BEHAVIORAL AND BRAIN SCIENCES (2000) **23,** 299–370
Printed in the United States of America

Merging information in speech recognition: Feedback is never necessary

Dennis Norris
Medical Research Council Cognition and Brain Sciences Unit, Cambridge, CB2 2EF, United Kingdom
dennis.norris@.mrc-cbu.cam.ac.uk
www.mrc-cbu.cam.ac.uk

James M. McQueen
Max-Planck-Institute for Psycholinguistics, 6525 XD Nijmegen, The Netherlands
james.mcqueen@mpi.nl www.mpi.nl

Anne Cutler
Max-Planck-Institute for Psycholinguistics, 6525 XD Nijmegen, The Netherlands
anne.cutler@mpi.nl www.mpi.nl

Abstract: Top-down feedback does not benefit speech recognition; on the contrary, it can hinder it. No experimental data imply that feedback loops are required for speech recognition. Feedback is accordingly unnecessary and spoken word recognition is modular. To defend this thesis, we analyse lexical involvement in phonemic decision making. TRACE (McClelland & Elman 1986), a model with feedback from the lexicon to prelexical processes, is unable to account for all the available data on phonemic decision making. The modular Race model (Cutler & Norris 1979) is likewise challenged by some recent results, however. We therefore present a new modular model of phonemic decision making, the Merge model. In Merge, information flows from prelexical processes to the lexicon without feedback. Because phonemic decisions are based on the merging of prelexical and lexical information, Merge correctly predicts lexical involvement in phonemic decisions in both words and nonwords. Computer simulations show how Merge is able to account for the data through a process of competition between lexical hypotheses. We discuss the issue of feedback in other areas of language processing and conclude that modular models are particularly well suited to the problems and constraints of speech recognition.

Keywords: computational modeling; feedback; lexical processing; modularity; phonemic decisions; reading; speech recognition; word recognition

1. Introduction

Psychological processing involves converting information from one form to another. In speech recognition – the focus of this target article – sounds uttered by a speaker are converted to a sequence of words recognized by a listener. The logic of the process requires information to flow in one direction: from sounds to words. This direction of information flow is unavoidable and necessary for a speech recognition model to function.

Our target article addresses the question of whether output from word recognition is fed back to earlier stages of processing, such as acoustic or phonemic analysis. Such feedback entails information flow in the opposite direction – from words to sounds. Information flow from word processing to these earlier stages is not required by the logic of speech recognition and cannot replace the necessary flow of information from sounds to words. Thus it could only be included in models of speech recognition as an additional component.

The principle of Occam's razor instructs theorists never to multiply entities unnecessarily. Applied to the design of processing models, this constraint excludes any feature that

DENNIS NORRIS is a member of the senior scientific staff of the Medical Research Council Cognition and Brain Sciences Unit, Cambridge, United Kingdom.
JAMES MCQUEEN is a member of the scientific staff of the Max-Planck-Institute for Psycholinguistics, Nijmegen, The Netherlands.
ANNE CUTLER is director (language comprehension) of the Max-Planck-Institute for Psycholinguistics and professor of comparative psycholinguistics at the University of Nijmegen.
All three work on the comprehension of spoken language, in particular spoken-word recognition. Between them they have written over 220 publications in this area, of which 49 are joint publications of at least two of them; this is the eleventh paper on which all three have collaborated.

© 2000 Cambridge University Press 0140-525X/00 $12.50

https://doi.org/10.1017/S0140525X00303247 Published online by Cambridge University Press

is not absolutely necessary in accounting for the observed data. We argue that models without the feedback we have just described can account for all the known data on speech recognition. Models with feedback from words to earlier processes therefore violate Occam's razor.

Nevertheless, many theorists have proposed such models; whether there is feedback from word recognition to earlier acoustic and phonemic processing has even become one of the central debates in the psychology of speech recognition. We consider the arguments that have been proposed in support of this feedback and show that they are ill-founded. We further examine the relevant empirical data – studies of how listeners make judgments about speech sounds – and show that the evidence from these studies is inconsistent with feedback and strongly supportive of models without this feature.

We argue, however, that no existing model, with or without feedback, adequately explains the available data. Therefore we also propose a new model of how listeners make judgments about speech sounds, demonstrating how information from word processing and from earlier processes can be merged in making such judgments without feedback from the later level to the earlier.

In the psychological literature on speech recognition, models without feedback are generally referred to as "autonomous," and models with feedback are termed "interactive." In autonomous models, each stage proceeds independent of the results of subsequent processing. In interactive models, feedback between any two stages makes them interdependent. (Strictly speaking, it is not models in their entirety, but stages of each model, to which the terms should be applied; Norris 1993. A model may have a mixture of autonomous and interactive stages.) Models with only autonomous stages have only feedforward flow of information, and this is also referred to as "bottom-up" processing, while feedback in interactive models is also referred to as "top-down" processing. Note that the only type of feedback at issue in the speech recognition literature, and in the present article, is feedback from later stages, which actually alters the way in which an earlier stage processes its input. The specific question is: Does information resulting from word (lexical) processing feed back to alter the immediate operation of prelexical processes (all the processes that intervene between the reception of the input and contact with lexical representations)?

2. Models of phonemic decision making

The debate about feedback in psychological models of speech recognition has centred on evidence from experimental tasks in which listeners are required to make phonemic decisions (judgments about categories of speech sounds). Such tasks include (1) phoneme monitoring, in which listeners detect a specified target phoneme in spoken input (e.g., respond as soon as they hear the sound /b/ in "brain"; see Connine & Titone 1996, for a review); (2) phonetic categorization, in which listeners assign spoken input to phoneme categories (see McQueen 1996, for a review); and (3) phoneme restoration, in which listeners hear speech input in which portions of the signal corresponding to individual phonemes have been masked or replaced by noise (see Samuel 1996a, for a review).

This section describes two example models in detail, to illustrate the structural differences we described above. The two models we have chosen represent the extreme positions of the feedback debate in this area. They are the interactive theory TRACE (McClelland & Elman 1986) and the autonomous Race model (Cutler & Norris 1979).

2.1. TRACE

The evidence on the relationship between lexical and prelexical processes seems at first glance to support the case for feedback. Many studies show convincingly that there are lexical influences on tasks involving phonemic decisions or phoneme identification.

In a wide range of tasks, phoneme identification is influenced by lexical context. For example, phoneme monitoring is faster in words than in nonwords (Cutler et al. 1987; Rubin et al. 1976). In sentences, phoneme monitoring is faster in words that are more predictable (Foss & Blank 1980; Mehler & Seguí 1987; Morton & Long 1976). Lexical biases are also observed in phonetic categorization tasks. Ganong (1980) generated ambiguous phonemes on a continuum between a word and a nonword (e.g., *type-dype*), and found that subjects are biased towards classifying phonemes in the middle of this range so as to be consistent with a word (*type*) rather than a nonword (*dype*). In phoneme restoration, listeners' ability to determine whether phonemes are replaced by noise or simply have noise added is worse in words than in nonwords (Samuel 1981a; 1987; 1996a).

An apparently simple and appealing explanation of these results is that lexical influences come about as a direct result of lexical processes exerting top-down control over a prior process of phonemic analysis. This is exactly what happens in the TRACE model of McClelland and Elman (1986). TRACE has three levels of processing. Activation spreads from the feature level to the phoneme level, and from there to the word level. In addition, activation of each word node feeds back to its constituent phonemes in the phoneme layer. (The relationship between lexical and phonemic processing in TRACE is thus directly analogous to the relationship between word and letter processing in Rumelhart and McClelland's [1982] Interactive Activation Model [IAM] of reading.)

Because of the top-down connections in TRACE, phonemes in words are part of a feedback loop that increases their activation faster than that of phonemes in nonwords. TRACE can thus readily account for the finding that phoneme monitoring responses to targets in words tend to be faster than to targets in nonwords. Phonemes in nonwords will also receive some feedback from words that they partially activate. Therefore, even phonemes in word-like nonwords will tend to be activated more quickly than phonemes in nonwords that are not similar to words. TRACE explains lexical effects in phonetic categorization in the same way. When listeners are presented with input containing ambiguous phonemes, top-down activation from the lexicon will act to bias the interpretation of the ambiguous phoneme so that it is consistent with a word rather than with a nonword, exactly as observed by Ganong (1980). Likewise, the top-down connections in TRACE can explain why phonemic restoration is more likely in words than in nonwords (Samuel 1996a).

An important feature of TRACE is that it is what Cutler et al. (1987) have described as a *single-outlet* model. The

https://doi.org/10.1017/S0140525X00303247 Published online by Cambridge University Press

only way TRACE can make a phoneme identification response is by read-out from a phoneme node in the phoneme layer. One immediate consequence of this single-outlet architecture is that when presented with a mispronunciation, TRACE is unable to identify the nature of this mispronunciation. Although a mispronounced word would activate the wrong set of input phonemes, there is no way that the error can be detected. The mispronunciation will reduce the overall activation of the word node, but the system will be unable to tell at the lexical level which phoneme was mispronounced because it has no independent representation of the expected phonological form against which the input can be compared. Because top-down feedback will act to correct the errorful information at the phoneme level, the only representation of the input phonology that is available in the model will also be removed. Indeed, a similar reduction in activation of a word's lexical node coupled with top-down activation of the phonemes corresponding to the word could arise from unclear articulation or noise in the input. Somewhat ironically, the feature that thus allows TRACE to fill in missing phoneme information is precisely what prevents it from detecting mispronunciations. In contrast, the Race model, to which we now turn, does have independent lexical representations of the phonological forms of words. These representations make it possible to detect mispronunciations, freeing the model from the need for feedback.

2.2. The Race model

The findings described above are consistent with the idea of interaction between lexical and prelexical processes. But although the assumption of interaction fits with many people's intuitions about the nature and complexity of the speech recognition process, it is certainly not forced by the data. In itself, the fact that lexical information can affect the speed of, say, phoneme detection is neutral with respect to the issue of whether lexical and prelexical processes interact; it is simply evidence that both lexical and phonemic information can influence a response. Information from two separate sources can influence a response without the processes delivering that information necessarily influencing each other.

This is exactly what happens in the Race model of Cutler and Norris (1979), and, as will be discussed later, in other models without top-down feedback. According to the Race model, there are two sources of information which can be used to identify phonemes: Identification can occur via a prelexical analysis of the input or, in the case of words, phonemic information can be read from the lexical entry. Thus, in contrast to the *single-outlet* assumption of TRACE, the Race model is a *multiple-outlet* model.

Responses in phoneme monitoring are the result of a race between these two processes. As in all first-past-the-post race models, the response is determined solely by the result of the first route to produce an output. The mean winning time of a race between two processes with overlapping distributions of completion times will always be faster than the mean of either process alone. So phoneme monitoring will be faster for targets in words than for targets in nonwords because responses made to targets in words benefit from the possibility that they will sometimes be made on the basis of the outcome of the lexical process, whereas for targets in nonwords, responses must always depend on the prelexical process. The Race model offers a similar explanation for the lexical bias observed in phonetic categorization. An ambiguous phoneme in a *type-dype* continuum, for example, will sometimes be identified by the lexical route and sometimes by the phonemic route. The contribution of the lexical route will lead to an overall bias to produce a lexically consistent response. Hence the repeated simple demonstration of lexical involvement in a range of phonemic judgment tasks does not distinguish between interactive and autonomous models; both TRACE and the Race model can account for all of the basic effects.

3. Is feedback beneficial?

Discussions of bottom-up versus interactive models frequently seem to draw on an implicit assumption that top-down interaction will help performance, but does it? In this section we examine this assumption, and conclude that the feedback in models we discuss is not beneficial.

3.1. Can feedback improve word recognition?

In models like TRACE (or Morton's 1969 logogen model), the main effect of feedback is simply to alter the tendency of the model to emit particular responses. The top-down flow of information does not help these models perform lexical processing more accurately.

Consider first how feedback from lexical to prelexical processes might facilitate word recognition. The best performance that could possibly be expected from a word recognition system is to reliably identify the word whose lexical representation best matches the input representation. This may sound trivially obvious, but it highlights the fact that a recognition system that simply matched the perceptual input against each lexical entry, and then selected the entry with the best fit, would provide an optimal means of performing isolated word recognition (independent of any higher-level contextual constraints), limited only by the accuracy of the representations. Adding activation feedback from lexical nodes to the input nodes (whether phonemic or featural) could not possibly improve recognition accuracy at the lexical level.

To benefit word recognition, feedback would have to enable the system to improve on the initial perceptual representation of the input. A better representation would in turn improve the accuracy of the matching process. For example, having isolated a set of candidate words, the lexical level might instruct the prelexical level to utilise a specialised, more accurate set of phoneme detectors rather than general-purpose detectors used in the first pass. But note that the plausibility of such a theory depends on the questionable assumption that the system performs a suboptimal prelexical analysis on the first pass. If it does the best it can the first time around, there is no need for feedback.

An interactive model that improved input representations would be analogous to the verification models that have been proposed for visual word recognition (e.g., Becker 1980; Paap et al. 1982). However, to our knowledge, no such model exists for spoken-word recognition. All models with feedback involve flow of activation from lexical to phoneme nodes, as in TRACE. In this kind of model, which we will call interactive bias models, interaction does not

https://doi.org/10.1017/S0140525X00303247 Published online by Cambridge University Press

have a general beneficial effect on word recognition, although it can influence phoneme recognition. This is confirmed by Frauenfelder and Peeters's (1998) TRACE simulations which showed that the performance of TRACE does not get worse when the top-down connections are removed (i.e., approximately as many words were recognized better after the connections were removed as were recognized less well).

3.2. Can feedback improve phoneme recognition?

In general, although interactive bias cannot assist word recognition, it can help phoneme recognition, especially when the input consists entirely of words. If the /n/ in the middle of *phoneme* cannot be distinguished clearly by the phoneme level alone, interactive bias from the lexical level can boost the activation for the /n/ and make it more likely that the phoneme will be identified. Of course, if the input were the nonword *phomeme* instead, the biasing effect would impair performance rather than help it, in that the mispronunciation would be overlooked. That is, interactive bias models run the risk of hallucinating. Particularly when the input is degraded, the information in the speech input will tend to be discarded and phonemic decisions may then be based mainly on lexical knowledge. This is because top-down activation can act to distort the prelexical representation of the speech input (Massaro 1989a). If there is strong top-down feedback, the evidence that there was a medial /m/ in *phomeme* may be lost as the lexicon causes the phoneme level to settle on a medial /n/ instead. In fact, repeated empirical tests have shown that mispronunciations are not overlooked, but have a measurable adverse effect on phoneme detection performance (Gaskell & Marslen-Wilson 1998; Koster 1987; Otake et al. 1996).

It is important to note that the biasing activation makes the lexically consistent phoneme more likely as a response. One might argue that this is a useful property of the model designed to deal with normal speech where the input does, of course, consist almost exclusively of words in the listener's lexicon. However, given that the interaction does not help word recognition, it is not clear what advantage is to be gained by making the prelexical representations concur with decisions already made at the lexical level. Once a decision has been reached at the word level, there is no obvious reason why the representations that served as input to the word level should then be modified. The only reason for feedback would be to improve explicit phoneme identification. But even this reason has little force because autonomous models offer an alternative way for lexical information to influence phonemic decisions, one that does not suffer from the disadvantages caused by introducing a feedback loop into the recognition process.

In interactive bias models, therefore, lexical information can sometimes improve phoneme identification and sometimes impair it. In verification models, on the other hand, lexical information should always improve phoneme identification in words. As Samuel (1996a) has demonstrated in the phoneme restoration task, subjects' ability to detect whether or not phonemes in noise are present is poorer in words than in nonwords. Moreover, this lexical disadvantage increases throughout the word. That is, in this task, lexical information introduces a bias that reduces the performance of the phoneme level. Samuel (1996a) points out that this finding is exactly what TRACE would predict. However, it is also exactly what the Race model would predict. According to the Race model, if the lexical route wins when the phoneme has been replaced by noise, this will produce an error, but the lexical route should never win for nonwords.

Despite the fact that this result is exactly what would be expected on the basis of the Race model, Samuel (1996a) interprets it as evidence against the Race model, arguing that lexical effects in the Race model can only be facilitatory. He reasons that because the lexical effect in the phoneme restoration study is the reduction of the discriminability of phonemes and noise, the lexical effect is not facilitatory, and hence contradicts the Race model. This is incorrect, however. Because lexical information races with phonemic information, lexical effects must certainly always have a facilitatory effect on phoneme monitoring latencies to targets in words, but the race will not facilitate all aspects of phoneme perception. If the lexical route produces a result even when the phoneme has been replaced by noise, the listener will have difficulty determining whether there really was a phoneme in the input, or just noise. The lexical route facilitates identification of the underlying phoneme, but this in turn impairs the listener's ability to discriminate the phoneme from the noise.

Hence Samuel's (1996a) study does not discriminate between TRACE and the Race model, or between interactive and autonomous models in general. The real significance of these restoration results is that they appear inconsistent with more active forms of interaction, such as the one discussed above, where feedback would act to improve input representations (as in the verification model of Becker, 1980). Such models incorrectly predict that lexical influences will always increase perceptibility. This in turn suggests that if the recognition system were interactive, it would be more likely to have the characteristics of an interactive bias model than of a verification model. As we have argued, however, interaction in bias models cannot improve word recognition and can cause misidentifications at the phonemic level.

3.3. Can feedback improve the sensitivity of phoneme identification?

We have argued that, in general, feedback of lexical activation can only bias phoneme identification, without actually improving sensitivity.[1] Indeed, Samuel's (1996b) phoneme restoration results were consistent with both TRACE and the Race model in showing that lexical information biased subjects towards lexically consistent responses instead of improving their sensitivity in discriminating between phonemes and noise. Nevertheless, if one could devise an experiment in which lexical information were shown to improve listeners' sensitivity in phoneme identification, this would prove particularly problematic for autonomous models. The standard way to investigate sensitivity and bias in perceptual experiments is to use Signal Detection Theory (SDT).

To many authors, SDT seems to offer a simple technique for distinguishing between interactive and autonomous theories. The decision to use SDT has generally been based on the idea that changes in sensitivity, as measured by d′, reflect changes in perceptual processes (cf. Farah 1989). For example, if context provided by lexical information influences sensitivity of phoneme identification, this is taken

https://doi.org/10.1017/S0140525X00303247 Published online by Cambridge University Press

as evidence that the contextual factor is interacting with the perceptual processes of phoneme identification. Although this is an appealing notion, applying SDT here is far from straightforward. In general, SDT studies of interaction have either used inappropriate analyses or drawn invalid conclusions from them.

One of the central pitfalls of applying SDT analyses is evident in work on the influence of context on visual word recognition. Rhodes et al. (1993) used SDT to study context effects in lexical identification. They applied standard unidimensional signal detection theory to data from visual word recognition and reported that semantic priming did indeed alter d′. From this they concluded that context was influencing the perceptual analysis of the words, in violation of modularity (Fodor 1983; 1985). Norris (1995), however, pointed out that the standard unidimensional SDT model was inappropriate under these circumstances because its assumptions do not map onto those of any current or even any plausible model of visual word recognition. Norris also showed that the unidimensional measure of sensitivity (d′) cannot even account for the basic effect of semantic priming and that a multidimensional version of SDT, embodying more plausible assumptions about visual word recognition (similar to Morton's 1969 logogen model or Norris's 1986 checking model), could account for the same data purely in terms of response bias, with no need for context to influence earlier perceptual processes. The lesson here is that the choice of SDT model must be guided by plausible psychological task models. If the underlying assumptions of SDT are not satisfied in the psychological model, the results of the SDT analysis will be meaningless.

Confusion over the interpretation of the results of SDT analysis can lead authors to make claims that are not justified by the data even when the technical application of SDT seems appropriate. In a study of assimilation effects in speech perception (e.g., *freight bearer* may be produced as *frayp bearer* but heard as *freight bearer*), Gaskell and Marslen-Wilson (1998) found that subjects were less able to perceive the assimilation in words than nonwords. From this they concluded that the effect was "perceptual"; their line of reasoning seems to be similar to Farah's argument that sensitivity effects must be due to perceptual processes. However, Gaskell and Marslen-Wilson (1998) also assume that the sensitivity effects tell them about the locus of the perceptual effect, namely, that it was not based on the output of lexical processing. Conclusions about the locus of the effect cannot be supported by this kind of data. The SDT analysis simply informs us that the discrimination performance of the system under observation is worse in words than in nonwords. Such data are perfectly consistent with the idea that the change in sensitivity arises after lexical processing. For example, if lexically derived phonemic information always determined responses when it was available, then detection of assimilation would always be much worse in words than nonwords even though the locus of the effect was at a late stage, when lexical and prelexical information were combined. That is, a late bias to respond lexically will be manifest as a decrease in sensitivity when the overall performance of the system is subject to SDT analysis.

A similar problem arises in a phonetic categorization study. Pitt (1995) used SDT to analyze the influence of lexical information on categorization. He concluded that lexical information influenced the perceptual analysis of phonemes and that his data supported interactive theories over autonomous ones. Pitt showed a lexical effect on the categorization of the phonemes /g/ and /k/ in the continua *gift-kift* and *giss-kiss* . He then transformed his data into d′ values by converting each subject's proportion of /g/ responses for each step of both continua into z scores and then calculating d′ by subtracting adjacent z scores. When plotted in this fashion the two continua differed in the location of their peak d′ score. Pitt concluded that because lexical information shifted the d' measure it must have been having an effect on phoneme perception, and that this was evidence of interaction.

This shift in the peak of the d′ function, however, is simply a direct reflection of the shift in the conventional identification function. Lexical information has not increased the observer's overall ability to discriminate between the two phonemes; it has just shifted the category boundary. As is usual in categorical perception studies, the category boundary corresponds to the peak in the discrimination function and the maximum slope of the identification function. Lexical information has not enabled the listener to extract more information from the signal; it has just shifted the point of maximum sensitivity. Lexical information has indeed altered the pattern of sensitivity, but it is the position – not the amount of sensitivity – that has been changed. Exactly this kind of lexically induced boundary shift can emerge from an autonomous bias model (as Pitt in fact admits). Indeed, Massaro and Oden (1995) showed that the autonomous Fuzzy Logical Model of Perception (FLMP; Massaro 1987; 1989b; 1998; Oden & Massaro 1978) could fit Pitt's data very accurately.

Pitt (1995) makes much of the fact that the change in identification functions induced by lexical information is different from that induced by a monetary payoff designed to bias subjects towards one response over the other. Following Connine and Clifton (1987), he argues that these differences could be evidence that the lexical effects are due to feedback from the lexicon. However, there are two quite distinct ways in which bias can influence phoneme identification. Monetary payoff and lexical information appear to operate in these different ways, but neither requires top-down feedback.

Monetary payoff tends to shift the functions vertically, whereas lexical bias produces a horizontal shift. The simple interpretation of the vertical shift is that subjects have a general bias to respond with the financially favored phoneme on some proportion of trials. Massaro and Cowan (1993) call this "decision bias" and distinguish it from "belief bias"; an example of the pattern can be clearly seen in the fast RTs in Pitt's Figure 6 (1995, p. 1046). The lexical shift, on the other hand, reflects a bias towards lexically favored responses on the basis of less bottom-up support than lexically unfavored responses. This leads to the horizontal shift of the boundary illustrated in Pitt's Figures 1 and 2 (1995, pp. 1040, 1041). These are two different patterns of data, but both are the result of biases.

To begin to make a case against autonomous models on the issue of sensitivity, one would need to demonstrate that lexical information could actually improve phoneme discriminability. That is, lexical information in a word-nonword continuum should result in improved discriminability (greater accuracy in paired-alternative discrimination or an increase in the number of Just Noticeable Differences [JNDs] between the two ends of the continuum) relative

https://doi.org/10.1017/S0140525X00303247 Published online by Cambridge University Press

to a nonword-nonword continuum. But no study has yet shown evidence that lexical information can produce any increase in the sensitivity of phoneme discrimination. So, although SDT may seem to offer a simple method of distinguishing between autonomous and interactive models, its use is far from straightforward and there are no SDT studies to date that allow us to distinguish between the models.

3.4. Can feedback improve the speed of recognition?

Even if lexical feedback cannot improve the accuracy of recognition, it might help speed the recognition process. But consider what would happen in a model like TRACE if we optimized all connections to perform both phoneme and word recognition as quickly and accurately as possible in the absence of top-down feedback. Feedback could definitely speed recognition of both words and phonemes (in exactly the same way as context or frequency could speed word recognition in the logogen model), but the effect of this speed-up would be for the system to respond on the basis of less perceptual information than before. Because feedback cannot improve the accuracy of word recognition, however, faster responses made on the basis of less perceptual information must also be less accurate. Also, following the arguments in the previous sections, it is only when the top-down information is completely reliable that there can be an increase in phoneme recognition speed without a decrease in accuracy. Furthermore, autonomous models such as FLMP (Massaro 1987) and Merge (which will be presented in sect. 5 below) can engage in a similar speed-accuracy trade-off by reducing the recognition criterion for word identification or choosing to place more emphasis on lexical than phonemic information when performing phoneme identification. Thus lexical information can speed recognition in interactive models, but no more than the same lexical information can speed recognition in bottom-up models.

3.5. How general is the case against feedback?

3.5.1. Visual word recognition. The case we have made against feedback in spoken word recognition has a direct parallel in visual word recognition. If we replace "phoneme" with "letter," then essentially the same arguments apply to reading as to speech. A review of empirical work in reading reveals a state of affairs closely analogous to that in speech: There is very solid evidence that lexical factors can influence letter recognition, but little evidence that this must be achieved via feedback. In visual word recognition research, the interaction debate has concentrated on the proper explanation of the Word Superiority Effect (WSE). With brief presentations and backward masking, letters can be more readily identified in words than in nonwords, and more readily identified in words than when presented alone (Reicher 1969). According to McClelland and Rumelhart's (1981) Interactive Activation Model (IAM), this lexical advantage can be explained by top-down feedback from lexical nodes to letter nodes. In this respect the explanation of the WSE given by the IAM is exactly the same as the explanation of lexical effects given by TRACE: Feedback from the lexical level activates letter nodes and makes letters that are consistent with words more readily identifiable. McClelland and Rumelhart's interactive account is also consistent with the results of many studies showing that letters in pronounceable nonwords are easier to identify than letters in unpronounceable nonwords (e.g., Aderman & Smith 1971) and that letters in pseudowords are easier to identify than letters alone (McClelland & Johnston 1977). Rumelhart and McClelland (1982) also found an advantage for letters in wordlike all-consonant strings like "SPCT" over nonwordlike strings such as "SLQJ." These data are consistent with interactive theories, but the WSE can also be explained without interaction. This has been clear for more than two decades (see, e.g., Massaro 1978).

Two more recent models have succeeded in capturing the crucial empirical findings on the WSE without recourse to feedback. First, the Activation-Verification Model (AVM) of Paap et al. (1982) has provision for a top-down verification process; however, its explanation of the WSE involves no feedback at all. In the AVM, visual input first activates a set of letters, which in turn activate a set of candidate words. With verification in operation, words in the candidate set are verified against the input. However, the verification process is assumed to be unable to operate in the case of the brief pattern-masked displays used to study the WSE. The AVM account of the WSE is therefore bottom-up. Under these circumstances, letter identification decisions can be made by pooling the letter identity information with information from any lexical candidates activated above a given threshold. If the total lexical information exceeds a second threshold, then letter identification decisions are made on the basis of the lexical information alone. As in the Race model, therefore, lexical effects on letter identification come about because letter information is available from the lexicon. However, in contrast to the Race model, the lexical information can be derived from more than one word candidate, and letter and word information can be pooled together. As in the Race model, the letter identity information read out from the lexicon is not fed back to the initial stage of letter identification. Because the decision process can pool lexical information from a number of candidates, the model can account for the fact that letters in pseudowords are better identified than in unpronounceable letter strings. A pseudoword is likely to activate words containing some of the same letters in the same positions. An unpronounceable letter string is unlikely to activate many words and those words are unlikely to contain the same letters in the same positions as the pseudoword.

Second, the Dual Read Out Model (DROM) of Grainger and Jacobs (1994) can also give a feedback-free account of the WSE. Architecturally, this model is similar to the IAM of Rumelhart and McClelland. Both models are IAMs with levels corresponding to features, letters, and words. The main difference is that, in the DROM, subjects can base their decisions either on letter-level activation (the only possibility in the IAM) or by reading out orthographic information from the most activated word (as opposed to a set of words in the AVM). Grainger and Jacobs examined the behaviour of their model both with and without top-down feedback. Without top-down feedback, the DROM is essentially the visual equivalent of the Race model, and there is none of the pooling of letter and lexical information that takes place in the AVM. Although Grainger and Jacobs suggest that the DROM slightly underestimates the size of the pseudoword advantage in the absence of feedback, they point out that this problem could be overcome if there were

https://doi.org/10.1017/S0140525X00303247 Published online by Cambridge University Press

an extra level of orthographic representation between the letter and word.

In contrast to the IAM, both the AVM and DROM are what Cutler et al. (1987) term "multiple-outlet models." Both lexical effects and pseudoword effects can be explained by permitting decisions to be made on the basis of either letter or lexical information, without any need for the processes delivering this evidence to interact. In the case of visual word recognition, there appears to be no sign of an imminent resolution of the interaction/autonomy debate. Although the WSE and related findings give the appearance of being evidence for feedback, as we have shown, a number of bottom-up explanations are also available. By Occam's principle, then, the bottom-up theories should be preferred to the interactive theories.

3.5.2. Syntactic-semantic interaction. We have argued that there is no need for the results of word recognition to be made known to earlier stages. Stages of phoneme or letter recognition simply do their best and pass that information on. Nothing they do later depends on whether their outputs agree with decisions reached at the lexical level. However, this relationship between levels does not necessarily hold throughout language processing; there may be cases in which feedback could indeed confer advantages. In research on syntactic processing, for example, there has been a lively debate as to whether syntactic analysis is independent of higher-level processes such as semantics, and this debate is by no means resolved. Note, however, that terminology in this area can differ from that used in word recognition. In parsing, there are theories referred to as "autonomous" that allow some feedback from semantics to syntax, and theories called "interactive" that maintain autonomous bottom-up generation of syntactic parses! As we shall show, however, examining the characteristics of the models makes it possible to compare them in the framework that we have been using throughout this discussion.

An important early model in this field was the "garden-path" theory of Frazier (1979; 1987). In Frazier's model, the parser generates a single syntactically determined parse. Initial choice of the parse is entirely unaffected by higher-level processes, and Frazier placed great emphasis on this aspect of her model: Syntax maintained autonomy from semantics. However, the model also needs to explain what will happen if this unique initial choice of parse turns out to be wrong. In a classical garden-path sentence like *Tom said that Bill will take the cleaning out yesterday,* for example, the initial parse leads to the semantically implausible interpretation that Bill is going to perform a future action in the past. Worse still, sentences like *The horse raced past the barn fell* can lead to the situation where no successful analysis at all is produced, although they are grammatical sentences of English. In this case, the syntactic processor simply fails to produce a complete output, in that *The horse raced past the barn* is assigned a full analysis, leaving no possible attachment for the subsequent word "fell."

Frazier's model assumes that in such cases the system will need to reanalyse the input and generate an alternative parse. But the fact that this must be a *different* parse from the one that was first generated compromises the autonomy of the model. That is, the parser needs to be told that the earlier interpretation was unsatisfactory and that another parse should be attempted, since otherwise it will simply produce the same output once again. To this end, information must be conveyed from the interpretive mechanism to the parser. This feedback simply takes the form of an error message: Produce another parse, but not the same one as last time. Higher-level processes still have no direct control over the internal operation of the parser. Nevertheless, in order to account for successful eventual resolution of garden paths, Frazier's "autonomous" model must incorporate some degree of informational feedback.

Alternative models of syntactic processing include fully interactive theories in which semantic or other higher-level information can directly constrain the operation of the syntactic parser (McClelland et al. 1989; Taraban & McClelland 1988). However, there is also the approach of Altmann, Steedman, and colleagues (Altmann & Steedman 1988; Crain & Steedman 1985; Steedman & Altmann 1989), which the authors term "weakly interactive." In this approach, the syntactic processor is held to generate potential parses in a fully autonomous manner, but in parallel: The alternative candidate parses are then evaluated, again in parallel, against the semantic or discourse context. The interpretations are constructed incrementally and continually revised and updated, so that most alternatives can be quickly discarded. Indeed, it was assumed that strict time limits applied on the maintenance of parallel candidates, and that these time limits explained why the wrong parse could triumph in a garden-path sentence. In Altmann and colleagues' approach, reanalysis of garden paths requires no constraining feedback from higher-level to syntactic processing, since it can be achieved by repeating the same syntactic generation of alternative parses but relaxing the time limits at the selection stage. Although this model is termed "interactive" by its authors, it does not allow feedback from higher-level processing to influence which parse is generated. This renders it effectively more autonomous than Frazier's model.

Probably the leading current models of syntactic processing are found among the class of constraint satisfaction models (e.g., Boland 1997; MacDonald et al. 1994; Trueswell & Tanenhaus 1994); these models differ in their details, but in general share with the "weak interaction" approach the feature that syntactic analyses are computed in parallel and that higher-level information, though it is used early in processing, constrains selection of syntactic structure but not initial generation.

Boland and Cutler (1996) compared the way the labels "autonomous" and "interactive" were used in the word recognition and parsing literature, and concluded that these terms were not adequate to capture the true dimensions of difference between the models. The two research areas differed, they pointed out, in whether debate about the influence of higher-level processing concerned principally the generation of outputs by the lower-level process, or selection between generated outputs. In word recognition, there is debate about the autonomy of the initial generation process, but relative unanimity about the availability of higher-level information to inform the final selection between generated candidates. In parsing, in contrast, there is comparative agreement among models that the initial generation of syntactic structure is autonomous, but lively debate about whether or not selection of the correct parse takes higher-level information into account. What is notable, however, is that to argue for the strictest autonomy of initial syntactic processing, with the processor producing only a single output, necessarily implies allowing for at least

https://doi.org/10.1017/S0140525X00303247 Published online by Cambridge University Press

a minimal form of feedback to account for the known facts about the processing of garden-path sentences.

Of course, a system that avoided all feedback between semantics and syntax could be achieved if the parser had no capacity limitations, so that it could pursue all parses in parallel. In this case, syntactic garden paths would never arise (for further discussion of this point, see Norris 1987); but they do arise, so this system cannot be the correct one. Here we begin to get a crucial insight into the factors that determine the value of feedback. Our model, like all models of word recognition, embodies the assumption that prelexical processing can consider the full set of prelexical units – phonemes or letters – in parallel. But consider what might happen if phoneme recognition were a serial process in which each of the 40-plus phonemes of English had to be tested against the input in sequence. In such circumstances, an advantage might accrue if lexical information were allowed to determine the order in which phonemes were tested, so that lexically more probable phonemes were tested first. Testing the most probable phoneme first could confer a considerable advantage on the speed with which that phoneme could be identified, at only a marginal cost to the recognition of other phonemes if the lexical information proved inaccurate. So, our argument against feedback in word recognition can now be seen to rest on the important assumption that phoneme recognition is a parallel process. Note that this assumption also covers our earlier comments about verification models. If the system is parallel, and not resource-limited, then all phonemes should be fully analysed to the best of the system's capability. That is, there is no advantage in producing an initial low-quality analysis that is then improved on instruction from higher levels.

