the recognition of phonemes. It seems plausible then that we might be more successful in observing decisional effects if we looked at a case where task difficulty relates to integration of perceptual and lexical knowledge. As a direct test of these ideas, Experiment 3 used the same phoneme detection task used in Experiments 1 and 2 but the task difficulty variable reflected not the nature of the cues to coarticulation but the lexical status of the spoken stimulus (word vs. non-word). The rationale was that if the integration of lexical and perceptual knowledge is located at a decisional (late) stage of processing then it should give rise to an additive effect with SOA. However if the effect of lexical status is located at an early stage of processing then it ought to produce an underadditive effect with SOA.

Experiment 3

An examination of the interaction of lexical status effects with SOA in the PRP paradigm relies on the existence of a strong lexical status effect in single-task conditions (comparable to the long SOA dual-task condition). Studies of the Ganong effect (Ganong, 1980) have often examined response times as well as categorization functions, and so response times to endpoint (i.e., phonetically unambiguous) stimuli in these experiments may provide a clue as to the existence of a reliable word advantage in a phoneme categorization task. In fact the evidence from these experiments is mixed (Connine and Clifton, 1987; McQueen, 1991), but effects seem to be strongest for categorizations involving the final phoneme of the stimulus (Pitt & Samuel, 1993), as in our case. However, Ganong experiments differ from the current situation in that only one of the two possible responses in the Ganong experiment is compatible with a word. In the current experiments, the two responses have been compatible with either words only (Experiment 1) or non-words only (Experiment 2). Subcategorical mismatch experiments (Marslen-Wilson & Warren, 1994; McQueen et al., 1999) are a little more comparable, and the no-mismatch conditions in these conditions appeared to show a small lexical RT advantage, but even in these cases one of the two alternatives in the phoneme decision tasks to non-words was compatible with a word. In all these cases, the direct response competition between a word and a non-word may have contributed to a lexical effect. Nonetheless, there was a modest between-participants difference (22 ms; ns in a combined analysis) in response times to the no-mismatch conditions of Experiments 1 and 2, which suggested that lexical status effects should emerge when the two types of stimulus were compared within participants in Experiment 3.

Method

Participants. Thirty-six undergraduate students (26 female and 10 male) from the University of York either were paid or received course credit for participation. The age of the participants ranged from 18 to 30 years (mean age = 20), and participants reported no speech, hearing, or visual disorders.

Materials. The Task 1 stimuli were the same as for Experiments 1 and 2. For Task 2, though, 108 word-non-word pairs were selected from those used in Experiments 1 and 2, e.g. *back-dack* (see Appendix A). Items were selected in order to match the words and non-words for duration, and to balance the proportions of stimuli with each pair of target phonemes. As before, the target phoneme was defined as the final phoneme (e.g. /k/) that occurred following either a word onset (e.g. *ba*) or a non-word onset (e.g. *da*). Stimuli were cross-spliced so that both the word and non-word from each pair contained the same physical stimulus as the final burst. This allowed the lexical status manipulation to be a within items factor. For half of the items, the final phoneme was taken from the word and for half of the items it was taken from the non-word. As in the previous experiments, the onset was always spliced from a separate recording regardless of whether the final phoneme came from the same word or non-word. The Task 2 response alternative for each stimulus was the same one used in the previous experiments, meaning that both alternatives completed either two words or two non-words. An additional 144 filler items were generated using the same cross-splicing method described previously.

Design. There were six conditions in the experiment, defined by the combination of two levels of lexical status (word vs. non-word) and three levels of SOA (100 ms, 200 ms and 1000 ms). The experimental items were assigned to six condition groups in a counterbalanced way, as in the earlier experiments.

As before, experimental items were assigned to experimental trials such that the S1 always occurred as the second shape in the sequence. On the filler trials, 132 were where S1 was the first shape in the sequence and in remaining 12, the S1 was the second shape in the sequence. Half of the fillers (72) appeared as words and half as non-words, and fillers were rotated through the SOA conditions.

In summary; every participant completed 108 experimental trials, 18 in each of the six conditions defined by the two levels of lexical status (word vs. non-word) and the three levels of the SOA factor (100 vs. 200 vs. 1000 ms). These experimental trials were randomly intermixed with an additional 144 filler trials and each participants received a different random order of the complete trial set.

Procedure. In the main, the procedure was identical to that described previously. For each participant, the experimental session was preceded by two practice sessions. The first lasted 30 trials, and required responses to S1 only. The second lasted 50 trials, and involved responses to both S1 and S2. The second practice session did not include any of the experimental items. The experimental session took around 30 minutes to complete and included one rest break.

Results

Responses faster than 100 ms were excluded from the analysis, as were responses slower than the cut-off time of 3000 ms (0.3% of trials). Trials in which an error was made in either Task 1 or Task 2 were also excluded from the RT analysis (6% of trials). Table 3 summarizes the remaining data.

Task 1 RTs. Separate analyses treated participants (F_1) and items (F_2) as random effects. Analyses of variance (ANOVAs) with lexical status (word vs. non-word) and SOA (100 vs. 200 vs. 1000 ms) revealed a statistically significant main effect of SOA by items only, $F_1(2,70)=1.35$, $p=.27$; $F_2(2,106)=3.58$, $p<.05$. This was due to a decrease in response times as the SOA increased. No other effects were statistically reliable.

Task 2 RTs. ANOVAs with lexical status (word vs. non-word) and SOA (100 vs. 200 vs. 1000 ms) revealed a statistically significant main effect of lexical status, $F_1(1,35)=24.08$, $p<.001$; $F_2(1,107)=20.40$, $p<.001$, with responses being 45 ms faster in response to words than in response to non-words. There was also a statistically reliable main effect of SOA, $F_1(2,34)=238.34$, $p<.001$; $F_2(2,106)=652.37$, $p<.001$, with RTs decreasing as the SOA increased. There was a 435 ms decrease in RT between the 100 and 1000 ms stimulus asynchronies, and a 111 ms decrease in RT between the 100 and 200 ms stimulus asynchronies.

Most crucially, there was also a statistically significant lexical status by SOA interaction, $F_1(2,34)=5.41$, $p<.01$; $F_2(2,106)=4.03$, $p<.05$ (see Figure 4). Planned interaction contrasts revealed that the lexical status effect at the 1000 ms SOA (84 ms) was significantly greater than at the 100 ms SOA (13 ms), $F_1(1,35)=10.89$, $p<.01$; $F_2(1,107)=7.70$, $p<.01$. Indeed, simple planned comparisons revealed that the 42 ms effect of lexical status at the 200 ms SOA was significant, $F_1(1,35)=5.35$, $p<.05$; $F_2(1,107)=4.24$, $p<.05$, as was the 84 ms lexical status effect the 1000 ms SOA, $F_1(1,35)=54.58$, $p<.001$; $F_2(1,107)=24.86$, $p<.001$, but that the 13 ms effect at the 100 ms SOA was not (both Fs<1).

Errors. Participants made an error in either Task 1 or Task 2 in a total of 239 cases (6% of all trials). Of these, 61% were errors in response to the phoneme decision task (e.g. a /t/ response when /k/ was correct), 27% were errors in responding to the color, and 16% were trials in which the participant made an error in response to both tasks.

The proportions of errors produced in all conditions are shown in Table 3. There were no statistically reliable effects in the accuracy analyses.

Discussion

The central finding of Experiment 3 was that the effect of lexical status revealed an underadditive interaction with SOA for Task 2. The strong advantage provided by the addition of lexical form information in the word condition at the long SOA (85 ms) was eliminated when decisional processes were postponed by the bottleneck associated with Task 1 processing at the shortest SOA. The underadditive interaction revealed in Experiment 3 differs slightly from the pattern of the previous experiments. In Experiments 1 and 2 there was no significant effect of task difficulty at the 200 ms SOA, whereas in Experiment 3, this effect was significant, although still only half the size of the long SOA effect. This result does not affect the validity of the argument that task difficulty has a pre-bottleneck locus, particularly as there was no effect of task difficulty at the shortest SOA. Indeed, the gradually diminishing effect of task difficulty as a function of the shortening of SOA, as seen in Experiment 3, is often thought of as a "textbook" indication of a pre-bottleneck variable (e.g., Carrier & Pashler, 1995, Figure 1). The extent to which pre-bottleneck task difficulty effects can be absorbed into slack time depends (in part) on the duration of the bottleneck process for Task 1 (see Figure 1). If this process lasts, say, 180 ms, then 180 ms is the maximum amount of time available for absorption of difficulty effects in slack time. Depending on the duration of this period, it could be that a complete absorption of difficulty at the shortest SOA becomes only a partial absorption at a slightly longer SOA. In reality, this process will vary between people and between trials; hence the weakened task difficulty effect at the intermediate SOA could reflect such variation. Indeed, the fact that this effect has emerged here, but not in Experiments 1 and 2 could conceivably be a consequence of the differential temporal extents of the two effects, with phonemic conflict resolved relatively quickly compared to more extended lexical effects, which may rely on build-up of lexical activation (e.g., McClelland & Elman, 1986). In any case, the crucial result is that the robust effect of task difficulty at the long SOA (84 ms) is nonsignificant at the shortest SOA.

This result is consistent with a pre-central or pre-decisional locus of the effect, and again is not a pattern of results that would be predicted by Merge (Norris et al., 2000). When considering Merge's behavior at the bottleneck or post-bottleneck decisional level in the current circumstances, activations at both pre-lexical and lexical levels need to be taken into account. At the long SOA, when decisional process are not disrupted by Task 1 processing, the word advantage in categorizing the final phoneme is easily accommodated because the model allows activations at the pre-lexical and lexical levels to be integrated at the decision level. So for the non-word *dack*, the /k/ node at the decision level is only activated by the /k/ node at the pre-lexical level, whereas for *back*, the extra activation from the respective word node allows the /k/ decision node to reach threshold more quickly. Once again, the problem for Merge comes when the short SOAs are considered, because Merge would predict that the lexical advantage remains at this point, because the lexical and phonemic activations can only be integrated once Task 1 has cleared the bottleneck, and at this point the word stimuli still have an advantage of two sources of information over one. Instead though the lexical advantage was reduced at the 200 ms SOA and eliminated at the shortest SOA.