From this point of view, it is clear that our argument against feedback in word recognition cannot necessarily be applied across the board to every relationship between different levels of language processing. The question of syntactic-semantic interaction has led to a different debate than the case of prelexical versus lexical processing; models both with and without feedback have again been proposed, but the role of feedback is not the same in all models. The precise function and the necessity of feedback can only be evaluated in the light of constraints specific to the type of processing involved.

3.6. Summary

We have argued that there are no good *a priori* reasons for favouring interactive models over autonomous models of spoken word recognition. Feedback in bias models like TRACE is not able to improve word recognition. Interaction of this type could improve phoneme recognition, but it does so at the cost of making phonemic decisions harder when the input is inconsistent with lexical knowledge, and at the cost of potential misperceptions (the perception of lexically consistent phonemes even when they did not occur in the speech input). Although feedback could potentially act to improve perceptual sensitivity, recent studies suggest that lexical context has a purely biasing effect on phoneme identification (Massaro & Oden 1995; Pitt 1995).

We have further argued that feedback is also not required in visual word recognition. Autonomous models of reading are to be preferred because there are no data that require top-down interaction. It is clear that modular models are particularly well-suited to the constraints of word recognition. Because the full set of prelexical units (phonemes or letters) can be considered in parallel, feedback cannot improve performance at either the lexical or prelexical level. In sentential processing, however, resource limitations that prevent the parallel examination of all possible parses could at least in principle make the use of feedback beneficial. However, even here the extent of the interaction remains an empirical issue. Adopting Occam's razor, we should still assume only the minimum degree of feedback required by the data. It is also noteworthy that although the constraints on the production of language might suggest a role for feedback loops in that process (e.g., as a control mechanism), it again appears that feedback is not required, and it is not incorporated into the latest model of the process (Levelt et al. 1999).

On these grounds alone, therefore, one should be tempted to conclude in favour of autonomous models. But such a conclusion cannot be adopted without examination of the available data, since it remains possible that there are data that can be accounted for by interactive but not by autonomous models. We therefore turn to an examination of this evidence, again focusing on lexical involvement in phonemic decision making. Although it has proved difficult to resolve the debate between interactive and autonomous models in the visual case, new data on spoken-word recognition, some of which take advantage of phenomena specific to speech, have provided evidence that strongly favours autonomous theories over interactive ones. We begin by looking specifically at data that have challenged either TRACE or the Race model, or both.

4. Challenges to TRACE and the Race model

4.1. Variability of lexical effects

Because the Race model and TRACE are both designed to account for the same general set of phenomena, few of the findings in the literature present an insurmountable problem for either model. However, there are results showing variability in lexical effects that appear to be more consistent with the underlying principles of the Race model than with TRACE. Cutler et al. (1987) characterized the Race model as a multiple-outlet model. Responses can be made via either a lexical or prelexical outlet. TRACE, on the other hand, has only a single outlet. All phoneme identification responses must be made by reading phonemic information from the phoneme nodes in TRACE. One consequence of this difference is that, according to the Race model, it should be possible to shift attention between the two outlets. That is, lexical effects should not be mandatory. To the extent that attention can be focused on the prelexical outlet, lexical effects should be minimized. Conversely, lexical effects should be at their greatest when attention is focused on the lexical outlet.

This is exactly the pattern of results that has been observed in a number of studies. Cutler et al. (1987) showed that the lexical effects found in monitoring for initial phonemes in monosyllabic targets were dependent on the composition of filler items in the experiment. Lexical effects were only present when filler items varied in syllabic length. There were no lexical effects with monosyllabic fillers. Cutler et al. argued that the monotonous nature of the monosyllabic filler condition led subjects to focus their attention at the prelexical outlet, with the effect that any potential in-

https://doi.org/10.1017/S0140525X00303247 Published online by Cambridge University Press

fluence of lexical information would be attenuated. This shift in attention between outlets is a natural consequence of the Race model architecture. However, to account for the same effect in TRACE would require the model to be able to modulate the overall weighting of the word-phoneme feedback connections. (A similar suggestion has been made for varying the weight of letter-to-word inhibition in response to experimental conditions in visual word recognition; Rumelhart & McClelland 1982.) But if word-phoneme feedback connections were important for the proper functioning of the speech recognition system, it is not clear why it should be either possible or desirable to reduce their effectiveness.

Further evidence that lexical effects in phoneme monitoring are volatile and depend on having listeners focus their attention at the lexical level comes from a set of experiments by Eimas et al. (1990) and Eimas and Nygaard (1992; see also Foss & Blank 1980; Foss & Gernsbacher 1983; Frauenfelder & Seguí 1989; Seguí & Frauenfelder 1986). Eimas et al. (1990) found that lexical effects on phoneme-monitoring targets in syllable-initial position in items in lists emerged only with the inclusion of a secondary task that oriented attention towards the lexical level. So, lexical effects emerged with a secondary task of either noun versus verb classification, or lexical decision, but not with a secondary length-judgment task. Eimas and Nygaard (1992) extended this work by showing that there were no lexical effects on target detection in sentences, even with secondary tasks. They suggested that when listening to sentences subjects could perform the secondary task by attending to a sentential (syntactic) level of representation. Attention would then be allocated to this level of processing, and phoneme monitoring would be based on prelexical codes. Their data are particularly puzzling from the interactive standpoint. If interaction is important in the normal process of sentence understanding, it is strange that this is exactly the situation where it is hardest to obtain evidence of lexical effects.

The idea that lexical effects have to be specially engineered also emerges from studies of phonetic categorization. Burton et al. (1989) found that lexical effects were present only in the absence of complete phonetic cues. McQueen (1991) studied lexical influences on categorization of word final fricatives. At the end of words, top-down effects in a model like TRACE should be at their maximum. Furthermore, the stimuli included fricatives that were ambiguous between /s/ and /ʃ/. With the input in this finely balanced state, these should have been the ideal conditions to observe the lexical influences that are predicted by a model like TRACE. However, McQueen found that lexical effects emerged only when the stimuli were low-pass filtered at 3 kHz. That is, stimuli had to be not only phonetically ambiguous, but perceptually degraded too.

A rather weaker conclusion about the importance of degradation in obtaining lexical effects was reached by Pitt and Samuel (1993) in their review of lexical effects in phonetic categorization. Although they concluded that degradation was not actually a necessary precondition for obtaining lexical effects, there seems to be little doubt that lexical effects in categorization are enhanced by degradation. In both phonetic categorization and phoneme monitoring, therefore, lexical effects are not as ubiquitous as might be expected from interactive models if such effects were due to a mechanism that could improve recognition performance.

4.2. Facilitation versus inhibition in phoneme monitoring

Other data that appear problematic for TRACE come from a phoneme monitoring experiment by Frauenfelder et al. (1990). In a study conducted in French, they had subjects perform generalized phoneme monitoring on three different kinds of target. Target phonemes could appear in words after the uniqueness point (e.g., /l/ in *vocabulaire*), in nonwords derived from the word by changing the target phoneme (/t/ in *vocabutaire*), or in control nonwords (*socabutaire*). They argued that TRACE should predict that targets in the derived nonwords should be identified more slowly than in control nonwords because the lexically expected phoneme should compete with the target owing to top-down facilitation. According to the Race model, however, lexical effects on phoneme identification can only be facilitatory. As predicted by the Race model, there was indeed no difference between the nonword conditions, though both were slower than the word condition.

Wurm and Samuel (1997) replicated the Frauenfelder et al. (1990) findings but raised the possibility that inhibitory effects might be masked because the nonwords in which inhibition might be expected were easier to process than the control nonwords. They presented results from a dual task study which were consistent with their view that the experimental and control nonwords were not equally difficult. Nevertheless, there is still no direct evidence for inhibitory lexical effects in phoneme monitoring. We should also bear in mind that the claim that TRACE predicts inhibition from the lexicon is specific to the particular implementation of TRACE rather than true of interactive models in general (Peeters et al. 1989). We will return to this issue later when discussing simulations of these results. For the moment we will simply note that TRACE could be modified to incorporate at the phoneme level a priority rule similar to Carpenter and Grossberg's (1987) "two-thirds rule." In the context of a simple interactive activation model, this would mean that top-down activation would only have an effect when at least some bottom-up activation was present. That is, feedback from lexical to phonemic nodes would be contingent on there being at least some perceptual support for the phoneme. The input *vocabutaire* would then not activate /l/ at all, and /l/ would therefore not inhibit /t/.

4.3. Compensation for coarticulation

A strong apparent challenge to autonomous models comes from an ingenious study by Elman and McClelland (1988). As mentioned above, a common criticism of models with feedback is that they run the risk of misperceiving speech. That is, if top-down information can actually determine which lower-level representations are activated, the system may perceive events that, although consistent with top-down expectation, are not actually present in the real world. In the case of TRACE, top-down activation feeding back from lexical nodes to phoneme nodes leads to activation of the phoneme nodes which is indistinguishable from activation produced by bottom-up input from featural information. Elman and McClelland took advantage of this property of TRACE to devise a test that would distinguish

https://doi.org/10.1017/S0140525X00303247 Published online by Cambridge University Press

between the predictions of interactive models like TRACE and of autonomous models like the Race model.

We have seen that lexical effects like the Ganong (1980) effect, in which an ambiguous stimulus on a *type-dype* continuum is more likely to be classified in accord with the word (*type*) than the nonword (*dype*), can be explained by both interactive and autonomous models. However, according to TRACE, the lexical bias will actually alter the activation of the component phonemes. An ambiguous phoneme /ʔ/ midway between /ʃ/ and /s/ will thus activate the /ʃ/ phoneme node in *fooliʔ* and the /s/ node in *christmaʔ*. Elman and McClelland (1988) harnessed the Ganong effect to a lower-level effect of compensation for coarticulation (Mann & Repp 1981), according to which the position of the boundary between /t/ and /k/ is closer to /k/ following /ʃ/ (i.e., there are more /t/ responses) and closer to /t/ following /s/ (more /k/ responses).

If the lexical bias in phonetic categorization has its locus only at the output of phonemic information from the lexicon, as suggested by the Race model, the ambiguous phonemes in *fooliʔ* and *christmaʔ* should behave in the same way at the prelexical level. The ambiguous phonemes are identical and should have identical effects on a following phoneme midway between /t/ and /k/. However, if TRACE is correct, the lexical contexts *foolish* and *christmas* will determine whether /ʃ/ or /s/ is activated, which should, in turn, produce an effect of compensation for coarticulation, just as if the listener had heard a real /ʃ/ or /s/. In line with the predictions of the interactive model, Elman and McClelland found evidence of compensation for coarticulation even with the ambiguous phoneme /ʔ/.

One possible way in which proponents of autonomous models could avoid accepting these data as evidence against autonomy is to suggest that the results are owing entirely to effects operating at the prelexical level. As an illustration of how this might be possible, Norris (1993) simulated Elman and McClelland's results using a simple recurrent network. In one of the simulations, the network had no word nodes at all. The network learned to use several phonemes of context in making decisions about phoneme identity. A similar simulation has also been reported by Cairns et al. (1995). One might assume, as in TRACE simulations, that if there is a bias to interpret /ʔ/ as /ʃ/ in the context of *fooli,* this must be because of top-down feedback from a node corresponding to *foolish* at the lexical level. But in the Norris (1993) and Cairns et al. (1995) simulations, the phoneme nodes themselves learned something about the statistical properties of the language, that is, which contexts they are most likely to appear in. It is this within-level statistical information that leads to apparent interactive effects in these simulations.

Cairns et al. (1995) showed on the basis of an analysis of a large corpus of spoken English that after /ə/, /s/ is more likely than /ʃ/, and after /ɪ/, /ʃ/ is more likely than /s/. All of Elman and McClelland's (1988) /s/-final words ended in /əs/ and all of their /ʃ/-final words ended /ɪʃ/. Their materials therefore contained sequential probability biases that could in principle be learned at the prelexical level. Elman and McClelland's results thus do not distinguish between interactive and autonomous models because they can be explained either by top-down processing or by a sequential probability mechanism operating prelexically.

Pitt and McQueen (1998) have tested these two competing explanations. They used nonword contexts ending with unambiguous or ambiguous fricatives. The contexts contained transitional probability biases; in one nonword /s/ was more likely than /ʃ/, whereas in the other /ʃ/ was more likely than /s/. These contexts were followed by a word-initial /t/-/k/ continuum. Categorization of the ambiguous fricative reflected the probability bias. There was also a shift in the identification function for the following /t/-/k/ continuum, suggesting that compensation for coarticulation was being triggered by the probability bias. These results lend support to the view that Elman and McClelland's (1988) results were owing to transitional probability biases rather than to the effects of specific words. The original results can therefore no longer be taken as support for interactive models.

The transitional probability effect is consistent with both autonomous models (where the probability bias is learned prelexically) and interactive models (where the bias could be due either to top-down connections from the lexicon or to a prelexical sensitivity to sequential probabilities). The compensation for coarticulation data presented so far therefore does not distinguish between TRACE and the Race model. But other conditions tested by Pitt and McQueen (1998) produced data that challenge interactive but not autonomous models. Two word contexts were used (*juice* and *bush*) where the transitional probabilities of /s/ and /ʃ/ were matched. There was no shift in the stop identification function following *juiʔ* and *buʔ,* suggesting that the compensation for coarticulation mechanism is immune to effects of specific lexical knowledge. Crucially, however, there were lexical effects in the identification of the ambiguous fricative (more /s/ responses to *juiʔ* than to *buʔ*).

These data are problematic for TRACE, since the model predicts that if the lexicon is acting top-down to bias fricative identification, the changes in activation levels of phoneme nodes produced by feedback should also trigger the compensation for coarticulation process. TRACE is therefore unable to handle the dissociation in the word contexts between the lexical effect observed in fricative labelling and the absence of a lexical effect in stop labelling. Furthermore, if TRACE were to explain both lexical effects in words and sequential probability effects in nonwords as the consequences of top-down connections, the model would be unable to handle the dissociation between the compensation effect in the nonword contexts and the lack of one in the word contexts. This latter dissociation therefore suggests that sensitivity to sequential probabilities should be modelled at the prelexical level in TRACE. Consistent with this view is a recent finding of Vitevitch and Luce (1998). They observed, in an auditory naming task, different sequential probability effects in words and nonwords. They argued that the facilitatory effects of high-probability sequences observed in nonwords were due to prelexical processes, whereas the inhibitory effects of high-probability sequences observed in words were due to the effects of competition among lexical neighbors sharing those (high-probability) sequences. But even if the compensation effect in nonword contexts could thus be explained in TRACE by postulating processes sensitive to sequential probabilities at the prelexical level, TRACE would remain unable to explain the dissociation in the word contexts between the lexical effect in fricative identification and the absence of lexical involvement in stop identification.

Pitt and McQueen's compensation data are, however, not problematic for the Race model. If the compensation for

https://doi.org/10.1017/S0140525X00303247 Published online by Cambridge University Press

coarticulation process is prelexical, and there is sensitivity at that level to sequential probabilities (as the results of Vitevitch & Luce, 1998, also suggest), the Race model can explain the nonword context results. Also in the Race model, fricative decisions in the word contexts can be based on output from the lexicon, but in line with the data, the model predicts that lexical knowledge cannot influence the prelexical compensation process. Clearly, however, the Race model would require development in order to give a full account of these data (specifically it requires the inclusion of prelexical processes that are sensitive to phoneme probabilities and a prelexical compensation mechanism as in the Norris and the Cairns et al. simulations). Nevertheless, Pitt and McQueen's nonword context results clearly undermine what had seemed to be the strongest piece of evidence for interaction. Furthermore, the word context results undermine models with feedback. The study also serves as a cautionary reminder that low-level statistical properties of the language can give rise to effects that can easily masquerade as top-down influences.

4.4. Phonemic restoration and selective adaptation

Samuel (1997) has recently reported data that he claims argue strongly for interaction. Using a phoneme restoration paradigm, he presented listeners with words in which a given phoneme (/b/ or /d/) had been replaced by noise, and showed that these words produced an adaptation effect: There was a shift in the identification of stimuli on a /bɪ/-/dɪ/ continuum relative to the pre-adaptation baseline. There was no such effect when the phonemes were replaced by silence. Samuel argued that the noise-replaced phonemes were being perceptually restored, and that these restored phonemes were producing selective adaptation, just as if the actual phonemes had been presented. In support of his claim that these adaptation effects have a perceptual locus, Samuel showed that the adaptation effect observed with intact adaptors was not influenced by lexical factors. However, the main problem in determining the implications of this study for the question of interaction is that, in contrast to the compensation for coarticulation effect studied by Elman and McClelland (1988), we do not know the locus of the adaptation effect in this situation. Although it is clear that there are low-level components of selective adaptation (e.g., Sawusch & Jusczyk 1981), recent results suggest that adaptation operates at a number of different levels in the recognition system (Samuel & Kat 1996), including more abstract levels (labelled "categorical" by Samuel & Kat).

It remains to be established what level(s) of processing are responsible for the adaptation observed in Samuel's (1997) restoration study. If this adaptation effect is not influencing prelexical processing, it would not inform us about interaction. Indeed, the model we will present below could account for these data by assuming that adaptation with restored phonemes has its main influence on an output process, where categorical decisions are made. Consistent with Samuel's experiments with intact adaptors, we would not expect to see lexical effects where phoneme categorization had been determined by a clear acoustic signal.

A further problem is that the pattern of adaptation produced by the restored phonemes differs somewhat from standard adaptation effects. Normal adaptation effects are usually found almost exclusively in the form of a boundary shift (e.g., Samuel 1986; Samuel 1997, Experiment 1). However, in the condition showing the greatest restored adaptation effect, the shift is practically as large at the continuum endpoints as at the boundary. The small shift that can be induced by a restored phoneme appears to take the form of an overall bias not to respond with the adapted phoneme.

Samuel's results contrast with those of Roberts and Summerfield (1981) and Saldaña and Rosenblum (1994), who used the McGurk effect (McGurk & McDonald 1976) to investigate whether adaptation was driven by the acoustic form of the input or by its phonemic percept. The Saldaña and Rosenblum study took advantage of the fact that an auditory /ba/ presented with a video of a spoken /va/ is perceived as /va/. However, adaptation was determined by the auditory stimulus and not by the phonemic percept. Even though the combination of auditory /ba/ and visual /va/ was perceived as /va/ all of the time by 9 out of 10 of their subjects, the effect of adaptation in the Auditory + Visual case was almost identical (in fact, marginally bigger in the A + V case) to that with an auditory /ba/ alone. There was no trace of any top-down effect of the percept on the adaptation caused by the auditory stimulus. (Although these studies show that adaptation does not depend on the phonemic *percept*, see Cheesman & Greenwood, 1995, for evidence that the locus of the effect can indeed be phonemic or phonetic rather than acoustic.) This might suggest that output or response processes cannot be adapted, however it is also consistent with the view that the primary locus of adaptation is acoustic/phonetic and adaptation at an output level can only be observed in the absence of acoustic/phonetic adaptation.

Thus to draw any firm conclusions about interaction from Samuel's restoration study we would need to establish both that adaptation was a genuinely prelexical effect and that the pattern of adaptation observed following restoration was identical to that produced by phonemic adaptation. Neither of these points has been properly established.

4.5. Lexical effects on phonemic decisions in nonwords

Overall, the Race model fares well in explaining why phoneme identification should be easier in words than nonwords. However, some recently reported studies also show lexical effects in the processing of nonwords which present severe problems for the Race model.

4.5.1. Phonetic categorization. Data from Newman et al. (1997) from the phonetic categorization task may present a challenge to the Race model. Their study employed a variant on the Ganong effect. Instead of comparing word-nonword continua, they examined nonword-nonword continua, where the nonwords at each continuum endpoint varied in their similarity to real words. For example, the continuum *gice-kice*, where *gice* has more lexical neighbors than *kice*, was compared with the continuum *gipe-kipe*, where the opposite endpoint, *kipe*, had the higher neighborhood density. Newman et al. (1997) found that there were more responses in the ambiguous region of the continuum consistent with the endpoint nonword with a denser lexical neighborhood (i.e., more /g/ responses to *gice-kice* and more /k/ responses to *gipe-kipe*). According to the Race model, there should be no lexical involvement in these nonword decisions.

However, as Vitevitch and Luce (1998) point out, there

https://doi.org/10.1017/S0140525X00303247 Published online by Cambridge University Press

is a high positive correlation between lexical neighborhood density and sequential probability: Common sequences of phonemes will tend to occur in many words, so nonwords with dense lexical neighborhoods will tend to contain high probability sequences. The results of both Pitt and McQueen (1998) and Vitevitch and Luce (1998) suggest that the prelexical level of processing is sensitive to sequential probabilities (see sect. 4.3). It is therefore possible that Newman et al.'s results may reflect this prelexical sensitivity. If so, they would not be inconsistent with the Race model. It remains to be established whether the apparent lexical involvement in nonword decisions observed by Newman et al. (1997) is due to the effects of lexical neighborhood density or to a prelexical sensitivity to sequential probability. Only in the former case would these results then pose a problem for the Race model. The results would, however, be compatible with TRACE no matter which level of processing proves responsible for the effect.

4.5.2. Phoneme monitoring. Connine et al. (1997) have shown that monitoring for phonemes occurring at the end of nonword targets is faster the more similar the nonwords are to real words. In their experiment, nonwords were derived from real words by altering the initial phonemes of those words by either one feature on average or by six features on average (creating, for example, *gabinet* and *mabinet* from *cabinet*). Monitoring latencies were faster in these derived nonwords than in control nonwords, and the single-feature-change nonwords led to faster responses than the multi-feature-change nonwords. Responses to targets in all nonwords were slower than those to targets in the real words from which they were derived. According to the Race model, the lexical route should not operate at all for targets in nonwords, so monitoring latencies should be unaffected by the similarity of the nonwords to real words. This means that the first-past-the-post Race model, in which responses must be determined by either the phonemic or the lexical route, can no longer be sustained. Interactive models like TRACE, however, predict increasing top-down involvement in nonwords and hence faster monitoring responses the more similar nonwords are to words.

Wurm and Samuel (1997) have also shown that phoneme monitoring is faster in nonwords that are more like real words than in nonwords that are less like real words. It is important, however, to point out that this effect is not the same as that found by Connine et al. (1997). In the latter study, the words and nonwords all shared the same target phoneme (e.g., the /t/ in *cabinet, gabinet,* and *mabinet*). In Wurm and Samuel's (1997) study, however, as in the study by Frauenfelder et al. (1990) on which it was based, the nonwords and the words on which they were based were designed to differ on the crucial target phoneme, that is, the target in the nonwords did not occur in the base word (e.g., the /t/ in both *vocabutary* and *socabutary* mismatches with the /l/ in *vocabulary*). In these cases, therefore, the lexical information specifying an /l/ could not possibly make detection of the /t/ easier; it could only hinder detection of the /t/ (but, as discussed in sect. 4.2, both Frauenfelder et al. 1990 and Wurm & Samuel 1997 failed to find this inhibition). The facilitation that Wurm and Samuel found (e.g., faster /t/ responses in *vocabutary* than in *socabutary*) is thus different from Connine et al.'s (1997) finding and is probably due, as argued by Wurm and Samuel, to an attentional effect that makes more wordlike strings easier to process.

4.6. Subcategorical mismatch

A study by Marslen-Wilson and Warren (1994) is particularly important because it provides evidence against both TRACE and the Race model. This study, based on earlier work by Streeter and Nigro (1979) and Whalen (1984; 1991), examined subcategorical phonetic mismatch (Whalen 1984) and the differential effects of that mismatch in words and nonwords. Streeter and Nigro cross-spliced the initial CV from a word like *faded* with the final syllable of a word like *fable*. The cross-splice creates conflicting phonetic cues to the identity of the medial consonant /b/; the transition from the first vowel provides cues appropriate to /d/ rather than /b/. A parallel set of stimuli was constructed from nonwords. Interestingly, the phonetic mismatch slowed auditory lexical decisions to the word stimuli, but not to the nonword stimuli. A similar, though nonsignificant, interaction between phonetic mismatch and lexical status was found by Whalen (1991). The design of Marslen-Wilson and Warren's study is rather more complicated, but it will be described in some detail since this will be essential for understanding the simulations presented later.

The critical stimuli used in Marslen-Wilson and Warren's experiments were based on matched pairs of words and nonwords like *job* and *smob*. Three experimental versions of these stimuli were constructed from each word and nonword by cross-splicing different initial portions of words and nonwords (up to and including the vowels) onto the final consonants of each critical item. These initial portions could either be from another token of the same word/nonword, from another word (*jog* or *smog*), or from another nonword (*jod* or *smod*). The design of the materials is shown in Table 1, which is based on Table 1 from Marslen-Wilson and Warren (1994). Marslen-Wilson and Warren performed experiments on these materials using lexical decision, gating, and phonetic categorization tasks. The important data come from the lexical decision and phonetic categorization experiments using materials where the critical final phonemes were voiced stops. In both of these tasks, the effect of the cross-splice on nonwords was much greater when the spliced material came from a word (W2N1) than from a nonword (N3N1), whereas the lexical status of the source of the cross-spliced material (W2W1 vs.

Table 1. *Experimental conditions in Marslen-Wilson and Warren (1994) and McQueen et al. (1999a)*

Item type	Notation	Example
Word		*job*
1. Word 1 + Word 1	W1W1	job + job
2. Word 2 + Word 1	W2W1	jog + job
3. Nonword 3 + Word 1	N3W1	jod + job
Nonword		*smob*
1. Nonword 1 + Nonword 1	N1N1	smob + smob
2. Word 2 + Nonword 1	W2N1	smog + smob
3. Nonword 3 + Nonword 1	N3N1	smod + smob

Note. Items were constructed by splicing together the underlined portions.

https://doi.org/10.1017/S0140525X00303247 Published online by Cambridge University Press

N3W1) had very little effect for words. Within cross-spliced nonwords, therefore, there was an inhibitory lexical effect: Performance was poorer when the cross-spliced material in the nonword came from a word than when it came from another nonword.

The implication of this result for the Race model should be clear. Phonemic decisions about nonword input can only be driven by the prelexical route. They should therefore be unaffected by the lexical status of the items from which the cross-spliced material is derived. But these results are also problematic for TRACE. Marslen-Wilson and Warren (1994) simulated their experiments in TRACE. They showed that the TRACE simulations deviated from the data in a number of important respects. The primary problem they found was that TRACE predicted a difference between the cross-spliced words (poorer performance on W2W1 than on N3W1) which was absent in the human data. TRACE also overestimated the size of the inhibitory lexical effect in the cross-spliced nonwords.

McQueen, Norris, and Cutler (1999a) reported four experiments using the same design as Marslen-Wilson and Warren (1994). Although these experiments were conducted in Dutch, the materials were modelled closely on those used by Marslen-Wilson and Warren.

McQueen et al. found that the interaction between lexical status and the inhibitory effect of competitors could be altered by subtle variations in experimental procedure. When they used a lexical decision task to emphasize lexical processing, there was a clear and reliable mismatch effect that interacted with the lexical status of the cross-spliced portions in nonwords but not in words, just as in Marslen-Wilson and Warren's experiment. However, in experiments using phoneme monitoring and phonetic categorization, respectively, McQueen et al. failed to replicate the mismatch effects. McQueen et al. noted that there were two differences between their categorization experiment and that of Marslen-Wilson and Warren. First, Marslen-Wilson and Warren had varied the assignment of responses to left and right hands from trial to trial, whereas McQueen et al. had kept the assignment constant throughout the experiment. A second difference was that McQueen et al. used only unvoiced stops (/p,t,k/) as final segments, whereas Marslen-Wilson and Warren had used both voiced and unvoiced stops. Both of these differences would have made the task harder in Marslen-Wilson and Warren's study. McQueen et al. therefore ran a further experiment in which they incorporated a wider range of targets and varied the response hand assignment. Under these conditions, McQueen et al. were able to produce an inhibitory lexical effect in cross-spliced nonwords. As in the Marslen-Wilson and Warren experiments, responses to targets in W2N1 nonwords were slower than responses to targets in N3N1 nonwords.

The inhibitory effect of lexical competitors on phonemic decisions in nonwords therefore follows a similar pattern to the facilitatory effects of lexical status seen in phoneme monitoring. Cutler et al. (1987), for example, showed that the size of the facilitatory lexical effect (faster responses to targets in words than in nonwords) can be modulated by task demands. McQueen et al. have shown that inhibitory lexical effects in cross-spliced nonwords can be modulated in a similar manner. As pointed out earlier, the variability of lexical involvement in phonemic decision making is problematic for TRACE. All lexical effects in phoneme decisions to targets in nonwords, even if those effects are variable, pose problems for the Race model.

4.7. Summary

This review demonstrates that neither TRACE nor the Race model is now tenable. TRACE is challenged by the findings showing variability in lexical effects, by the lack of inhibitory effects in nonwords in Frauenfelder et al. (1990), by the latest data on compensation for coarticulation, and by the data on subcategorical mismatch (Marslen-Wilson & Warren 1994; McQueen et al. 1999a). Marslen-Wilson and Warren explicitly attempted to simulate their results in TRACE, without success. The Race model, similarly, is challenged by the demonstrations of lexical involvement in phonemic decisions on nonwords. Three recent studies (Connine et al. 1997; Marslen-Wilson & Warren 1994; McQueen et al. 1999a) all show lexical effects in decisions made to segments in nonwords (as also may Newman et al. 1997). Such effects are incompatible with the Race model's architecture whereby the lexical route can only influence decisions to segments in words.

There would thus appear to be no available theory that can give a full account of the known empirical findings in phonemic decision making. In the following section we will show, however, that it is possible to account for the data in a model that reflects the current state of knowledge about prelexical processing in spoken-word recognition, and we will further demonstrate that our proposed new model successfully accounts for a wide range of results. Moreover, we will argue that this new model – the Merge model – remains faithful to the basic principles of autonomy.

5. The Merge model

5.1. The model's architecture

The models we have contrasted above represent extreme positions with regard to the relationship between lexical and prelexical processing in phonemic processing. TRACE has an architecture in which the lexical level is directly linked via hardwired connections to the prelexical level, and responses must be susceptible to whatever lexical information is available. The Race model has an architecture in which responses via the lexical or the prelexical level are completely independent. Both architectures have now been found wanting.

What is required is, instead, a model in which lexical and prelexical information can jointly determine phonemic identification responses. Such a model must be able to allow for variability in the availability of lexical information, such that responses to a given phoneme in a given input may be susceptible or not to lexical influence, as a function of other factors. In other words, the model must not mandate lexical influence by fixed connections from the lexical to the prelexical level, but neither must it avoid all possibility of lexical and prelexical codetermination of responses by making lexical information only available via successful word recognition. The model must moreover capture this variability in just such a way as to be able to predict how strong lexical influences should be in different situations.

The required class of models consists of those in which the outputs of processes that are themselves fully autonomous can be integrated to determine a response. Such

https://doi.org/10.1017/S0140525X00303247 Published online by Cambridge University Press

models have been proposed in several areas. A general model of perception incorporating such an approach, for instance, is the FLMP (Massaro 1987; 1998), in which multiple sources of information are simultaneously but independently evaluated, and continuous truth values are assigned to each source of information as a result of the evaluation process. Specifically with respect to lexical and prelexical processing, an example of an integration model is Norris's (1994a) model of the transformation of spelling to sound in the pronunciation of written words, or the Activation Verification Model of reading (Paap et al. 1982).

Applied to the issue of phonemic decision making, the integration approach allows prelexical processing to proceed independently of lexical processing but allows the two processes to provide information that can be merged at the decision stage. In the Merge model, prelexical processing provides continuous information (in a strictly bottom-up fashion) to the lexical level, allowing activation of compatible lexical candidates. At the same time, this information is available for explicit phonemic decision making. The decision stage, however, also continuously accepts input from the lexical level and can merge the two sources of information. Specifically, activation from the nodes at both the phoneme level and the lexical level is fed into a set of phoneme-decision units responsible for deciding which phonemes are actually present in the input. These phoneme decision units are thus directly susceptible to facilitatory influences from the lexicon, and by virtue of competition between decision units, to inhibitory effects also.

In Merge there are no inhibitory connections between phoneme nodes at the prelexical level. The absence of inhibitory connections between phonemes at this level is essential in a bottom-up system. Inhibition at this level would have the effect of producing categorical decisions that would be difficult for other levels to overturn; information vital for the optimal selection of a lexical candidate could be lost. If a phoneme is genuinely ambiguous, that ambiguity should be preserved to ensure that the word that most closely matches the input can be selected at the lexical level. For example, if *phoneme* is pronounced as /ʔonim/ where /ʔ/ is slightly nearer a /v/ than /f/, then inhibition between phonemes would leave the representation /vonim/ to be matched against the lexicon. This would be a much worse match to *phoneme* than a representation that preserved the input as giving partial support to both /v/ and /f/. There is, however, between-unit inhibition at the lexical level and in the decision units. The lexical-level inhibition is required to model spoken-word recognition correctly, and the decision-level inhibition is needed for the model to reach unambiguous phoneme decisions when the task demands them.

This need to have between-unit inhibition at the decision level, but not at the level of perceptual processing itself, is in itself an important motivation for the Merge architecture. Perceptual processing and decision making have different requirements and therefore cannot be performed effectively by the same units. So even if the question of interaction were not an issue, any account of phonemic decision making should separate phonemic decision from phonemic processing. The structure of Merge thus seems essential for maximum efficiency.

The empirical demonstration of lexical effects on nonwords does not constitute a problem for the Merge model, since on this account nonwords activate lexical representations to the extent that they are similar to existing words; this additional activation can facilitate phoneme detection. In the Merge model it is not necessary to wait (as in the Race model) until one of the two routes has produced a clear answer, since the output of those routes is continuously combined. However it is also not necessary to compromise the integrity of prelexical processing (as in TRACE) by allowing it to be influenced by lexical information. Further, Merge is prevented from hallucinating by the incorporation of a bottom-up priority rule. This rule, following the suggestion of Carpenter and Grossberg (1987), prevents decision nodes from becoming active at all in the absence of bottom-up support, thus ensuring that phonemic decisions are never based on lexical information alone. The Merge model therefore allows responses to be sensitive to both prelexical and lexical information, but preserves the essential feature of autonomous models – independence of prelexical processing from direct higher-level influence.

One might argue that the addition of phoneme decision nodes opens the Merge model to attack from Occam's razor. Why have separate representations of the same information at both the prelexical and the decision levels – that is, unnecessary multiplication of entities? The addition of phoneme decision nodes certainly makes Merge more complex than the Race model, but their addition is necessary in order to account for the data that the Race model cannot explain. As we have already argued, bottom-up flow of information from prelexical to lexical levels is logically required for the process of spoken word recognition; furthermore, the decision units, which should have between-unit inhibition, must be separate from the prelexical units, which should have no inhibition. Merge's decision nodes, and the connections from both prelexical and lexical levels to these nodes, are the simplest additions to the logically demanded structures which allow one to describe the available data adequately. The Merge architecture is thus bottom-up and also optimally efficient.

5.2. Merge simulations

In order to study the behaviour of the Merge model, we constructed a simple network that could be readily manipulated and understood.

The Merge network is a simple competition-activation network with the same basic dynamics as Shortlist (Norris 1994b). The network has no intrinsic noise; we are interested in modelling RT in tasks that are performed with high levels of accuracy. We acknowledge that a noise-free network is unlikely to be suitable for modelling choice behaviour with lower levels of accuracy, but this is independent of the main architectural issues being dealt with here (McClelland 1991). As Figure 1 shows, both the word and phoneme nodes are connected by facilitatory links to the appropriate decision nodes. But there is no feedback from the word nodes to the prelexical phoneme nodes. In the simulations of subcategorical mismatch, the network was handcrafted with only 14 nodes: six input phoneme nodes corresponding to /dʒ/, /p/, /g/, /b/, /v/, and /z/; four phoneme decision nodes; and four possible word nodes: *job*, *jog*, *jov*, and *joz*. The latter two word nodes simply represent notional words ending in phonemes unrelated to either /b/ or /g/.

The input in these simulations is *job*. The different experimental conditions are created by varying the set of word nodes that are enabled. By enabling or disabling the word node for *job*, the same input can be made to represent ei-

https://doi.org/10.1017/S0140525X00303247 Published online by Cambridge University Press

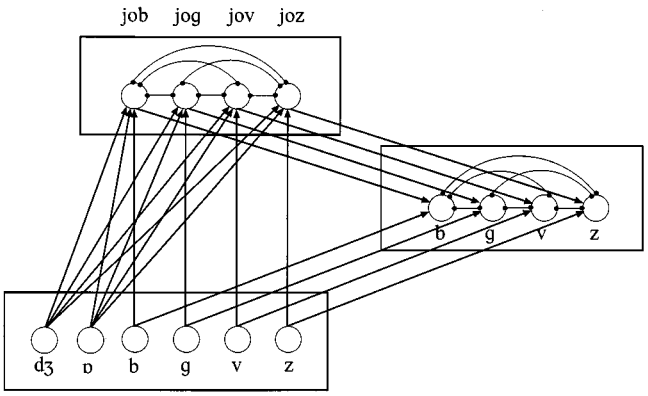


Figure 1. *The Merge model.* The basic architecture is shown, together with the connectivity patterns for the node types used in the simulations. Activation spreads from the input nodes to the lexical nodes and to the phoneme decision nodes, and from the lexical nodes to the phoneme decision nodes; inhibitory competition operates at the lexical and phoneme decision levels. Excitatory connections, shown with bold lines and arrows, are unidirectional; inhibitory connections, shown with fine lines and closed circles, are bidirectional.

ther a word or a nonword, and by altering whether *jog* is enabled the same stimulus can be given a lexical competitor or not.

We will present only the activation levels of the units in the network, rather than attempting to model explicitly the mapping of these activations onto response latencies. A problem that arises in simulating reaction-time data in models like Shortlist or TRACE is that activation levels can never be translated *directly* into latencies. The simplest solution is to assume that responses are made whenever activations pass a given threshold. In presenting their simulations, Marslen-Wilson and Warren (1994) plotted response probabilities derived from the Luce choice rule (Luce 1959) rather than the underlying activations. Response times can then be derived by thresholding these probabilities.

There are a number of problems associated with the use of the Luce choice probabilities in the decision mechanism of a model of continuous speech recognition. One simple problem is that the Luce rule introduces a second "competition" process into the recognition process in addition to the inhibitory competition already present. In the Luce calculations, other active candidates compete with the most active candidate and reduce its response probability. The extent of this competition is dependent on the exponent used in the Luce choice rule. As the exponent increases (this is equivalent to decreasing the amount of noise in the system), so accuracy is increased and the influence of competitors decreases. In Marslen-Wilson and Warren's (1994) TRACE simulations of lexical decision, the error rates for W1W1 stimuli, for example (in Fig. 12, p. 669), are about ten times greater than in the human data. A more appropriate choice of exponent would have greatly reduced the competition effect introduced by the Luce rule.

The tasks we are simulating are performed with a high level of accuracy that varies little between conditions. The crucial effects are all reflected in differences in latency rather than accuracy. A simple response threshold, therefore, provides the most straightforward means of deriving a response from variations in activation level, without the additional complexity and assumptions of the Luce rule. Note that although it is easy to make the long-term average behaviour of a connectionist network follow the Luce choice rule by adding noise and then selecting the node with the largest activation (Page 2000; McClelland 1991), there is no such obvious way to derive response probabilities directly from network activations on a single trial. Probabilities can be calculated and used to determine latency (cf. Massaro & Cohen 1991), but this would add a very complicated additional mechanism to any connectionist model. Furthermore, as Shortlist simulations presented in Norris (1994b) show, activations for words in continuous speech often rise transiently to quite high levels before being suppressed by other candidates. In a complete account of the decision process, activation should therefore be integrated over time. However, for present purposes, the simple threshold can serve as a measure of recognition point, or of YES responses in lexical decision. Negative responses in lexical decision are slightly more problematic. Here we adopt the procedure proposed by Grainger and Jacobs (1996) for visual lexical decision. Grainger and Jacobs suggest that NO responses are made after some deadline has elapsed, but that the deadline is extended in proportion to the overall level of lexical activity. Ideally we would have a quantitative implementation of the decision process so that we could fit the complete model directly to the data. However, the decision component of the model would, in itself, require several extra parameters. We therefore present only the activations from the network and show that, given the assumption that decisions are made when activations pass a threshold (or, in the case of NO responses, when a deadline is reached), the patterns are qualitatively consistent with the experimental data.

We present detailed simulation results from the theoretically most critical sets of data described above: the subcategorical mismatch findings of Marslen-Wilson and Warren (1994) and McQueen et al. (1999a), and the phoneme monitoring results of Connine et al. (1997) and Frauenfelder et al. (1990). These studies provide very highly constraining data against which to evaluate Merge. The subcategorical mismatch data have already led us to reject the Race model and, according to Marslen-Wilson and Warren's simulations, have eliminated TRACE too. It is thus crucial to establish whether Merge can simulate the effects of mismatch, lexical status, and their interaction observed in those studies, and also the dependency of the inhibitory lexical effect on task demands. Likewise, the data of Connine et al. (1997) are important to simulate because they also led us to reject the Race model. This simulation will allow us to establish whether Merge can provide a bottom-up account for graded lexical effects in nonwords. The results of Frauenfelder et al. (1990) are similarly theoretically important because they offer a crucial challenge to TRACE and allow comparison of facilitatory and inhibitory effects in phoneme monitoring.

5.2.1. Subcategorical mismatch simulation. As stated, the input was always *job*. In these simulations, only two of the word nodes were ever enabled in a given condition. In the W1W1 and the W2W1 conditions, the nodes *jog* and *job* were enabled. In the N3W1 condition, the nodes *jov* and *job* were enabled. For both N1N1 and N3N1, *jov* and *joz* were enabled, and for W2N1, *jog* and *joz* were enabled. The nodes *jov* and *joz* acted to balance the overall potential for

https://doi.org/10.1017/S0140525X00303247 Published online by Cambridge University Press

lexical competition in the different simulations. They reflect the fact that, in addition to the matched experimental words, the language will usually contain other words beginning with the same two phonemes.

Input to the network consisted of four vectors representing the total for each phoneme node for each time slice. Under normal, unspliced conditions, the input to each phoneme built up over three time slices: 0.25, 0.5, and 1.0. It then remained at its peak value through the remaining time slices. The first phoneme began at time slice 1, the second at time slice 4, and the third at slice 7. This form of input is analogous to that used in Shortlist, where each phoneme produces the same bottom-up activation regardless of its position in the input sequence. Simulations in which activation reached a peak and then decayed symmetrically were also carried out. However, although this kind of input required different parameters, it made little difference to the qualitative nature of the simulations.

For the cross-splice conditions we assumed that the total support from the competing /b/ and /g/ phonemes remained 1.0. At slice 7 the input for /g/ was 0.15, where it stayed for the remainder of the input. The input for /b/ had 0.15 subtracted from all slices from 7 onwards. So, according to this scheme, a /b/ in a cross-spliced condition reached a final activation value of only 0.85 instead of 1.0, whereas there was an input of 0.15 to the competing /g/ phoneme from the cross-splice onwards. The main aspect of the input representations that alters the outcome of the simulations is the magnitude of the support for the competing phoneme in the cross-splice condition. If the cross-splice support for the /g/ is weighted more heavily relative to the /b/ (e.g., 0.25 vs. 0.75), the simulations will be more likely to show equivalent effects of the splice for words and nonwords across the board. With too small a cross-splice the effects will disappear altogether. However, there is a large range of parameters in between where the general behaviour of the model remains similar.

As noted above, any effective decision mechanism operating on the lexical output of Shortlist would need to perform some integration over time. However, although we did not introduce an integrating decision mechanism at the lexical level in Merge, we did find it necessary to add some integration, or momentum, at the phoneme decision nodes. The decision nodes were run in the Shortlist-style reset-mode (Norris et al. 1995). That is, the activation levels for these nodes were reset to zero at the start of each new time slice. On its own, this generally leads to a rapid and almost complete recovery from the effect of the cross-splice. In order to make the decision process sensitive to the activation history, some proportion of the final level of activation on each slice was fed in along with the input to the next slice. This "momentum" term controlled the extent to which the decision nodes integrated their input over time. In the simulations both the word and decision levels cycled through 15 iterations of the network for each time-slice of input. However, because the input phoneme units had no between-phoneme inhibition this level did not cycle. Phoneme activations for each slice were calculated by multiplying the previous activation by the phoneme level decay and then adding in the new input.

The set of parameters that produced the simulations shown in Figures 2 and 3 is listed in the Appendix. The most difficult aspect of the simulations was to find parameters that would give a good account of the phonetic categorization data. The basic pattern for the lexical decision data was robust and could be reproduced by a wide range of parameters. Correct adjustment of the momentum term in the decision units proved to be critical for these simulations. Note that the simulations never produced the opposite pattern from that observed in the data, but often the phonetic categorization responses to N3N1 and W2N1 did not differ significantly (as in McQueen et al.'s experimental data).

Figure 2 provides simulations of the lexical decision data. Figure 2A shows the activation functions for lexical nodes given unspliced *job* as input; Figure 2B shows lexical acti-

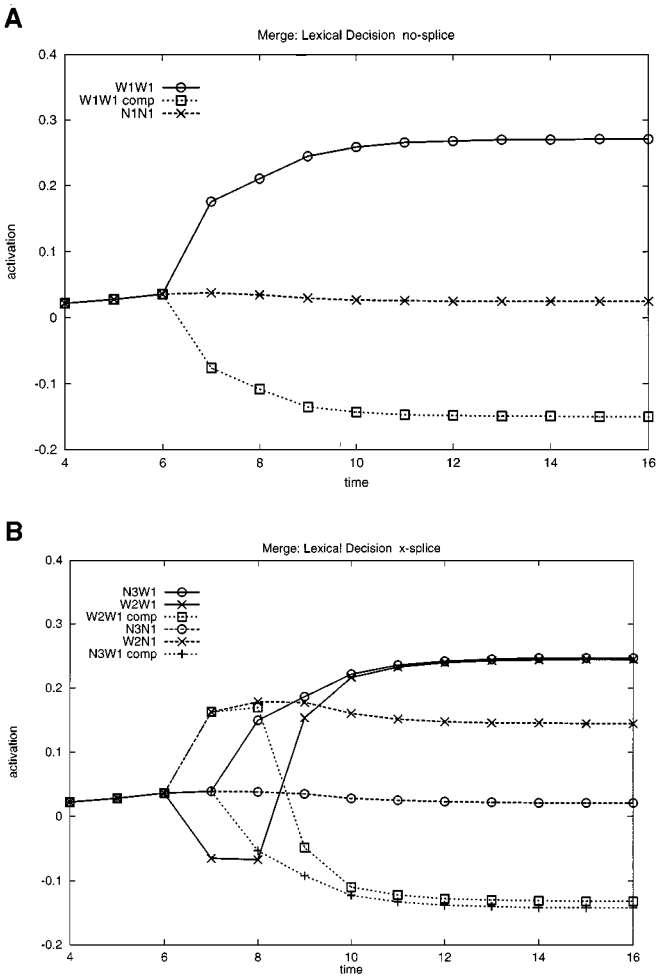


Figure 2. Simulation of lexical decisions in the subcategorical mismatch experiments. In all cases, the labels refer to the conditions used in those experiments, as shown in Table 1. Figure 2A shows the activation levels for lexical nodes given unspliced *job* as input. W1W1 shows the activation of the *job*-node when it was switched on as a possible word, and W1W1 comp shows the activation of the node for the lexical competitor *jog*, which was also switched on in this simulation. N1N1 shows the activation of the *jov*-node when neither *job* nor *jog* was a word. Figure 2B shows the activation levels for lexical nodes given cross-spliced *job* as input (i.e., with information in the vowel consistent with a following /g/). W2W1 and W2W1 comp show the activation levels of the *job*- and *jog*-nodes, respectively, when both were switched on as words. N3W1 shows the activation of the *job*-node when *job* was a word and *jog* was a nonword. N3W1 comp thus shows the activation of the other activated word in this condition, that of *jov*. W2N1 shows the activation of the *jog*-node when *jog* was switched on as a word, but *job* was not. Finally, N3N1 shows the activation of the *jov*-node when neither *job* nor *jog* was a word.

https://doi.org/10.1017/S0140525X00303247 Published online by Cambridge University Press

vation given cross-spliced *job* as input (i.e., a token containing information in the vowel specifying a /g/ instead of a /b/). In unspliced W1W1, *job* is a word in the Merge lexicon, and its activation rises quickly to an asymptote near 0.25. If we assume a response threshold of 0.2, lexical decisions should be faster in this condition than with the cross-spliced words W2W1 and N3W1. This reflects the basic mismatch effect observed for words in the human data, as shown in Table 2. With the same response threshold, as also shown in Table 2, there will be almost no difference in the response times to words in the two cross-spliced conditions, again as in the human data. In the nonword conditions, where *job* is not enabled as a word in the lexicon, there is effectively no activation of any lexical nodes for both unspliced N1N1 and cross-spliced N3N1; the model thus captures the fact that there was little effect of mismatching information in the nonword data when the cross-splice involved another nonword. In the W2N1 condition, however, the activation of W2 (*jog*) remains high throughout the target. According to the Grainger and Jacobs (1996) decision rule, the increased activation in the W2N1 case will delay the deadline and lead to slower responding, exactly as seen in the data.

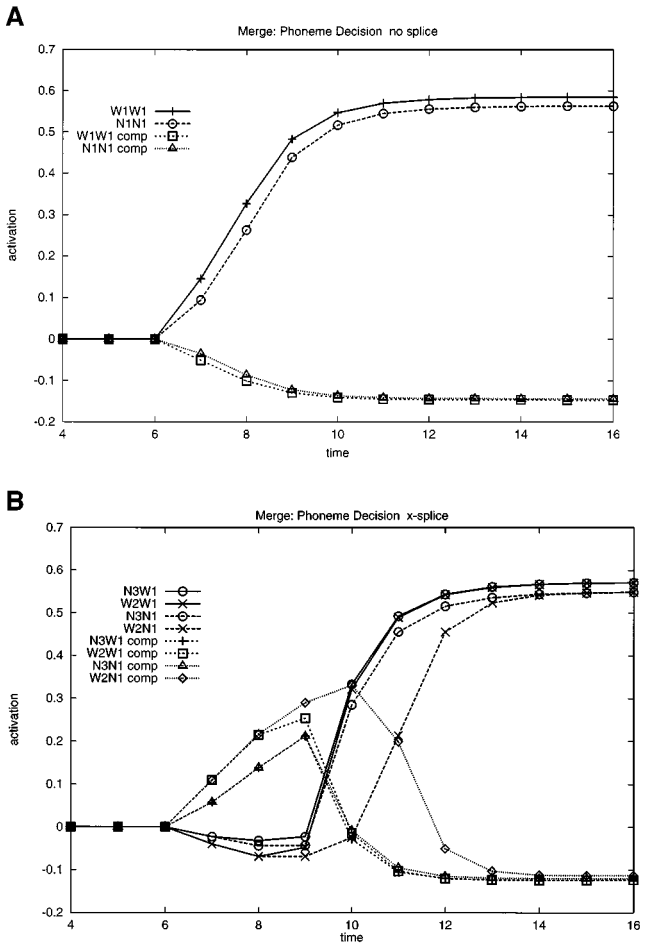


Figure 3. Simulation of phonemic decisions in the subcategorical mismatch experiments. In all cases, the labels refer to the conditions used in those experiments, as shown in Table 1. Figure 3A shows the activation levels for the /b/ and /g/ phoneme decision nodes given unspliced *job* as input. W1W1 shows the activation of the /b/-node when *job* was switched on as a possible word in the lexicon, and W1W1 comp shows the activation of the /g/-node, corresponding to the lexical competitor *jog*, which was also switched on in this simulation. N1N1 and N1N1 comp show the activations of the /b/- and /g/-nodes, respectively, when neither *job* nor *jog* was a word. Figure 3B shows the activation levels for /b/ and /g/ phoneme decision nodes given cross-spliced *job* as input (i.e., with information in the vowel consistent with a following /g/). W2W1 and W2W1 comp show the activations of the /b/- and /g/-nodes, respectively, when both *job* and *jog* were switched on as words. N3W1 shows the activation of the /b/-node when *job* was a word and *jog* was a nonword. N3W1 comp shows the activation of the /g/-node in this condition. W2N1 shows the activation of the /b/-node when *jog* was switched on as a word, but *job* was not, while W2N1 comp shows the activation of the /g/-node in this condition. Finally, N3N1 and N3N1 comp show the activation levels of the /b/- and /g/-nodes when neither *job* nor *jog* was a word.

Table 2. *Times at which lexical and phoneme node activations reach threshold in the Merge model, compared with the subcategorical mismatch data*

Lexical Decision	Lexical node threshold reached at	MWW Expt.1	MNC Expt. 3
Word			
W1W1	7.7	487	340
W2W1	9.7	609	478
N3W1	9.4	610	470
Phonetic Decision	**Phoneme decision node threshold reached at**	**MWW Expt.1**	**MNC Expt. 3**
Word			
W1W1	8.4	497	668
W2W1	10.4	610	804
N3W1	10.4	588	802
Nonword			
N1N1	8.8	521	706
W2N1	11.8	654	821
N3N1	10.7	590	794

Note. The first column shows the time (time-slice numbers estimated by interpolation from Figs. 2 and 3) at which nodes attained the criterial threshold of 0.2 (lexical node for YES lexical decisions) or 0.4 (phonemic decision node for phonemic decisions) in each condition in the Merge model simulations of the subcategorical mismatch data. The RT data (mean RT in msec), from both Marslen-Wilson and Warren (1994, Experiments 1 and 3, voiced stops) and McQueen et al. (1999a, Experiments 3 and 4) are shown in columns two (MWW) and three (MNC) respectively. No data are given for the NO lexical decisions to nonwords, since these decisions are based not on activation thresholds, but on a response deadline being reached (see text for details).

Figure 3 shows the activation functions for the phoneme decision nodes, in the simulation of the phonetic categorization data. Figure 3A shows that in the unspliced items there is only a relatively weak lexical effect. Activation of /b/ rises somewhat more rapidly and attains a higher asymptote when *job* is a word in the Merge lexicon (W1W1) than when it is not a word in the lexicon (N1N1). As shown in Table 2, this small facilitative lexical effect is in line with the experimental data; the effect was significant in Marslen-Wilson and Warren (1994) but not in McQueen et al. (1999a). In the cross-spliced phonetic categorization simulations (Figure 3B and Table 2), a threshold set at 0.4 would

https://doi.org/10.1017/S0140525X00303247 Published online by Cambridge University Press

result in almost no differences between W2W1, N3W1, and N3N1 response times. The activation functions for these conditions are almost identical at this point. But the activation of /b/ reaches 0.4 in all three of these conditions later than in the unspliced conditions (Fig. 3A and Table 2); this is the basic mismatch effect observed in phonetic categorization in both words and nonwords. The model therefore correctly predicts no difference between the two types of cross-spliced word; it also correctly predicts an inhibitory lexical effect in the cross-spliced nonwords. The activation of /b/ in W2N1 grows more slowly than do the others; this is because of the activation of the /g/ node (W2N1 comp), which receives support from the lexical node for W2 (*jog;* its activation is plotted in Fig. 2B). Thus, only the nonwords show an effect of lexical competition. The model therefore gives an accurate characterization of the competition effects in both the lexical decision and the phonetic categorization tasks, and provides an account of why competition effects are only observed in the case of nonwords.

Given the architecture of the network, any factor that either reduces lexical activation or the strength of the connections from the lexical to the decision nodes will clearly reduce the size of the lexical effects. Merge thus copes naturally with data showing that lexical effects vary with changes in the task, both those on subcategorical mismatch (McQueen et al. 1999a) and the other effects of variability reviewed in section 4.1. When task demands discourage the use of lexical knowledge, decisions will be made on the basis of the prelexical route and lexical activation will simply not contribute to decision node activation.

Although the network might be thought of as permanently connected (as in TRACE), we prefer to view Merge as having the same architecture as Shortlist (Norris 1994b), in which the lexical network is created dynamically as required. This means that the word nodes cannot be permanently connected to the decision nodes. Instead, the 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 Merge model, therefore, the demands of the experimental situation determine how the listener chooses to perform the task. If task demands encourage the use of lexical knowledge, connections will be built from both prelexical and lexical levels to the decision nodes. But if the use of lexical knowledge is discouraged, only connections from the prelexical level will be constructed. Task demands could similarly result in decision nodes only being employed when they correspond to possible experimental responses. In a standard phoneme monitoring experiment, with only a single target, there might only be a single decision node. If so, there could never be an inhibitory effect in phoneme monitoring because there could never be competing responses.

5.2.2. Phoneme monitoring simulation. Connine et al. (1997) showed that phoneme monitoring responses to final targets in nonwords that were more like words (e.g., /t/ in *gabinet*, which differs from *cabinet* only in the voicing feature of the initial stop) were faster than those to targets in nonwords that were less like words (e.g., *mabinet*, with a larger featural mismatch in initial position), which in turn were faster than those to targets in control nonwords (not close to any real word, e.g., *shuffinet*). This graded lexical involvement in phoneme monitoring in nonwords cannot be explained by the Race model. We have already seen that Merge can simulate a simple lexical advantage; can it also simulate graded effects?

As most of Connine et al.'s stimuli were several phonemes long we added two more phonemes to the network so that we could simulate processing words and nonwords that were five phonemes in length. To simulate Connine et al.'s word condition (*cabinet*) the lexicon contained a single five-phoneme word and that word was presented as input. For the multi-feature-change nonword condition (*mabinet*), the input simply had a different initial phoneme that did not activate the initial phoneme of the word at all. That is, there was no perceptual similarity between the initial phoneme of the word and the multi-feature-change nonword. For the control nonword (*shuffinet*), the input was actually identical to that used in the word condition, but there were no words in the lexicon. Figure 4 shows the results of this simulation, which uses exactly the same parameters as the previous simulation.

It can be seen that activation of the final target phoneme rises more slowly in the multi-feature-change nonword than in the word, but most slowly of all in the control nonword. Note that we simulated only the multi-feature-change nonwords, as the positioning of the single-feature-change nonwords (between words and multi-feature-change nonwords) depends almost entirely on how similar we make the input representations of the initial phonemes and the initial phonemes of the words. Note also that the exact amount of lexical benefit in these experiments should further depend on the pattern of lexical competitors. To the extent that nonwords elicit more competitors not sharing the target phoneme, facilitation should decrease.

Because lexical activation is a function of the goodness of match between lexical representations and the input, and because lexical activation is fed continuously to the phoneme decision nodes, Merge can thus explain Connine et al.'s (1997) data. Nonwords that are more like words tend to activate lexical representations more than nonwords that are less like words, and this increases word-node to decision-node excitation, so that phoneme decisions to targets in more wordlike nonwords will tend to be faster than those to targets in less wordlike nonwords.

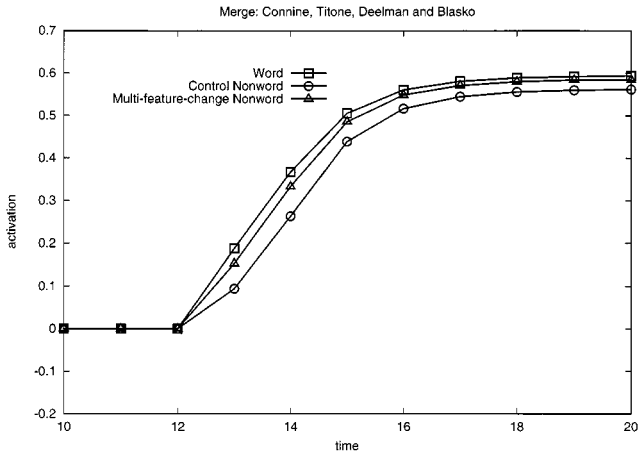


Figure 4. Simulation of Connine et al. (1997). The activation of the phoneme decision node for /b/ is shown in three conditions corresponding to the original study: Word, Multi-feature-change Nonword, and Control Nonword. In all three conditions, /b/ was the final phoneme of the input (i.e., the target phoneme).

https://doi.org/10.1017/S0140525X00303247 Published online by Cambridge University Press

We also tested Merge's ability to account for the phoneme monitoring data from Frauenfelder et al. (1990). These findings allow us to examine Merge's ability to simulate both facilitatory lexical effects in phoneme monitoring, and the absence of inhibitory effects when materials contain no subcategorical mismatches. Even though Merge combines lexical and prelexical information, the use of the bottom-up priority rule means that Merge correctly accounts for the results of the Frauenfelder et al. study. Frauenfelder et al. found no difference in phoneme monitoring latency between the /t/ targets in items like *vocabutaire* and *socabutaire*. They had argued that TRACE should predict that *vocabutaire* should activate *vocabulaire*, which should activate /l/, which in turn should inhibit /t/. In Merge there are no inhibitory connections between lexical and decision units so activation of *vocabulaire* by *vocabutaire* does not directly inhibit the decision node for /t/. The /t/ decision node can only be inhibited by another decision node, but the /l/ decision node is never activated because it gets no bottom-up support. Consequently, activations for /t/ in *vocabutaire* and /t/ in *socabutaire* are identical. Merge simulations confirm that activations of targets in items like these are both identical to the activations of the control nonwords in the plot of the Connine et al. simulations shown in Figure 4. Lexical facilitation, however, is still observed in this situation where there is no lexical inhibition. Words like *vocabulaire* are equivalent to Connine et al.'s word condition, and, as Figure 4 shows, activation rises faster there than in nonwords.

One might wonder why Merge shows no inhibition in simulating Frauenfelder et al.'s study but does show inhibition in the subcategorical mismatch simulations. The answer is that in the subcategorical case inhibition arises because the cross-spliced coarticulatory information provides bottom-up support for one phoneme (e.g., /g/), which then competes at the decision level with the activation of the final phoneme (e.g., /b/). That is, there is perceptual support for two competing phonemes. With perceptual support the bottom-up priority rule is satisfied and the activation of the corresponding decision units can be modulated by lexical activation. Note that removing the bottom-up priority rule makes no difference to the subcategorical mismatch simulations. Abandoning the bottom-up priority rule introduces some inhibition into the Frauenfelder et al. simulations, but the exact amount depends on such factors as the length of the words and the setting of the response criterion.

This simulation makes an important point that also applies to a lesser extent to models not using the bottom-up priority rule. The existence of inhibitory lexical influence (whether on decision nodes or phoneme nodes themselves) may often not be apparent when there is unambiguous bottom-up support for a single phoneme. Lexical effects will be most visible when there is some ambiguity in the input, such as when a subcategorical mismatch creates support for two possible phonemes.

5.2.3. Merge and lexical competition. Merge is fully compatible with Shortlist; in fact, in the present simulations the operation of the lexical level of Merge is a simplified version of the lexical level of Shortlist. Indeed, the ability of Merge to account for the phonetic categorization data hinges on several key assumptions that are required for Shortlist to be able to account for continuous speech recognition. The central claim of Shortlist is that word recognition is based on a process of competition between activated lexical hypotheses. There is now a considerable body of empirical evidence for competitive lexical processes from a variety of tasks (Cluff & Luce 1990; Goldinger et al. 1992; Marslen-Wilson et al. 1996; McQueen et al. 1994; Norris et al. 1995; Slowiaczek & Hamburger 1992; Vroomen & de Gelder 1995), and Shortlist is able to account for these data (Norris et al. 1995; 1997). Merge therefore also incorporates lexical competition. In fact, the ability of Merge to account for the subcategorical mismatch data depends crucially on lexical competition. Detecting phoneme targets in nonwords cross-spliced with words (W2N1), and making lexical decisions to these items, is harder than with nonwords cross-spliced with other nonwords (N3N1) precisely because of the inhibitory effects of the cross-spliced word (W2 in W2N1).

The lack of an inhibitory effect in the cross-spliced words (W2W1 versus N3W1) is also a consequence of lexical competition. At first glance, this may appear counterintuitive. Indeed, the lack of an inhibitory effect in the cross-spliced words led Marslen-Wilson and Warren (1994) to reject lexical competition. The explanation for this apparent contradiction is important because it relies on another aspect of Shortlist's account of continuous speech recognition. In Shortlist, and therefore Merge, lexical competition involves multiple interactive activation cycles (15 in the present simulations) on each phoneme (or time slice in Merge), followed by a reset of lexical activation levels before the next phoneme is processed. Both the reset and the cycles of multiple activation play important roles in helping Shortlist achieve an optimal interpretation of the input (Norris et al. 1995).

Without a reset, interactive activation models become insensitive to later-arriving information. Such models tend to settle on a single interpretation of a given stretch of speech, which they are unable to revise in the light of later input. Word nodes consistent with that interpretation maintain a high level of activation that tends to suppress the emergence of any new and more appropriate interpretations. This behaviour prevents such models from being able to account for data showing effects on word recognition due to competition from later-arriving information (e.g., Norris et al. 1995; Vroomen & de Gelder 1995). Like the reset, repeated cycles of activation on each time slice are required for optimization. The network must be allowed to iterate through a number of cycles in order to settle into an optimal interpretation of the input.

The reason competition acts to prevent an inhibitory effect in the cross-spliced words is therefore that the multiple cycles of activation at the lexical level in Merge allow the network to settle on the optimal interpretation of the input, which is of course the base word (W1), whether the word is cross-spliced with another word (W2W1) or with a nonword (N3W1). In the former case, competition certainly operates between W2 and W1, but W1 wins the competition over repeated cycles and thus dominates the activation pattern as much as it does in the absence of a strong lexical competitor (the latter case; see Fig. 2B). The lexical level thus operates in a winner-take-all fashion. As soon as evidence for the final phoneme arrives, the word target *job* has far more evidence in its favor than a competitor word like *jog* (even in W2W1), especially if the system is sensitive to the mismatch between the /b/ and the /g/ in *jog*, as in Merge, Shortlist, and the Cohort model (Marslen-Wilson

https://doi.org/10.1017/S0140525X00303247 Published online by Cambridge University Press

1993). Note also that in all of these models we would expect a mismatch at the segmental level (i.e., of a complete phoneme) to count more than a subcategorical mismatch. In Merge, the fact that /b/ mismatches the /ɡ/ in *jog* will completely counteract the effects of the match on the first two phonemes, and *jog* will actually have a negative amount of perceptual evidence in its favor (mismatching phonemes are weighted three times as heavily as matching phonemes). By the final phoneme in *job*, there will therefore be no competition from *jog*.

The exact behaviour of a network will depend on whether it is set up like Merge, where a number of cycles are computed and the network is allowed to settle to a near asymptotic state for each new input segment, or like TRACE, where only a single network cycle is computed for each new input slice. If, as in Merge, the network is allowed to run to asymptote with each new segment, it will behave in a more winner-take-all fashion. The word with the largest activation can eventually suppress the activation of its competitors until they have a very small or even negative activation. At this point the competitors will no longer have any inhibitory influence on the target, and the target's activation will be largely determined by the degree of bottom-up input it receives. At asymptote, therefore, it is possible that inhibitory effects will only be observed when competitors actually have at least as much bottom-up support as the target. Inhibition of nonword responses still occurs, however, because the competing word (in W2N1) is misleadingly activated but there is no correct competitor word available to cause the incorrect word to be inhibited.

If the network operates like TRACE, however, and runs for only one or two cycles on each time slice, all competing words will inhibit one another to some degree and the activation of the target will be influenced by the presence of competitors in both words and nonwords. This seems to be the view of competition that Marslen-Wilson and Warren (1994) have in mind when they criticize lateral inhibition. As we show below, it seems to be this feature that prevents TRACE from simulating the subcategorical mismatch data.

Marslen-Wilson and Warren (1994) also argued against lexical competition on the basis of data reported in Marslen-Wilson (1993) and, in greater detail, in Marslen-Wilson et al. (1996). In Experiment 1 in Marslen-Wilson et al. (1996), however, there was a competition effect of 31 msec. "This effect, although significant on a post hoc test . . . , is not backed up by any broader statistical support" (pp. 1381–82). Furthermore, Experiment 2 of Marslen-Wilson et al. (1996) also shows a 31 msec competition effect (see Fig. 2, p. 1388). The authors indeed admit "that the presence of priming in the no-competitor condition is indeed a competitor effect" (p. 1388) and suggest that a competition model like Shortlist (Norris 1994b) could account for their results. These data therefore appear to support rather than challenge the claim that spoken word recognition is based on lexical competition.

A common factor in many demonstrations of inhibitory competition is that the input is fully compatible with both the target word and its competitors. For example, in McQueen et al. (1994) subjects found it harder to detect *mess* in /dəmɛs/ than in /nəmɛs/. Right until the end of the stimulus /dəmɛs/, the input is fully compatible with both *mess* and competitors such as *domestic.* In general, inhibitory effects on the target are going to be greatest when the bottom-up support for the competitor is greater than that for the target. Whether competition effects will be observable (as in McQueen et al. 1994, and in the cross-spliced nonwords in the subcategorical mismatch experiments) or not (as in the word materials in the mismatch experiments) depends on the fine balance of informational support for different candidate words. In both cases, however, the data support the claim that spoken word recognition is based on an active process of lexical competition, as instantiated in both Shortlist and Merge.

5.3. Summary

The simulations show very clearly that an autonomous model with a decision process that combines the lexical and phonemic sources of information gives a very simple and readily interpretable account of the data of Marslen-Wilson and Warren (1994), McQueen et al. (1999a), Connine et al. (1997), and Frauenfelder et al. (1990). It explains inhibitory effects of competition in nonwords with subcategorical mismatches, facilitatory effects in nonwords that are more like real words relative to those that are less like real words, and lack of inhibitory effects in nonwords that diverge from real words near their end. It also explains how both facilitatory and inhibitory effects come and go according to task demands.

It is interesting to note that these simulations undermine one of the main conclusions Marslen-Wilson and Warren (1994) drew from their data. They argued that their data were evidence against phonemic representations at the prelexical level. But the present simulations use phonemic representations at this level. We could have replaced the phonemic input layer with a featural layer and achieved exactly the same ends. Either form of representation can be arranged to deliver identical inputs to the lexical and decision layers. The subcategorical mismatch data are therefore completely neutral with respect to the existence of a phonemic prelexical level of representation.

We have chosen to simulate these three studies both because they provide some of the evidence that most strongly constrains models of phonemic decision making and because we felt that quantative analysis was required to establish whether Merge could explain the detailed and interlocking pattern of results presented by these studies. It should be clear without report of further simulations, however, that Merge can also explain the other basic lexical effects observed in the literature. The explanation for lexical effects in phonetic categorization is the same as that provided for effects in phoneme monitoring: Lexical node activation can bias phoneme decision-node activation such that an ambiguous phoneme in a word-nonword continuum like *type-dype* will tend to be labelled in a lexically consistent manner (i.e., as /t/ here). Likewise, lexical activation will act to boost phoneme decision-node activation so that there tend to be more phonemic restorations in words than in nonwords.