At this point we should return to the caveat mentioned in the discussion of Experiment 3: Our characterization of Merge has assumed that perceptual information becomes available at the pre-lexical and lexical levels all at once. This is a reasonable assumption when considering responses at the short SOAs, because the bottleneck in processing will have allowed time for activations to settle at these levels. However, at the long SOA with no such bottleneck we need to consider the fact that speech information accumulates and lexical activations settle over a reasonably extended period of time. The information that provides the manipulation of task difficulty (pre-burst coarticulatory cues in Experiments 1 and 2, lexical information in Experiment 3) will in fact build up over several hundred milliseconds. This information will not be sufficient to trigger a response, but the critical information in the release of the final consonant is also extended over a non-negligible period of time. It is difficult to know exactly how this factor might affect processing in the Merge model. On the one hand, lexical activations are likely to gain strength as time goes on, which means that Merge might predict a larger lexical advantage for short SOAs, where the lexical network has had longer to settle. This possibility would result in an overadditive pattern and clearly does not fit our results.

On the other hand, if the task difficulty variable has substantially influenced activations prior to the onset of the critical release burst (e.g., through the differential activation of a lexical node), then at long SOAs (where no bottleneck is evident) the limiting factor in making a response may be the speed at which burst information accumulates. In the case where a lexical node provides a secondary source of information consistent with the release, then less perceptual information is needed to reach a response threshold. Thus, the task difficulty effect at a long SOA may be more to do with the accumulation of information at the perceptual level than the integration of this information at the decisional level. For the short SOA conditions (where a bottleneck delays decisional processing), activations will have built up and are merely sitting in wait for decisional mechanisms to be freed up, and so this argument does not apply. Instead, any task difficulty effect must relate to decisional integration, which might mean that the short SOA effect is reduced as compared to the long SOA.

It is important to note that none of these scenarios would predict the observed pattern in the first three experiments: a complete loss of the task difficulty effect at the shortest SOA; but the latter scenario does suggest that an underadditive pattern is possible even with integration of phonemic and lexical information at a decisional level. We decided to address this uncertainty empirically in the final experiment by introducing a task difficulty variable that we expected to have a bottleneck or post-bottleneck locus. To this end a variant of the task used in Experiment 3 was developed in which one of the key press responses was biased: on every trial one response was designated as being over 3.5 times more likely to be correct (cf. Luck, 1998). Over the course of the experiment, this bias should lead to faster responses for the favored response key and slower responses for the disfavored key. Importantly, when one considers an individual trial, this biasing information is readily available by the time the target release is presented. This means that we can look at the degree to which an underadditive pattern is possible with a decisional variable in the precise case we are studying, where the target burst information accumulates over time. If this variable also shows underadditivity with SOA then we would need to rethink our interpretations of the first three underadditive patterns. On the other hand, if this decisional variable is additive with SOA then we could be confident that the task difficulty variables that have shown an underadditive pattern are truly pre-bottleneck.

Experiment 4

Method

Participants. Forty-eight undergraduate students (42 female and 6 male) from the University of York either were paid £4 or received course credit for participation. Participants' ages range from 18 to 22 years (mean age = 19), and participants reported no speech, hearing, or visual disorders.

Materials. S1 and S2 stimuli were the same as in Experiment 3.

Design. There were 12 conditions in the experiment, defined by the combination of two levels of lexical status (word vs. non-word), two levels of response bias (congruent vs. incongruent) and three levels of SOA (100 ms vs. 200 ms vs. 1000 ms).

Lexical status and SOA were manipulated as in Experiment 3. The response bias factor was driven by the fillers. Within each list, all filler items were assigned to the same response button. As a consequence, in total 198 responses were assigned to one key (144 fillers and 54 experimental items) and only 54 to the other (all experimental items). A test trial in which the correct response was the same as all the filler trials was labeled congruent and an item where the correct response was different to the filler trials was labeled incongruent.

The experimental items were assigned to twelve condition groups for counterbalancing purposes. In each group, nine appeared in the congruent word condition, nine in the incongruent word condition, nine in the congruent non-word condition and nine in the incongruent non-word condition. The SOA conditions (100, 200 and 1000 ms) were rotated across the condition groups such that twelve participant lists were created. Each participant list contained all items, with an equal number of items in the word and non-word condition, an equal number in the congruent and incongruent conditions, and an equal number in each SOA condition. Every item appeared four times in every SOA condition, six times in both lexical status conditions, and six times in both response bias condition across the twelve participant lists. 108 experimental items appeared to each participant.

The bias-congruent filler response key was rotated through two sets of the participant lists so that both the upper and lower buttons appeared as congruent and incongruent responses. This was so that any difference in RTs to congruent and incongruent items could not be attributed to the positioning of the buttons.

Every participant completed 108 experimental items, 9 in each of the twelve conditions defined by the two levels of lexical status (word vs. non-word), two levels of response bias (congruent vs. incongruent) and the three levels of SOA (100 vs. 200 vs. 1000 ms). The order of presentation was randomized for each participant.

Procedure. The procedure was as for Experiment 3.

Results.

Responses faster than 100 ms were excluded from the analysis, as were responses slower than the cut-off time of 3000 ms (0.1% of trials). Trials in which an error was made in either Task 1 or Task 2 were also excluded from the RT analysis (8% of trials). Table 4 summarizes the remaining data.

Task 1 RTs. ANOVAs with lexical status (word vs. non-word), response bias (congruent vs. incongruent) and SOA (100 vs. 200 vs. 1000 ms) revealed a statistically significant main effect of response bias, $F_1(1,47)=5.29$, $p<.05$ $F_2(1,107)=4.11$, $p<.05$, with response times 16 ms slower for bias-incongruent than congruent trials. There was also a statistically significant main effect of SOA by items only, $F_1(2,94)=1.84$, $p=.16$; $F_2(2,214)=4.16$, $p<.05$, with responses slowing as the SOA increased. No other effects were statistically significant.

Task 2 RTs. ANOVAs with lexical status (word vs. non-word), response bias (congruent vs. incongruent) and SOA (100 vs. 200 vs. 1000 ms) revealed a significant effect of lexical status, $F_1(1,47)=12.91$, $p<.01$; $F_2(1,107)=7.03$, $p<.01$, with RTs 28 ms shorter to words than non-words. There was also a statistically significant effect of response bias, $F_1(1,47)=50.13$, $p<.001$; $F_2(1,107)=73.29$, $p<.001$, with responses 70 ms faster when the item was in the bias-congruent condition than when it was in the bias-incongruent condition. As expected, there was a statistically significant effect of SOA, $F_1(2,94)=458.22$, $p<.001$; $F_2(2,214)=686.87$, $p<.001$, with RTs decreasing as SOA increased. RTs were 406 ms shorter at the 1000 ms SOA than at the 100 ms SOA, and 122 ms shorter at the 200 ms SOA than at the 100 ms SOA.

The interaction of lexical status by SOA failed to reach statistical significance but a more detailed examination of the overall pattern of results using planned comparisons indicated the basic underadditive pattern of responding akin to that reported in Experiment 3 (see Figure 5, lower graph). At the 100 ms SOA, the 12 ms effect of lexical status was non-significant (both Fs<1), at the 200 ms SOA there was a marginally significant 27 ms lexical status effect, $F_1(1,47)=4.14$, $p<.05$; $F_2(1,107)=3.34$, $p=.071$, and at the 1000 ms effect there was a robust 45 ms lexical status effect, $F_1(1,47)=14.16$, $p<.001$; $F_2(1,107)=15.23$, $p<.001$. Indeed, marginally significant planned interaction contrasts suggested that the lexical status effect at the 1000 ms SOA was greater than at the 100 ms SOA, $F_1(1,47)=4.02$, $p=.051$; $F_2(1,107)=3.00$, $p=.086$.

More importantly, there was no interaction of response bias by SOA (see Figure 5, upper graph). Planned comparisons revealed that the effect of response bias was statistically significant at all stimulus asynchronies. At the 100 ms SOA there was a 71 ms effect, $F_1(1,47)=15.05$, $p<.001$; $F_2(1,107)=23.37$, $p<.001$, at the 200 ms SOA there was a 72 ms effect $F_1(1,47)=26.32$, $p<.001$; $F_2(1,107)=18.34$, $p<.001$, and at the 1000 ms SOA there was a 69 ms effect, $F_1(1,47)=54.41$, $p<.001$; $F_2(1,107)=48.07$, $p<.001$. In other words, response times to bias-congruent items were shorter than to bias-incongruent items at all stimulus asynchronies, and the effects were of equivalent size (within 4 ms of each other) at all SOAs.

Error analysis. Participants made an error in either Task 1 or Task 2 in a total of 426 cases (8% of all trials). Of these 68% were errors in response to the phoneme decision task (e.g. a /t/ response when /k/ was correct), 22% were errors in responding to the color, and 10% were trials in which the participant made an error in response to both tasks.

ANOVAs with lexical status (word vs. non-word), response bias (congruent vs. incongruent) and SOA (100 vs. 200 vs. 1000 ms) revealed a statistically reliable effect of response bias, $F_1(1,47)=28.89$, $p<.001$; $F_2(1,107)=51.05$, $p<.001$. This was because participants made more errors in the bias-incongruent condition (mean .11 proportion error responses) than the congruent condition (mean .05 proportion error responses). There was an effect of SOA by items only, $F_2(2,214)=3.76$, $p<.05$. This was due to the increase in the proportions of errors produced at the 1000 ms SOA compared to the 100 ms and 200 ms SOA, however it was non-significant by participants. No other effects were statistically reliable.

Discussion

The key finding here was the additive pattern of response times linked to the manipulation of response bias. At the long SOA, with no bottleneck in processing, this factor showed a 69 ms effect. This effect is comparable in strength to the lexical status and mismatch effects at the long SOA in earlier experiments. However, in contrast to previous experiments this effect was just as strong at the 100 and 200 ms SOAs. This additive pattern is exactly in line with the prediction that variables that affect the mechanisms responsible for producing a response are located at a post-bottleneck stage of processing. The manipulation of response bias tapped into a post-bottleneck stage of processing, exerting its influence at a later stage of processing than the other psycholinguistic variables that we have examined in our research; namely word frequency (Cleland, et al., 2006), cues to co-articulation, and lexical status.

The underadditive pattern shown relative to the manipulation of lexical status was clearly less robust in this experiment than in Experiment 3. However, the task constraints in Experiment 4 were considerable—participants not only had to bear in mind the phoneme-to-response-key mapping on every trial but they also had to remember which of the alternatives was most likely to occur. Under more straightforward conditions (as in Experiment 3) the underadditive pattern was clear cut. Furthermore, an analysis across Experiments 3 and 4 (collapsing across response bias in Experiment 4) provided very clear evidence for an underadditive effect of lexical status in the PRP paradigm, with the word advantage at the long SOA diminishing to nothing at the shortest SOA (lexicality by SOA interaction: $F1(2,164)=7.01$, $p<.005$; $F2(2,428)=5.24$, $p<.01$). This effect did not differ between experiments (lexicality by SOA by experiment interaction: $Fs < 1$).

These results clarify the interpretation of the previous experiments, and rule out a bottleneck or post-bottleneck explanation of the task difficulty effects in the previous experiments (see Experiment 3 Discussion). If the underadditive effects of subcategorical mismatch and lexical status are explainable in terms of a decisional locus of integration, but with stronger effects of task difficulty at longer SOAs because of the drawn-out nature of the target bursts, then exactly the same pattern should be observed for the bias manipulation. In the absence of underadditivity here, we can be confident that the integration of phonemic knowledge has a pre-bottleneck locus.

General Discussion

Several recent lines of argument suggest that a categorical pre-lexical phonemic representation of speech is inadequate for spoken word recognition. These have led researchers to question the involvement of phonemes in speech perception, with, for example, Warren (2000, p. 350) arguing that “there is no phonemic level employed in the perceptual processing of speech”. Nonetheless, our experiments showed that resolution of phonemic ambiguity and the integration of phonemic knowledge takes place during the normal course of speech perception without any reliance on a later decisional mechanism. Experiments 1 and 2 demonstrated this behavior in the context of subcategorical mismatches in the place of articulation of syllable-final consonants. In such cases, where coarticulatory and burst cues mismatch, the perceptual representation initially reflects these mismatches but then over time forms a representation that appears more categorical, with these subtleties lost. Experiment 3 demonstrated similar behavior in the context of integrating lexical biases with perceptual information about final consonant. Here, lexical information provides an initial advantage in the processing of a word, but when decisions are postponed, the disadvantage in processing a final phoneme of a non-word dissipates rapidly. Experiment 4 then demonstrated that such integration of two sources of information is not possible in pre-bottleneck "slack time" when one of the factors is a decisional rather than a perceptual variable.

In the case of both subcategorical mismatch and lexical status effects, the critical conclusion is that a categorical representation of speech is formed at some level other than a central decisional one during the course of spoken word recognition. What, then, are the consequences of our findings for models of spoken word recognition?

We have already fleshed out the implications of these results for the Merge model (Norris et al., 2000), which makes distinctive predictions about the decisional nature of phonemic integration. Two properties of Merge cause problems in the accommodation of our data. First, Merge makes use of a veridical representation of speech at input (i.e. subphonemic conflict remains crystallized in the sublexical representation of speech). Second, lexical and perceptual cues to the identity of phonemes are only integrated at an explicitly decisional level. These factors combined mean that according to Merge, phonemic integration in either case studied here can only occur at a decision level, predicting that such integration would be postponed by the imposition of a dual-task bottleneck. Therefore, Merge predicts that the lexical status and subcategorical mismatch should combine additively with SOA.

The contrast between the task difficulty variables examined in Experiments 1-3 and the effect of response bias in Experiment 4 emphasizes the problem for Merge. Why should it be that integration of perceptual information with a response bias cannot take place prior to bottleneck processing, and yet integration of mismatching perceptual information or of perceptual and lexical information can? Perhaps one could argue that at short SOAs (when the bottleneck delays a decision) the perceptual information has had sufficient time to build up to a level at which the lexical bias is irrelevant, hence underadditivity. But if this was the true for lexical biases, it should be equally true for response key biases. In both cases, Merge predicts that a response network should be set up prior to or during the course of the experiment that integrates the biasing information with perceptual information. If phonemic integration is truly decisional then both types of variable should exert their effect at or after a lexical bottleneck. We are forced to conclude that phonemic integration is not a decisional process.

A decisional model of phonemic integration might be viable if it could be shown that the late decisional level is nonetheless prior to the dual-task bottleneck. However, such an argument goes against a large body of theoretical and empirical research relating to bottleneck effects. As mentioned, Sigman and Dehaene (2005, p. 334) focused very clearly on "decision based on noisy integration of evidence" as the stage in task processing subject to bottleneck constraints, and this is explicitly the level of processing identified by Norris et al. (2000) as being the stage at which listeners integrate phonemic knowledge. Empirically, there is now a wide range of variables that have shown additive (i.e., bottleneck or post-bottleneck) effects when combined with manipulation of SOA, supporting the notion of a decisional or response selection bottleneck. These include Stroop interference (Fagot & Pashler, 1992), the color version of the Simon effect (Magen & Cohen, 2005), visual image foreshortening (Lawson, Humphreys & Jolicoeur, 2000) and word frequency

(McCann et al., 2000). Although Cleland et al. (2006) found an underadditive effect of visual and spoken word frequency in the PRP paradigm, the effects in their experiments were very different from the complete loss of task difficulty effects at short SOA found here (frequency effects at the shortest SOA were 70-80 ms). This partial underadditive pattern of responses suggests that at least part of the word frequency effect is located at or after the dual-task bottleneck.

Johnston and McCann (2006) provided a particularly compelling example of additive bottleneck effects. In one experiment they showed that manipulating the difficulty of judging the left/right position of a cross in a circle (by altering the extent to which the cross was off-centre) was near-additive with SOA, and three more experiments manipulating the difficulty of a box-width judgment showed a purely additive pattern of results. Just like the current experiments, the manipulations were intended to make the participants' binary judgments more or less easy, but unlike the current experiment, this decisional difficulty manipulation could not be absorbed into slack time at short SOAs. Johnston and McCann (2006) contrasted their results with those of Johnston, McCann and Remington (1995), who found that distortions of the shape of letters produced an underadditive interaction with SOA, implying that letter identification occurs prior to the processing bottleneck. Their suggestion was that letter categorization is sufficiently practiced and important to become an automatic component of visual word recognition. The current results imply that, despite the good reasons for needing to encode fine-grained detail in the speech signal (e.g., Salverda & Tanenhaus, 2006), the same is true for phonemes.

According to the Merge model, making decisions about phonemes should be equivalent to judging box widths or cross positions. Norris et al. (2000; p. 324) stated that "The phoneme nodes are not part of everyday speech comprehension, and in experimental situations they are built on the fly to cope with the specific task demands." with the number of decision nodes "assumed to be determined by the phonemic decision task at hand". Norris et al. (2000; p.355) also suggest that this same mechanism should be able to deal with the near-limitless range of tasks that a listener might be able to perform (e.g., monitoring for word-initial phonemes in animal words when a signal light turns green). The separation of this "flexible decision process" (Norris et al., 2000; p. 355) from normal speech perception means that Merge cannot explain why the difficulty of judging the width of a box or the position of a cross leads to additive combination with SOA, whereas the difficulty of judging the identity of a phoneme leads to underadditivity.

To accommodate our results, we suggest that Merge would need to be adapted, such that phonemic integration becomes part of normal speech perception, in which case activation could flow from pre-lexical phoneme nodes and lexical nodes to a phonemic category level without interference from a central or decisional bottleneck. However, the explicit decisional aspect of this level was one of the underlying principles behind Merge, and such an alteration would not be trivial. Adding a nondecisional, dedicated categorical phoneme level would reduce the parsimony of Merge's explanation of lexical effects on phoneme recognition, and would reduce the distinctiveness of the Merge model, as compared to other models. It might also reduce the flexibility of Merge to deal with different response requirements in different situations (see Mirman, McClelland & Holt, 2006, for an alternative encapsulation of attentional factors in lexical processing). Both of these were principal arguments put forward by Norris et al. (2000) for the original Merge architecture.

Our data may also be problematic for episodic models of spoken word recognition, particularly in their more extreme form. Goldinger (1998) examined a "pure" episodic model, testing the hypothesis that lexical representations of words consisted of multiple exemplars of detailed episodic memory traces in which aspects such as speech rate and aspects of the speaker's voice are preserved. Insofar as these models do not generate abstract categorical representations of phonemes they predict that any ambiguities in the speech signal should not diminish over time. Thus they might explain the long SOA cost of subcategorical mismatch found in Experiments 1 and 2, but would be unable to capture the lack of a cost at short SOAs in our experiments (see also McQueen, Cutler & Norris, 2006). Of course intermediate positions that retain some features of episodic and abstractionist models exist, and the extent to which they can explain our data would depend on how abstract phonological representations are accommodated. For example, Pierrehumbert (2002) presented an exemplar model of perception in which spoken words are represented in terms of detailed phonetic encoding. However, Pierrehumbert also assumed that phoneme decisions could be made on the basis of the decisional mechanism outlined in Merge. Clearly, this kind of compromise position is weakened by our data for the very same reasons that the Merge model is.

Other models of spoken word recognition do not rely on decisional mechanisms for phonemic integration, and so at least superficially they might appear better able to accommodate the results we report here. For example, the observed pattern might be a consequence of slow "settling" behavior at a pre-lexical phonemic level. This description is most obviously compatible with TRACE (McClelland & Elman, 1986), for which interactive activation and competition induce a settling process at the pre-lexical phonemic level. Appendix B describes the results of simulations using jTRACE (Strauss, Harris & Magnuson, in press) illustrating TRACE's behavior in these circumstances. Broadly speaking, TRACE appears to provide a good account of the time-course of the subcategorical mismatch effects, with immediate effects early on in processing (equivalent to the mismatch effect at the long SOA) and reduced/eliminated effects later on (equivalent to the lack of mismatch at the shortest SOA).

However, it is worth noting that a major contributor to the variation in the effect of mismatch at different delays in TRACE is the Luce decision rule, which operates on the raw activations, presumably at a decisional level. When we plotted these raw phoneme activations in the subcategorical mismatch simulation, there remained a cost associated with the mismatch condition even at the later processing cycles. However, using the Luce choice rule to transform activations into response probabilities either reduced this late effect or eliminated, depending on the choice of constant. Thus, it may be that the extra processing of speech that goes on during bottleneck slack time is not producing completely categorical phoneme representations, just pushing activation values up to levels at which a categorical decision becomes easy. Because decisions are not explicitly modeled in TRACE, and because different constants have been applied for different datasets, it is unclear exactly how much raw activation differences will lead to processing costs at a decision level.