Merge can also account for the results of Newman et al. (1997). As we discussed above, these results could be due either to differences in lexical activation, depending on lexical neighborhood, or to prelexical sensitivities to transitional probabilities between phonemes. If these neighborhood effects in categorization prove to be due to differences in degree of lexical activation, then Merge could explain those data in the same way that it can account for Connine et al.'s (1997) results. For Merge to be able to explain these

https://doi.org/10.1017/S0140525X00303247 Published online by Cambridge University Press

data – should they prove instead to be due to transitional probabilities – then, as with the Race model, a mechanism sensitive to sequential dependencies would be needed at the prelexical level. But it is clear that there is nothing about the architecture or the processing assumptions of Merge that would prevent it from accounting for such data.

Similarly, there is no need for simulations to show that Merge can account for the results of Pitt and McQueen (1998). As just discussed, the influence of transitional probabilities on compensation for coarticulation could be modelled by adding a process sensitive to these probabilities at the prelexical level – that is, at the level of processing where compensation for coarticulation is also thought to operate. Merge also has the correct architecture to explain the dissociation in lexical involvement observed by Pitt and McQueen. The lexicon can influence decisions to ambiguous fricatives at the ends of words (i.e., more /s/ responses to *jui?* than to *bu?*) via flow of activation from the lexical level to the phoneme decision nodes. But, because there is no feedback from the lexicon to the prelexical level, this lexical involvement cannot influence the compensation for coarticulation process, and thus, as observed, there is no lexical influence on the stop identifications in spite of the lexical effect in the fricative decisions. Merge can thus account for all of this evidence on lexical involvement in phonemic decision making. As we argued above, without clear evidence about the locus of the adaptation effect, it is currently not possible to determine whether the results of Samuel (1997) are problematic for Merge.

6. Comparing Merge with other models

6.1. TRACE

Marslen-Wilson and Warren (1994) showed in simulations that TRACE was unable to account for their data. First, in the response probabilities at the lexical level (used to simulate lexical decision), TRACE shows a large effect of the lexical status of the cross-splice (W2W1 vs. N3W1) for words as well as nonwords (W2N1 vs. N3N1). Second, in the nonword stimuli, the W2N1 items produce response probabilities that are as large as the probabilities for any of the word stimuli. The W2N1 nonwords should therefore have been systematically misclassified as words. The phonetic categorization results were simulated by using response probabilities calculated at the phoneme level. As in the lexical decision simulations, TRACE incorrectly shows a large effect of the lexical status of the cross-splice (W2W1 vs. N3W1) for words as well as nonwords (W2N1 vs. N3N1).

Marslen-Wilson and Warren suggest that the failure of TRACE to simulate the data is attributable to its use of lateral inhibition and top-down feedback, and to the fact that it does not use mismatch information. Furthermore, they argue that because TRACE is "the only viable candidate of the classical representational type" (p. 673), by which they mean a model using phonemes rather than features as prelexical representations, these results argue against models in which phonemes play a role as prelexical representations. We have already shown that these data are completely neutral with respect to the existence of prelexical phonemic representations and that a model instantiating lexical competition can account for the pattern of data. Why then can TRACE not account for these data? We believe that the primary reason why TRACE is unable to account for Marslen-Wilson and Warren's data is that, unlike in Merge, lexical level processes are not allowed to cycle to asymptote on a fast enough time scale. As discussed above, the model therefore incorrectly predicts competition effects in the W2W1 items. It is probable that this behaviour also causes the model to overestimate the inhibitory effect in the cross-spliced nonwords.

To test this, we constructed an interactive model with no decision nodes. The model had word-to-phoneme feedback, and, to enable unambiguous decisions at the phoneme layer, phoneme-to-phoneme inhibitory connections. It proved impossible to set parameters for this model by hand (of course, it is always extraordinarily difficult to set parameters by hand in an interactive model because any adjustment at one level interacts with other levels). However, via an optimization procedure (Powell's conjugate gradient descent method; Press et al. 1986), we did eventually find parameters that could produce activation patterns similar to those of Merge; these parameters are listed in Table 3. The interactive model then produced an acceptable simulation of the subcategorical mismatch data of Marslen-Wilson and Warren (1994), as long as it was allowed to use 15 cycles per time slice. But it was not possible to find any set of parameters that would produce a plausible simulation using only a single cycle per slice (as in TRACE). Within a single cycle the winning lexical candidate cannot completely suppress its competitors, so that the competitor effect for words is not eliminated.

This exercise clearly established that it was not the presence of word-to-phoneme feedback that caused the interactive model to fail at simulating the subcategorical mismatch data; failure only occurred when the model was, like TRACE, restricted to a single cycle. This then is presumably the explanation for Marslen-Wilson and Warren's failure to achieve simulation of their data in TRACE. With Merge-like dynamics, an interactive model could approximate the correct data pattern.

However, we did note that even the best simulation that we could achieve of the subcategorical mismatch data with

Table 3. *Parameter values used in the simulations by the Merge model and the Interactive Model (IM)*

	Merge	IM
phoneme excitation	0.311	0.173
phoneme to word excitation	0.024	0.097
phoneme to word inhibition	0.021	0.0
phoneme to phoneme inhibition	0.0	0.453
phoneme to decision excitation	0.826	0.0
phoneme decay	0.230	0.810
phoneme momentum	0.0	0.118
word to decision excitation	0.235	0.0
word to phoneme excitation	0.0	0.186
word to word inhibition	0.537	0.536
word decay	0.093	0.133
word momentum	0.0	0.002
decision to decision inhibition	0.870	0.0
decision unit decay	0.622	0.0
decision momentum	0.581	0.0
cycles per input slice	15	15

https://doi.org/10.1017/S0140525X00303247 Published online by Cambridge University Press

this interactive model was less than optimal, and certainly not equal to the fit given by the best set of parameters in Merge. Although the interactive-model simulation with 15 cycles per time slice reproduced the observed interaction between inhibition and lexical status in the cross-spliced items, it showed very large word-nonword differences not observed in the human data. With the same interactive model, with 15 cycles, we then also attempted to simulate the Frauenfelder et al. (1990) data, which appear to offer a crucial challenge to TRACE. Interestingly, although the interactive model did show a great deal of inhibition with five-phoneme words, with three-phoneme words the network produced facilitation with very little inhibition. The balance of inhibition and facilitation depends on the amount of lexical support for the target phoneme or its competitor.

The simulations using both Merge and the interactive network described above have addressed sets of data that present major problems for existing models of word recognition. The subcategorical mismatch evidence of Marslen-Wilson and Warren (1994) and McQueen et al. (1999a) is incompatible with TRACE and the Race model. We have shown that Merge can give a detailed account of the pattern of results from both the lexical decision and phonetic categorization tasks used in these studies. In simulations using an interactive model we have also shown that one of the most important factors that enables models to account for these data is the use of Shortlist/Merge-like dynamics in which the optimal result of the competition process is achieved as quickly as possible. Data from Connine et al. (1997) and Frauenfelder et al. (1990) variously challenged both the Race model and TRACE. Again, Merge accurately simulated the observed results.

The fact that the interactive model could be induced to simulate some of the results should, however, offer little consolation to proponents of interactive models. Such models have already been shown to be inconsistent with the compensation for coarticulation results of Pitt and McQueen (1998). The simulations have also emphasised the point that interactive models that attempt to combine phoneme and decision nodes are suboptimal and unstable. Small changes in parameter values lead to large changes in model behaviour. Note that although a bottom-up priority rule can in fact prevent TRACE from predicting inhibition in phoneme monitoring, it cannot help to reconcile TRACE with the data from Pitt and McQueen. The TRACE predictions in that case follow directly from TRACE's account of the Ganong (1980) effect. Top-down activation biases the interpretation of an ambiguous phoneme. If the ambiguous phoneme did not activate both alternative phonemes and hence satisfy the rule, there could be no lexical bias at all. In the Pitt and McQueen study, the interpretation of the ambiguous final phoneme of the first word is indeed subject to a lexical bias. According to TRACE, the bias must be reflected in a change in activation at the phoneme level that should induce compensation for coarticulation. But Pitt and McQueen found no compensation for coarticulation when transition probability was held constant. Therefore TRACE cannot account for the result.

Interactive models like TRACE also suffer from the limitation first raised by Cutler et al. (1987) in the context of the facilitatory effects of lexical knowledge on phoneme identification. McQueen et al. (1999a) showed that the subcategorical mismatch effect comes and goes according to the nature of the task and the stimulus materials. Although one can stipulate that the top-down connections in TRACE be modulated according to the experimental situation, it is far from clear why any interactive model should sometimes choose to forgo the supposed benefits of top-down feedback. The Merge model, on the other hand, remains fundamentally a dual-outlet model in which the decision mechanism can selectively emphasize either phonemic or lexical knowledge without in any way altering the bottom-up nature of the word recognition process itself. The major achievement of the Merge simulations is thus to show that the pattern of lexical effects in phoneme identification is fully consistent with modular models – that is, that feedback is not required in speech recognition.

6.2. A post-lexical model

Marslen-Wilson and Warren's (1994) explanation of their data proposes that decisions about phoneme identity are made on the basis of lexical representations. To account for the ability to identify phonemes in nonwords, they propose that nonwords "are perceived though the lexicon, in an analogical fashion" (p. 673). Gaskell and Marslen-Wilson (1995; 1997) present a connectionist implementation of Marslen-Wilson and Warren's theory. They tried to simulate the data in a simple recurrent network that produced both a semantic and a phonemic output. This network is identical in structure to the one described by Norris (1993) to simulate the Elman and McClelland (1988) study, apart from the fact that the output targets in that network involved just a single node being activated, and that the output that Norris labelled "lexical" is labelled "semantic" by Gaskell and Marslen-Wilson (1997). However, Gaskell and Marslen-Wilson's simulation suffers from problems similar to those of the TRACE simulation presented by Marslen-Wilson and Warren.

In their simulations of phonetic categorization decisions, Gaskell and Marslen-Wilson (1997) plot an index of the evidence for the correct phoneme that ranges from 1.0 to −1.0. The index is 1.0 when the correct phoneme is fully activated and the nearest competitor is not active, and −1.0 when the competitor is fully active but the correct response is not. The sign of this measure could not possibly be available to the decision mechanism because it is determined entirely by the experimenter's notion of what constitutes the correct response. The decision mechanism could only plausibly be expected to have available the unsigned measure of the difference in evidence between the most active phoneme and its nearest competitor. It could not additionally know which was supposed to be the correct response. However, Gaskell and Marslen-Wilson suggest that a simple threshold on this signed index produces the correct pattern of results. But the use of the signed index attributes the model with the psychic ability to know which phoneme represents the correct response on a given trial. Armed with that knowledge, the model should be able to respond before the trial even begins! If they had instead used a threshold on the appropriate *unsigned* index, the model would consistently have made the wrong response on W2N1 trials. In the W2N1 condition there was much more evidence for a competitor phoneme (presumably the phoneme in W2) than there ever was for the correct phoneme. No matter where an unsigned threshold were set, this model would always respond with the competitor phoneme rather then the correct one.

https://doi.org/10.1017/S0140525X00303247 Published online by Cambridge University Press

So, the specific simulation of the subcategorical mismatch data offered by Gaskell and Marslen-Wilson (1997) is clearly unsatisfactory. However, even the general principles of the Marslen-Wilson and Warren model are challenged by data on phonemic decision making, including those of McQueen et al. (1999a). Marslen-Wilson and Warren's suggestion that phoneme identification is a postlexical process immediately faces a problem in accounting for data showing that the magnitude of lexical effects in phoneme monitoring can be made to vary depending on the exact nature of the task (Eimas et al. 1990; Eimas & Nygaard 1992) or even the nature of filler items used in the experiment (Cutler et al. 1987). The results of McQueen et al. (1999a) show that the inhibitory effects that Marslen-Wilson and Warren themselves reported also come and go with changes in experimental procedure. Cutler et al. (1987) suggested that their results were problematic for TRACE because in TRACE the lexical effect can only be modulated by varying the strength of all of the phoneme-to-word feedback connections. However, if phoneme identification operates by the lexical analogy process suggested by Marslen-Wilson and Warren, then lexical identification and phoneme identification are inextricably locked together. Phoneme identification depends on lexical activation. Reduction in the level of lexical activation might slow phoneme monitoring overall, but it could not possibly eliminate the lexical advantage. Phonemes in words should always benefit from the presence of matching lexical representations. Similarly, inhibitory effects could not be eliminated without removing the lexicon, which would make phoneme identification impossible.

Gaskell and Marslen-Wilson's (1997) model also links the phonemic and lexical levels in such a way that "lexical" and competitive effects are not open to strategic control. Lexical effects in their model would remain even if the semantic nodes were eliminated completely after training. As there is no feedback from the semantic nodes to the phoneme nodes, the semantic nodes can be removed completely without affecting the performance of the phoneme nodes in any way. The only common level of representation is that of the hidden units. The hidden units and their connections to the phoneme units could not be altered without changing phoneme identification performance on both words and nonwords. It is worth noting also that in Norris's (1993) simulations of "lexical" effects in a recurrent network, these effects remained even when the network was no longer trained to identify words. The effects were simply due to the network learning the statistical regularities in the input. It is therefore not even clear that the behaviour of the Gaskell and Marslen-Wilson model at the phonemic level is in any way a reflection of a true lexical influence.

6.3. FLMP

In the FLMP (Massaro 1987; 1989b; 1998; Oden & Massaro 1978), perceptual decisions are based on three processes: (1) evaluation of the degree to which the input supports stored prototypes, where the evaluation of each source of information is independent of all others; (2) integration of the outputs of these evaluations; and (3) decision making based on the relative goodness of match of each response alternative. FLMP is a generic model of perceptual decision making, and has been applied in a wide variety of different perceptual domains, most notably in the domain of multimodal speech perception (e.g., the integration of auditory and visual speech information). FLMP is a mathematical model tested by measuring its ability to account for experimental data through parameter estimation procedures. The FLMP has been applied in this way to the question of lexical involvement in phonemic decision making (Massaro 1996; Massaro & Oden 1995).

FLMP has several parallels with Merge. In FLMP, as in Merge, lexical and phonemic sources are integrated to produce a decision. Furthermore, both models assume that decisions are based on continuous rather than categorical representations (activation values in Merge's decision nodes, continuous truth values in FLMP), and that decisions are based on goodness of match (via inhibition between decision nodes in Merge, via a Relative Goodness Rule [RGR] in FLMP). Finally, both models are autonomous, since neither allows a feedback loop from the lexicon to the prelexical level. (Although Massaro, 1996, characterizes the FLMP as nonautonomous, it is certainly autonomous in the sense of the term used here because the model excludes top-down interactions; Massaro argues that FLMP is not autonomous because it involves integration, but Merge is autonomous and has integration).

In spite of these close similarities, Merge and FLMP have two fundamental differences. Both of these differences relate to how the models are seen in the wider context of spoken language understanding. The first is that FLMP has independent evaluation of lexical and phonemic sources of information. However, in one sense at least, Merge is not independent.

Although the concept of independence is central to FLMP, Massaro appears to use the term in two slightly different ways. In one sense, independence refers to the property of a system whereby stimulus and contextual information are combined so that context has an effect only on bias and not sensitivity (e.g., Massaro 1978; 1989a). In an excellent review of these issues, McClelland (1991) describes models with this mathematical property as "classical" accounts of context effects. Such models have their origins in Signal Detection Theory and Luce's (1963) theory of choice. However, Massaro also uses the term independence to refer to architectural properties of the system. According to Massaro (e.g., Massaro & Oden 1995), the basic perceptual *processes* (e.g., phoneme and word recognition) are also independent. Sometimes systems that are independent in the classical sense are also independent in the architectural sense, but this need not always be so. For example, as Massaro argues, the combination of acoustic-phonetic and phonotactic information is independent in both of these senses. Massaro (1989a) and Massaro and Cohen (1983) studied the influence of phonotactic context on the perception of stimuli that were ambiguous between /r/ and /l/. There were more /l/ responses in the context /s*i/ than in the context /t*i/, where /*/ represents the ambiguous phoneme. Here the context and the phoneme are genuinely independent in both senses. Perception of the ambiguous phoneme should have no effect on the representation of the preceding phonological context. These two architecturally independent sources of information (context and ambiguous phoneme) are then combined independently as specified by classical models of perception. Massaro showed that FLMP can fit these data very well, whereas a nonindependent (in both senses) model like TRACE cannot.

However, the architectural and signal-detection versions

https://doi.org/10.1017/S0140525X00303247 Published online by Cambridge University Press

of independence do not always pattern together. McClelland (1991) showed how the interactive activation model could be modified by the addition of noise (the Stochastic Interactive Activation Model or SIAC) so as to generate independence in the same manner as classical models, even though processing of stimulus and context remained architecturally nonindependent (note that in TRACE the phonotactic context delivering the bias is derived from the lexicon rather than from a direct sensitivity to transition-probability within the phoneme level).

A more serious problem for FLMP is that the assumption of architectural independence becomes impossible to sustain when we consider how to account for lexical involvement in phonemic decision making. In any model of word recognition, the degree of support for a lexical hypothesis must be some function of the degree of support for its component segments: If there is good perceptual evidence for a /g/, for example, there is also good perceptual evidence for words containing /g/. But this is not so in FLMP. In FLMP, support for /g/ in a lexical context (e.g., the extent of support for /g/ due to the word *gift* given the string /ʔɪft) does not depend on whether the string begins with an unambiguous /g/, an ambiguous phoneme, or an unambiguous /k/ (Massaro & Oden 1995). Since this evaluation of contextual support for /g/ is made through comparison of the input with the stored representation of *gift*, this implies that information consistent (or inconsistent) with the initial /g/ does not influence the goodness of match of the input to the word. In other words, architectural independence depends on the remarkable assumption that the support for a word has nothing to do with the perceptual evidence for that word. Note that if architectural independence is abandoned in FLMP, it becomes difficult to test classical independence: We can no longer assume that the level of contextual bias should remain constant across a phonetic continuum, as it does in all FLMP simulations, even though a given amount of lexical and phonemic information may still be combined classically.

Thus, although FLMP can fit the data on lexical involvement in phonetic categorization with materials such as /ʔɪft/ (Massaro & Oden 1995), it does so while making an indefensible assumption about word recognition. In contrast to FLMP, the Merge account of word recognition necessarily violates architectural independence: A lexical node's activation depends on the activation of the prelexical nodes of its constituent phonemes.

The second difference between the models concerns the extent to which they make claims about processes. FLMP is not a model of perception in the same way that Merge and TRACE are. Whereas Merge and TRACE attempt to model the mechanisms leading to activation of particular words or phonemes, FLMP takes as its starting point the probabilities of giving particular words or phonemes as responses in an experiment. FLMP makes no claims about how the independent processes of phoneme and word recognition actually operate. In FLMP, there is no specification of how words are selected and recognized given a speech input; in Merge, word recognition depends on competition between lexical hypotheses. In fact, as we have argued above, the dynamics of competition between candidate words are an essential feature of Merge's account of the subcategorical mismatch data. Without a competition-based account of lexical access to determine the appropriate inputs to the decision process, FLMP has difficulty making detailed predictions about the outcome of the subcategorical mismatch experiments. More generally, FLMP as it currently stands offers no explanation for performance in any lexical decision task. Note that Massaro (1987, p. 281) has argued that there should be no inhibition within any processing level in FLMP. As we have pointed out above, inhibition between phoneme decision nodes in Merge has the same function as the RGR in the FLMP. But Merge also uses inhibition at the lexical level. Word segmentation and recognition cannot be achieved by a simple mechanism like the RGR (McQueen et al. 1995). Not only does FLMP not have an account of word recognition, the nature of the computations required for this task appear to be beyond the scope of the FLMP framework.

These two differences between FLMP and Merge reflect two major problems with how lexical information is processed in FLMP. A third problem with FLMP concerns the way the model has sought to account for the compensation for coarticulation data of Elman and McClelland (1988). FLMP successfully fits the stop identification data in both the unambiguous and ambiguous fricative contexts (Massaro 1996). But it does so because the compensation process (the modulation of interpretation of a stop consonant contingent on the preceding fricative) operates at integration, not during evaluation. The preceding context (with an unambiguous or ambiguous fricative) provides a bias that is integrated with the evidence for the following stop only after evaluation of the acoustic-phonetic evidence for the stop has been completed. This account is highly implausible, since it suggests that the compensation process operates only to modify phonemic decisions. How can it then be of value in lexical access? Special mechanisms that allow the perceptual system to deal more effectively with coarticulation are unlikely to have developed unless they were of value in word recognition. But in the FLMP account, the contextual bias due to word-final fricatives is unable to influence the recognition of the following stop-initial word, and can only influence decisions about the stop. It would appear that the only way in which this bias could influence word recognition would be if the output of the phonemic decision process were in turn used as input to the word recognition process for the stop-initial word. But this would go against a fundamental assumption of FLMP, by making it interactive: Word recognition would depend on a top-down feedback loop involving the preceding word. In any case, Pitt and McQueen's (1998) results suggest that contextual influence on fricative-stop compensation is due to sensitivity to transitional probability and not to lexical involvement, and that the process of compensation has a prelexical locus; both of these findings run counter to the FLMP account.

All three of these problems with FLMP have arisen because phonemic decision making has not been considered in the light of the constraints of everyday language processing. As we will describe below, Merge has been developed explicitly within such a framework; the account of phonemic decision making that Merge offers is one that supports the Shortlist model's account of spoken word recognition. These fundamental problems with FLMP lead us to reject it as a plausible account of lexical involvement in phonemic decision making.

https://doi.org/10.1017/S0140525X00303247 Published online by Cambridge University Press

7. Phonemic decisions and language processing

In the above sections we have demonstrated – in the specific context of performance in phonemic decision-making tasks – how different sources of information can be merged to influence a decision response, without any need to create feedback loops in the process-model architecture. We now return our attention to the more general theoretical questions at issue in this paper, and to the relation of phonemic decision making to everyday language processing.

First, consider again whether, by assuming that phonemic decision making is accomplished by a mechanism separate from the processing sequence in spoken-word recognition, we might have violated Occam's precept by constructing a model with a more complicated architecture than is strictly necessary. As we have argued above, however, such an accusation cannot be maintained. The available experimental evidence on phonemic decision making is incompatible with models that incorporate feedback between lexical and prelexical processing, but also incompatible with strictly feed-forward models that allow no lexical influence on phonemic decisions other than in words. The Merge architecture is thus precisely in conformity with Occam's principle: It is the simplest architecture compatible with the empirically derived data.

Furthermore, it is a natural architecture in that phonemic decisions are of course not part of normal speech recognition. As we argued in describing the Merge model, the requirements of prelexical nodes and decision nodes are very different. A decision process must select some interpretation of the input, but prelexical processing itself must preserve any ambiguity in the input. As the utility and popularity of phonemic decision-making tasks in this area of psycholinguistics attest, phonemic decisions are something that listeners can readily do. It is clear that the ability to make phonemic decisions does not develop prior to the ability to recognize words, however, and indeed does not necessarily develop as a consequence of word-recognition abilities either. Young listeners acquire spoken-word recognition abilities from the first year of life and onwards (Jusczyk 1997), but the ability to make phonemic decisions is learned much later. Learning to read in an alphabetic orthography certainly facilitates the capacity to make decisions about phonemic structure, since there is evidence that illiterates find such decision making much more difficult than literates do (Morais et al. 1979; 1986). Speakers of languages that are written without an alphabet also have difficulty with some phonemic tasks (Mann 1986a; Read et al. 1986). However, it is also the case that children's demonstrated capacity to engage in phonemic manipulation tasks is, in languages with a phonological structure that encourages phonemic awareness, a sign of readiness to learn to read (Bradley & Bryant 1983; Liberman 1973). Further, both illiterates and speakers of languages with nonalphabetic orthographies can, with appropriate training, easily perform phonemic decisions (Cutler & Otake 1994; Morais 1985). Outside the laboratory, phonemic decision making may be called upon in playing language games, in writing unfamiliar names, in teaching children to read, or in retrieving words that are "on the tip of the tongue." The phonemic decision making mechanism is, in other words, a natural part of mature language processing abilities, but it is clear that it is developed separately from, and later than, spoken-word recognition. In the terms of the modelling framework for language processing adopted in the present work, Merge is separate from and developed later than Shortlist.

Shortlist can be conceived of as explaining normal word recognition, whereas Merge explains how listeners access the prelexical information present in the normal bottom-up speech recognition system. Between them, they explain how a bottom-up system can account for the complex pattern of data about spoken-language processing derived from phonemic decision tasks. Thus, we see Merge as an integral part of the language processing model that incorporates Shortlist. We assume, moreover, that it embodies the main structural features of Shortlist. Since phonemic decisions are in general not part of spoken-word recognition, it seems plausible to assume, as in Merge, that phoneme decision nodes are built in response to the need to make such decisions. That is, we view the Merge network as, like the Shortlist network, being created dynamically as required.

This enables the decision nodes in Merge to be set up in response to a particular experimental situation. For example, we suggested that perhaps only one phoneme decision node is set up when listeners are monitoring for one particular phoneme. In the subcategorical mismatch experiments, and in other phonemic categorization tasks, only the two response alternatives may need to be set up. They may be set up when the subject in an experiment is given instructions, or perhaps they develop over the course of the first part of an experimental session. In either case, they are set up in response to the need to make an explicit phonemic decision. As with the lexical nodes in Shortlist, therefore, the decision nodes in Merge are built on the fly. For all common phonemic decision tasks, irrespective of which phonemes from the full inventory are targets in a given situation, the behaviour of the decision nodes should be equivalent to that observed in the present simulations.

In other words, the performance of Merge is not dependent on its size as instantiated in our present simulations. An instantiation with a larger lexicon and a full phoneme inventory would simulate the experimental data as well as the demonstration version presented here. The nodes at the prelexical input level are not connected to one another, so the number of representations at this level should have little impact on performance. It is possible that with a full phoneme inventory, partial activation of similar-sounding phonemes could occur if the prelexical input representations were sufficiently finely tuned. Thus, for example, if /d/ were added to the phoneme inventory for the present simulations, it could be weakly activated whenever *job* or *jog* were presented as input. But this activation would have little effect on processing at the lexical and decision stages, because the activation it passed on would be rapidly suppressed by competition from the more highly activated word or phoneme nodes.

The lexical level of Merge can be viewed as a simplification of Shortlist's lexical level. In Shortlist, only the most highly activated words (from a passive lexicon of over 25,000 words) make it into the shortlist to compete actively for recognition. The shortlist is a small competitive network that is built anew for every fresh input sequence. Thus there is a strict limit on the number of words in this lexical workspace in Shortlist, and therefore, by extension, in Merge. Simulations reported in Norris (1994b) show that Shortlist performs well even when only two words are allowed to compete with each other for a given stretch of input (as in the present Merge simulations). Varying the size limit on the shortlist has little effect on recognition perfor-

https://doi.org/10.1017/S0140525X00303247 Published online by Cambridge University Press

mance. Thus we assume that Merge would work in a very similar way even with a very large passive lexicon, because the number of words activated and taking part in a phonemic decision at any moment in time would remain very small. Since, as we described above, the number of decision nodes is also assumed to be determined by the phonemic decision task at hand, we are therefore quite confident that the Merge model would perform similarly even if it were scaled up at all three of its levels of processing.

Finally, note that although the decision nodes in Merge must represent phonemes in order to do, for example, phoneme monitoring, we can in fact remain agnostic as to the exact form of the prelexical input representations feeding both Merge and Shortlist. We have already discussed the possibility that these representations could, for instance, be featural and need not represent phonemes explicitly. In that case, part of the task of Merge would be to translate the representations used for lexical access into the representations more suited to the performance of phonemic decision tasks.

We believe that the Merge account of phonemic decision making is a natural account that accords with the known role of such decision making in everyday language processing, and at the same time the most theoretically economical account compatible with the existing evidence. The phoneme decision nodes in Merge are not a part of everyday speech comprehension, and in experimental situations they are built on the fly to cope with the specific task demands. The tasks that require the decision nodes, however, are a valuable source of data about continuous speech recognition. As we have argued above, for example, data from these tasks suggest that speech recognition is based on competition between candidate words and that it is an autonomous, data-driven process. If we are to answer questions about the prelexical and lexical levels of speech processing, then phonemic judgment tasks provide us with an extremely powerful tool. But we require an explicit and testable model of task performance, such as Merge, if we want to use the data from such tasks to establish how normal speech comprehension is achieved.

8. Conclusions

The task of recognizing speech is surely so difficult that one might expect the human speech recognition system to have evolved to take full advantage of all possible evidence that would be of assistance in recognizing spoken words. We think it has done so, solving the problem in the most economical way, by allowing no feedback loops to exist, by passing information in one direction only, and by delegating decision making about phonemic representations to a separate mechanism.

In this target article we have marshalled arguments against interactive models involving feedback loops in the context of the hotly contested theoretical issue of modelling phonemic decision making in spoken-language recognition. Our attack has been waged on three main fronts. First we have shown that the widely assumed benefits of interaction are illusory. Modular theories can perform as well as, or even better than, interactive theories, and thus there is simply no need for interaction. Second, the data are inconsistent with the model that is the main standard bearer of interaction. Finally, we have shown how we can account for the data with a new modular model.

We have shown that, given a prelexical processor that is performing at its full efficiency, word recognition cannot possibly benefit from lexical feedback. Although interactive theories appear to be based on an assumption that interaction will necessarily produce more effective processing, we have shown that this assumption is mistaken: There is in fact no benefit to be gained from allowing lexical knowledge to interact with prelexical processing. As Frauenfelder and Peeters' (1998) simulations show, lexical feedback in a model like TRACE improves recognition for some words, but impairs performance on just as many other words.

Lexical feedback might offer some advantages in phoneme identification by making phoneme decisions line up with decisions made at the lexical level. This might be helpful if the input consists of real words, but exactly the same ends can be achieved with noninteractive models too. It is not interaction that offers the benefit, but the process of combining two different kinds of information – the lexical and the prelexical. A model like Merge therefore exploits the advantages of interaction but avoids the disadvantages. On grounds of parsimony, if models with and without feedback can both account for a finding, the model without feedback should be chosen.

These theoretical arguments in favor of strictly bottom-up models would carry little weight if there were reliable empirical support for interaction. However, as we demonstrated, results that had appeared to provide strong evidence for interaction turn out to have been artifactual. Much other evidence also weighs against models incorporating feedback. None of this completely eliminates the possibility that some interactive model might be constructed that would be consistent with the data. For example, the data from Pitt and McQueen (1998) argue against the specific details of TRACE, but one could presumably formulate a model in which lexical and phonemic information interacted yet compensation for coarticulation was driven from a earlier autonomous featural level. Such revisions are always possible; and, more fundamentally, as Forster (1979) pointed out many years ago, a general claim that processes interact cannot be refuted. We cannot prove the null hypothesis that no interaction takes place. In our view, however, this remains one of the best arguments for adopting autonomous theories as the default option in this field: Occam's razor dictates that we do so.

The Merge model is, we believe, the most economical explanation compatible with the currently available data on phonemic decision making, while also being a natural account in the framework of our knowledge about spoken-language processing in general. But it, too, remains a model whose purpose is to be put to empirical test. The empirical challenge of testing theories can only be met in the context of specific models. Newell (1973) warned us that we "can't play 20 questions with nature and win." To test general claims about interaction in the absence of particular theories is to join in that game of twenty questions. However, the two models we used as the extreme examples in our demonstration, TRACE and the Race model, have both been specific enough to be tested empirically and have been found wanting. Merge is also specific. Whether it passes future empirical challenges remains to be seen. Its role in the present paper has been to serve as an existence proof of a noninteractive account of the empirical data, and hence an argument in favor of our central claim about models of speech recognition: Feedback is never necessary.

https://doi.org/10.1017/S0140525X00303247 Published online by Cambridge University Press

ACKNOWLEDGMENTS

We thank Steve Levinson for suggesting the name of the Merge model to us. We also thank Peter Jusczyk, Antje Meyer, and Joanne Miller for very helpful comments on the original version of this paper, and Dom Massaro, Arthur Samuel, Mark Pitt and two anonymous reviewers for later comments. Please address correspondence to: Dennis Norris, MRC Cognition and Brain Sciences Unit, 15 Chaucer Road, Cambridge, CB2 2EF, UK; E-mail: dennis.norris@mrc-cbu.cam.ac.uk.

NOTE

1. In fact, in TRACE, this is not strictly true. Feedback in TRACE can actually alter the pattern of sensitivity (Massaro 1989a), but this is a consequence of the simplifying assumptions made in implementing it. All processing is performed without noise, and response probabilities are derived by applying the Luce choice rule to the resulting activations. The noise simulated by this rule is independent of all other processing. Top-down feedback can alter the signal (arbitrarily) during processing without altering noise. That is, feedback can alter sensitivity. When noise is added to these systems at input, or during processing, and the Luce rule is replaced by the Best One Wins decision rule, feedback will alter both signal and noise together, leaving sensitivity unchanged (McClelland 1991). So, the tendency of TRACE to show effects of feedback on sensitivity is really the result of the simplified, but unrealistic, noise-free processing assumption. This problem can be remedied by replacing the IAM with a Stochastic Interactive Activation Model (McClelland 1991).

Open Peer Commentary

Commentary submitted by the qualified professional readership of this journal will be considered for publication in a later issue as Continuing Commentary on this article. Integrative overviews and syntheses are especially encouraged.

Merging information versus speech recognition

Irene Appelbaum[1]

Department of Philosophy, University of Montana, Missoula, MT 59812.
appel@selway.umt.edu

Abstract: Norris, McQueen & Cutler claim that all known speech recognition data can be accounted for with their autonomous model, "Merge." But this claim is doubly misleading. (1) Although speech recognition is autonomous in their view, the Merge model is not. (2) The body of data which the Merge model accounts for, is not, in their view, speech recognition data.

Norris, McQueen & Cutler claim that all known speech recognition data can be accounted for with an autonomous model ("Merge") that merges lexical and prelexical information. This claim elicits an immediate objection. Autonomous models, by definition, are non-interactive. Merging, on the face of it, is a form of interaction. How can a model that features interaction be defended as one that does not? In response, Norris et al. waver between denying that merging is interaction and denying that it is problematic, if it is. In particular, they imply that the part of Merge that does what they call "phonemic decision-making" need not be autonomous, because phonemic decision-making is not part of ordinary speech recognition. But this response gives rise to a second difficulty. If phonemic decision-making is distinct from speech recognition, then the data Merge accounts for – phoneme decision data – is not speech recognition data, making their central claim misleading.

Four features distinguish Merge from TRACE. (1) Merge consists of two phonemic stages: a first "processing" stage and a second "decision-making" stage. (2) Because the first (processing) stage lacks intra-node inhibition, its output preserves phonemic ambiguity. (3) Connections between lexical and phonemic information are flexible – built "on the fly" in response to task demands and attentional shifts. And crucially: (4) Lexical information does not affect the first (processing) stage; instead, lexical information and the output of the first phonemic (processing) stage are merged in the second (decision) stage, which outputs a determinate phonemic identification.