Turning to the lexical status effects in Experiments 3 and 4, our simulations (Appendix B) suggested that there should be little or no difference between the word and non-word conditions at long SOAs, whereas Experiment 3 revealed an 84 ms effect and Experiment 4 showed a 46 ms effect. Doubling the normal level of word-phoneme feedback did allow a small difference between the word and non-word conditions to emerge, but generally speaking it seems that the standard version of TRACE does not predict lexical effects on phoneme recognition with our stimuli. This result fits in with other accounts of TRACE's operation. For example, the original TRACE paper (McClelland & Elman, 1986) stressed that lexical effects on phoneme processing should only be expected if there is sufficient time for lexical activations to build up prior to the processing of the target phoneme.

Similarly, the Marslen-Wilson and Warren (1994) simulations of their experiments showed only very minor differences between the processing of syllable-final control stimuli differing in lexical status (see Marslen-Wilson & Warren, 1994, Figures 14 & 15, p. 670). In general, TRACE predicts that lexical effects on phoneme processing should only be observable towards the end of polysyllabic words with early uniqueness points, or in cases where there is real ambiguity as to the identity of a key phoneme (as in Ganong, 1980). In the current case, the words were monosyllabic, were phonemically unambiguous, and were not uniquely identifiable prior to offset. Furthermore, for our word stimuli we ensured that both phoneme response options were compatible with the speech prior to the final phoneme, meaning that any lexical feedback prior to the final phoneme would be facilitative of both response options. Thus, although TRACE provides an obvious means for combining lexical and phonemic information prior to a decisional level, it does not predict a strong lexical effect for the materials in our experiments.

Although our research was motivated by the clear predictions that can be made for Merge with respect to decisional costs, it seems that TRACE does not accommodate our data either. We should emphasize that our data do not speak directly to the broader debate over top-down feedback in spoken word recognition. It could be that the data can be captured in a model without feedback provided that categorical phonemic representations become available for words and non-words alike as part of the normal process of speech perception, rather than "on the fly". Thus, models in which phonemic representations for words and non-words emerge in parallel to lexical access (e.g., Gaskell & Marslen-Wilson, 1997) or postlexically (Marslen-Wilson & Warren, 1994) could also potentially accommodate our data, provided that they allow lexical information to influence phonological representations. Indeed, Marslen-Wilson and Warren's original interpretation of their own results was in terms of a postlexical phonemic representation, which was lexically stored for words and built by analogy to the lexicon for non-words.

These arguments were part of the inspiration for Gaskell & Marslen-Wilson's (1997) Distributed Cohort Model (DCM). This model avoids any categorization into phoneme units prior to lexical access, but posits a level of representation that concurrently codes both semantic and phonological information in separate banks of units. The phonological units are further sub-divided into three groups that respectively encode syllable onset, nucleus and coda. Phoneme decisions are therefore based on the activation of particular phoneme units at this level of representation. So in terms of the present experiments, target phoneme decisions would be based on the most activated unit in the coda group. DCM can utilize fine-grained detail to distinguish between lexical items differing nonphonemically (e.g., in terms of syllable boundary position or vowel duration), but offers the possibility of providing a phonological representation as part of the recognition process for words and non-words alike (Gaskell & Marslen-Wilson, 1997). Gaskell and Marslen-Wilson (1997) showed that DCM predicts costs associated with subcategorical mismatch. Unlike TRACE, DCM also shows a lexical advantage at the phonological output level in the kinds of conditions studied here (see Gaskell & Marslen-Wilson, 1997, Figures 4 and 5; Gaskell, 2000, Figure 1). The reason that the DCM shows lexical effects in these circumstances is that producing a phonological representation for a non-word in a backpropagation network is an example of generalization from trained forms (words) to untrained forms (non-words). Although with sufficient training the network can learn to carry out this generalization reasonably well, the phonological representations are always somewhat less categorical for non-words than for words (particularly so for Gaskell & Marslen-Wilson, 1997; less so for Gaskell, 2000). In other words, the lexical status effect we have found can be explained as a detrimental effect of making a judgment on an unlearnt form.

DCM therefore deals well with the long SOA effects of subcategorical mismatch and lexical status. However, this model does not employ a clean-up mechanism for the phonological units at the lexical level. Thus, the model would predict the observed subcategorical mismatch costs and lexical status effects on the activation of phonological units at long SOAs, but these effects would not dissipate over time, meaning that the underadditivity at the short SOAs would not be predicted. As a minimum, a revised model with a settling network at this level would need to be built in order to allow DCM to accommodate these data (see Rodd, Gaskell, & Marslen-Wilson, 2004, for one implementation along these lines).

It seems that none of the current models of spoken word recognition can capture the full range of effects demonstrated in our experiments as they stand. Merge (Norris et al., 2000) does not make use of settling behavior at the sublexical phonemic level, and uses an explicitly decisional level to integrate lexical and phonemic representations. TRACE (McClelland & Elman, 1986) has the potential to allow this integration as part of speech perception, but fails to show lexical effects for short monosyllabic stimuli and only shows settling behavior given phonemic mismatch after the application of a decisional transformation. Finally, DCM (Gaskell & Marslen-Wilson, 1997) shows swift lexical and subcategorical mismatch effects in these circumstances, but also does not incorporate settling at the phonemic output level to allow these effects to dissipate over time.

The precise nature of these categorical representations is unclear. As has already been described, these representations could be prelexical, as in an adapted form of TRACE, with settling behavior gradually resolving ambiguities and feedback from the lexical level providing a word advantage in the short term. Equally though, they could be lexical or post-lexical, as in an adapted form of Merge or DCM. At the beginning of this article we discussed data (e.g., Davis et al., 2002; McQueen et al., 2003; Salverda et al., 2003; Salverda & Tanenhaus, 2006) suggesting that a cascaded phoneme level is insufficiently fine-grained for spoken word recognition. Such a representation is unable to distinguish between for example the short /ae/ in *captain* and the longer /ae/ in *cap*, whereas listeners do show some sensitivity to these factors. For this reason the optimal solution in future models might retain or enhance the settling behavior of TRACE, but with DCM and Merge's preservation of subphonemic detail pre-lexically, meaning that categorical phoneme representations are formed during the course of normal speech perception, but alongside the lexical access process.

Such representations could be involved in the production system, allowing listeners to generate articulatory plans for words and nonwords, or they could provide a starting point for phonological short-term memory. An interesting alternative is that these representations are related to learning to read an alphabetic language, and form the basis of literate listeners' phoneme awareness (Castles & Coltheart, 2004). Much recent research points to the automaticity of retrieval of orthographic information during speech perception (e.g., Chereau, Gaskell & Dumay, 2007; Ziegler & Ferrand, 1998; Frost & Ziegler, in press). If this categorical representation emerges as a consequence of, or in tandem to, literacy, then it is feasible that dyslexic readers would not show such effects in speech perception because they only have access to a more veridical pre-lexical representation of speech (cf. Serniclaes, Sprenger-Charolles, Carré & Demonet, 2001; Serniclaes, Ventura, Morais, & Kolinsky, 2005). That is, whereas normal readers would show the above pattern in the PRP paradigm, dyslexics might form categorical representations only as a consequence of a decisional mechanism, leading to an additive pattern of results in a PRP experiment.

Regardless of how these data might be modeled or characterized, it seems that the perceptual system for speech is performing a neat trick. Accumulated evidence points to a recognition system that is exquisitely sensitive to the fine-grained nature of speech, but our current data show that this veridical representation of speech is quickly lost (or at least obscured), with a more categorical representation of words and non-words emerging. The categorical representation of speech eliminates subphonemic conflict and integrates lexical and pre-lexical representations, all without the involvement of decisional processes. Providing a representational system that can incorporate both of these properties should be a significant challenge for current and future models.

References

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language, 38,* 419-439.

Andruski, J. E., Blumenstein, S. E., & Burton, M. (1994). The effects of subphonetic differences on lexical access. *Cognition, 52,* 163-187.

Baayen, R. H., Piepenbrock, R., & van Rijn, H. (1993). *The CELEX lexical database (CD-ROM)*. Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania.

Bertelson, P., & de Gelder, B. (1991). The emergence of phonological awareness: Comparative approaches. In I.G. Mattingly & M. Studdert-Kennedy (Eds.), *Modularity and the motor theory of speech perception*. Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Castles, A., & Coltheart, M. (2004). Is there a causal link from phonological awareness to success in learning to read? *Cognition, 91*, 77-111.

Chereau, C., Gaskell, M. G., & Dumay, N. (2007). Reading spoken words: Orthographic effects in auditory priming. *Cognition, 102*, 341-360.

Cleland, A. A., Gaskell, M. G., Quinlan, P. T., & Tamminen, J. (2006). Frequency effects in spoken and visual word recognition: Evidence from dual-task methodologies. *Journal of Experimental Psychology: Human Perception and Performance, 32,* 104-119.

Connine, C. M., & Clifton, C. (1987). Interactive use of lexical information in speech-perception. *Journal of Experimental Psychology-Human Perception and Performance, 13*, 291-299.

Cutler, A., & Norris, D. (1979). Monitoring sentence comprehension. In W. E. Cooper & E. C. T. Walker (Eds.), *Sentence Processing: Psycholinguistic Studies Presented to Merrill Garrett*. Hillsdale, NJ: Erlbaum.

Dahan, D., Magnuson, J. S., & Tanenhaus, M. K. (2001). Time course of frequency effects in spoken-word recognition: Evidence from eye movements. *Cognitive Psychology, 42*, 317-367.

Dahan, D., Magnuson, J. S., Tanenhaus, M. K., & Hogan, E. M. (2001). Subcategorical mismatches and the time course of lexical access: Evidence for lexical competition. *Language and Cognitive Processes, 16,* 507-534.

Davis, M. H., Marslen-Wilson, W. D., & Gaskell, M. G. (2002). Leading up the lexical garden path: Segmentation and ambiguity in spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 28,* 218-244.

Fagot, C., & Pashler, H. (1992). Making 2 responses to a single object: implications for the central attentional bottleneck. *Journal of Experimental Psychology: Human Perception and Performance, 18*, 1058-1079.

Ferreira, V. S., & Pashler, H. (2002). Central bottleneck influences on the processing stages of word production. *Journal of Experimental Psychology: Learning Memory and Cognition, 28*, 1187-1199.

Frost, R. & Ziegler, J. (in press). Speech and spelling interaction: The interdependence of visual and auditory word recognition. In M. G. Gaskell (Ed.). Oxford Handbook of Psycholinguistics. Oxford: Oxford University Press.

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

Gaskell, M. G., & Marslen-Wilson, W. D. (1997). Integrating form and meaning: A distributed model of speech perception. *Language and Cognitive Processes, 12,* 613-656.

Gaskell, M. G. (2000). Modeling lexical effects on phonetic categorization and semantic effects on word recognition. *Behavioral and Brain Sciences, 23*, 329-330.

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review, 105*, 251-279.

Hazeltine, E., Ruthruff, E., & Remington, R. W. (2006). The role of input and output modality pairings in dual-task performance: Evidence for content-dependent central interference. *Cognitive Psychology, 52*, 291-345.

Johnston, J. C., McCann, R. S., & Remington, R. W. (1995). Chronometric evidence for two types of attention. *Psychological Science, 6*, 365-369.

Johnston, J. C., & McCann, R. S. (2006). On the locus of dual-task interference: Is there a bottleneck at the stimulus classification stage? *Quarterly Journal of Experimental Psychology, 59*, 694-719.

Luce, R. D. (1959). *Individual choice behavior.* New York: Wiley.

Luck, S. J. (1998). Sources of dual-task interference: Evidence from human electrophysiology. *Psychological Science, 9*, 223-227.

Magen, H., & Cohen, A. (2005). Location specificity in response selection processes for visual stimuli. *Psychonomic Bulletin & Review, 12*, 541-548.

Magnuson, J. S., McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2003). Lexical effects on compensation for coarticulation: the ghost of Christmash past. *Cognitive Science, 27*, 285-298.

Marslen-Wilson, W. D., & Warren, P. (1994). Levels of perceptual representation and process in lexical access: Words, phonemes, and features. *Psychological Review, 101,* 653-675.

Marslen-Wilson, W. D., Moss, H. E., & van Halen, S. (1996). Perceptual distance and competition in lexical access. *Journal of Experimental Psychology: Human Perception and Performance, 22,* 1376-1392.

McCann, R. S., Remington, R. W., & Van Selst, M. (2000). A dual-task investigation of automaticity in visual word processing. *Journal of Experimental Psychology: Human Perception and Performance, 26,* 1352-1370.

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

McClelland, J. L., Mirman, D., & Holt, L. L. (2006). Are there interactive processes in speech perception? *Trends in Cognitive Sciences, 10*, 363-369.

McQueen, J. M. (1991). The influence of the lexicon on phonetic categorization: stimulus quality in word-final ambiguity. *Journal of Experimental Psychology: Human Perception and Performance, 17*, 433-443.

McQueen, J. M. (2003). The ghost of Christmas future: didn't Scrooge learn to be good? - Commentary on Magnuson, McMurray, Tanenhaus, and Aslin (2003). *Cognitive Science, 27*, 795-799.

McQueen, J. M., Cutler, A., & Norris, D. (2003). Flow of information in the spoken word recognition system. *Speech Communication, 41*, 257-270.

McQueen, J. M., Cutler, A., & Norris, D. (2006). Phonological abstraction in the mental lexicon. *Cognitive Science, 30*, 1113-1126.

McQueen, J. M., Dahan, D., & Cutler, A. (2003). Continuity and gradedness in speech processing. In N. O. Schiller & A. S. Meyer (Eds.), *Phonetics and Phonology in Language Comprehension and Production: Differences and Similarities* (pp. 39-78). Berlin: Mouton de Gruyter.

McQueen, J. M., Norris, D., & Cutler, A. (1999). Lexical influence in phonetic decision making: Evidence from subcategorical mismatches. *Journal of Experimental Psychology: Human Perception and Performance, 25*, 1363-1389.

Meyer, D. E., & Kieras, D. E. (1997). A computational theory of executive cognitive processes and multiple task performance: Part 2. Accounts of psychological refractory-period phenomena. *Psychological Review, 104,* 749-791.

Mirman, D., McClelland, J.L., and Holt, L.L. (2006). Attentional Modulation of Lexical Effects on Speech Perception: Computational and Behavioral Experiments. *Proceedings of the 28th Annual Cognitive Science Society Meeting* (pp. 591-596), Mahwah, NJ: Lawrence Erlbaum Associates.

Morais, J., & Kolinsky, R. (1994). Perception and awareness in phonological processing - the case of the phoneme. *Cognition, 50*, 287-297.

Navon, D., & Miller, J. (2002). Queuing or sharing? A critical evaluation of the single-bottleneck notion. *Cognitive Psychology, 44,* 193-251.

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

Norris, D., McQueen, J. M., & Cutler, A. (2000). Merging information in spoken word recognition: Feedback is never necessary. *Behavioral and Brain Sciences, 23,* 299-370.

Pashler, H. (1994). Dual-task interference in simple tasks: Data and theory. *Psychological Bulletin, 116,* 220-244.

Pashler, H. E. (1998). *The psychology of attention.* Cambridge, MA: The MIT Press.

Pierrehumbert, J. B. (2002).Word-specific phonetics. In C. Gussenhoven & N.Warner (Eds.), Laboratory phonology 7 (pp. 101–139). Berlin: de Gruyter.

Pitt, M. A., & Samuel, A. G. (1993). An empirical and metaanalytic evaluation of the phoneme identification task. *Journal of Experimental Psychology: Human Perception and Performance, 19*, 699-725.

Read, C., Zhang, Y., Nie, H., & Ding, B. (1986). The ability to manipulate speech sounds depends on knowing alphabetic writing. *Cognition, 24*, 31-44.

Reynolds, M., & Besner, D. (2006). Reading aloud is not automatic: Processing capacity is required to generate a phonological code from print. *Journal of Experimental Psychology: Human Perception and Performance, 32*, 1303-1323.

Rodd, J. M., Gaskell, M. G., & Marslen-Wilson, W. D. (2004). Modelling the effects of semantic ambiguity in word recognition. *Cognitive Science, 28,* 89-104.

Rubin, P., Turvey, M. T., & Gelder, P. V. (1976). Initial phonemes are detected faster in spoken words than non-words. *Perception and Psychophysics, 19*, 394-398.

Salverda, A. P., Dahan, D., & McQueen J. M. (2003). The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension. *Cognition, 90,* 51-89.

Salverda, A.P., & Tanenhaus, M. K. (2006, August). *Lexical activation is affected by durational information established with respect to preceding phonetic context.* Paper presented at the AMLaP Conference, Nijmegen (the Netherlands).

Samuel, A. G. (1981). The role of bottom-up confirmation in the phonemic restoration illusion. *Journal of Experimental Psychology: Human Perception and Performance, 7*, 1124-1131.

Samuel, A. G., & Pitt, M. A. (2003). Lexical activation (and other factors) can mediate compensation for coarticulation. *Journal of Memory and Language, 48*, 416-434.

Scharenborg, O., Norris, D., ten Bosch, L., & McQueen, J. M. (2005). How should a speech recognizer work? *Cognitive Science, 29,* 867-918.

Serniclaes, W., Sprenger-Charolles, L., Carré, R., & Demonet, J. F. (2001). Perceptual discrimination of speech sounds in developmental dyslexia. *Journal of Speech Language and Hearing Research, 44*, 384-399.

Serniclaes, W., Ventura, P., Morais, J., & Kolinsky, R. (2005). Categorical perception of speech sounds in illiterate adults. *Cognition, 98*, B35-B44.

Sigman, M., & Dehaene, S. (2005). Parsing a cognitive task: A characterization of the mind's bottleneck. *Plos Biology, 3*, 334-349.

Strauss, T. J., Harris, H. D. & Magnuson, J. S. (2007). jTRACE: A reimplementation and extension of the TRACE model of speech perception and spoken word recognition. *Behavior Research Methods, 39,* 19-30.

Streeter, L. A., & Nigro, G. N. (1979). The role of medial consonant transitions in word perception. *Journal of the Acoustical Society of America, 65*, 1533-1541.

Tombu, M., & Jolicoeur, P. (2003). A central capacity sharing model of dual-task performance. *Journal of Experimental Psychology: Human Perception and Performance, 29,* 3-18.

Tombu, M., & Jolicoeur, P. (2005). Testing the predictions of the central capacity sharing model. *Journal of Experimental Psychology: Human Perception and Performance, 31,* 790-802.

Van Galen, G. P., & ten Hoopen, G. (1976). Speech Control and Single Channelness. *Acta Psychologica, 40*, 245-255.

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

Warren, R. M. (2000). Phonemic organization does not occur: Hence no feedback. *Behavioral and Brain Sciences, 23,* 350-351.

Whalen, D. H. (1984). Subcategorical phonetic mismatches slow phonetic judgements. *Perception and Psychophysics, 35*, 49-64.

Ziegler, J. C., & Ferrand, L. (1998). Orthography shapes the perception of speech: The consistency effect in auditory word recognition. *Psychonomic Bulletin & Review, 5,* 683-689.

Appendix A

Table A1. Words used in Experiment 1 (in each word pair, the higher frequency member is listed first), and non-words used in Experiment 2 (in each non-word pair, the non-word used in the matching condition is listed first.) Words marked with an asterisk were used in Experiments 3 and 4, along with their non-word counterparts (e.g., bat-dat). Quadruplet used in the TRACE simulation are marked (†).