So: the first (processing) stage gets no feedback from either the lexicon or the second (decision-making) stage; the lexicon gets input only from the first (processing) stage; and the second (decision-making) stage gets information from both the first (processing) stage and the lexicon. Given this architecture, the first (processing) stage does seem to meet normal standards of autonomy. The question is what to say about the second (decision-making) stage. Is it also autonomous? It would seem not to be since "phoneme decision units are . . . directly susceptible to facilitatory influences from the lexicon" (sect. 5.1, para. 4). Yet if the second stage is not autonomous, then neither is the Merge model overall, and Norris et al.'s claim to account for all speech recognition data with an autonomous model fails.

Faced with this objection, one strategy is to argue that the second, decision-making stage is autonomous after all. At times Norris et al. seem tempted by this strategy. Though not stated explicitly, this view is implied by their contrasting "merging," "combining," and "integrating" (sect. 5) lexical and prelexical information with interactive processing. For example: "It is not interaction that offers the benefit, but the process of combining two different kinds of information – the lexical and the prelexical" (sect. 8, para. 4).

How can merging or integration be considered a form of autonomy, rather than interaction? Their idea seems to be that the decisive criterion for distinguishing autonomous from interactive processes is *feedback.* That is, Norris et al. seem to consider a system directly influenced by higher-level information to still be autonomous as long as the intra-modular information flow involves no feedback loops. But this is a problematic move. Reinterpreting the interactive/autonomous distinction in terms of a feedback/non-feedback distinction artificially eclipses the theoretical space for other (i.e., non-feedback) kinds of interactive processing. Moreover, to automatically classify all non-feedback top-down processes as autonomous is contrary to normal usage. However, although Norris et al. gesture in this direction, I do not think it is (or should be) their considered view.

Instead, Norris et al. seem to endorse a second strategy: acknowledge that decision-making is not autonomous, but deny that this threatens the autonomy of speech recognition. Norris et al.'s crucial move in defense of this view is to deny that the second, decision-making stage is part of ordinary speech recognition.

One might worry that Norris et al. are simply stipulating that phonemic decision-making is outside speech recognition in order to justify their autonomy claim. It would certainly appear this way if phonemic decision-making were a necessary final stage of the overall phonemic task. But an extremely interesting and consequential claim of their target article is that it is not. In their view, the ambiguity-preserving output of the first phonemic stage is as complete as phonemic processing gets in *ordinary speech recognition.* The explicit phoneme decision tasks that subjects perform in the lab, they claim, are distinct from the kind of phonemic processing that goes on in ordinary speech recognition. Real-world speech recognition requires only the first stage; only explicit phoneme decision tasks (of the sort performed in laboratories) require both.

And sure enough, if second-stage phonemic decision-making is

https://doi.org/10.1017/S0140525X00303247 Published online by Cambridge University Press

not part of speech recognition, then it does not matter whether its merging processes are classified as autonomous or interactive. Either way, speech recognition remains autonomous, because first-stage phonemic processing – the only stage it requires – unproblematically is.

Nevertheless, this line of response creates its own complications. For one thing, even if speech recognition is autonomous, the Merge model is not, because whether or not phonemic decision-making is part of speech recognition, it is still part of Merge. The system that accounts for the data (i.e., Merge) is not autonomous; and the system which is autonomous (i.e., phonemic processing) does not account for the data. This leads to an even more consequential difficulty. It is no longer clear how the body of data that Merge is advertised as accounting for bears on the processes of speech recognition. For this body of data is derived from exactly the sorts of experimental phoneme decision tasks that Norris et al. claim are different from ordinary speech recognition.

Norris et al. are aware of this difficulty. They respond by claiming that from the data on explicit phoneme decision tasks we can infer what ordinary speech recognition is like, since the model that accounts for the explicit data (theirs) contains an autonomous speech recognition component. This is an interesting claim. Even if it is their considered view, however, their overarching claim that Merge is an autonomous model that accounts for all the data from speech recognition would still need to be qualified. Merge is not autonomous, even if one of its components is; and the data that Merge accounts for is not (in Norris et al.'s view) speech recognition data, even if it may shed indirect light on the process of speech recognition.

ACKNOWLEDGMENTS

I would like to thank Brian C. Smith for discussion and comments, and the Center for the Study of Language and Information, Stanford University, for supporting this research.

NOTE

1. Author is also affiliated with the Center for the Study of Language and Information, Stanford University, Stanford, CA 94305, irenea@csli.stanford.edu.

Lexical biases are useful

José R. Benkí

Program in Linguistics, University of Michigan, Ann Arbor, MI 48109-1285.
benki@umich.edu www-personal.umich.edu/~benki/

Abstract: I present two criticisms: (1) Merge is a model of performance in speech perception experiments but not an ecologically valid model integrating both word recognition and speech perception, and (2) Merge's implementation of the variability of lexical effects must be made more precise or the model is indistinguishable from its alternatives.

Norris, McQueen & Cutler have increased the empirical coverage of the Shortlist/Race models by adding a phoneme decision network in which lexical and prelexical information are combined while maintaining independence between lexical and phonemic processing. I focus my critique here on the shortcomings of Merge that arise from it being presented as a model of performance in speech perception experiments rather than as an ecological valid model integrating both word recognition and speech perception.

Following a convincing argument against feedback in speech perception, Norris et al. propose a severely constrained role for top-down information in word recognition and speech perception. As a result, Merge merely describes certain experimental data on phoneme decision-making rather than provide a teleologically explanatory basis for such data. In addition, while the design of the phoneme decision network allows Merge to explain lexical effects which the antecedent Shortlist/Race models cannot explain, the plasticity of some of the connections in the model make Merge potentially indistinguishable from other either interactive or autonomous models of speech perception such as TRACE and Shortlist/Race.

The first criticism of the Merge model concerns the purpose of the phoneme decision network as presented by Norris et al., which seems to do little else than to explain some very subtle and specific effects of the lexicon on the processing of nonwords. This criticism is inconsequential if these data are mere idiosyncrasies of the way humans perform in speech perception experiments. However, if these lexical effects represent important qualities of the system, then the Merge model is a description instead of an explanation.

Norris et al. present a powerful and convincing attack on the need for feedback in speech perception. However, feedback represents only one way that top-down information could affect phonemic processing, and the impressive critique of feedback obscures the very real and useful nature of other sorts of top-down information for word recognition and speech perception. What possible benefit to the speech perceiver could lexical effects on the processing of nonwords represent? In general, top-down information in any pattern recognition system limits the possible interpretations available to the perceiver, making the process more robust and efficient in a potentially noisy environment. In this light, lexical effects on the processing of nonwords can be seen as manifestations of a general lexical bias which makes lexical access more efficient and accurate given a priori knowledge of the relative likelihoods of different linguistic interpretations of utterances.

That this bias is under active control by the listener is evidence that top-down information is used by speech perceivers to optimize their task. Classic findings by Miller et al. (1951) on speech intelligibility show that more limited response alternatives, sentential context, and repetition can increase the intelligibility of speech in noise, in effect increasing particular lexical biases. Recent computational modeling by Nearey (forthcoming) is consistent with a lexical bias explanation for Boothroyd and Nittrouer's (1988) finding that words are more perceptible in noise than nonwords. The effects of lexical biases can also be actively shifted in the other direction, as observed by Cutler et al. (1987), when listener expectations are shifted from words toward monosyllabic nonwords.

Is the phoneme decision network in the Merge model the appropriate mechanism for implementing what seem to be listener strategies to make their task more efficient and accurate? As it stands, the phoneme decision network models listener behavior in a particular class of speech perception experiments. The network could be made to represent the output of a module that computes the phonological structure of the incoming signal, combining lexical and prelexical information in the process. This module (or equivalent function) is necessary in any theory that seeks to explain the perception of both words and nonwords. Such an implementation seems consistent with Norris's (1986) proposal that criterion bias models can explain both context and frequency effects.

In the spirit of parsimony, a model with a single mechanism for explaining lexical bias effects in nonwords would be superior to a model with multiple mechanisms. The simulation data presented by Norris et al. show that Merge is capable of explaining the other basic lexical effects observed in the literature, as the authors note (sect. 5.3, para. 5), but the plasticity of the mechanisms in Merge makes the model too powerful. Given that the phenomena at hand are variable, plasticity in the model is necessary. However, if the model can account for any type of inhibitory or facilitatory effects, then we are left with the situation that currently exists for interactive and autonomous models, that they largely predict the same data and are therefore indistinguishable. The specific details of how the Merge model can account for all of the data should be worked out and evaluated. To their credit, Norris et al. suggest one type of data which would falsify autonomous models in general (and the Merge model in particular): increased sensitivity in pho-

https://doi.org/10.1017/S0140525X00303247 Published online by Cambridge University Press

neme discrimination in words over nonwords as diagnosed by signal detection theory (sect. 3.3, para. 9).

At the same time, many questions remain about the extent and details of the influence of the lexicon on the processing of nonwords. Task demands, stimulus uncertainty (Pitt 1995; Pitt & Samuel 1993), and context have been cited as factors in modulating these effects – not to mention other phenomena, such as sequential probabilities, which can have similar effects (Cairns et al. 1995; Norris 1993; Vitevitch & Luce 1998). Some of these variables are under the control of the listener while other effects seem to persist, and a comprehensive theory must explain this variability adequately. The more precise the hypotheses that are under consideration are, the more useful the experimental results are likely to be toward increasing our understanding of speech perception.

Merging auditory and visual phonetic information: A critical test for feedback?

Lawrence Brancazio[a] and Carol A. Fowler[b]

[a]*Department of Psychology, Northeastern University, Boston, MA 02115.* [b]*Haskins Laboratories, New Haven, CT 06511; Department of Psychology, University of Connecticut, Storrs, CT 06269; Yale University, New Haven, CT 06520.* **brancazio@neu.edu** **fowler@haskins.yale.edu**

Abstract: The present description of the Merge model addresses only auditory, not audiovisual, speech perception. However, recent findings in the audiovisual domain are relevant to the model. We outline a test that we are conducting of the adequacy of Merge, modified to accept visual information about articulation.

Norris, McQueen & Cutler have made a provocative contribution to the debate on the relationship between phoneme identification and lexical access with their contention that an additional decision stage is more economical and therefore preferable to feedback. However, Merge, like most models of speech perception (including TRACE, Race, Shortlist, and Marlsen-Wilson & Warren's [1994] model, but excepting FLMP) fails to accommodate the effects of visible phonetic information on phoneme and word identification. Perceivers do use this information, though, so the models are incomplete. We will suggest that coupling questions about the ways in which lexical knowledge and visual articulatory information affect phone identification can be informative. Researchers have recently been considering visual contributions to speech perception with reference to lexical access (Brancazio 1998; 1999; Iverson et al. 1998), and we propose a useful extension of this research line.

It is now well known that visible phonetic information from a speaker's articulating mouth has a marked effect on perception. This is most clearly demonstrated by the "McGurk effect" (McGurk & MacDonald 1976): When an unambiguous auditory signal specifying one syllable (for example, /ba/) is dubbed with a face producing a different syllable (such as /ga/), the resulting percept is often changed (to /da/). We can ask, as we do of lexical knowledge, where this visual contribution occurs; evidence suggests that audiovisual integration occurs early in the perceptual process (e.g., Green & Miller 1985).

For example, Fowler et al. (in press) demonstrated compensation for coarticulation (the phenomenon exploited by Pitt & McQueen [1998] in their phonotactics study; sect. 4.3) using stimuli in which the coarticulating contexts were distinguished optically but the compensations were applied to segments distinguished acoustically. That is, visible phonetic information induced the compensation effect. Fowler et al. used stimuli in which use of relevant transitional probabilities between phonemes would not have given rise to the compensations that were found and therefore cannot underlie the compensation effect. Accordingly, they concluded that their compensation effect must arise in perceivers' remarkable sensitivity to talkers' coarticulatory behavior.

In the framework of Merge, visual contributions to phoneme perception, by virtue of their interaction with the compensation for coarticulation mechanism, must be prelexical and thus be unaffected by lexical knowledge. Therefore, evidence of lexical influences on audiovisual integration would pose a strong challenge to Merge.

In fact, Brancazio (1998; 1999) has demonstrated that phoneme identification of audiovisually discrepant stimuli is affected by lexical status. That is, visually-influenced responses are more frequent when they form words than nonwords. Moreover, this effect parallels the auditory Ganong (1980) effect in its basic properties. One parallel is the finding of a "neighborhood" effect on phoneme identification in nonwords like that of Newman et al. (1997). This effect cannot be due to sensitivity to transitional probabilities, contrary to Norris et al.'s suggestion (sects. 4.5.1; 5.3). Consider a set from Newman et al.'s stimuli, "bowth-powth" and "bowsh-powsh." "Bowth" has a larger neighborhood than "powth," but "powsh" has a larger neighborhood than "bowsh," and, accordingly, "p" responses are more frequent for a "powsh-bowsh" than a "bowth-powth" continuum. However, the sequential probabilities of p(/b/ | /aU/) and p(/p/ | /aU/) cancel out across the stimuli, and the higher-order sequential probabilities p(/b/ | /aUʃ/), p(/p/ | /aUʃ/), p(/b/ | /aUθ/) and p(/p/ | /aUθ/) are all zero.[1] Therefore, in Merge, these "neighborhood" effects can only arise via the lexical, not the prelexical, pathway. (This may ultimately prove relevant for evaluation of Merge as our understanding of neighborhood effects improves.)

Brancazio's finding of a lexical influence on the McGurk effect might be due to lexical feedback on audiovisual integration. However, the outcome is also consistent with an autonomous model such as Merge in which audiovisual integration occurs prelexically, and lexical influence emerges at a later decision stage, affecting the interpretation of ambiguous outputs from the prelexical level of processing.

We are currently conducting experiments to pin down the proper interpretation of Brancazio's findings, and thereby provide an audiovisual test of Merge. To achieve this, we are exploiting the lexically-induced shift in the McGurk effect (more word-forming than nonword-forming visually-influenced responses) and the McGurk effect-induced compensation for coarticulation (visual determination of an acoustically ambiguous syllable-final consonant as /l/ or /r/, which influences the identification of a following syllable as /da/ or /ga/). We are testing whether a lexically-induced "strengthening" of the McGurk effect will increase visually-induced compensation for coarticulation. If visual contributions to phoneme identification are prelexical as Fowler et al. have shown (see also Green & Miller 1985) and, if, as Norris et al. claim, the lexicon does not affect prelexical processing, then there should be no second-order lexical effect (via the McGurk effect and compensation for coarticulation) on following stop identification. That is, a lexically induced increase in identifications of the syllable-final consonant as /l/ should not increase /ga/ over /da/ identifications. However, a positive outcome would suggest that lexical feedback occurs.

Finally, attending to the audiovisual domain highlights a further distinction between Merge and Massaro's (1987) FLMP, which, as Norris et al. point out, are computationally quite similar. In the FLMP, all sources of information are integrated at a single stage. Therefore, in the audiovisual Merge, but not in the FLMP, audiovisual integration precedes and is unaffected by lexical processing. Thus, concern for audiovisual speech perception in the lexical domain offers new means of distinguishing contemporary models. We hope that other researchers will exploit this potential.

ACKNOWLEDGMENTS

Preparation of this commentary was supported by NIH grants HD01994 to Haskins Laboratories and DC00373 and DC00130 to Northeastern University.

NOTE

1. Additionally, C_1:C_2 probabilities, ignoring the identity of the intervening vowel – p(/b/ | /ʃ/) vs. p(/p/ | /ʃ/) and p(/b/ | /θ/) vs. p(/p/ | /θ/) – can

https://doi.org/10.1017/S0140525X00303247 Published online by Cambridge University Press

be ruled out: Newman et al. also used a "beyth-peyth-beysh-peysh" series, whose stimuli have identical C1:C2 probabilities to the "bowth" set. Critically, "beyth" is a high-neighborhood nonword but "bowth" is low, and the two stimulus sets produced neighborhood effects in opposite directions.

Inhibition

Cynthia M. Connine and Paul C. LoCasto

Department of Psychology, State University of New York at Binghamton, Binghamton, NY 13902. **connine@binghamton.edu**
psychology.binghamton.edu/index.htm

Abstract: We consider the motivation for the principle of bottom-up priority and its consequence for information flow in Merge. The relationship between the bottom-up priority rule and inhibitory effects is also discussed, along with data that demonstrate inhibitory influences in phoneme monitoring.

At a basic level, the Merge model bifurcates perceptual and decision processes during spoken word recognition. In doing so, Merge permits lexical knowledge to influence the phoneme decisions while perceptual processes operate autonomously according to a bottom-up priority principle. The explicitness of the model and its development hand in hand with empirical findings render it an important counterpoint to interactive models such as TRACE.

We focus here on the bottom-up priority rule and its relationship to the sequence of processing in the phoneme decision nodes. The bottom-up priority rule at its essence directs lexical knowledge as to when it can inform phoneme decisions. The time course of lexical involvement is constrained so that it is invoked only when there is bottom-up evidence for a given phoneme. As a consequence, activation of a phoneme decision node cannot be driven solely by lexical knowledge. A primary empirical motivation for the bottom-up priority rule is the lack of evidence for inhibitory effects for phoneme detections when the phoneme mismatches an otherwise lexical carrier (Frauenfelder et al. 1990). Some results from our laboratory investigating inhibitory effects suggest that the conclusions of Frauenfelder et al. may be premature. In our experiments, a nonword control that had little resemblance to a real word was included to serve as a baseline comparison against the mismatch condition (see also Connine 1994). A nonword control that is very similar sounding to a real word (as in Frauenfelder et al.) may inadvertently produce inhibition similar in kind to a word carrier, thus masking any inhibitory effects on the mismatching target phoneme. Our second innovation was to include two kinds of mismatching phonemes: one condition where the mismatching phoneme was very similar to the lexically consistent phoneme (e.g., *chorus – chorush*) and a second condition where the mismatching phoneme was very dissimilar (*chorus – chorum*). The similar mismatch target is predicted to show inhibitory effects while the dissimilar mismatching condition may benefit from additional attentional resources (following Wurm & Samuel 1997) that override inhibition. Note that either an inhibitory effect in the similar mismatch compared with the nonword control (where phoneme target and carrier stimulus is held constant) or against the dissimilar mismatch (where phoneme target is held constant but carrier stimulus is varied) would count as evidence for lexical inhibition. Here we examine the former prediction and avoid a comparison across different target types. Participants monitored for the 'sh' or the 'm' (stimulus presentation was blocked). Timing marks were placed at the onset of the target segment. The similar mismatch condition showed a small (14 msec) but significant inhibitory effect relative to the nonword control ($t(50) = 6.1$, $SD = 52$, $p < .01$; *chorush* vs. *golush,* 488 msec vs. 474 msec). In contrast, the dissimilar mismatch condition showed a facilitatory effect relative to the nonword control (45 msec, $t(50) = 3.5$, $SD = 93$, $p < .01$; *chorum* vs. *golum,* 636 msec vs. 681 msec). Similar to Wurm and Samuel (1997), we attribute the facilitatory effect to recovery from inhibition as a result of focused attentional resources. Neither TRACE nor Merge currently incorporates cognitive factors such as attention, and our results highlight the important role attentional processes play in language processing.

Could Merge be modified to model correctly our observed inhibitory effects? As Norris et al. explicitly state (sect. 5.2.2, para. 6), removing the bottom-up priority rule does result in a small degree of inhibition. However, this would effectively deny one major motivation of the model. The principle of bottom-up priority tips the balance of influence in favor of bottom-up input and against lexical knowledge. In doing so, this permits phoneme nodes (the speech signal) to drive initial activation of phoneme decision nodes and any subsequent lexical effects. Norris et al. argue that this characteristic prevents the occurrence of "hallucinations" in processing speech and that this is a desirable property for spoken language processing. The motivation for this claim is that the system should be designed to be error-free and to optimize behavior. But is error-free, maximally optimal processing an accurate characterization of processing continuous speech? We do not believe so. In continuous speech, the bottom-up input is not always deterministic with respect to conscious experience or segmental identity and in some instances is missing information entirely. Errors in processing can occur along with re-computation and recovery. Another domain that makes the same point is the strong evidence for syntactic preferences in parsing locally-ambiguous sentences. The parser has a built-in error-producing mechanism – garden path sentences are difficult to process precisely because the strategy (or frequency-based bias) produces errors (or conflicts) in parsing. The resulting mechanism can produce efficient, errorless representations as well as sub-optimal processing (requiring a re-computation). In spoken word recognition, interactive mechanisms may maximize processing for some classes of words while making no difference or even having detrimental effects on others.

The implications for model development are clear. Lexical information provides a powerful source of constraint with behavioral consequences that are consistent with interactive architectures such as TRACE. As argued over a decade ago, the representational compatibility of lexical and segmental levels render them ideal for interaction (Connine & Clifton 1987).

ACKNOWLEDGMENTS

Support was provided by National Institutes of Health (DC02134). We thank Tom Deelman (now at Southwest Bell Communications) for helpful discussion.

Features and feedback

Tobey L. Doeleman,[a] Joan A. Sereno,[b] Allard Jongman,[b] and Sara C. Sereno[c]

[a]*Department of Psychology, University of Washington, Seattle, WA 98195;* [b]*Department of Linguistics, University of Kansas, Lawrence, KS 66045-2140;* [c]*Department of Psychology, University of Glasgow, Glasgow G12 8QF Scotland, United Kingdom.* **doeleman@u.washington.edu**
faculty.washington.edu/doeleman **{sereno; jongman}@ukans.edu**
www.linguistics.ukans.edu/Dr_{Sereno; Jongman}.html
ssereno@psy.gla.ac.uk
www.gla.ac.uk/departments/psychology/staff/Sereno.html

Abstract: Our commentary outlines a number of arguments questioning an autonomous model of word recognition without feedback. Arguments are presented against the need for a phonemic decision stage and in support of a featural level in a model including feedback.

Norris, McQueen & Cutler make a strong claim against the inclusion of feedback in models of word recognition, and describe an autonomous model in which prelexical phonemic information is merged with lexical information at a phonemic decision stage. The target article clearly describes the defining characteristics of au-

https://doi.org/10.1017/S0140525X00303247 Published online by Cambridge University Press

tonomous and interactive models of word recognition and presents a concrete model with testable hypotheses, challenging researchers to empirically evaluate its predictions. Although the authors have shown that the model can account for an impressive number of experimental findings, we identify four aspects requiring additional support. The following comments invite the authors to fine-tune their arguments.

The first point concerns Norris et al.'s justification of the independence of the prelexical level and the phonemic decision stage. They claim that these levels are necessarily independent because perceptual processing and decision making have different requirements and therefore cannot be performed effectively by the same units. If this were the case, then one might argue for an additional lexical decision stage to be connected to the lexical processing level. The argument for independence of processing and decision making is expanded in the discussion of phonemic decisions and language processing in section 7, where the authors claim that phonemic decision making is not a part of normal speech recognition. They support this claim with evidence of difficulties in the ability of subjects to make phonemic decisions due to illiteracy or non-alphabetic orthographies. The authors conclude that the ability to make phonemic decisions is a learned process, separate from word recognition abilities. But this conclusion ignores a whole body of research on infant speech perception, where the head-turn paradigm and high-amplitude sucking procedure are widely accepted techniques used to demonstrate phoneme discrimination in infants. If these infants do not possess a phonemic decision making mechanism, what guides their behavioral responses?

The simplicity of Merge is further compromised by the fact that the prelexical phonemic level is, in essence, duplicated at the phonemic decision stage. Thus, the flow of information from the lexical level to the phonemic decision stage is not clearly different from what might be found in a model with feedback from the lexical level to a prelexical phonemic level. Although it may be argued that bottom-up theories are more parsimonious than interactive theories, if bottom-up theories must posit an additional decision level that mimics the phoneme level, then interactive theories including feedback may be preferable. Moreover, Norris et al. seem to ignore the architecture of the mechanism they are trying to model. Within the well-studied visual or auditory systems of monkeys, for example, there are as many feedback as feedforward connections between different areas.

The second point involves the nature of the prelexical level in Merge. The authors most commonly refer to this level as phonemic, and use phoneme nodes in their simulations. The Merge model is explicitly designed so that lexical information does not influence prelexical phonemic information. However, research in phonetics, phonology, and spoken word recognition suggests that features may be the more critical units in speech perception and word recognition. Recall that the inability to recover from mispronunciations in models with interacting lexical and phonemic levels is a major motivation for the autonomous architecture adopted in the Merge model. If a featural level were added, lexical and phonemic information could interact while the featural level would ensure recovery from mispronunciations. Lexical information can influence phonemic decisions without information in the speech input being discarded or modified. If input is inconsistent with lexical knowledge, there will not be a risk of misperceiving speech while at the same time retaining all top-down lexical effects. No phonemic decision unit would need to be postulated. A model which allows featural, phonemic, and lexical levels, with feedback between the phonemic and lexical levels would then seem more parsimonious.

Third, although Norris et al. leave open the question of whether featural information may be represented, they explicitly state that the prelexical level includes sequential probability information (which could be argued to be a duplication of information that is necessarily contained at the lexical level). Storing sequential probability information at the prelexical level is necessary in order for Merge to account for data showing that phonotactic probabilities affect nonwords as well as words. It also allows the authors to further their argument that Merge differs significantly from models with feedback, in part because Merge allows for a dissociation of effects due to compensation for coarticulation and lexical bias. The former is said to be a result of prelexical processes, while the latter results from lexical involvement. Norris et al. argue that an interactive model such as TRACE is not capable of such a dissociation, since lexical bias in TRACE should produce a change in activation at the phoneme level thus inducing compensation for coarticulation. But since Pitt and McQueen (1998) showed that coarticulatory compensation effects are most likely due to transitional probabilities, the TRACE simulation of compensation for coarticulation must have simulated the network's learned knowledge of sequential patterns at the lexical level rather than the compensation process intended at the prelexical level. While it is true that compensation for coarticulation (or transitional probability) and lexical bias effects may both stem from information stored at the lexical level in TRACE, this does not imply that the former effect will necessarily induce the latter nor that the network is incapable of dissociating these effects.

Finally, the empirical data that are modeled in Merge derive from phoneme monitoring, phonetic categorization, and phoneme restoration tasks. While these tasks have provided much information about speech processing, they impose an artificial situation on normal listening and require subjects to make decisions they would normally not make, introducing issues of response bias and ecological validity. Real-time data from ERPs and localization data from fMRI, as well as eye movement data during reading, may be able to elucidate the stages/loci of processing. Results across a variety of methodologies should allow an optimal, biologically plausible model of word recognition to e-Merge.

Modeling lexical effects on phonetic categorization and semantic effects on word recognition

M. Gareth Gaskell

Department of Psychology, University of York, Heslington, York YO10 5DD United Kingdom. **g.gaskell@psych.york.ac.uk**
www-users.york.ac.uk/~mgg5/

Abstract: I respond to Norris et al.'s criticism of Gaskell and Marslen-Wilson (1997). When the latter's network is tested in circumstances comparable to the Merge simulations in the target article, it produces the desired pattern of results. In another area of potential feedback in spoken word processing, aspects of lexical content influence word recognition and our network provides a simple explanation of why such effects emerge. It is unclear how such effects would be accommodated by Merge.

The Marslen-Wilson and Warren (1994) data on subphonemic mismatch are central to the arguments in the target article. Norris et al. criticize the Gaskell and Marslen-Wilson simulations of these data on a number of counts. Most importantly, they argue that our simulation of the phonetic categorization data is insufficient, because it uses a signed measure of the target-competitor activation difference to predict human performance. The reason we preserved the sign of the measure was because we wished to interpolate between our data points in order to derive predicted response times. Transforming our data into an unsigned measure too early would lead to incorrect predictions in cases where a time-sequence of difference scores crossed the zero line. However, Norris et al. are correct in noting that a suprathreshold deflection either side of the zero line must be counted as triggering a response, which means that our simulations would predict false positives or no responses for at least some of the nonword conditions in this experiment, depending on where the threshold is set. This is another example (one we originally missed) of the unsatisfactory

https://doi.org/10.1017/S0140525X00303247 Published online by Cambridge University Press

phonological representation of nonwords in the network, discussed on pp. 634–36. In essence, the network was too strongly influenced by the phonological patterns of existing words learnt during training.

Note that the Merge lexical decision simulation comes dangerously close to having a similar problem. The activation threshold of 0.2 is carefully chosen (see target article Fig. 2). A threshold below 0.18 would predict false positives in the W2N1 condition, whereas a threshold above 0.24 would predict misses in the spliced word conditions. It seems plausible that this narrow window would close entirely if any system noise or variability between different lexical items was incorporated, or if competitor environments were larger than the maximum of two lexical items used.

The Gaskell and Marslen-Wilson simulations attempted to recreate the details of the time-course and competitor environment involved in the perception of these stimuli. It is interesting to note that employing some of the simplifications used in Merge allows simulation of the original data without the problem isolated by Norris et al. I retrained our original network using a competitor environment similar to Norris et al. (each test item paired with 0–2 competitors of equal frequency, intermixed with a large number of lower frequency monosyllabic words). Subcategorical mismatch was simulated by replacing the final segment of a word or nonword with a mixture of phonemes on a 0.8:0.2 ratio (Norris et al. used 0.85:0.15, but again a range of weightings was used, with little qualitative difference between the results). The model then simulated the experimental results without predicting false positives (see Fig. 1). Applying appropriate thresholds to the difference score graphs provides a reasonable fit to the relevant experimental data for the phonetic decision and "yes" lexical decision data. Likewise, the subthreshold activity for the lexical decision "no" responses is consistent with the experimental results.

It would seem that both models can accommodate the data from Marslen-Wilson and Warren (1994). This should not be too surprising, since both models use a level of phonological representation in which acoustic information can interact with lexical knowledge without feedback. Norris et al. also question whether the effects in the Gaskell and Marslen-Wilson model are truly lexical, pointing out that connectionist networks can demonstrate "lexical" effects in the absence of a lexical identification task in training by taking advantage of the statistical regularities in the training sequence. The boundary between these two explanations is becoming increasingly blurred, since much research has shown that a good way to identify word boundaries and hence isolate lexical units is through the use of statistical regularities (Cairns et al. 1997). In any case, our model does use explicit training on word identity, and so regardless of the lack of feedback from semantic units during test, the network can be thought of as representing words implicitly at the hidden unit level.

The Gaskell and Marslen-Wilson model does clearly fail to simulate any variability of lexical effects on phoneme categorization. McQueen et al. (1999a) argue that the lexical effects observed by Marslen-Wilson and Warren (1994) disappear when the experimental task is simple enough to perform without engaging the lexicon. Presumably normal speech perception is not as simple as this, and so it seems reasonable to retain a focus on processes in speech perception that do engage lexical processing (as does, for example, the Shortlist model of Norris [1994b]). Merge provides an excellent model of phonemic decision making in cases where the listener focuses on the low-level form of speech.

Another interesting case of potential feedback, not discussed in the target article, is in the interaction between lexical semantic information and word recognition. In auditory and visual word recognition, aspects of a word's meaning often influence its recognition in isolation. The concreteness of a word's meaning, the number of meanings associated with a word form, and even the relationship between those meanings can all affect recognition performance (Azuma & Van Orden 1997; Rodd et al. 1999). For a model that treats lexical decision as an operation feeding off a distributed representation of lexical content, these influences are to be expected in certain circumstances. Functionally, lexical content interacts with word recognition despite the preservation of bottom-up processing. However, a model like Merge, where lexical identification depends on activation of abstract nodes, would seem to require a set of feedback links from the level of lexical content to the word recognition level in order to incorporate these effects within a single recognition level. In some sense, then, our model is an even stronger proponent of the no feedback principle. It provides another demonstration of how multiple sources of information can be integrated in word recognition without recourse to feedback.

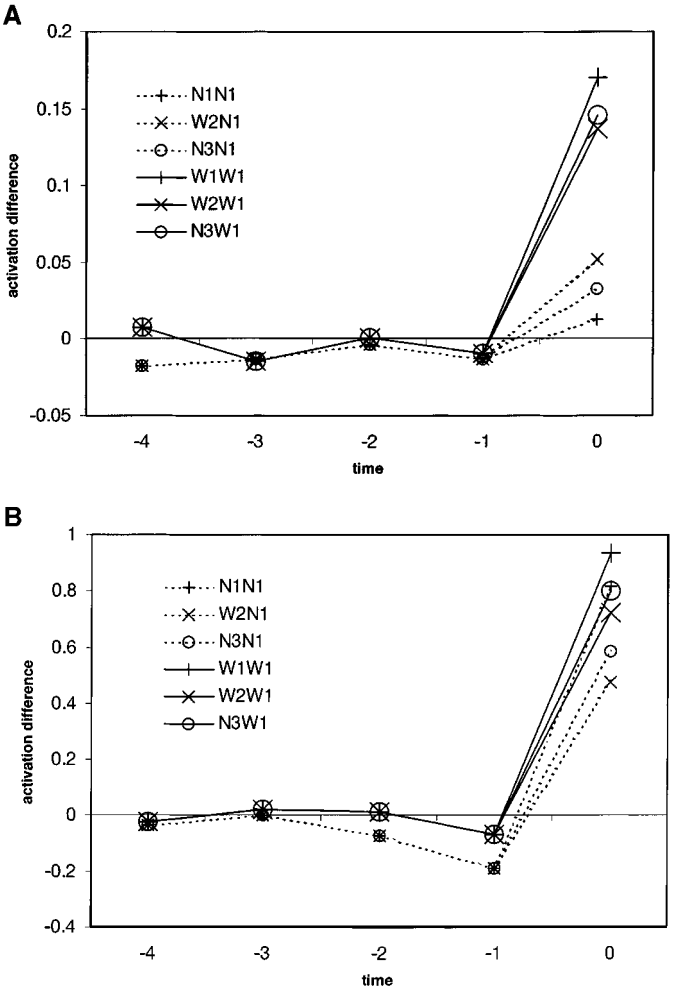


Figure 1 (Gaskell). Simulation of lexical decision (Fig. 1a) and phonetic decision (Fig. 1b) experiments. In each case, the measure is the difference between the target activation and the nearest competitor. The x-axis marks the time course of presentation of cross-spliced stimuli, with zero being the final segment, containing potentially mismatching information about two consonants. In the lexical decision simulation the activation difference is derived from distances in multidimensional space, whereas in the phonetic decision simulation it is derived from the activations of the relevant phoneme nodes in the Gaskell and Marslen-Wilson network. In each simulation, reaching a threshold activation difference is assumed to trigger a response. The conditions in the key are explained in the Norris et al. target article section 4.6 and their Figure 2.

https://doi.org/10.1017/S0140525X00303247 Published online by Cambridge University Press

One phonemic representation should suffice

David W. Gow[1]

Neuropsychology Laboratory, Massachusetts General Hospital, Boston, MA 02114. **gow@helix.mgh.harvard.edu**

Abstract: The Merge model suggests that lexical effects in phonemic processing reflect the activation of post-lexical phonemic representations that are distinct from prelexical phonemic input representations. This distinction seems to be unmotivated; the phoneme fails to capture the richness of prelexical representation. Increasing the information content of input representations minimizes the potential necessity for top-down processes.