Words	*Non-words*	*Words*	*Non-words*
back – bat*	dack - dat	map – mat*	smap - smat
bait – bake	thait - thake	make* – mate*	chake - chate
bed* – beg*	ged - geg	mud* – mug	wud - wug
big – bid	flig - flid	neck – net*	sheck - shet
bite* – bike	gite - gike	pack* – pat*	drack - drat
bob* – bog*	pob - pog†	part – park*	rart - rark†
bug* – bud*	nug - nud	peak* – peat	deak - deat
cart* – carp*	sart - sarp†	pick* – pit	vick - vit
chap – chat	wap - wat	pipe* – pike*	thipe - thike
cheap – cheat*	theap - theat	pot – pop	smot - smop
chip* – chick*	fip - fick	port* – pork*	jort - jork
click* – clip*	bick - bip	quick – quit	swick - swit
coat – cope	foat - fope	rat* – rack*	gat - gack
creep* – creek*	treep - treek†	rib* – rig*	chib - chig
curb* – curd*	lurb - lurd	right* – ripe*	dight - dipe
cut* – cup*	lut - lup†	rock* – rot	thock - thot
dark* – dart*	jark - jart	rub* – rug*	flub - flug
debt – deck*	kebt - keck	seat* – seek*	jeat - jeek
dock – dot*	vock - vot	shake* – shape*	pake - pape
flat* – flap*	flet - flep	shop* – shock*	shup - shuck
flick* – flip*	smick - smip	sit* – sick	drit - drick
flock* – flop*	snock - snop	slap – slack	spap - spack
heat* – heap*	sneat - sneap	slip – slit	stip - stit
herb* – herd*	jerb - jerd	snap* – snack*	grap - grack
hit – hip	het - hep	soap* – soak*	boap - boak
hot – hop*	brot - brop	spite* – spike	stite - stike
job* – jog*	chob - chog	stab* – stag*	clab - clag
kick* – kit*	glick - glit	state* – stake*	smate - smake
knot* – knock	fot - fock	stop* – stock*	trop - trock†
lab* – lag*	thab - thag	street – streak	breet - breek
lack* – lap*	fack - fap	stud* - stub*	brud - brub†
lake* – late*	loke - lote	sweet – sweep*	queet - queep
lead – league*	jead - jeague	take* – tape*	spake - spape
leap – leak*	neap - neak	thud – thug	gud - gug
leg – lead	cheg - ched	tip* – tick	tep - teck
lit* – lick*	jit - jick	track* – trap*	brack - brap
like* – light	jike - jight	trip* – trick*	plip - plick
lot – lock*	bot - bock	wait* – wake*	nait - nake
loop* – loot*	doop - doot†	week – wheat	keek - keat

Appendix B—TRACE Simulations

The jTRACE simulator (Strauss et al., in press) was used to evaluate the extent to which TRACE (McClelland & Elman, 1986) is compatible with our current data. We did not consider it necessary to provide a full exploration of the parameter space of TRACE, nor did we attempt to generate detailed predictions relating to the dual-task aspects of the current experiments. Instead, we investigated TRACE more broadly, examining whether it could exhibit the kind of settling behavior at the phoneme level that our data imply.

Our simulations made use of the standard TRACE parameters except where stated otherwise (see McClelland & Elman, p. 22), and the "biglex901" lexicon of 901 words supplied with jTRACE. One word from our materials ("bog") was added to this lexicon to increase the number of stimuli available for the lexical status simulation. The network's predicted response in a phoneme categorisation task was measured using the jTRACE forced-choice option, which used the Luce (1959) choice rule to convert activation levels into response probabilities given a restricted set of two legal responses (the phoneme response options in the experiments).

Subcategorical Mismatch

We investigated the impact of subcategorical mismatch using a single word/non-word pair of stimuli (*creep* and *treep*), based on the assumption that lexical feedback effects would be relatively unimportant (indeed the difference between the word and the non-word cases was slight, and turning off feedback made very little difference). For these stimuli, a 9-point continuum was created for the final segment from /p/ to /k/, with the stimuli near the /p/ end (points 2, 3 and 4) intended to represent stimuli containing mismatching coarticulatory cues to the final phoneme.

The Luce choice rule makes use of a constant, *k*, as follows:

$$p(resp_i) = \frac{e^{ka_i}}{\sum_j e^{ka_j}}$$

where a_i is the activation of phoneme *i* in the set of *j* phonemes under consideration (in this case, two). Whereas most simulations using TRACE have used the same standard parameters, the value of *k* has varied. For example, Dahan, Magnuson and Tanenhaus (2001) and Allopenna, Magnuson & Tanenhaus (1998) used $k = 7$, whereas Mirman et al. (2006) used $k = 15$. Figure A1 shows the effect of mismatch over time, averaged over the word and non-word context, for both of these *k* values. In these figures, the heavy line marked 1 plots the response function for the no mismatch condition, and the other three lines plot the same function for varying degrees of mismatch. The critical phoneme is centred on cycle 30, and a difference in response probability between the different mismatch conditions is established soon after that. Responses in the long SOA condition should reflect the state of the network soon after this, at a point where mismatch has a strong effect on response speed. Later on in processing the effects of mismatch are reduced or eliminated (particularly for $k = 15$), suggesting that responses in the short SOA condition, where the network has had 300-400 ms extra processing time, should show a weakened or eliminated cost of subcategorical mismatch. Thus, the behavior of TRACE shows a reasonably good fit with the pattern of results found Experiments 1 and 2.

Lexical status effects

The simulation of lexical status effects made use of the 8 sets of 4 stimuli (i.e., 16 words and 16 non-words) that could be fully transcribed into TRACE's limited phoneme set (see Appendix A). These were a mixture of 3 and 4 phoneme words, and so the target phoneme response probabilities were temporally aligned with respect to the centre of the target phoneme. The response functions were plotted using the Luce choice rule with $k = 7$. The upper graph of Figure A2 shows the behavior of TRACE using standard parameters. Essentially, the pattern of responses was no different to the case where word to phoneme feedback was switched off (not shown), with no discernable lexical effect on phoneme identification. This suggests that TRACE in its normal form would not predict the lexical effects found in Experiments 3 and 4 at long SOA. The lower graph in Figure A2 plots the same function when the word to phoneme feedback setting was doubled from .03 to .06. In this case, the lexical effect was discernable, although still quite small compared to the subcategorical mismatch effects in Figure A1. There was also little evidence of a differential effect of lexical status between early and late stages of processing (although increasing the parameter k would have reduced the lexical effect at the longer cycle times). Further increases in the word-to-phoneme feedback level showed stronger lexical status effects, but at a cost of generally poorer discrimination of target phonemes (the response function peaked at $p = .84$ and tailed off at later cycles).

Table 1
Response Times and Error Rates across all Conditions for Experiment 1.

Response time and error rate	SOA 100 ms	SOA 200 ms	SOA 1000 ms
Task 1 Error rates (proportion)			
Matching cues	3.63 (0.94)	2.35 (0.74)	2.56 (0.81)
Mismatching cues	2.35 (0.67)	2.56 (0.81)	4.06 (1.17)
Task 2 Error rates (proportion)			
Matching cues	3.42 (1.04)	4.27 (0.99)	5.77 (1.16)
Mismatching cues	5.56 (1.17)	4.91 (1.34)	5.77 (1.16)
Task 1 response times (ms)			
Matching cues	650 (30)	674 (30)	629 (31)
Mismatching cues	677 (33)	681 (33)	631 (31)
Task 2 response times (ms)			
Matching cues	1007 (37)	926 (35)	556 (19)
Mismatching cues	1025 (33)	937 (36)	641 (17)

Note. Standard error values are reported in brackets.

Table 2
Response Times and Error Rates across all Conditions for Experiment 2.

Response time and error rate	SOA 100 ms	SOA 200 ms	SOA 1000 ms
Task 1 Error rates (proportion)			
Matching cues	2.99 (1.07)	5.13 (1.26)	3.42 (0.89)
Mismatching cues	1.92 (0.64)	3.63 (0.94)	2.99 (0.77)
Task 2 Error rates (proportion)			
Matching cues	6.62 (1.34)	5.34 (1.14)	7.27 (1.47)
Mismatching cues	5.98 (1.27)	5.56 (0.95)	8.97 (1.69)
Task 1 response times (ms)			
Matching cues	666 (33)	669 (33)	630 (44)
Mismatching cues	639 (29)	671 (34)	643 (39)
Task 2 response times (ms)			
Matching cues	971 (33)	857 (39)	534 (22)
Mismatching cues	940 (34)	873 (35)	580 (19)

Note. Standard error values are reported in brackets.

Table 3
Response Times and Error Rates across all Conditions for Experiment 3.

Response time and error rate	SOA		
	100 ms	200 ms	1000 ms
Error rates (proportion)			
Word	0.05 (.009)	0.07 (.010)	0.04 (.008)
Non-word	0.07 (.016)	0.06 (.012)	0.07 (.010)
Task 1 response times (ms)			
Combined word and non-word means	710 (36)	701 (36)	678 (44)
Word	713 (39)	685 (35)	682 (44)
Non-word	706 (35)	716 (38)	674 (46)
Task 2 response times (ms)			
Combined word and non-word means	986 (34)	875 (32)	551 (22)
Word	980 (36)	854 (33)	509 (19)
Non-word	993 (34)	896 (35)	593 (25)

Note. Standard error values are reported in brackets.

Table 4
Response Times and Error Rates across all Conditions for Experiment 4.

Response time and error rate	SOA 100 ms	200 ms	1000 ms
Error rates (proportion)			
Biased			
Word	0.06 (.015)	0.03 (.008)	0.06 (.014)
Non-word	0.05 (.010)	0.06 (.013)	0.06 (.012)
Non-biased			
Word	0.09 (.020)	0.09 (.016)	0.12 (.020)
Non-word	0.11 (.021)	0.10 (.016)	0.14 (.017)
Task 1 response times (ms)			
Combined across lexical status and bias	660 (23)	651 (20)	625 (30)
Combined across lexical status			
Biased	649 (23)	637 (19)	628 (30)
Non-biased	672 (25)	666 (22)	623 (30)
Combined across response bias			
Word	658 (23)	648 (21)	620 (29)
Non-word	663 (24)	654 (20)	630 (32)
Across all conditions			
Biased			
Word	648 (24)	625 (22)	628 (29)
Non-word	649 (23)	648 (20)	627 (34)
Non-biased			
Word	668 (26)	671 (24)	612 (31)
Non-word	677 (28)	661 (23)	633 (32)
Task 2 response times (ms)			
Combined across lexical status and bias	946 (27)	824 (25)	540 (18)
Combined across lexical status			
Biased	911 (28)	788 (27)	506 (19)
Non-biased	981 (28)	860 (25)	574 (19)
Combined across response bias			
Word	940 (27)	810 (25)	517 (17)
Non-word	952 (29)	837 (27)	563 (21)
Across all conditions			
Biased			
Word	905 (29)	766 (28)	485 (17)
Non-word	916 (30)	810 (29)	526 (22)
Non-biased			
Word	974 (29)	855 (26)	549 (19)
Non-word	988 (32)	865 (27)	599 (22)

Note. Standard error values are reported in brackets.

Figure 1. Schematic illustration of the locus-of-slack methodology. Effects of task difficulty at short SOA (panel A) and long SOA (panel B). Tasks are divided into three components: pre-bottleneck (Pre BN), bottleneck (BN; shaded) and post-bottleneck (Post BN). For both SOAs, the effects of task difficulty are illustrated for pre-bottleneck and at-bottleneck loci. Manipulation of task difficulty at the post-bottleneck stage is equivalent to a bottleneck locus. Dashed lines mark SOAs and RTs; the thicker dashed line marks the point in time at which bottleneck processes become available for Task 2. Gaps between pre-bottleneck and bottleneck boxes indicate slack time. See text for further explanation.

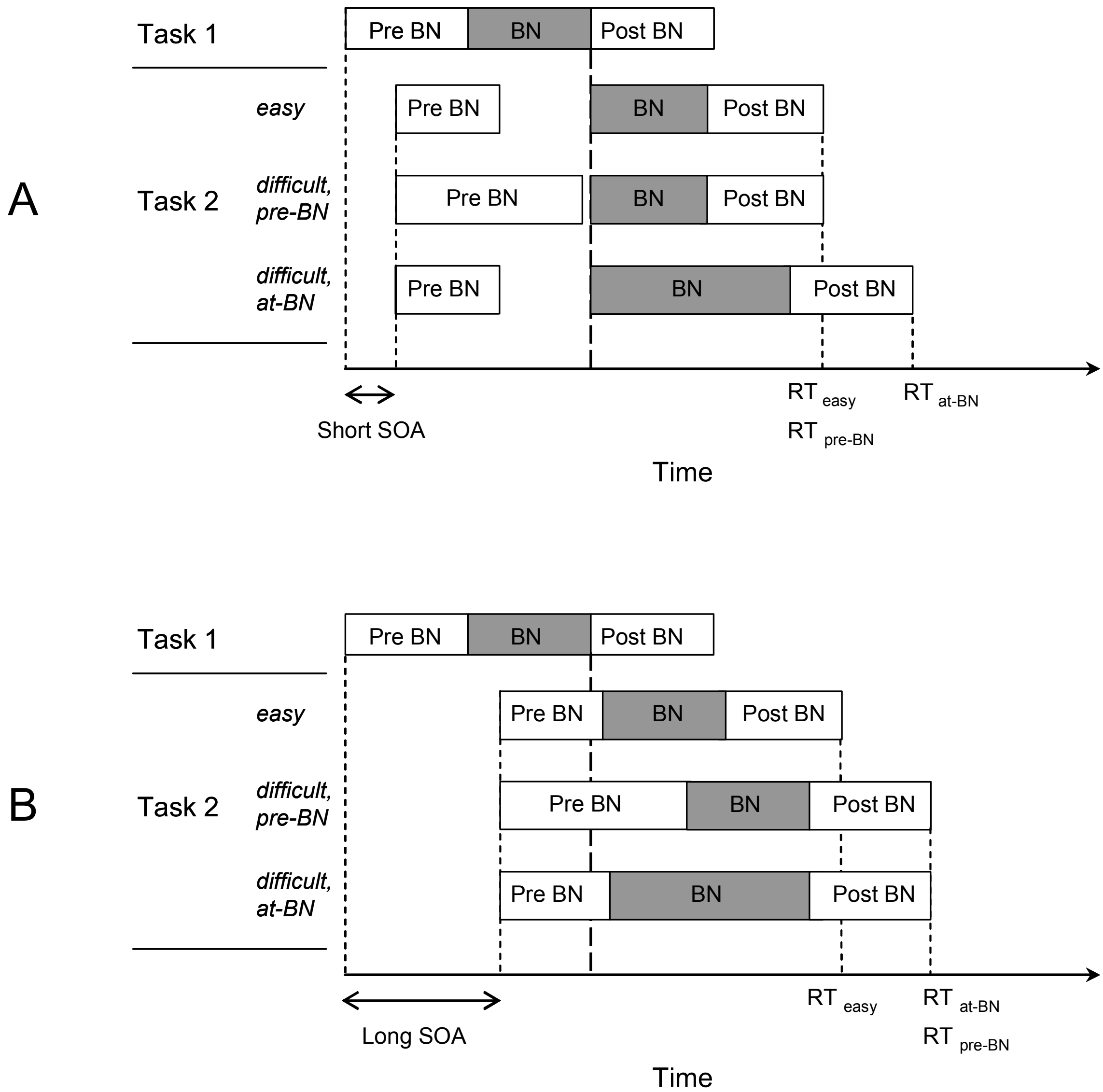


Figure 2. Response times in Experiment 1 across SOA for all conditions. RT1: response time to Task 1; RT2: response time to Task 2. Error bars represent standard error.

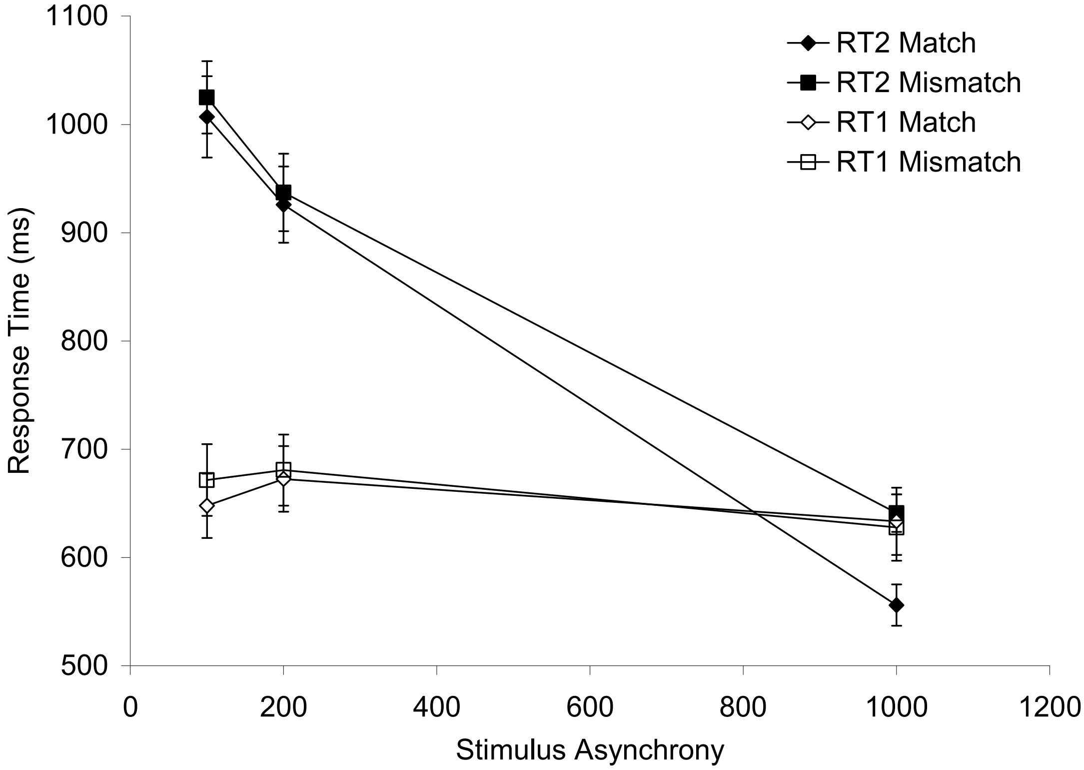


Figure 3. Response times in Experiment 2 across SOA for all conditions. RT1: response time to Task 1; RT2: response time to Task 2. Error bars represent standard error.

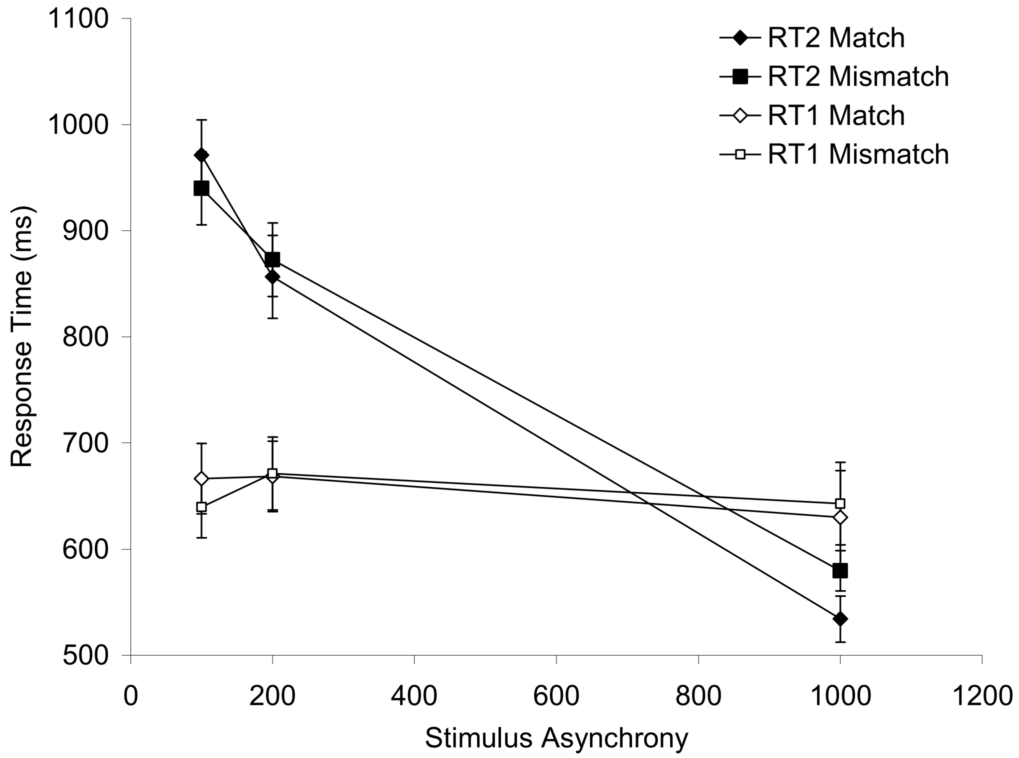


Figure 4. Response times in Experiment 3 across SOA for all conditions. RT1: response time to Task 1; RT2: response time to Task 2. Error bars represent standard error.