Norris et al. offer the Merge model as an existence proof that a completely feedforward model can account for apparent top-down effects in phoneme processing. The key to Merge's performance is a processing reorganization that multiplies the phoneme into a prelexical input representation and a quasi-post-lexical representation that receives some activation from the lexicon. The authors thus skirt the necessity for top-down feedback by asserting that the critical phonemic representation needed to explain lexical effects is post-lexical, not prelexical, as is frequently assumed. To some, this may sound like the equivalent of blaming one's misdeeds on a heretofore unknown evil twin. This is unfortunate, because Norris et al.'s suggestion that lexical effects in phoneme processing reflect the phoneme's postlexical status, while not entirely novel (Gaskell & Marslen-Wilson 1997), is an interesting one. I would like to explore a slightly different approach to the story that perhaps avoids the distraction of unnecessarily multiplying phonemic representations, and addresses the ultimate utility of lexical influences on speech perception.

There is no need for two phonemic representations. The claim that two phonemic representations are needed is not well-motivated. Norris et al. base the distinction on three arguments. The first is that listeners hearing mispronounced words show a disconnect between word recognition and phoneme recognition. The authors oversell the point. Evidence from a variety of paradigms (cf. Marslen-Wilson & Welsh 1978) including shadowing, mispronunciation monitoring, and monitoring for the surface forms of phonologically modified segments, suggests that listeners may fail to appreciate segmental alterations while occupied with the business of recognizing words. Furthermore, even a single phonemic representation can account for a disconnect between word recognition and phoneme recognition, because over the course of processing, a fully interactive model allows for initial bottom-up activation of phonemes which may accurately reflect input. Only over time do lexical factors swamp this activation. Thus, the same set of nodes can reflect different patterns of activation when sampled at different points in time.

The second motivation is the observation that earlier autonomous models with only one phonemic level fail to accurately model the full range of lexical effects in phoneme processing. This is a misuse of Occam's razor, because it restricts the possible range of models to ones premised on the unproven doctrine of autonomy. This is a particularly circular argument given that their defense of autonomy is based on the claim that autonomous models are inherently simpler than interactive ones.

Finally, Norris et al. justify the division based on the notion that the two types of representations must show different activation dynamics. They suggest that phoneme decision nodes show lateral inhibition because output must always be unambiguous. Conversely, they argue that prelexical representations cannot show lateral inhibition because this could miscategorize some ambiguous signals and derail lexical access. Neither post-lexical competition nor the absence of competition in prelexical representations is ultimately indispensable at this stage in our understanding of word recognition. First, it is unclear how ambiguous the real speech signal is. As I will argue in a moment, the speech signal does not offer the kind of ambiguity that motivates interactive accounts. Furthermore, there is no principled reason why phonemic decision nodes have to produce unambiguous output. The use of confidence ratings, goodness judgments, and reaction times in phonemic categorization tasks reveals listeners are not always confident identifying the phonemes they hear (cf Miller 1994).

Input representations make a difference. Granting that the postlexical representation of phonemes may be a notion worth pursuing, we are left in need of an appropriate prelexical representation to serve as input into the system. On this point Norris et al. are quite flexible, suggesting that the exact prelexical representation one chooses does not affect the dynamics of the model as a whole. I would argue that the nature of input representations is crucial to the performance of any model. Specifically, I would like to suggest that the necessity for lexical influences on perceptual processes is contingent on ambiguity in the input representation.

Consider two potential ambiguities that occur in normal speech. English place assimilation may cause an apparent shift in place of articulation that makes the phrase *right berries* sound like the phrase *ripe berries,* or something halfway between the two as judged by listeners in an offline perceptual task. However, in phonological priming tasks, listeners show access of *right,* but not *ripe,* even after hearing strongly modified tokens (Gow, submitted a). Moreover, they appear to use this perceived modification to anticipate the labial place of the onset of *berries* (Gow, submitted b). Acoustic analysis of these tokens suggests that assimilation produces place cues intermediate between those associated with the coronal /t/ and the labial /p/. In this case, listeners have access to more information than traditional phonemic categories provide.

Similarly, Manuel (1992) examined natural tokens of the word *support* in which the first vowel appears to have been completely deleted, leaving no vocalic pitch periods between the voiceless /s/ and /p/. Listeners hearing such tokens might be expected to report having heard the word *sport,* but instead report having heard *support.* Analysis of these tokens reveals that the /p/ is strongly aspirated, marking it as a syllable onset, rather than part of a consonant cluster. In both cases, listeners make use of acoustic information that does not correspond to traditional phonemic categories to recover information that appears to be lost from the signal. In the case of complete vowel neutralization this information even serves to correct a potential misalignment between lexical and input representations that would block recognition of *support* by existing phoneme-based models.

Word recognition largely consists of the computational problem of forming the right mapping between input representations and lexical representations. Top-down processes may facilitate this process when input is perceptually ambiguous. However, if the input representation is sufficiently rich and unambiguous, there may be little need for such processes. I would suggest that a deeper consideration of the nature of prelexical representation may shed additional light on the issue of autonomy in spoken word recognition.

NOTE

1. The author is also affiliated with the Department of Psychology, Salem State College.

The trouble with Merge: Modeling speeded target detection

Jonathan Grainger

Laboratoire de Psychologie Cognitive, Université de Provence, 13621 Aix-en-Provence, France. **grainger@up.univ-mrs.fr**

Abstract: The model of phoneme monitoring proposed by Norris et al. is implausible when implemented in a localist connectionist network. Lexical representations mysteriously inform phoneme decision nodes as to the presence or absence of a target phoneme.

Norris, McQueen & Cutler provide a further example of how cascaded activation networks with localist representations can be

https://doi.org/10.1017/S0140525X00303247 Published online by Cambridge University Press

used to describe the mechanics of human cognition (see Grainger & Jacobs 1998; and Page 2000 for further discussion of localist connectionism). However, their proposed extension of the Shortlist model of spoken word recognition (Norris 1994b) to the question of modeling phoneme detection latencies is seriously flawed. In this commentary I will explain why the particular solution adopted by Norris et al. cannot be correct when applied within the more general localist connectionist framework that they adopt.

The problems arise as soon as one attempts to provide a detailed, computationally explicit, account of performance in response-limited laboratory tasks, where RT (response time) is the main dependent measure. Norris et al. are to be congratulated for even wanting to take this step, so often avoided by cognitive modelers. In response-limited laboratory tasks the participant must somehow create an artificial stimulus category-response category mapping. One way to do this is to map the activation in some relevant and identifiable[1] dimension of cognitive activity generated by the stimulus onto the appropriate motor response. A criterion can then be set on the activation level of that specific dimension (or on activity in the appropriate motor response units) in order to tell the system that there is sufficient evidence to justify response output.

The goal then for the modeler of speeded target detection tasks is to find some psychologically plausible dimension of activity that could be the basis of this task-specific mapping process. For the lexical decision task (speeded word/nonword classification), for example, Grainger and Jacobs (1996) proposed that two dimensions of activity could simultaneously map onto a positive response in that task: unit activity in whole-word representations, and global lexical activity (summed across all positively activated word representations). Global lexical activity is also used to shift a negative response deadline, so that the more a nonword resembles a real word, the longer it takes to respond negatively to that stimulus. Norris et al. apply the same general approach in their simulations of lexical decision performance.

For the phoneme monitoring task, the only plausible dimension is activity in phoneme representations (or some configuration of activation of feature units that correspond to phonemes). Activity in whole-word representations cannot be used since this activity does not provide direct information about whether a given phoneme is present or not. This is the basic problem faced by Norris et al.: how to model lexical influences on phoneme monitoring without word-phoneme feedback in a localist connectionist network? The solution offered by Norris et al. is both psychologically and computationally implausible. On each trial of a phoneme detection task the decision nodes must be connected on the fly to all whole-word representations that contain that phoneme. Admittedly, using Shortlist as the underlying framework implies that this mapping will only be made for a small number of words on any given trial. However, this solution does not avoid the central question here: just how does the system that is building these connections on the fly know that a given word contains a given phoneme? The only way to have access to that knowledge in a localist connectionist model like Merge or Shortlist or TRACE, is to look at the activity in phoneme representations. Whole-word representations only provide information about their constituent phonemes to the extent that they are connected to phoneme representations that are simultaneously active with the whole-word representation. The modeler of course knows which words contain the target phoneme, but Norris et al. are now giving Merge that "psychic ability" for which they quite rightly accuse the model of Gaskell and Marslen-Wilson (1997).

Finally, it should be noted that this problem is specific to response-limited paradigms where a response is prepared on-line during target processing. Data-limited paradigms with untimed responses (e.g., the Reicher-Wheeler task) do not offer the same challenge. In the latter, various sources of information derived from the target can be combined at leisure to formulate a response. For example, if a participant believes that she perceived the word TABLE when tested in the Reicher-Wheeler paradigm, then she can use knowledge of word spellings to infer that the letter T and not the letter C must have been present in the first position (Grainger & Jacobs 1994). It is for this reason, and only for this reason, that Grainger and Jacobs found that they could dispense with word-letter feedback in providing an account of word superiority effects observed with this paradigm. Future research should examine the same range of effects in speeded letter detection, the visual analog of phoneme monitoring.

NOTE

1. Identifiability of the relevant dimension is not a trivial problem for a homunculus-free cognitive system. One solution is for the system to monitor activity during presentation of the specified target in order to isolate target-specific brain activity.

Brain feedback and adaptive resonance in speech perception

Stephen Grossberg

Department of Cognitive and Neural Systems, Center for Adaptive Systems, Boston University, Boston, MA 02215. **steve@cns.bu.edu**
www.cns.bu.edu/Profiles/Grossberg

Abstract: The brain contains ubiquitous reciprocal bottom-up and top-down intercortical and thalamocortical pathways. These resonating feedback pathways may be essential for stable learning of speech and language codes and for context-sensitive selection and completion of noisy speech sounds and word groupings. Context-sensitive speech data, notably inter-word backward effects in time, have been quantitatively modeled using these concepts but not with purely feedforward models.

Norris et al. argue that "top-down feedback does not benefit speech recognition" and that "no experimental data imply that feedback loops are required for speech recognition. Feedback is accordingly unnecessary" (Abstract). They carry this position perhaps as far as it can go, and nicely describe how their feedforward Merge model can explain some data at least as well as the feedback TRACE model and the feedforward Race model. They focus on TRACE as a representative feedback model because it is "the main standard bearer of interaction" (sect. 8). This is a debatable assumption because TRACE has major conceptual and data-predictive problems that are not shared by other feedback models (Grossberg et al. 1997a). On the conceptual side, TRACE is not a real-time model, cannot self-organize, and experiences a major combinatorial explosion. On the data side, TRACE cannot explain a host of data in which backward effects contextually alter speech percepts. FLMP also has such problems. Norris et al. are also selective in their choice of psychological and neural data with which to support their thesis, and underplay serious conceptual problems with their own model that feedback models have already overcome.

Massive and selective feedback processes exist in every cortical and thalamic region (Felleman & Van Essen 1991). Norris et al. are claiming that these processes play no role in speech recognition. In fact, neural models have recently suggested how the laminar circuits of neocortex merge feedforward, horizontal, and feedback pathways to elegantly achieve three goals: (1) stable development of cortical connections and adult learning; (2) seamless fusion of bottom-up and top-down processing, whereby top-down feedback modulates, matches, and attentively selects bottom-up data that are consistent with learned top-down hypotheses; and (3) a synthesis of analog representation and coherent binding of distributed information that is called *analog coherence* (Grossberg 1999a; Grossberg et al. 1997b).

Norris et al. do not explain how a feedforward model can explain classical phonemic restoration data: Let a listener hear a broad-band noise followed rapidly by the words "eel is on the. . . ." If this word string is followed by "orange," then "noise-eel" sounds like "peel"; if by "wagon," it sound s like "wheel"; if by "shoe," it

https://doi.org/10.1017/S0140525X00303247 Published online by Cambridge University Press

sounds like "heel" (Warren 1984; Warren & Sherman 1974). If some formants of the expected sound are missing from the noise, then only a partial reconstruction is heard (Samuel 1981a; 1981b). If silence replaces the noise, then only silence is heard, and the sentence meaning changes, for example, consider "eel is on the shoe." These results strongly argue that the feedforward signal is not what is consciously heard. Instead, contextual feedback from the meaning of the entire sentence "feeds backwards in time" across several words to select those noise formants that are consistent with a contextually sensitive top-down expectation. This top-down matching process cannot, however, "create something out of nothing." It can only select and focus attention on what is already in the feedforward data stream. This attentive process can take from 100 to 200 msec. to generate a conscious percept. It demonstrates an intimate interaction between lexical and prelexical processes.

Adaptive resonance theory (ART) models explain such data as properties of brain resonances that focus attention upon important bottom-up data while stabilizing brain development and learning (e.g., Boardman et al. 1997; Cohen & Grossberg 1986; Grossberg 1978; 1980; 1986; 1995; 1999b; Grossberg & Stone 1986). The time scale of conscious speech is identified with the time needed for interacting bottom-up and top-down processes to achieve resonance. The matching properties help to stabilize brain development and learning.

There are many other examples of backward effects in time. Repp (1980) studied categorical perception of VC-CV syllables. He varied the silence interval between the VC and CV syllables in [ib]-[ga] and [ib]-[ba]. If the silence is short enough, then [ib]-[ga] sounds like [iga] and [ib]-[ba] sounds like [iba]. Remarkably, the transition from [iba] to [ib]-[ba] occurs after 100–150 msec more silence than the transition from [iga] to [ib]-[ga]. This is a very long interval for a feedforward model to bridge. Moreover, whether fusion or separation occurs at a given silence interval is context-sensitive. These data have been quantitatively explained by resonant fusion in the case of [iba] and resonant reset in the case of [iga] (Grossberg et al. 1997a). They illustrate the ART hypotheses that "conscious speech is a resonant wave" and that "silence is a temporal discontinuity in the rate with which resonance evolves."

Repp et al. (1978) varied the silence interval between the words GRAY CHIP and the fricative noise duration in CH. They hereby generated percepts of GREAT CHIP, GRAY SHIP, and GREAT SHIP. Remarkably, increasing silence duration transforms GRAY CHIP into a percept of GREAT CHIP, and increasing noise duration can transform it into a percept of GREAT SHIP. Why should more silence or more noise in a future word convert a past word GRAY into GREAT? Why should more noise remove the CH from CHIP and attach it to GRAY to form GREAT, leaping over a silent interval to do so, and becoming detached from its contiguous word? These effects have also been quantitatively simulated by ART (Grossberg & Myers 1999).

The Merge model shares some key processes with ART, such as competition between activated lexical hypotheses, multiple interactive activation cycles, and reset events (Grossberg 1980; Grossberg & Stone 1986). But Merge also has serious weaknesses due to its feedforward structure. It keeps lexical and prelexical computations independent until they are merged at the decision stage. How this scheme can naturally explain the backwards-in-time data above is unclear. Merge's feedforward decision stage is, moreover, not a real-time physical model: "the word nodes cannot be permanently connected to the decision nodes . . . the connections . . . must be built on the fly, when the listener is required to make phonemic decisions (sect. 5.2.1) . . . decision nodes . . . set up in response to a particular experimental situation" (sect. 7). This cannot be how the brain works. In addition, the Merge decision stage represents both phonemic and lexical information in a way that can "translate the representations used for lexical access into the representations more suited to . . . phonemic decision tasks." How and why this should happen is left unclear.

ART naturally overcomes these problems using evolving spatial patterns of activation across working memory items that resonate with a level of list chunks. The list chunks that are learned in this way can include phonemic, syllabic, and word representations. The resonant context determines which chunks are competitively selected and learned. A Masking Field architecture was introduced to represent list chunks of variable length. It explains how phonemic, syllabic, and lexical information can coexist at the list chunk level, and how the speech context determines whether phonemic or lexical information will dominate (Cohen & Grossberg, 1986; Grossberg 1978). Thus, there is no need to generate connections on the fly. This property helps to explain the Magic Number 7, word length and superiority effects, the GRAY CHIP percepts, and why phonemic decisions may not develop prior to word recognition, among other data.

In summary, the feedforward Merge model has not yet solved core problems for which the feedback ART model has proposed real-time, neurally-supported, self-organizing, and data-predictive solutions.

What sort of model could account for an early autonomy and a late interaction revealed by ERPs?

Frédéric Isel

Max Planck Institute of Cognitive Neuroscience, D-04103 Leipzig, Germany.
isel@cns.mpg.de **www.cns.mpg.de**

Abstract: Norris, McQueen & Cutler demonstrated that feedback is never necessary during lexical access and proposed a new autonomous model, that is, the Merge model, taking into account the known behavioral data on word recognition. For sentence processing, recent event-related brain potentials (ERPs) data suggest that interactions can occur but only after an initial autonomous stage of processing. Thus at this level too, there is no evidence in favor of feedback.

This comment focuses on Norris et al.'s proposal that there may be cases in which feedback from higher to lower level processes could confer advantages. According to all the known data on speech recognition, it appears that the more valid approach to account for the different processes engaged during lexical access is an autonomous approach. In the target paper, the authors demonstrate that no feedback is necessary from the lexical to the prelexical level and they propose the Merge model in which no feedback exists, thus following Occam's razor's instructions not to multiply unnecessary entities.

However, as mentioned by Norris et al. in research on sentence processing, it is more difficult to describe a system which could take into account all the known data. Indeed, behavioral studies report findings in favor of both autonomous models and interactive models. Certain autonomous theories allow some feedback from semantics to syntax, while certain interactive theories propose that the first stage of parsing is totally autonomous. But how can autonomous models allow for the presence of interactions which is normally one of the functional bases of interactive systems? And how can interactive models justify the presence of an encapsulated module for parsing in their architecture? This apparent contradiction may only reflect the difficulty inherent in RT studies of tapping on-line into different phases of the comprehension processes, thus presenting only a partial view of the temporal coordination of the different subcomponents responsible for sentence processing. Indeed, using a higher online technique like event-related electroencephalography (EEG), which allows a diagnosis of the behavior of the system millisecond-by-millisecond, one can imagine describing a processing system that presents an early autonomy and a late interaction.

In my commentary, I will argue on the basis of recent event-related brain potentials (ERPs) data obtained by Gunter et al. (1998), which support a model in which syntactic and semantic

https://doi.org/10.1017/S0140525X00303247 Published online by Cambridge University Press

processes are encapsulated and run in parallel during a first autonomous phase and interact together in a later phase.

Until now, three different ERP[1] components have been identified as correlating with different aspects of language comprehension [see Donchin & Coles: "Is the P300 Component a Manifestation of Context Updating?" *BBS* 11(3) 1988]. The N400 which appears to reflect lexical integration (Kutas & Hillyard 1983; Kutas & Van Petten 1988; Van Petten & Kutas 1991), the left anterior negativity (LAN) present between 300 and 500 msec for morphosyntactic errors (Coulson et al. 1998; Gunter et al. 1997; Münte et al. 1993; Osterhout & Mobley 1995), and for violation of verb's argument structure (Rösler et al. 1993), as being present between 100 and 200 msec for phrase structure violations (Friederici et al. 1993; Hahne & Friederici 1999; Neville et al. 1991), and a "late" centro-parietal positivity (P600) present between 350 to 800 msec after a critical element which violates either a structural preference or a structural expectancy (Mecklinger et al. 1995; Osterhout & Holcomb 1992; 1993). The late positive component is assumed to be correlated with a revision process while the left anterior negativity, in particular the early one, has been taken to reflect initial structuring processes (Friederici 1995).

Gunter et al. (1998) collected ERP data which shed a new light on the debate on the time course of syntactic and semantic sources of information. Indeed, they investigated semantic and syntactic processes by manipulating semantic expectancy of a target noun given a preceding sentential context (high cloze vs. low cloze) and gender agreement between this target noun and its article (correct vs. incorrect). They showed a significant semantic expectancy main effect between 300 and 450 msec (i.e., high cloze target nouns gave rise to a smaller N400 than low cloze target nouns) and a significant gender main effect between 350 and 450 msec (i.e., the LAN was significantly smaller for the syntactically correct sentences than for the syntactically incorrect sentences). Moreover, the lack of a significant interaction between gender and semantics expectancy for both N400 and LAN suggest that the N400 is independent of syntax and that the LAN is independent of semantics. Although these two components were elicited in the same time window, their scalp distributions were different: The LAN was mostly observed at the left anterior electrodes whereas the N400 was much more broadly distributed. A second significant main effect of gender between 550 and 950 msec was identified as the P600. For this latter component, the interaction between gender and semantic expectancy was significant (i.e., there was a large reduction of the P600 in the low cloze sentences). Taken together, these data suggest that semantics and syntax work in parallel during an early stage of processing (between 300 and 450 msec after the presentation of the target word) whereas they interact during a later stage of processing (between 550 and 950 msec).

The pattern of results obtained by Gunter et al. (1998) reveals an early autonomy of syntactic and semantic processes and a late interaction between these two processes. Thus, the high on-line properties of the event-related potentials give a new picture of the sentence processing system which appears to be constituted of two sequential entities in which the implied processes work either in parallel (early stage), or in interaction (late stage). However, the open question is whether in a stage occurring after the stage in which the processes of structural reanalysis and repair took place, the semantic and syntactic information interact or work in parallel. Like behavioral studies on word recognition, research on sentence processing using event-related brain potentials (ERPs) do not show evidence in favour of feedback.

ACKNOWLEDGMENTS

I thank Angela Friederici and Thomas Gunter for their fruitful comments. My work was supported by a grant from the Max Planck Society (Germany).

NOTE

1. The event-related brain potential (ERP) represents the electrical activity of the brain correlated with a particular stimulus event.

Some implications from language development for merge

Peter W. Jusczyk and Elizabeth K. Johnson

Department of Psychology, Johns Hopkins University, Baltimore, MD 21218-2686. **{jusczyk;zab}@jhu.edu www.psy.jhu.edu/~jusczyk**

Abstract: Recent investigations indicate that, around 7-months-of-age, infants begin to show some ability to recognize words in fluent speech. In segmenting and recognizing words, infants rely on information available in the speech signal. We consider what implications these findings have for adult word recognition models in general, and for Merge, in particular.

One issue rarely mentioned in discussions of models of adult word recognition processes is how these processes developed and what implications this has for understanding the mechanisms that support them. In the past five years, we have begun to learn a lot about the early stages in the development of word recognition abilities. In what follows, we consider some relevant findings and what they imply about the plausibility of a model such as Merge.

Infants begin to display some ability to recognize the sound patterns of certain words shortly after 7-months-of-age. Jusczyk and Aslin (1995) demonstrated this in experiments in which they familiarized infants with a pair of words, such as "cup" and "dog," and then tested whether infants subsequently recognized these items when they occurred in another context. Regardless of whether they were familiarized with the words spoken in isolation, or in fluent speech contexts, the infants displayed recognition of these words. By comparison, 6-month-olds did not show evidence of recognizing these words when tested under the same circumstances.

A number of questions follow from this first demonstration of word recognition by infants. For example, how precise are infants' representations of the sound patterns of these words? Evidence from Jusczyk and Aslin's investigation and a subsequent one by Tincoff and Jusczyk (1996) suggest that infants' early word representations contain considerable detail about the phonetic properties of these words. Thus, infants familiarized with highly similar phonetic items, such as "tup" or "cut" did not subsequently treat these items as instances of "cup."

Another important question about these early abilities concerns how infants are successful in recognizing these words in fluent speech. Although several potential word segmentation cues exist for English (e.g., prosodic stress, context-sensitive allophones, phonotactic constraints, and statistical regularities), none of these cues is completely foolproof. Recent data suggest that English-learners may initially rely on stress-based cues to begin segmenting words from fluent speech (Jusczyk et al. 1999). At 7.5 months, infants can segment words with the predominant, strong/weak, stress pattern (e.g., "kingdom"). However, at the same age, infants fail to segment words with the less frequent, weak/strong, stress pattern (e.g., "surprise"). It is not until about 10.5-months that English-learners correctly segment words with weak/strong stress patterns. What allows these older infants to succeed in segmenting weak/strong words? The obvious answer is that they have also learned to rely on other potential cues to word boundaries in fluent speech. Indeed, 8-month-olds can draw on statistical regularities, such as transitional probabilities about the likelihood that one syllable follows another, to infer potential word boundaries in fluent speech (Saffran et al. 1996). By 9 months, English-learners seem to recognize which types of phonotactic sequences (co-occurrences of phonetic segments) are more likely to occur between two words as opposed to within a particular word (Mattys et al. 1999) and to use this information in segmenting words (Mattys & Jusczyk, submitted). By 10.5 months, English-learners show a similar sensitivity to how context-sensitive allophones typically line up with word boundaries, for example, the variants of /t/ and /r/ that occur in words such as "nitrates" and "night rates" (Jusczyk et al. 1999).

The point is that word recognition abilities begin in the second

https://doi.org/10.1017/S0140525X00303247 Published online by Cambridge University Press

half of the first year, and at least during the early stages, infants rely heavily on information in the speech signal itself to find words in fluent speech. Note that in the studies reviewed above, most words that infants were tested on were not ones that they are likely to be previously familiar with. Thus, the words had no real meaning for the infants, they were just familiar sound patterns that infants could recognize in fluent speech. Hence, during these early stages, word recognition processes in infants are necessarily bottom-up. If word recognition processes in adults are interactive (contra Norris et al.), then some kind of reorganization must take place during the course of development. If the interactive view is correct, it should be possible to identify when this reorganization occurs in development. However, our own view is that if Merge is correct, then one might also expect to see evidence of some reorganization in development.

Merge assumes a phonetic decision layer. But is such a layer necessary for language learners to recognize words in fluent speech? We suspect not. There is no necessary reason why one would need to have access to an explicit phonetic description in order to recognize words. However, there are other tasks that language learners typically face that may require accessing an explicit phonetic representation, namely, learning to read. If this assumption is correct, we would expect to see some changes in word recognition processes occurring around the time that children are learning to read. Some of the empirical phenomena that have driven the formulation of Merge should not be evident in the responses of children tested before this point (or in illiterates, for that matter).

ACKNOWLEDGMENTS

Preparation of this paper was facilitated by grants from NICHD (15795) and NIMH (01490) to PWJ.

Most but not all bottom-up interactions between signal properties improve categorization

John Kingston

Linguistics Department, University of Massachusetts, Amherst, MA 01003.
jkingston@linguist.umass.edu **www.umass.edu/linguist**

Abstract: The massive acoustic redundancy of minimally contrasting speech sounds, coupled with the auditory integration of psychoacoustically similar acoustic properties produces a highly invariant percept, which cannot be improved by top-down feedback from the lexicon. Contextual effects are also bottom-up but not all entirely auditory and may thus differ in whether they affect sensitivity or only response bias.

The great redundancy in the speech signal and frequent degradation or even complete loss of parts of that signal encourage the common assumption that listeners use top-down feedback continually in extracting the linguistic message from the signal. However, Norris et al. show that lexical feedback is in fact ordinarily not used by listeners, even when the signal is so degraded that it conveys ambiguous or even conflicting information about a phoneme's identity, unless the task's nature or difficulty demands listeners turn to this resource. Most of the time, listeners need not do so because the phonological content is extracted so well from the signal prelexically that lexical feedback adds little or nothing. How can the prelexical processing do so well? Part of the answer lies in the phonetic redundancy of the signal itself, and the rest in the effects of that redundancy on auditory processing.

All minimal contrasts between speech sounds are conveyed by many covarying acoustic properties. Some covary because a single articulation has more than one acoustic consequence, others because most minimal contrasts differ in more than one independently controlled articulation (Kingston 1991; Kingston & Diehl 1994). So the signal itself is massively redundant acoustically. Often, the acoustic consequences of the various articulations that differ between minimally contrasting sounds are also very similar, so much so that they may integrate psychoacoustically in a way that enhances the contrast (Kingston & Diehl 1995; Kingston & Macmillan 1995; Kingston et al. 1997; Macmillan et al. 1999). So not only do minimally contrasting sounds differ from one another acoustically in many ways, but many of those differences produce similar enhancing perceptual effects in the listener. The signal and its auditory processing therefore make the listener very likely to extract the same phonological value from signals in which the phoneme or feature was produced in various ways or was degraded during transmission. The richness of these bottom-up sources of redundancy leaves little room for any top-down sources to improve the phonological yield.

However, phoneme or feature judgments are not always unperturbed by variation in the signal: in the context of different neighboring sounds, listeners evaluate the same acoustic (or psychoacoustic) stuff quite differently in making these judgments (Repp 1982). Context effects are uniformly contrastive in that the listener shifts the criterion for evaluating the members of an acoustic continuum property toward the context's value for that property, and thus judges more of the continuum to belong to the category with the opposite value. These contrastive shifts occur when that acoustic property can be heard in the context, that is, when the context is a speech sound (Mann 1980; Mann & Repp 1981) or a nonspeech sound with the right acoustic property (Lotto & Kluender 1998), and when it can be predicted, that is, when the context is a speech sound that is auditorily but not visually ambiguous (Fowler et al. 1999) or that is auditorily ambiguous but predictable from transitional probabilities (Pitt & McQueen 1998) or stimulus blocking (Bradlow & Kingston 1990; Ohala & Feder 1994).

It is the simplest hypothesis that the first two contextual effects arise from strictly auditory interactions between the acoustic properties of the adjacent signal intervals, but the last two instead arise from integration of visual, statistical, or inferred information with the output of auditory processing. Therefore, the first two contextual effects may alter the listener's sensitivity as well as response bias to differences in the property that determines the judgment, whereas the last two should only shift the decision boundary along the continuum of that property's values, that is, alter response bias alone. Macmillan et al. (1999) show how listeners' accuracy in classifying stimuli differing along multiple dimensions can be used to derive a perceptual representation for those stimuli. (The minimal stimulus set for such an experiment is a 2×2 array in which the stimuli vary orthogonally for Low and High values of two dimensions. Classification consists of sorting all six possible pairs of stimuli drawn from this array in separate blocks of trials.) This perceptual representation can in turn be used to predict quantitatively how much the perceptual interaction between those dimensions changes response bias and sensitivity when listeners categorize those stimuli. (In the minimal experiment just described, categorization would consist of sorting all four stimuli as Low vs. High for one dimension, while ignoring their differences along the other.) Classification performance can thus be used to confirm the hypothesis that the first two context effects are located in the auditory processing stage but the last two in the stage in which auditory information is integrated with other kinds of information. And classification performance can be used to predict changes in sensitivity in categorizing the stimuli caused by strictly auditory context effects. As sensitivity may be reduced by the auditory interaction between the context and the target interval, the intended phonological value may be extracted less reliably than otherwise. On the other hand, no change in sensitivity, for better or worse, is expected from interactions with nonauditory contexts, so that phonological value should be extracted just as reliably as when the context was absent.

Two bottom-up interactions between signal properties, acoustic redundancy and auditory integration, should thus make the extraction of the intended phonological value from the signal nearly

https://doi.org/10.1017/S0140525X00303247 Published online by Cambridge University Press

perfect, but auditory context effects may prevent that perfection from being achieved. Other, non-auditory context effects will not, because they affect only bias and not sensitivity.

It's good . . . but is it ART?

Paul A. Luce,[a] Stephen D. Goldinger,[b] and Michael S. Vitevitch[c]

[a]*Department of Psychology and Center for Cognitive Science, University at Buffalo, Buffalo, NY 14260;* [b]*Department of Psychology, Arizona State University, Tempe, AZ 85287;* [c]*Speech Research Laboratory, Indiana University, Bloomington, IN 47405.* **paul@deuro.fss.buffalo.edu wings.buffalo.edu/~luce goldinger@asu.edu mvitevit@indiana.edu www.indiana.edu/~srlweb**

Abstract: We applaud Norris et al.'s critical review of the literature on lexical effects in phoneme decision making, and we sympathize with their attempt to reconcile autonomous models of word recognition with current research. However, we suggest that adaptive resonance theory (ART) may provide a coherent account of the data while preserving limited inhibitory feedback among certain lexical and sublexical representations.

Norris, McQueen & Cutler deserve praise for a provocative proposal. In a detailed analysis of previous interactive and modular accounts of spoken word recognition, they correctly find the models wanting: Neither the standard-bearers for autonomy nor interactionism fully explain lexical effects on phoneme decision making. However, despite their laudable treatment of the available evidence, Norris et al. take a step that may be premature. Abhorring the vacuum left by the discredited models, and invoking Occam's razor, Norris et al. reject the notion of feedback between lexical and sublexical levels of representation. Born is a presumably simpler model that merges the outputs of two autonomous stages at a new phoneme decision stage.

Although sympathetic to the authors' endeavor, we question the need for another contender in a crowded field of models. But more than this, we wonder if Occam's razor necessitates a model that rejects the notion of feedback outright and proposes a new set of task-specific decision nodes with connections configured on the fly. We suggest that a potentially more elegant – and perhaps more parsimonious – theoretical framework already exists in which the problems of lexical and sublexical interaction may find solutions, namely Grossberg's adaptive resonance theory (ART; Grossberg 1986; Grossberg et al. 1997a; Grossberg & Stone 1986; see also Van Orden & Goldinger 1994; Vitevitch & Luce 1999).

The problem of deciding between modular and interactive word recognition systems, and the consequent debate over feedback, stems from the presumption of distinct tiered levels of representation corresponding to words and pieces of words. ART provides an alternative architecture, allowing a different view of feedback. In ART, speech input activates *items* composed of feature clusters. Items in turn activate *list chunks* in short-term memory that correspond to possible groupings of features, such as segments, syllables, and words. Chunks are not fixed representations relegated to levels, as in models like TRACE or Race, but instead represent attractors of varying size (Grossberg et al. 1997a).

Once items make contact with matching list chunks, they establish *resonances* – stable feedback loops that momentarily bind the respective parts into a coherent entity. Attention is drawn to such resonant states, making them the basis of conscious experience. In typical resonance, longer chunks (e.g., words) *mask* smaller chunks (e.g., phonemes), so the largest coherent unit constitutes the natural focus of attention (McNeill & Lindig 1973). However, in procedures like phoneme identification, attention can be directed to attractors that may not represent the normally strongest resonance in the system (Grossberg & Stone 1986). Nonetheless, in this framework, responses are based on resonances between chunks and items, rather than on specific nodes arranged in a hierarchy.

ART captures the modular nature of Merge in that lexical chunks themselves do not directly facilitate sublexical chunks (hence avoiding the pitfalls of facilitative feedback discussed by Norris et al.). But ART's limited inhibitory feedback between larger and smaller chunks enables it to account for, among other things, the differential effects of subphonetic mismatch as a function of lexicality (Marslen-Wilson & Warren 1994; Whalen 1984). Briefly, when lexical chunks are strongly activated (as in W2W1 and N3W1; see Norris et al.), they dominate responding while simultaneously inhibiting their component sublexical chunks, thus attenuating effects of mismatch among the smaller chunks. However, when no lexical chunks achieve resonance (as in W2N1 and N3N1), responses will reflect the most predictive sublexical chunks. In the case of W2N1, however, masking from the weakly activated lexical chunk (W2) will slightly inhibit its component sublexical chunks, resulting in differential processing of W2N1 and N3N1. The effects of task demands and attentional focus reported by McQueen et al. (1999a) are also accommodated in the ART framework, given its facility for selective allocation of attention to chunks of various grains.

ART provides similar accounts of the data reported by Frauenfelder et al. (1990) and by Connine et al. (1997). In so doing, the adaptive resonance framework constitutes a truly parsimonious approach to the problem of lexical-sublexical interaction by eliminating hierarchical levels and by avoiding the construction of task-specific architectures. In short, Grossberg's ART is uniquely suited to accommodate the data reviewed by Norris et al., and many other data in speech-language processing. Moreover, it makes fundamentally different assumptions compared to models such as TRACE and Shortlist, which allows it to sidestep the points of contention raised by the target article. But most appealing, ART is an almost unifying theory, with applications to learning, visual perception, memory, attention, and many other domains. Unlike Merge, which casts speech perception as an insular system, segregated from general cognition, ART provides a broad framework, linking speech perception to other cognitive domains.