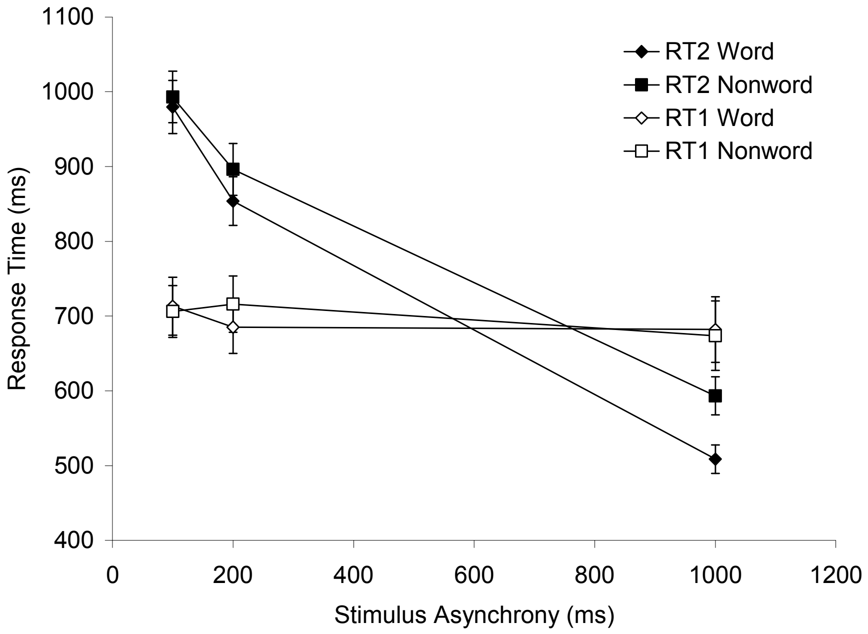
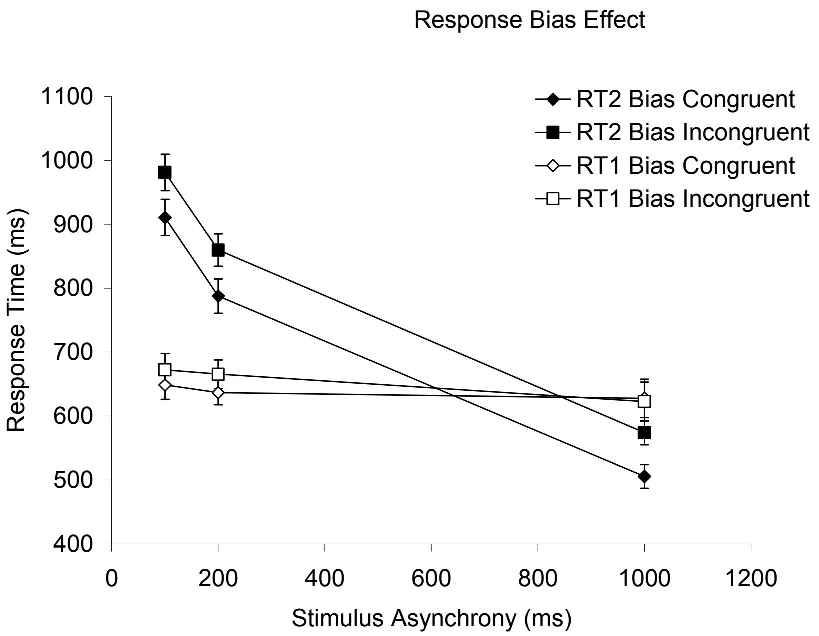


Figure 5. Response times in Experiment 4 across SOA. The upper graph illustrates the manipulation of response bias collapsing across lexical status. The lower graph illustrates the manipulation of lexical status collapsing across response bias. RT1: response time to Task 1; RT2: response time to Task 2. Error bars represent standard error.



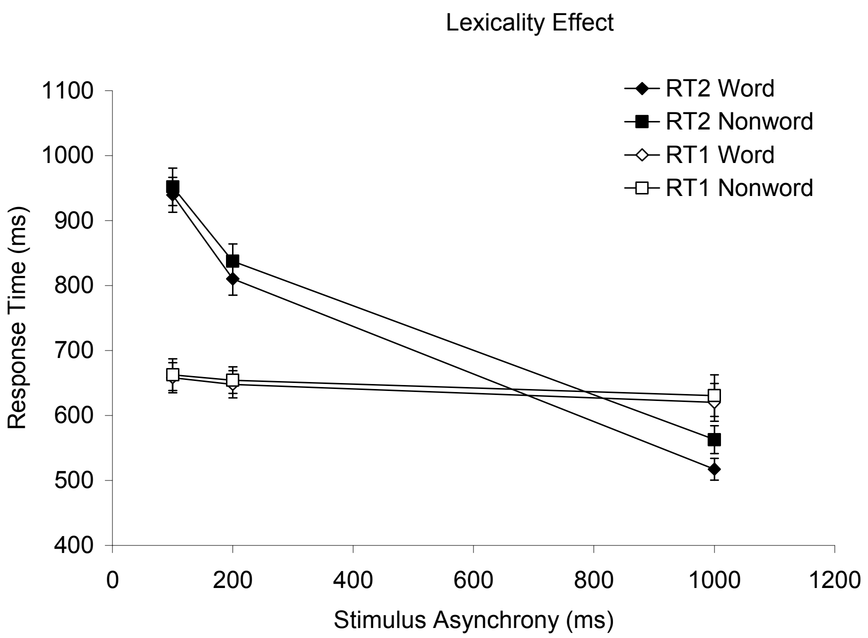


Figure A1. Simulation of subcategorical mismatch effects in TRACE. The figure plots the probability of a correct response in a phoneme categorization task as a function of time averaged across the two sample stimuli. The four curves plot the data for the control condition with a normal final consonant (1) and three levels of mismatch on the final consonant (2-4). The upper graph makes use of a Luce choice rule with $k = 7$, and the lower graph uses $k = 15$.

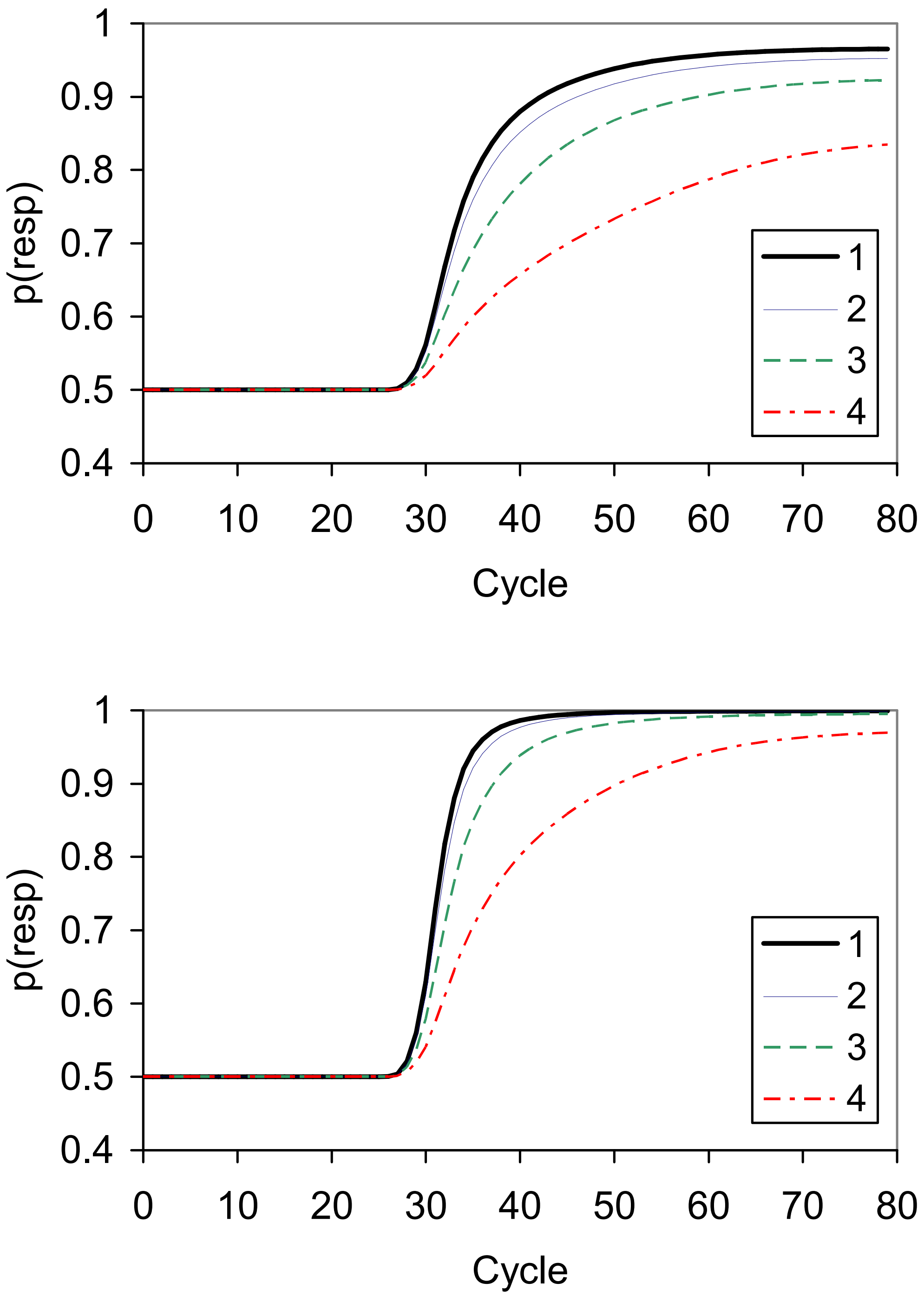


Figure A2. Simulation of lexical status effects in TRACE. The figure plots the probability of a correct response in a phoneme categorization task as a function of time averaged across 16 sample stimuli. The solid line plots the data for the non-word condition, and the dashed line plots the data for the corresponding words. The upper graph plots the data using TRACE's standard parameters, whereas for the lower graph word phoneme feedback has an higher weighting (increased from .03 to .06).