In the true spirit of Occam's razor, we should tolerate local complexity, such as lexical to sublexical feedback, in the interest of global simplicity. In other words, broadly appropriate constructs should be broadly applied, keeping theories consistent across the span of cognition. Feedback may be a good candidate for such inclusion. It is well-known that the brain is designed for feedback; cortical areas are reciprocally connected in complex maps, supporting resonant dynamics (Freeman 1991; Luria 1973). More important, feedback processes are central to theories across cognition, including general perception, learning, and memory. In all these domains, theorists have found feedback systems highly beneficial, and often necessary. For example, global memory models are typically cast as parallel systems, in which inputs establish resonance with prior knowledge (Goldinger 1998; Hintzman 1986; Shepard 1984; Van Orden & Goldinger 1994).

Because the "feedback hypothesis" is a centerpiece of modern cognitive psychology, perhaps Occam's injunction should lead us not to excise feedback altogether, but encourage us to explore architectures in which it functions more elegantly. We suggest that the adaptive resonance framework is such an architecture.

ACKNOWLEDGMENTS

Support provided by grants R01 DC 0265801 (UB), R29-DC02629 (ASU), R01 DC 00111 (IU), and DC 00012 (IU) from the National Institute on Deafness and Other Communication Disorders, National Institutes of Health.

https://doi.org/10.1017/S0140525X00303247 Published online by Cambridge University Press

What phonetic decision making does not tell us about lexical architecture

William D. Marslen-Wilson

MRC Cognition and Brain Sciences Unit, Cambridge CB2 2EF, United Kingdom. **william.marslen-wilson@mrc-cbu.cam.ac.uk**

Abstract: Norris et al. argue against using evidence from phonetic decision making to support top-down feedback in lexical access on the grounds that phonetic decision relies on processes outside the normal access sequence. This leaves open the possibility that bottom-up connectionist models, with some contextual constraints built into the access process, are still preferred models of spoken-word recognition.

The Norris et al. target article is a stimulating but somewhat paradoxical treatment of the architecture of the speech processing system, arguing that information at the lexical level cannot modulate in any way the processes of pre-lexical analysis that project the speech input onto these lexical representations. What seems paradoxical about this article is that it turns out not to be about these processes at all. The experimental results it reflects, and the novel content of the model, are about the quite different process of making overt phonemic decisions. This can be tapped into in a variety of ways, ranging from phoneme monitoring to forced choice phonemic discrimination tasks. The crucial move that Merge makes is to claim that all of these tasks operate on a form of representation (phonemic decision nodes) that is not part of the normal process of spoken word-recognition, and where these nodes only come into existence when required by the use of phonemic decision tasks.

In earlier days, of course – and still very much so in speech research – these tasks were used on the assumption that they tapped directly into a pre-lexical stage of processing. This was an assumption that I exploited in my earlier work with Warren (Marslen-Wilson & Warren 1994, henceforth MWW,94), using the phonemic judgment task with sub-categorical mismatch material to ask whether phonemic judgments in nonwords were affected by the lexical status of mismatching sub-categorical cues to phonemic identity. The fact that we did find such effects – now replicated under certain conditions by McQueen et al. (1999a) – seemed to present severe problems for theories that claimed that there was an autonomous level of pre-lexical speech processing, and that tasks like phonetic decision tapped directly into events at this level (MWW,94 p. 671).

There are a number of ways out of this. One of them, which both Merge and MWW,94 join in rejecting, is to give up the autonomy assumption and to allow, as in TRACE, for top-down effects from the lexical level onto acoustic-phonetic pre-lexical levels. A second option, initially sketched in MWW,94 and given considerably more substance in Gaskell & Marslen-Wilson (1995; 1997; 1999), maintains a form of the autonomy assumption, but moves the generation of phonemic decisions out of the direct processing path to the lexicon. It does so in the context of a distributed connectionist model, which I will refer to as the Distributed Cohort Model (DCM). The third option, embodied in Merge, maintains the classical autonomy assumption, and also moves the phonemic decision task away from the pre-lexical processing path into the lexicon, using instead the special purpose decision units mentioned above. Merge adopts a very different processing framework from the DCM, and Norris et al. make a number of criticisms of the DCM approach. I will consider these points later in this note.

The move to recognise that phonemic decision is not a direct route into the heart of the speech processing system, but is instead an essentially meta-linguistic process, heavily influenced by literacy, and open to a variety of strategic influences, is one that I would agree with. But it does have the paradoxical outcome, from the perspective of a theory of spoken word-recognition, that performance on phonemic decision becomes relatively uninformative about the properties of primary pre-lexical speech analysis. In the Merge account, not only are phonemic decision nodes architecturally separated from pre-lexical processes of acoustic-phonetic analysis, with variable connections to lexical level phonological representations, but also they operate on a significantly different processing substrate. Lateral inhibition between nodes is an important component of the phonemic decision simulations, but this is excluded at the pre-lexical level in order to prevent premature categorical assignments.

The net effect of all this is to limit the scope of the claims that are made in the target article. Norris et al. begin with a number of persuasive general arguments against the use of feedback in perceptual processing systems. But the actual case of feedback they concern themselves with – lexical influences on phonemic decisions – is argued not to be a perceptual process at all but an ad hoc decision process. What Norris et al. have demonstrated, therefore, is not the general point that pre-lexical processing is strictly autonomous, but rather the more limited point that we cannot infer that it is not autonomous from the results of experiments using phonemic judgement tasks. This seems to leave open the possibility that the perceptual processing of the speech input may not be completely independent of the higher-order linguistic environment in which it occurs. And this, in effect, is what we argue for in our alternative account of spoken word recognition and phonemic decision-making.

Bottom-up connectionist models of language processing. The original research by MWW,94 was not primarily driven by issues of autonomy and interaction in the language processing system, but by the even older problem of what distinct levels of representation need to be postulated and what are the properties of these levels (Marslen-Wilson & Tyler 1980). We focused on pre-lexical speech processing, and used subcategorical mismatch stimuli to probe whether featural cues to phoneme identity were indeed integrated at a pre-lexical level, as required by conventional views of the speech processing system. The finding of lexical effects for subcategorical mismatches embedded in nonwords was clearly a problem for this view. On this basis, and given the failure of our TRACE simulations to capture the data adequately, we rejected the notion of a pre-lexical level of phonemic representation, and looked to a quite different kind of account.

The properties of this account were conditioned by the need to meet three basic challenges. The first requirement was to account for the results of the MWW,94 experiments, with the particular mixture of dependencies they revealed between the lexical, phonemic, and featural properties of speech. The second was to explain how phonemic judgements can be made in the absence of a specifically phonemic level of sub-lexical representation. The third was to provide a uniform account of the perceptual basis for phonetic and lexical decisions, both for words and nonwords.

With respect to this last point, one of the attractions of the standard phonemic theory is that it seems to offer a consistent account of the perception of both words and nonwords. In each case, the listener's perceptual experience is assumed to be built up out of common sub-lexical elements, with words differing from nonwords only in the presence of an additional representation at the lexical level. This is linked to the widespread methodological assumption, mentioned above, that the listener has direct access to the products of phonetic processing, so that speech perception can indeed be studied as an autonomous sub-lexical level of the language system. Given that we had rejected the notion of a sub-lexical phonemic level, this required us to provide an alternative account of how nonwords are perceived.

It was chiefly for this reason that we proposed a distributed computational model, along familiar lines (e.g., Seidenberg & McClelland 1989), where the input to the network is some preliminary analysis of the speech stream, and where the output is a phonological representation. This output, we speculated, could form the basis for the listener's perceptual experience of speech, as well as for meta-linguistic operations such as phoneme decision. Because such a system would be trained in the context of the existing words of the language, the network would learn to encode

https://doi.org/10.1017/S0140525X00303247 Published online by Cambridge University Press

the underlying regularities in the mappings between sequences of speech sounds and lexical phonological representations. When nonwords were presented, their representation at the output would depend on this lexically conditioned system, and would share the properties of the lexical items with which they overlap, just as is claimed for nonwords processed in the pronunciation nets of Seidenberg and McClelland (1989) and their successors.

The distributed cohort model. The subsequent implementation of this approach, in the form of the DCM (Gaskell & Marslen-Wilson 1997; 1999) diverged significantly from this original sketch, through the use of a recurrent network and of a dual output layer, where the network is trained to link phonetic featural inputs to a joint semantic and phonological output vector, which combined a 50-unit distributed "semantic" representation with a 52-unit localist phonological representation. Although considerably larger in scale, these DCM networks are structurally similar to some earlier proposals by Norris (1993), where recurrent nets projected simultaneously onto phoneme nodes and word nodes.

The DCM models, at least in principle, seem able to meet the three requirements we originally specified – of being able to model the MWW,94 data, of providing a basis for phonemic judgements, and a uniform processing substrate for the perception of words and nonwords. In addition, as Norris himself argues in his important 1993 chapter, learning models with this kind of architecture allow us to model apparent top-down feedback effects under conditions where "there can be no top-down flow of information because the network architecture simply does not permit it" (Norris 1993, p. 229). The network learns to encode the mapping between input features and phonological representations in the statistical context of the lexical sequences it is exposed to during training, and this encoded contextual information affects subsequent processing of inputs, even when those inputs are novel sequences (nonwords) as far as the network is concerned.

Given these many strengths of distributed connectionist models of language processing, and the additional attraction that they acquire many of their properties through the learning process rather than having them directly specified by the modeller, it is perhaps surprising that Norris et al. are so critical both of the original sketch in MWW,94 and of the DCM implementation. There are a number of reasons for this. One of these, though not discussed in Norris et al., is the argument that connectionist networks of this type are fundamentally flawed because of their inability to recognise onset-embedded words – although this may turn out to have been unduly pessimistic (Davis et al. 1997a; 1997b).

In Norris et al., two main points are made. One of these concerns a specific flaw in the DCM simulation of the MWW,94 results. This seems easy to fix, as the commentary by Gaskell demonstrates. One subtext of the Merge paper, in fact, seems to be the short shelf-life of arguments based on specific simulations of complex data sets. Given the number of parameters involved, and the range of assumptions that need to be made to fit model performance to different types of experimental data, it is hard to know how much explanatory weight to assign to a given model's successes or failures. The MWW,94 TRACE simulations are a case in point.

The second point concerns the lability of the lexical effects in the McQueen et al. (1999a) replications of the MWW,94 experiments. If nonwords and words share the same processing pathway, and if this pathway is a bottom-up connectionist network trained in the manner specified, then the perceptual representation of nonwords must always be affected by the statistical properties of the lexical environment in the context of which the system learned the mapping from input to output. Such a model, on its own, could not account for an apparent absence of "lexical" effects under the appropriate strategic conditions.

It is, however, quite premature to conclude that the variability in the McQueen et al. results forces us to abandon DCM-type architectures. First, the empirical basis for this – the absence of lexical effects in two experiments – is still very limited. McQueen et al. present a contrast between two testing environments; one where the phonemic decision task is made maximally predictable, both in terms of stimulus variation and in terms of response choice, and a second where both factors are much less restrictive. Until this contrast has been explored in more detail, it is hard to say what the critical features are, and how they generate variations in the strength of the crucial effects (i.e., the reduction of mismatch for pure nonwords). In particular, we need positive evidence that the decisions in the *non-lexical* cases are genuinely drawing on a different source of information analogous to the prelexical route in Merge, rather than using the same information differently. We also need evidence that the decisions reflecting lexical constraints are indeed lexical in the direct sense assumed by Merge – that is, through feedback from an explicit lexical phonological representation, as opposed to the more implicit lexical influences modelled in the DCM, where these are encoded into the underlying network. In this connection, it should be clear that the DCM is not strictly speaking a post-lexical model in the sense that MWW,94 originally seemed to suggest.

Second, it is always open to the DCM to postulate additional mechanisms to augment the basic word-recognition model, in the same way that Merge was called into being to augment the operations of Shortlist. Both models need to postulate additional decision processes, operating on the phonological output of the system, to carry out post-perceptual operations such as phonemic decision. The question at issue, as I noted above, is whether this decision process has two kinds of information available to it, one of which is not affected by the higher-order properties of the speech processing environment, and, if so, what is the nature of this second information source.

In the Merge/Shortlist account, this second source is the output of the pre-lexical processor. In the DCM architecture this is not possible, since there is no distinct prelexical level in the conventional sense. An alternative option is that the secondary source is an auditory, nonspeech representation of the speech input. When the phonemic decision task is made sufficiently predictable, subjects may learn to attend to acoustic differences between types of phonemes, so that attention is directed away from the phonological properties of speech and towards some specific acoustic cue. Of course, these are only speculations, and need further investigation. But until these investigations have been carried out, there really is not the empirical evidence to mediate between the Merge/Shortlist and DCM approaches.

In summary, there is a great deal to agree with in the target article, and in the phenomena it is trying to model. Nonetheless I believe that it is premature to reject bottom-up connectionist architectures of the type exemplified by the DCM. The path suggested by Norris (1993) may yet prove to be more promising than the direction he and his colleagues have taken here.

The horse race to language understanding: FLMP was first out of the gate, and has yet to be overtaken

Dominic W. Massaro

Department of Psychology, Social Sciences II, University of California - Santa Cruz, Santa Cruz, CA 95064. **massaro@fuzzy.ucsc.edu**
mambo.ucsc.edu/psl/dwm/

Abstract: Our long-standing hypothesis has been that feedforward information flow is sufficient for speech perception, reading, and sentence (syntactic and semantic) processing more generally. We are encouraged by the target article's argument for the same hypothesis, but caution that more precise quantitative predictions will be necessary to advance the field.

Given the *Zeitgeist* of interactive activation models, Norris, McQueen & Cutler are to be applauded in articulating the bald hypothesis that feedback is never necessary in speech recognition.

https://doi.org/10.1017/S0140525X00303247 Published online by Cambridge University Press

An overwhelming amount of research during the last two decades has accumulated evidence for this hypothesis in a variety of domains of language processing. These results have been adequately described by the fuzzy logical model of perception (FLMP). Because they share this basic hypothesis, it was necessary that Norris et al. contrast their Merge Model with the FLMP. Most importantly, they adopt our long-term assumption (Massaro 1973; 1975; 1998) of the integration of multiple sources of continuous information represented at different levels (e.g., phonemic and lexical). Although there are many other parallels between the two models, the differences they emphasize are of more interest than the similarities. One putative similarity, however, might be a significant difference. They view inhibition between decision nodes in Merge as analogous to the Relative Goodness Rule (RGR) in the FLMP. We believe that the optimal processing strategy for language processing is to maintain continuous information at different levels for as long as possible. This continuous information is necessarily lost in Merge because of inhibition at the level of the decision nodes in their model. This inhibition in their model accomplishes exactly the outcome that the authors criticize in their paper: that two-way flow of information can only bias or distort resolution of a sensory input.

Their first contrast claims that the FLMP has independent evaluation, which somehow is not representative of Merge. In normal communication situations, perceivers evaluate and integrate multiple sources of information from multiple levels to impose understanding at the highest level possible. This general principle must be operationalized for specific experimental tasks, such as the task in which participants must identify the first segment of a speech token. The first segment is sampled from a speech continuum between /g/ and /k/ and the following context can be /ɪft/ or /ɪs/. As cleared detailed by Oden (this issue), our account of context effects in this task in no way requires the assumption that "the basic perceptual *processes* (e.g., phoneme and word recognition) are also independent" (sect. 6.3, para.4). In our feed-forward model, featural support for phonemes will also provide support for the words that make them up. Thus, we do not disagree with the statement that "a lexical node's activation depends on the activation of the prelexical nodes of its constituent phonemes" (sect. 6.3, para. 7). Our story has not changed since 1973 when we stated, "A string of letters can be correctly identified given partial visual information, if the letters conform to definite spelling rules that are well learned and utilized by the reader." (Massaro 1973, p. 353). Pursuing this thesis, quantitative model tests led to the conclusion that "Any assumption of orthographic context overriding and changing the nature of feature analysis is unwarranted" (Massaro 1979, p. 608).

Thus the implementation of the model does not violate the obvious fact that "the degree of support for a lexical hypothesis must be some function of the degree of support for its component segments" (sect. 6.3, para. 6). In typical situations, the support for a lexical item would be (1) a function of all the segments making up the phonetic string and (2) the degree to which the lexical item is favored by linguistic or situation context. In the experimental situation, perceivers are asked to report the identity of the initial segment. Both the speech quality of the segment and its context are independent contributions to its identification. Norris et al. want the lexical context to change with changes along the phonetic continuum; however, we implement the FLMP in terms of two independent sources coming from the initial segmental information and the following context.

Norris et al. criticize our previous account of coarticulation data (Elman & McClelland 1988), because Pitt and McQueen's (1998) results pinpointed the context effect as one of transition probability rather than coarticulation. Their criticism is only valid in terms of what we identified as the additional source of information, not our formalization of the FLMP's quantitative description of the results. We treated the preceding segment as an additional source of information for identification of the following stop consonant. This formalization is tantamount to treating transition probability as an additional source of information. We have, in fact, predicted many studies in which higher-order constraints such as transition probability influence segment and word identification (Massaro & Cohen 1983b). Thus, our published mathematical fit (Massaro 1996) still holds but the additional source of information is now happily acknowledged as transition probability rather than coarticulation.

I have argued over the years that quantitative models are necessary to distinguish among theoretical alternatives in psychological inquiry. There has been a resurgence of interest in model testing and selection, with exciting new developments in evaluating the falsifiability and flexibility of models (Massaro et al., submitted; Myung & Pitt 1997). Merge is formulated in terms of a miniature neural network that predicts activation levels that are qualitatively compared to empirical measures of RT. The network requires something between 12 and 16 free parameters to predict the desired outcomes, which basically involve the qualitative differences among a few experimental conditions. In an unheeded paper, I demonstrated that neural networks with hidden units were probably not falsifiable (Massaro 1988), which was later substantiated in a more formal proof (Hornik et al. 1989). I'm worried that mini-models may have the same degree of flexibility, and mislead investigators down a path of limited understanding.

Finally, for once and for all, we would appreciate it if the field would stop claiming that somehow these mini-neural networks are modeling the "mechanisms leading to activation" (sect. 6.3, para. 8), whereas the FLMP is doing something less. The authors claim that "FLMP is not a model of perception in the same way that Merge and TRACE are" (ibid.). One might similarly criticize Sir Isaac Newton's Law of Universal Gravitation, which simply states that the gravitational force FG between any two bodies of mass *m* and M, separated by a distance *r,* is directly proportional to the product of the masses and inversely with the square of their distance. As any "dynamic mechanistic" model should, we have formalized, within the FLMP, the time course of perceptual processing and have made correct predictions about the nature and accuracy of performance across the growth of the percept (Massaro 1979; 1998, Ch. 9; Massaro & Cohen 1991).

Merging speech perception and production

Antje S. Meyer and Willem J. M. Levelt

Max Planck Institute for Psycholinguistics, NL 6500 AH Nijmegen, The Netherlands. **{asmeyer;pim}@mpi.nl**

Abstract: A comparison of Merge, a model of comprehension, and WEAVER, a model of production, raises five issues: (1) merging models of comprehension and production necessarily creates feedback; (2) neither model is a comprehensive account of word processing; (3) the models are incomplete in different ways; (4) the models differ in their handling of competition; (5) as opposed to WEAVER, Merge is a model of metalinguistic behavior.

In their commentary on our recent *BBS* target article on lexical access in speech production (Levelt et al. 1999), Cutler and Norris (1999) praised our rigorous application of Ockham's razor, that is, our effort to design the simplest possible model of lexical access that would be consistent with the available evidence. We proposed a model minimizing inter-level feedback. Our commentary on the target article by Norris, McQueen & Cutler is an obvious place to return the compliment. We are pleased to see them propose a model of spoken word recognition in which there is no feedback from higher to lower level processing units. As Norris et al. point out, the functions of language production and comprehension are intimately related, and the corresponding models should be compatible in their general architecture.

1. Our first comment is that it is precisely this intimate relation between perception and production that forces us to assume some

https://doi.org/10.1017/S0140525X00303247 Published online by Cambridge University Press

feedback in the system. In our target article we proposed that the representational levels of lemmas (syntactic words) and lexical concepts are shared between perception and production. Hence, there should be bi-directional activation spreading from concepts to lemmas (in production) and from lemmas to concepts (in perception), that is, full feedback. Uni-directionality of processing (i.e., non-feedback) can only be claimed for those parts of the system that are not shared between perception and production. These are the prelexical and word levels in Merge, and the morphophonological and phonetic levels in WEAVER. (We leave undiscussed here the issue of self-monitoring, which involves the perceptual system in still another way.)

So far there seems to exist perfect complementarity between WEAVER and Merge. Still the two models do not yet fit together like pieces of a jigsaw puzzle, forming a comprehensive and consistent picture of spoken word processing. Rather, each research team has extensively studied certain areas of language processing, leaving others largely uncharted. That is the topic of our next two comments.

2. Some core areas of language processing have not been systematically incorporated in either model. For instance, both models were designed as computational accounts of single word processing. But a comprehensive picture of word processing must include a computational account of the processing of words in their multiword syntactic and semantic contexts. It is not predictable how such a comprehensive account will ultimately affect our partial models of single word processing.

3. There are some areas that have received ample attention in modeling production, but not in modeling comprehension, or vice versa. For instance, the model proposed by Levelt et al. and its computational implementation (WEAVER) include specific assumptions about the mapping from lexical concepts to lemmas and from lemmas to phonological forms. The model proposed by Norris et al. concerns the mapping of the speech input onto lexical forms; the activation of syntactic properties and meanings of words are not part of the model. Precisely where the two systems may be shared (see [1]), no modeling of comprehension is available. On the other hand, Shortlist, which is part of Merge, provides a more detailed treatment of word processing in context than does WEAVER in its account of phonological word formation. What is worse, there are clear mismatches between the Merge and WEAVER:

4. WEAVER has opted for a Luce rule treatment of competition, but Norris et al. opt for a threshold treatment. One argument for the latter approach is that adding a Luce treatment would involve unnecessary reduplication, because Merge already has inhibitory connections among word nodes. There are no inhibitory connections in WEAVER; it fares very well without. We cannot judge whether Merge (or Shortlist for that matter) could be made to run successfully without inhibitory connections, only using Luce's rule (which is somewhat like asking a diesel owner to drive her automobile on gas). There is, however, a crucial point in the background: WEAVER is a model of reaction times (speech onset latencies), whereas Merge is a model of activation levels; it suffices for Merge to display the correct monotonic relation between activation levels and (lexical or phoneme decision) reaction times. This brings us to our final comment:

5. Merge is a model of metalinguistic judgment, whereas WEAVER models the primary word naming process. This reflects marked differences in the production and comprehension research traditions. The major empirical methods in comprehension research have been metalinguistic: phoneme decision, lexical decision, word spotting, and so on. There is nothing wrong with this as long as modeling these tasks involves as a core component the primary process of word recognition. That is the case for Merge, which essentially incorporates Shortlist. One should only start worrying when a different, ad hoc core component is designed for every metalinguistic task to be modeled. In production research the tradition has been to model the chronometry of the primary process of word production, or alternatively the distribution of speech errors. Metalinguistic tasks, for instance lexical decision, gender decision or production phoneme monitoring, are occasionally used in studying word production, but they lead a marginal existence. Granting the significance of these research traditions, we would still see it as an advantage if more primary tasks were used in word comprehension research. The eye scanning paradigm (Tanenhaus et al. 1995) is one possibility, picture/word verification may be another one.

Feedback: A general mechanism in the brain

Marie Montant[1]

Department of Psychology, Carnegie Mellon University, Pittsburgh PA 15213-3890.[1] **montant@lnf.cnrs-mrs.fr**

Abstract: Norris, McQueen & Cutler argue that there is no need for feedback in word recognition. Given the accumulating evidence in favor of feedback as a general mechanism in the brain, I will question the utility of a model that is at odds with such a general principle.

In the neuroscience literature, a large body of evidence suggests that feedback is used by the brain for various aspects of perception, action, language, and attention. In the visual system, for instance, feedback connections are not the exception but the rule. Many anatomical studies have shown that most connections between cortical areas are reciprocal, and, in some cases, like the ventral occipito-temporal pathway (the "what" pathway), feedback connections are predominant (for a review, see Salin & Bullier 1995).

Feedback connections seem to have various roles. They can filter the visual input and improve its quality by changing the sensitivity of the afferent pathways to some aspects of the stimulation. This is done by modifying the balance of excitation and inhibition in lower order cortical areas and in the subcortical nuclei of the afferent pathways (Alonso et al. 1993; Deschenes & Hu 1990; Marrocco et al. 1982; Sillito et al. 1993; Ullman 1995).

Feedback connections are suspected to play an important role in figure-ground separation because they convey top-down expectations about the visual environment that make it possible to segment the visual scene (Hupé et al. 1998). By conveying top-down expectations, feedback connections are also involved in attention-driven modulation of visual processing (e.g., Luck et al. 1997) and visual word recognition (e.g., Nobre et al. 1998).

Feedback is also crucial in the synchronization of adjacent populations of neurons in the cortex, a phenomenon that is considered as the neural mechanism of "feature binding." That is, the binding of the different features of an object for the construction of a unified percept (Bullier et al. 1993; Finkel & Edelman 1989; Freeman 1991; Roelfsema et al. 1996; Singer 1995; Tononi et al. 1992; Ullman 1995). Synchronous neuronal activity would reflect recurrent bottom-up and top-down activation of neural assemblies that code different aspects of a same object during the process of recognition (Tallon-Baudry et al. 1997).

The above described feedback mechanisms are not specific to visual perception. Similar mechanisms have been described in vestibular perception (e.g., Mergner et al. 1997), in auditory perception (e.g., Hines 1999; Slaney 1998; Turner & Doherty 1997), and also in the domain of sensorimotor integration and action (see MacKay 1997, for a review).

If feedback mechanisms are used in such crucial aspects of perception, in most sensory modalities, why should it be that they are not used in speech perception? The authors argue that this is the case because (1) the system can perform an optimal bottom-up analysis of the auditory input on the first pass, (2) feedback cannot improve the quality of the input but, on the contrary, can make the system hallucinate, and (3) feedback solutions are less parsimonious than pure bottom-up solutions. However, these three assumptions are highly questionable.

https://doi.org/10.1017/S0140525X00303247 Published online by Cambridge University Press

First, there is tremendous ambiguity in the speech signal, which makes it very unlikely that a pure bottom-up analysis can be efficient. Words have to be identified against a background of noise, reverberation, and the voices of other talkers. The speech signal is further distorted by coarticulatory effects and segmental reductions and deletions. Such distortions "produce considerable ambiguities in the speech signal, making a strictly content-addressable word recognition system based on phonetic encoding unrealistic." (Luce & Pisoni 1998, p. 2).

Second, feedback can improve the quality of the visual input without running the danger of hallucinating. This has been formally proven for many years in the context of adaptive resonance theory (ART, Grossberg 1980). In ART, learned top-down expectations are matched against bottom-up data. Perception occurs when bottom-up and top-down processes reach an attentive consensus (a resonant state) between what is expected and what is there in the outside world. The matching process improves the input by amplifying the expected features in the bottom-up signal and suppressing irrelevant features. ART successfully captured a number of auditory effects (e.g., variable speech rate perception) that seem to be out of reach for models having no concepts like feedback and resonance to count on (Grossberg et al. 1997a). Contrary to Norris et al.'s claim, bottom-up/top-down matching will not lead to hallucinations because feedback in the absence of bottom-up stimulation is not able to activate cells or cell assemblies above threshold (Salin & Bullier 1995). Feedback cannot create something out of nothing (Grossberg 1999b). In fact, studies on schizophrenia have shown that auditory hallucinations result from the *absence* of feedback rather than the opposite (e.g., Frith & Dolan 1997; Silverstein et al. 1996). However, feedback can be responsible for (existing) illusions when top-down expectations are high and bottom-up input is partially degraded. The auditory continuity illusion (Warren 1984) is a good case in point.

Finally, it may be that feedback solutions are less parsimonious than pure bottom-up solutions in certain word recognition models. However, this may not be true in biological systems like the brain where the most parsimonious solution is the one that is most commonly used. Nature uses similar solutions for similar problems. If feedback is the brain's solution in a number of perceptual and cognitive domains, then it is hard to see how speech perception could be the exception.

NOTE

1. Current address is Laboratoire de Neurosciences Intégratives et Adaptatives, Université de Provence, 13397 Marseille, cedex 13, France.

Interaction versus autonomy: A close shave

Wayne S. Murray

Department of Psychology, University of Dundee, Dundee DD1 4HN, Scotland. **w.s.murray@dundee.ac.uk**
www.dundee.ac.uk/psychology/wsmurray/

Abstract: Approaches to model evaluation in terms of Occam's razor or principles of parsimony cannot avoid judgements about the relative importance of aspects of the models. Assumptions about "core processing" are usually considered more important than those related to decision components, but when the decision is related to a central feature of the processing, it becomes extremely difficult to tease apart the influences of core and decision components and to draw sensible conclusions about underlying architecture. It is preferable, where possible, to use experimental procedures that avoid the necessity for subject decisions related to critical aspects of the underlying process.

Norris, McQueen & Cutler make considerable use of Occam's razor in their target article "Merging information in speech recognition: Feedback is never necessary," but by its final use (sect. 8, para. 4 and 5), I suspect that it may have become a little blunt. They say at this point that "if models with and without feedback can both account for a finding, the model without feedback should be chosen." This claim would be unexceptionable if they had used the phrase "*equivalent* models" (and this is, I presume, their intention), but it is in the nature of this "equivalence" that problems arise.

In section 3, Norris et al. argue that if an autonomous model – Race – can account for the data as well as an interactive model, such as TRACE, then the autonomous model is to be preferred on grounds of parsimony. This would be true, if the models were equivalent in all other respects. Where they are not, an evaluation must be made concerning the relative "cost" of one assumption as compared to another. TRACE indeed includes an extra parameter in its central core, but compared to Race, it also makes one simplifying assumption – it assumes a single output mechanism. Which addition would offend Occam least? I certainly don't know.

There seems to be an assumption here that parameters central to the language processing functions of the model (see sect. 7) are more important or more weighty than those related to a "decision component." At first blush this seems entirely defensible, but it is as well to consider what the model is evaluated against – behavioural data. In accounting for those data, it would be nonsense to say that we do not care how many assumptions are made about the decision component, otherwise an "optimal" model would have little or no "central processing" but a decision component as large or as complex as you like. Norris et al. are correct, however, in pointing out that the "business" of the model is to optimally describe its core function – language processing – not the way in which particular, perhaps fairly artificial, decisions are made.

The point is that, Occam notwithstanding, there is no simple answer to this problem. Turning to the Merge model and its comparison to the enhanced Interactive Model, it can be seen that Merge employs 12 active parameters, while the interactive model does a fairly good job of accounting for most of the critical findings with only 10. Could it equal or better Merge with the addition of one or two more? I hope by now it will be apparent that this is probably a nonsensical question. What counts as more complex depends upon purely subjective judgements about what is really important in the architecture of the process.

There are perhaps better arguments for preferring autonomous over interactive models when both have approximately equivalent complexity. Fodor (1983) does a good job of covering most of these and Norris et al. focus on one particularly important argument here – the desire (perhaps indeed the need) to avoid hallucination. There are very good reasons why we need to be able to see or hear what is really out there, rather than what we anticipate. I would prefer to remain within the category of individuals who hallucinate only a minority of the time (usually when asleep). Those who hallucinate more frequently tend not to function quite so adequately. To this might be added the fact that we simply find serial, autonomous, models easier to understand. We are better able to derive (stable) predictions from them and to understand the implications of their architecture. Small changes in interactive models, on the other hand, can have profound, and often difficult to comprehend, consequences. Depending upon one's particular view of the Philosophy of Science, this alone is an important consideration.

It might be suggested, therefore, that the particular value of this paper is in demonstrating that a plausible modular architecture can do as well (or better) than interactive alternatives in accounting for the overall pattern of data across many experiments. Since modular models are simply interactive models with very low feedback, it is clear that no modular model could ever do better than all interactive models, and this is as good as it can possibly get. What Norris et al. have done, however, is to move the "interaction" out of the core of the model and into a decision component. This may well be the correct thing to do, but there are also costs to this type of approach. One of Forster's (1979) reasons for proposing a modular language processing architecture was that this was intrinsically more testable, because evidence for interac-

https://doi.org/10.1017/S0140525X00303247 Published online by Cambridge University Press

tion would invalidate the model, whereas an interactive model could accommodate data which either did or did not show interaction (see also Norris et al., sect. 8, para. 5). However, it rapidly became apparent in the sentence processing literature that the addition of a General Processing System incorporating a decision component made it very difficult to test the core predictions of the model. To the extent that any apparently interactive effect could be laid at the door of this component, the model was untestable. This is not to say that in that case, or in the case of Merge, that such a decision component is implausible or undesirable. It is simply, unfortunately, the case that the more "intelligent" the task specific component of the model, the less testable are its core assumptions.

The problem lies not in the nature of the model, but in the methodologies we use for testing its core processes. When we ask subjects to make decisions that are related to the nature of the processing we are interested in, we cannot avoid the potential duplication of effects of that parameter in both a core and a decision process. It then becomes rather difficult to determine which of these we are really testing. As I have argued elsewhere (Murray & Rowan 1998, p. 5, based on an original quotation from J. D. Fodor and colleagues), "appropriate paradigms for measuring the timing of subprocesses during comprehension are ones which require no judgement (conscious or unconscious) about the sentence." The same logic applies with "sentence" replaced by "word."

In fact, in the current sentence processing literature, the technology of choice is probably eye movement recording. The reason is that it provides a sensitive on-line index of underlying processes without (apparently) requiring the subject to participate in decision making – certainly usually not decisions directly related to the type of processing under consideration. Results from studies requiring subjects to make decisions related to the nature of the particular manipulation, although not absent, tend to be treated rather more cautiously – at least until corroborated by evidence from "taskless tasks" such as eye movement records.

It is certainly clear that tasks and their precise nature can make a big difference to the results of auditory word processing studies. The same is true of studies of auditory syntactic parsing (see, for example, Watt & Murray 1996). No doubt it is the case that some fundamental questions about auditory language processing cannot be answered without recourse to tasks which require subjects to make judgements about particular aspects of the signal, but I suspect that questions of architecture are going to be particularly difficult to resolve in the absence of data derived from studies in which subjects are not required to make these types of judgements. Unfortunately, sensitive on-line measures of auditory language processing appear to be particularly hard to find.

Some concerns about the phoneme-like inputs to Merge

Terrance M. Nearey

Department of Linguistics, University of Alberta, Edmonton, AB, T6G 2E7, Canada. **t.nearey@ualberta.ca** **www.arts.ualberta.ca/~linguist/**

Abstract: As a proponent of phoneme-like units in speech perception, I am very sympathetic to Merge's use of phoneme-oriented input. However, in the absence of any known way to provide input in exactly the form assumed, further consideration needs to be given to how the variation in the details of the specification of this input might affect Merge's (and Shortlist's) overall behavior.

I am very impressed by what Norris et al. have accomplished with Merge. Contrary to claims of other models, Merge demonstrates that a wide range of results from lexical access experiments can be simulated using phoneme-like input with no lexicon-to-phoneme interaction. My own work on categorization (e.g., Nearey 1997; forthcoming) has espoused bottom-up, phoneme-like units in speech perception. Not surprisingly, I am favorably disposed toward the general shape of Merge.

Although I have no specific criticism of model-internal aspects of Merge, I do have some reservations about the form of its input and how it might interact with other aspects of the system. Currently, the input to Merge is made-up. This is a traditional and understandable simplification in modeling lexical access: Merge, Shortlist, and TRACE II (McClelland & Elman 1986) all use constructed input that can roughly be described as a fuzzy transcription matrix. Thus, as Norris et al. clearly acknowledge, some prior signal-to-phonemic (or distinctive-feature) transduction is assumed to have taken place. In Shortlist (Norris 1994b, pp. 208–209), it is suggested that this could be accomplished by a phoneme recognition component implemented in a recurrent neural network. However, Norris may be radically oversimplifying when he states: "This is clearly a very practical proposition because similar networks are already in use in automatic speech recognition systems."

Granted, recurrent neural networks have been fairly successful in some artificial tasks involving dilation and translation of the time axis (as Norris 1994b notes). However, no existing nets can perform the phonetic transduction tasks even for simple CVC syllables at anything approaching the level of human performance.[1] Furthermore, Bourlard and Nelson (1994, p. 99) suggest that recurrent neural networks in themselves perform quite poorly (by ASR [automatic speech recognition] standards) with all but very short stretches of speech. However, Bourlard and Nelson do describe some fairly successful (by ASR, not human, standards) systems where artificial neural nets (ANN), both recurrent and pure feed-forward, serve to estimate phoneme probabilities in *hybrid* ANN/HMMs (hidden Markov models).

Norris (1994b, p. 199) notes that the lexical competition as implemented in Shortlist (and thus in a full Merge) subsumes some of the tasks of dynamic programming (a crucial part of HMMs) in finding an optimal path through a lexical lattice, thus segmenting an input stream into words. However, in HMMs (including hybrids), dynamic programming does not only apply to word segmentation (as the lexical networks of Merge and Shortlist can). It also *simultaneously coordinates* (by embedding within the word segmentation) both temporal dilation of phoneme elements and the alignment of those elements with the input signal. In the process, the relative probabilities of multiple alignments of *strings of phonemes* with the signal are estimated. It is by no means clear how MERGE/Shortlist-style lexical competition, combined with a free-running recurrent neural network as input, could be successfully substituted for the dynamic programming mechanisms and still achieve the modest level of success of current ANN/HMM systems.

Some additional simulations might give some insight into the extent of potential problems. There are in effect two clocks in Merge. The slower clock ticks are at input frame updates, set *phoneme-synchronously* at three frames per phoneme. The faster clock drives within-frame competition and operates at 15 times the frame clock. How would desynchronization by time dilation of the input strings affect Merge (or Shortlist)? Suppose additional time slices were interpolated between some or all of the input frames. A Shortlist-like lexical decoder might well encounter such temporal variability if it were hooked up to any currently feasible front-end. Perhaps the current small fragment of a full MERGE network, with its preselected lexicon and predefined between-word inhibition weights, would prove to be fairly robust to such changes. However, phoneme-synchronous input plays a crucial role in Shortlist, because such input is central to both the lexical competition and the "fast-match" process that determines the admission of word candidates. Clearly this issue deserves careful consideration. Must temporal regularization take place in a prior network? If so, we need an existence proof of networks that can actually produce a well-behaved phoneme-synchronous sequence from an asynchronously streaming input signal. Otherwise, some accommodation of variations in the duration of input patterns must be added to Merge.

https://doi.org/10.1017/S0140525X00303247 Published online by Cambridge University Press

Finally, note that any possible weakness of Merge and Shortlist suggested above also applies to all psychological models (to my knowledge) of lexical access and speech perception. We all finesse the input in one way or another.[2] Unlike the rest of us, Norris, McQueen & Cutler have at least offered a promissory note on how their evolving inputs might be derived from real speech signals. Furthermore, they have, in this and previous work, shown an admirable willingness to test the robustness of aspects of their models by "wiggling" those models' architectures as well as their parameters. I think much would be learned by additional experimentation incorporating a wider range of input specifications.

NOTES

1. Human performance on this task is excellent. For example, Allen (1994), citing work from several sources, showing English CVC nonsense syllables can be identified at better than 98% correct by human listeners. Even in conditions of high noise, phonemes in nonsense CVCs are identified at rates no worse than about 10 percentage points less than real words (Boothroyd & Nittrouer 1988). Thus, "the human front-end" (i.e., human phonetic processing with no lexical support) is capable of doing a remarkably good job, far better than *any* current ASR system (Allen 1994).

2. For example, in my models, instead of temporally evolving acoustic waveforms, I start with neatly packaged, pre-parsed "cues" which feed just the right nodes in my models in just the right way. Furthermore, my models are simple static-pattern recognizers where all evidence is presented instantaneously, rather than emerging through time.

Not all neighborhood effects are created equal

Rochelle S. Newman

Department of Psychology, University of Iowa, Iowa City, IA 52242.
rochelle-newman@uiowa.edu
www.psychology.uiowa.edu/Faculty/Newman/Newman.html

Abstract: Norris, McQueen & Cutler provide two possible explanations for neighborhood effects. The first suggests that nonwords that are more similar to words tend to activate those words more than do less similar nonwords, and the second is based on sequential probabilities between phonemes. Unfortunately, neither explanation is sufficient to explain all reported neighborhood effects.

Norris et al. have put together an impressive model of word recognition, one that explains a large number of seemingly contradictory results. My commentary focuses on their explanations of neighborhood effects on phonemic perception, and the ways in which these explanations point to areas for future research.

Neighborhood effects occur when the number and/or frequency of lexical items similar to a particular sequence influences perception of that sequence. For example, Newman et al. (1997) presented listeners with pairs of nonword-nonword series, in which one endpoint of each series was similar to more real words in English than was the other endpoint. Listeners labeled ambiguous items as being whichever endpoint made them similar to more real words – the number of similar words influenced perception of nonword items.

Norris et al. provide two possible explanations for such neighborhood effects. They suggest that "nonwords which are more like words tend to activate lexical representations more than nonwords which are less like words" (sect. 5.2.2, para. 4). This can explain situations in which the degree of similarity to a single neighbor influences perception. However, it does not explain true "ganging" effects, in which the number of neighbors (or the density of lexical space in which an item resides) has an effect. In these cases, the critical issue is not that one nonword activates a word *more* than does another, but that one nonword activates *more words* than another. There is a distinction between neighborhood effects caused by the degree of similarity to a single word, and neighborhood effects caused by the number of words to which a nonword is (slightly) similar, as pointed out by Bard and Shillcock (1993). The simulation of Connine et al.'s (1997) results demonstrates that MERGE can account for the one type of neighborhood effect, but further work is necessary to demonstrate whether it can also explain the other type. Indeed, Norris et al. state explicitly that the number of words that receive activation in MERGE at any moment in time is quite small (see sect. 7, para. 7). This suggests that such ganging effects, caused by the simultaneous activation of large numbers of words, might not fall out of the model. This could be a serious limitation, if effects of neighborhood density can be clearly shown to influence nonword perception.

However, Norris et al. provide a second possible account for these neighborhood effects, based on sequential probabilities between phonemes. In an approach similar to Vitevitch and Luce (1998), they suggest that neighborhood effects, such as those reported by Newman et al., may really be caused by transitional phonotactics (the likelihood of adjacent phonemes co-occurring in the language), rather than by true lexical knowledge. Indeed, Newman et al.'s effect was facilitatory, rather than inhibitory, supporting this idea (see Vitevitch & Luce 1998). However, it could *not* have been caused by *sequential* phonotactics (the probabilities of the CV and VC sequences), as Norris et al. suggest. As an example, one pair of series ranged from "gice" to "kice," and "gipe" to "kipe." Both series contain the same CV sequences (/gaɪ/ and /kaɪ/). If one of these were higher in probability than the other, it could cause an overall bias towards that endpoint. But this bias would occur across both of the two series, and thus could not cause a differential bias between the series, as Newman et al. found. Similarly, a higher probability for /aɪp/ (or /aɪs/) could also exist – but this would be present for both endpoints of the relevant series. Such a bias would influence overall performance on a series, but not the location of a category boundary within the series. In fact, there are no sequential probabilities that could explain a bias for /g/ in "gice kice" and a simultaneous bias for /k/ in "gipe kipe."

This is not to say that statistical probabilities could not play a role. Indeed, the effects may be caused by such probabilistic phonotactics, rather than being a top-down effect of neighborhoods. But the effect would have to be based on higher-order phonotactics, such as the co-occurrence relations between the two nonadjacent consonants, rather than by the type of sequential probabilities the authors describe here. Sensitivity to this type of discontinuous probability could be added to the Merge model, as there is nothing specific to the architecture of the model that would forbid such an enhancement. However the implications of such a change need further testing, and might lead to different predictions for other studies. For example, in Pitt and McQueen (1998), the word contexts "juice" and "bush" were selected to match the transitional probabilities between the vowels and the final /s/ or /ʃ/. However, the probabilities between the initial and final consonants were not matched, and the inclusion of discontinuous probabilities in Merge might lead the model to make different predictions than were actually found.

Merge can clearly account for some types of neighborhood effects (those caused by the degree of perceptual similarity to a single neighbor, and those caused by sequential phonotactics). However, whether it can truly explain all of the reported effects of lexical knowledge on phonemic perception remains to be shown.

https://doi.org/10.1017/S0140525X00303247 Published online by Cambridge University Press

Implausibility versus misinterpretation of the FLMP

Gregg C. Oden

Departments of Psychology and Computer Science, University of Iowa, Iowa City, IA 52242. **gregg-oden@uiowa.edu** **www.cs.uiowa.edu/~oden**

Abstract: The case for the independence of featural processing supports Merge and FLMP alike. The target article's criticisms of the latter model are founded on misunderstanding its application to natural language processing. In fact, the main difference in the functional properties of the models is the FLMP's ability to account for graded perceptual experience.

It is startling to be informed that beliefs you have held for a quarter of a century are indefensible and implausible. So it comes as a relief to discover that it was all just a big misunderstanding.

Norris, McQueen & Cutler make a compelling case for the immunity of featural processing from lexical influences in speech perception. As the authors correctly observe, this principle is at the heart of not only their new Merge model but also of the twenty-some year old Fuzzy Logical Model of Perception (FLMP). So it is natural that they should try to establish how these two models that are alike in this most fundamental way are different in other respects. Unfortunately, in so doing, they mischaracterize the FLMP and neglect the most important difference between it and Merge.

Before addressing the main concerns of my commentary, let me briefly make a couple of observations. First, it is wrong to say that the FLMP is not a process model. Claims of independence and data flow are all about process and naturally accommodate classical notions of information accrual and its time course as Massaro and Cohen (1991) have shown. Second, the FLMP can readily incorporate several forms of conditionality. For example, Oden and Massaro (1978) argue for the syllable as the unit of speech perception, and that paper and its sequel (Massaro & Oden 1980a) provide dramatic demonstrations of how the model can exhibit another kind of configurality without compromising the independence of evaluation processes.

Most important, it must be clearly stated that contrary to the target article's claims, the FLMP was, indeed, developed "in the light of the constraints of everyday language processing." One of its immediate precursors was a model of the role of semantic constraints in syntactic ambiguity resolution (Oden 1978; 1983), and it has been repeatedly applied to natural language processing tasks involving information at and across different linguistic levels (see, for example, Massaro & Oden 1980b; Oden et al. 1991; Rueckl & Oden 1986). The misunderstanding of this fact seems to be due to misinterpretations of components of the FLMP and that, in turn, is the basis for other misstatements in the article.

As a feedforward model, the FLMP postulates perceptual units that are composed of subunits that are composed of subsubunits and so on back to sensory primitives. All candidate identities of a unit at any level are presumed to be evaluated as to degree of fit to the input independently and in parallel. The evaluation of a unit at one level results from the integration of the relevant continuous information from its subunits. Thus, the evaluation of a unit *depends* on that of its subunits but is independent of everything else going on. The FLMP gets its top-down power despite having no top-down dataflow by deferring decisions as long as possible (see Oden 1983 for a discussion of how long that is likely to be in language processing). Thus, it is misleading to ponder whether phoneme and word recognition are independent in the FLMP as if there were separate decisions made.

A manifestation of this sort of misunderstanding in the target article is the claim that the FLMP would have to allow feedback in order to accommodate a biasing effect of the properties of one word on the recognition of the next. On the contrary, such influences would be presumed to occur in the subsequent integration stage where the information about the candidate identity of the sequence of words up to that point in the utterance would be integrated with the independent evaluation for the word in question.

Another such misunderstanding leads to the most serious and peculiar mischaracterization of the FLMP as assuming "that the support for a word has nothing to do with the perceptual evidence for that word" (sect. 6.3, para. 6). As we have just seen this statement is absolutely counter to the essential nature of the model. This error appears to have resulted from a confusion between the support for a word and the support that part of a word provides for the identification of its initial phoneme. In the experimental situation discussed in Massaro and Oden (1995), the degree to which a stimulus, S_{ij}, is "gift" would, according to the FLMP, be specifiable as

$$t(\text{gift} \mid S_{ij}) = g_i \times f_j \times w$$

where g_i is the degree to which the initial phoneme is /g/, f_j is the degree to which the rest of the stimulus matches /ɪft/, and w is the degree to which words are favored over nonwords. The likelihood of identifying the stimulus as "gift" would be its degree of match relative to that of the alternatives. The likelihood of identifying the initial phoneme to be /g/ would correspond to the likelihood of identifying the whole stimulus as either "gift" or "giss." Then c_j, the contextual support for /g/ in this case, would be given by

$$c_j = f_j \times w + (1 - f_j) \times (1 - w)$$

Clearly all of the terms in this expression, and thus c_j itself, would be independent of g_i but $t(\text{gift} \mid S_{ij})$, the degree of support for the word "gift," would definitely *not* be independent of g_i.

Finally, the target article ignores the most important difference between the models, that the mutual inhibition in the decision stages of Merge will tend to produce all-or-none values even for ambiguous inputs, in contrast to the FLMP with its relative goodness rule. Thus, Merge makes the implausible claim that people should never experience unclear or murky perceptions and, as a result, would seem incapable of accounting for the systematically intermediate perceptual experiences people report in rating the degree to which a stimulus is one thing versus another (Massaro & Cohen 1983a). More recent evidence involving the perception of handwritten words (Oden & McDowell, in preparation) makes a strong case for such ratings directly reflecting the perceptual processes that underlie identification. It will be a challenge for Merge to handle this.

Model evaluation and data interpretation

Mark Pitt

Department of Psychology, The Ohio State University, Columbus, OH 43210.
pitt.2@osu.edu **lpl.psy.ohio-state.edu**

Abstract: Norris et al. present a sufficiency case for Merge, but not for autonomy. The simulations make clear that there is little reason to favor Merge over TRACE. The slanted presentation of the empirical evidence gives the illusion that the autonomous position is stronger than it really is.

With the introduction of Merge, autonomous and interactive models of phoneme decision making become more similar architecturally, because in both model classes information from multiple sources (lexical and phonemic levels) is combined. "Autonomy" no longer applies to strictly feed forward models and "interactivity" narrows to meaning only direct top-down feedback to lower levels. While these distinctions are not necessarily new (e.g., FLMP), they are new in the context of a model of word recognition.

Model evaluation. A central issue in model development in any field is model selection: How does one choose between competing models of a cognitive process, in this case phoneme decision making? The criteria used to select a model are varied (e.g., de-

https://doi.org/10.1017/S0140525X00303247 Published online by Cambridge University Press

scriptive adequacy, falsifiability, generality), and reflect the difficulty of the problem. The problem can increase in complexity with the introduction of computational instantiations of a model, for in this case a model can have two forms, a verbal (i.e., box and arrow) description of its basic architecture and information flow and a computational implementation of that architecture. Merge's superiority over TRACE depends on which of these two criteria is used to evaluate its performance. If we restrict ourselves to their verbal forms, then Merge can come out ahead if one buys into three nontrivial claims: (1) the model is autonomous; (2) splitting the phoneme levels in two is justified; (3) Norris et al.'s assessment of the literature is accurate. In these circumstances, Merge is the better model (sect. 7, para. 2): Integration is necessary (eliminating Race) but not to the extent of interactivity (eliminating TRACE).

When the computational versions of Merge and Trace are compared, however, Merge does not reign supreme. Descriptive adequacy (i.e., can the model simulate the data?) is the criterion on which the models are evaluated. The TRACE-like interactive model simulates the mismatch data almost as well as Merge, and does so with two fewer parameters than Merge (nine instead of eleven). Because additional parameters increase the ability of a model to fit data (see Myung & Pitt 1998; Myung, in press), one could just as easily argue that the trade-off in fit quality for number of parameters is justified, and hence favor the simpler model, TRACE. On the basis of simulated data alone, then, there is little reason to favor one model over the other. We will have to await future empirical data to decide between the two.

In their summary of the simulation results (sect. 5.3), Norris et al. state that Merge should have no difficulty simulating lexical effects, such as listeners' biases to categorize perceptually ambiguous phonemes as forming the onset of a word rather than a nonword (Ganong 1980). Although the model in its present form could probably simulate this basic effect, it may have difficulty simulating the time course of this effect, particularly its reversal as a function of phoneme position in a word. When the ambiguous phoneme occurs word-initially, lexical influences increase in magnitude over time, being smallest at fast response rates and largest at slow response rates. Just the opposite is found when the ambiguous phoneme occurs word-finally: Lexical influences are largest at the fastest reaction times and smallest at the slowest reaction times (McQueen 1991). There appears to be no mechanism in MERGE that can modulate the strength of lexical influences over time in a manner necessary to capture this pattern of results.

Data interpretation. Because there is little reason to favor the computational versions of Merge over TRACE, Norris et al. try to strengthen their case for autonomy by playing up findings that are problematic for interactivity. Throughout the paper, Norris et al. tout the Pitt and McQueen (1998) compensation for coarticulation data as strong evidence against interactivity and for an autonomous model like Merge. This is necessary, as it is the only strong evidence they have. To make the case for autonomy even stronger, Norris et al. try to neutralize the data that are suggestive of interactivity, dismissing *all* such evidence. Using questionable criteria, they find fault with methodologies that have been used to demonstrate interactivity. They raise the standard of evidence of interactivity in order to discount data across many studies as being insufficiently persuasive.

Their skepticism of this literature contrasts with their confidence in interpreting findings using other methodologies (e.g., phoneme monitoring and identification) and types of data (e.g., reaction time and labeling). For example, the loci of monitoring effects are assumed to be clear-cut, and attention is freely invoked to account for patterns of data across studies. Compensation for coarticulation is assumed to be due only to a prelexical process sensitive to statistical properties of the language; this strong assertion is made without evidence that exhausts other alternatives. The results offered to champion autonomy are no less fragile than those presented in support of interactivity. If the field is to move forward on the defining issue of information flow, a consensus needs to emerge on what counts as evidence of interactivity.

In sum, neither the modeling data nor the empirical evidence make a strong case for autonomy or Merge. Norris et al. say as much when they conclude in the last paragraph of the paper that Merge is no more than a demonstration that a model with its particular design, which could just as easily be considered mildly interactive rather than truly autonomous, can account for a portion of the experimental data.

Merge: Contorted architecture, distorted facts, and purported autonomy

Arthur G. Samuel

Department of Psychology, SUNY Stony Brook, Stony Brook, NY 11794-2500. **asamuel@psych1.psy.sunysb.edu**
www.psy.sunysb.edu/asamuel/

Abstract: Norris, McQueen & Cutler claim that Merge is an autonomous model, superior to the interactive TRACE model and the autonomous Race model. Merge is actually an interactive model, despite claims to the contrary. The presentation of the literature seriously distorts many findings, in order to advocate autonomy. It is Merge's interactivity that allows it to simulate findings in the literature.

Merge, the latest model devised by distinguished scientists Norris, McQueen & Cutler, is extraordinarily clever and misguided. In order to portray Merge as autonomous, Norris et al. must contort the model's architecture and distort many findings in the literature. The effort ultimately fails because lexical feedback to phonemic processing *is* sometimes necessary.

Pretending that Merge is autonomous requires two contortions of the model's architecture. First, Norris et al. label the two lower levels of the model as "phonemic input" and "phoneme decision." They argue that dividing phonemic processing into two separate levels is reasonable and necessary. However, if one examines what these two levels are doing, it becomes clear that these levels really correspond directly to the "feature" and "phoneme" levels in TRACE and other models. Indeed, Norris et al. explicitly acknowledge that "We could have replaced the phonemic input layer with a featural layer and achieved exactly the same ends" (sect. 5.3, para. 2).

What, then, makes Merge's phoneme level different from the phoneme level in other models? The second contortion. In Merge, these units have no feedforward to the lexical level – they are processing dead-ends. But it is absolutely essential to note that *there are facilitative links from the lexical nodes down to the phonemic nodes.* By any normal definition of top-down processing, if lexical nodes influence the activation of phonemic codes, a model is *not* autonomous. Norris et al. offer several reasons for including such top-down connections, and they are exactly correct: Top-down lexical influences are necessary.

If Merge unambiguously employs top-down processing, how can Norris et al. maintain that "feedback is never necessary"? The only option is to invent a new definition of autonomy in which feedback cannot exist if feedforward does not. This accounts for the second contortion of the model: In order to avoid a feedback loop (which would violate Merge's autonomy), Norris et al. stranded the outputs of the phonemic level. It is ironic that the method of achieving "autonomy" was to leave in the top-down lexical-phonemic links, and to eliminate the bottom-up ones. Of course, the "input phonemic" units feed the relevant information forward anyway, so that even with a bizarre definition of autonomy Merge is not autonomous.

In order to maintain their claim of autonomy, Norris et al. present a view of the literature that is breathtakingly skewed. I will only mention two examples that involve work from my own laboratory, but the distortion is quite pervasive.

https://doi.org/10.1017/S0140525X00303247 Published online by Cambridge University Press

Wurm and Samuel (1997) followed up Frauenfelder et al.'s (1990) important finding of a lack of lexical inhibition, a problematic result for TRACE. Norris et al. report

> Wurm and Samuel (1997) replicated the Frauenfelder et al. findings but raised the possibility that inhibitory effects might be masked because the nonwords in which inhibition might be expected were easier to process than the control nonwords. They presented results from a dual task study which were consistent with their view. (sect. 4.2, para. 2)

This is a fair summary. However, Norris et al. then say "Nevertheless, there is still no direct evidence for inhibitory lexical effects in phoneme monitoring," and for the remainder of the paper they present the lack of lexical inhibition as a fact. In this case, Norris et al. reported the contradictory result, but then simply ignored it.

The second example is more egregious because it involves a clear misconstrual. Samuel (1997) reported that lexically-defined phonemes produced significant adaptation effects. This result, like Elman and McClelland's (1988) compensation for coarticulation effect, is potentially fatal for Merge, because both results involve top-down lexical influences that *cannot* be attributed to dead-end phonemic decision units. Norris et al.'s solution is to criticize the restoration study on two grounds. First, they assert that "normal adaptation effects are usually found almost exclusively in the form of a boundary shift," whereas the effect here "is practically as large at the continuum endpoints as at the boundary" (sect. 4.4, para. 2). This criticism is simply factually false: Although adaptation effects are usually larger at the boundary, this depends on many factors, and there are many adaptation studies in the literature showing shifts across the continuum.

Norris et al.'s second critique is to suggest that the adaptation effect might be occurring at the lexical level itself, in which case there would be no top-down effect. Norris et al. cite Samuel and Kat's (1996) description of multiple levels of adaptation to support their speculation, focusing particularly on what they call the "categorical" level. This criticism drastically distorts the facts in two ways. First, Samuel explicitly addressed the possibility of adaptation occurring directly at the lexical level, and provided *four* sets of data that refute that possibility (Fig. 11, p. 120). Second, the "categorical" level of Samuel and Kat is unambiguously described as sublexical in that paper. In fact, it is repeatedly equated with the phonetic level in other studies. Suggesting that Samuel and Kat's results support a lexical interpretation of the restoration effect is simply misleading.

These examples illustrate the distortion of the literature required to advocate a model that is truly autonomous. Fortunately, Merge is not such a model, given its lexical influences on phonemic perception. It is precisely these connections that allow it to successfully simulate Connine et al.'s (1997) phoneme monitoring results. The other two simulations, while elegant, are similarly nondiagnostic. The subcategorical mismatch results in the literature are so varied that any simulation must be extremely parameter-dependent (Marslen-Wilson & Warren [1994] only got their effect in about a fourth of their original stimuli, and McQueen et al. [1999a] report additional variability). Wurm and Samuel's results must qualify any simulation of Frauenfelder et al.'s data, and in fact, Norris et al. report that Merge actually does produce lexical inhibition with certain parameter choices. In short, the simulations do not in any way speak to the issue of autonomy versus interactivity.

I am actually quite encouraged by Merge, as its true nature is very much in accord with the "partially interactive" (Samuel 1981) approach that I believe the data require. This convergence gives one hope that our field is closing in on the correct description of the remarkable system that accomplishes speech perception.

ACKNOWLEDGMENT

Support for this work is provided by NIMH, grant No. R01MH51663.

Interaction, function words, and the wider goals of speech perception

Richard Shillcock

Institute for Adaptive and Neural Computation, Division of Informatics, and Department of Psychology, University of Edinburgh, Edinburgh EH8 9LW United Kingdom. **rcs@cogsci.ed.ac.uk** **www.cogsci.ed.ac.uk/~rcs/**

Abstract: We urge caution in generalising from content words to function words, in which lexical-to-phonemic feedback might be more likely. Speech perception involves more than word recognition; feedback might be outside the narrow logic of word identification but still be present for other purposes. Finally, we raise the issue of evidence from imaging studies of auditory hallucination.

First, most research on lexical-to-phonemic interaction in speech processing is carried out on content words or contentive-like nonwords. New information is typically conveyed by content words; in contrast, function words are often highly predictable. Phenomenally, the two word types are equally real to the listener, yet function words are disproportionately more poorly specified phonetically. We might expect the perception of function words to be the best place to look for lexical-to-phonemic effects. Indeed, the perception of function words has already been shown to be susceptible to context effects that are primarily syntactic (Shillcock & Bard 1993); in contrast, the initial access of content words seems to be impervious to such syntactic contexts (see, e.g., Tanenhaus & Donnenwerth-Nolan 1984). There may be similar processing differences between the two word types, in which the listener augments the impoverished signal corresponding to functor morphemes; note, in this respect, that function word processing and phonological processing are both typically lateralised towards the left hemisphere. Unfortunately, the function words do not necessarily lend themselves to every experimental paradigm that has been used with content words: multiword contexts are often required to establish a word's functor nature, and the phonemic content of (English) function words is less diverse than that of content words. It may be that other languages will yield more opportunities than English for resolving the issue of lexical-to-phonemic feedback and functors. Function word processing is not a minor aspect of speech processing; in English conversation, function word tokens typically outnumber content word tokens, and the imbalance is even greater when bound functors are taken into account. Until we know more about their processing, it might be best to qualify conclusions reached about lexical-to-phonemic interaction with a clause that such conclusions are only assumed to generalise to function words.

Second, Norris et al. refer to Occam's razor, in part because they argue for the null hypothesis, the non-existence of lexical-to-phonemic interaction. They assume that word recognition is normally the only goal of speech perception. However, speech processing also has, at least, the goals of creating a unified conscious perception of meaningful speech, and a phonological record of the speech, such as might be useful in vocabulary learning. It is not difficult to imagine a role for lexical-to-phonemic interaction in both of these activities (although in reality we know little enough about either). Logically, lexical-to-phonemic interaction might exist as a proper part of a psychologically realistic model of speech perception even though it is not required by the narrow logic of word identification.

Finally, Norris et al. refer only to hallucination as a risk inherent in interactive models. It is tempting to think that pathological auditory hallucination might constitute an interesting control in which the complete absence of sensory input is accompanied by apparently realistic speech perception. David et al. (1996) report an imaging study of a schizophrenic subject, showing activation of the relevant sensory and association cortex during hallucinated speech. Even though we understand little about what such activation is actually representing, it is a nod in the direction of "perception as controlled hallucination."

https://doi.org/10.1017/S0140525X00303247 Published online by Cambridge University Press

Hesitations and clarifications on a model to abandon feedback

Louisa M. Slowiaczek

Department of Psychology, Bowdoin College, Brunswick, ME 04011.
lslowiac@bowdoin.edu www.bowdoin.edu

Abstract: Hesitations about accepting whole-heartedly Norris et al.'s suggestion to abandon feedback in speech processing models concern (1) whether accounting for all available data justifies additional layers of complexity in the model and (2) whether characterizing Merge as noninteractive is valid. Spoken word recognition studies support the nature of Merge's lexical level and suggest that phonemes should comprise the prelexical level.

Norris, McQueen & Cutler have set out to accomplish two goals: They argue that feedback is not necessary in speech processing and they propose an autonomous model (Merge) that can account for the available speech processing data. Their arguments are well-organized and persuasive.

Regarding their first goal, I must admit to being one of those for whom "the assumption of interaction fits with many people's intuitions about the nature and complexity of the speech recognition process" (sect. 2.2, para. 1). In light of Norris et al.'s presentation I have been forced to re-evaluate my position. They have managed to chip away at the support for a widely accepted theoretical position. They dismiss the prevailing assumption that feedback is beneficial and dismantle the evidence that supports that assumption. In the process, serious doubt has been raised for the case of feedback in speech processing and I cannot as readily call myself an interactionist. Moreover, their no-feedback position will have tremendous implications for the way in which researchers regard language processing models.

Although I am swayed by their well-constructed arguments, two nagging doubts block whole-hearted acceptance of their Merge model. The first concerns the nature of our theorizing about psychological processes. A good theory is one that achieves the appropriate balance between comprehensiveness and parsimony while stimulating theoretical exploration. Accounting for the data, at the same time losing sight of the broader process or parsimoniously modeling a process without accounting for the data would be equivalent anathemas. It is true that Merge is better able than TRACE or Race to account for the available data, but it has done so at the cost of postulating an additional level of information, possibly compromising the modeling of the speech process. Despite Norris et al.'s position that "The phonemic decision-making mechanism is . . . a natural part of mature language processing abilities" (sect. 7, para. 3), the added complexity in Merge is in response to data that may be more related to our experimental procedures than to normal speech processing. We may be building models that can account for a variety of experimental findings (some "artifactual" and some not) at the expense of models that capture the psychological processes we hope to understand. One test of this concern will be whether the Merge model, like TRACE and Race, is able to generate serious questions about spoken language processing, for example, will it stimulate theoretical exploration?

A second concern focusses on what may be a semantic quibble. Norris et al. suggest that Merge is not an interactive model. This claim is based on a somewhat narrow definition of interaction. It is true the Merge does not include feedback from the lexical level to the prelexical level, but one might argue that the merging of lexical and prelexical information at the phonemic decision level is a type of interaction. In that way, the postulation of a phonemic decision level and claims to have eliminated lexical feedback may be a theoretical sleight of hand.

With these philosophical concerns aired, we can turn to a more concrete assessment of the model itself. I am in agreement with Norris et al.'s characterization of the lexical level. A growing body of evidence has accumulated that suggests that the processing of spoken words is influenced by phonological overlap between those words and preceding words. Primes and targets that share initial phonemes slow response times to the target word (Goldinger 1999; Hamburger & Slowiaczek 1996; 1999; Monsell & Hirsh 1998; Radeau & Colin 1996; Slowiaczek & Hamburger 1992.) These effects have been obtained with the shadowing and the lexical decision tasks suggesting that the effect is perceptual and not solely the result of production processes (though results with lexical decision are somewhat more variable; see Monsell & Hirsh 1998 for a discussion of confounding factors in lexical decision). Such inhibition has been found for words, but not for nonwords (Radeau et al. 1989; Slowiaczek & Hamburger 1992) suggesting that the inhibition occurs between lexical items. Also, similar inhibitory effects have been reported in the visual word recognition literature (e.g., Colombo 1986; Lupker & Colombo 1994) suggesting that this effect is not specific to speech processing. As a result, the inhibitory connections between the lexical representations in Merge (and Shortlist) are consistent with the empirical findings in the phonological priming and spoken word recognition literature, as well as the speech processing data outlined by Norris et al.

Although Norris et al. provide arguments to support the absence of feedback from the lexical to the prelexical level, my reaction to the prelexical level would be more enthusiastic if they were more specific about the nature of the prelexical representations. In section 5.1 of their paper, they refer to the prelexical level as the phoneme level, but in section 5.3 they suggest that they "could have replaced the phonemic input layer with a featural layer and achieved exactly the same ends" (para. 2). Marslen-Wilson and Warren (1994) suggest that the nature of prelexical representations is not phonemic and studies that require explicit phonological judgments (e.g., phoneme monitoring, phonetic categorization) make it difficult to determine the nature of these representations. There are data from tasks that do not rely on explicit phonological judgements (i.e., shadowing and lexical decision in phonological priming) however, that suggest the prelexical level should be comprised of phonemic representations (Slowiaczek et al. 2000).

A number of studies have found that primes and targets that share word-final phonemes (e.g., SLACK–BLACK) decrease response times to the target item (Burton 1992; Cutler & Chen 1995; Dumay & Radeau 1997; Emmorey 1989; Radeau 1995; Radeau et al. 1995). This word-final phonological facilitation was obtained using shadowing and lexical decision tasks (i.e., tasks that do not involve explicit phonological judgement; Radeau & Colin 1996; Radeau et al. 1994; Slowiaczek et al. 1997; Slowiaczek et al. 1999.) Obtaining the effect with different tasks suggests that this facilitation is perceptual (rather than solely due to production.) The fact that it has been found for word as well as nonword stimuli suggest that it is prelexical (Dumay & Radeau 1997; Radeau et al. 1995; Slowiaczek et al. 1997; Slowiaczek et al. 2000). The fact that it is found for spoken stimuli but not necessarily for visual stimuli suggests that it is specific to speech processing (Dumay & Radeau 1997; Radeau et al. 1994). The most important finding with regard to the nature of prelexical representations is that the effects are based on whether or not the prime and target in the experiments rhymed and/or shared final phonemes (Slowiaczek et al. 2000). As Slowiaczek et al. (2000) argue, these data provide support for a phonemic prelexical representation.

ACKNOWLEDGMENTS

My sincere thanks to Richard A. Kreiner and Maria L. Slowiaczek for comments that helped to clarify issues raised in this paper.

https://doi.org/10.1017/S0140525X00303247 Published online by Cambridge University Press

Recognition of continuous speech requires top-down processing

Kenneth N. Stevens

Department of Electrical Engineering and Computer Science and Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA 02139. **stevens@speech.mit.edu**

Abstract: The proposition that feedback is never necessary in speech recognition is examined for utterances consisting of sequences of words. In running speech the features near word boundaries are often modified according to language-dependent rules. Application of these rules during word recognition requires top-down processing. Because isolated words are not usually modified by rules, their recognition could be achieved by bottom-up processing only.

In this commentary, I will address a question that is related to the problem under discussion here, but is somewhat more general: Does lexical access during running speech utilize top-down information from hypothesized lexical units to influence the processing of the speech signal at the sublexical level? The evidence in the target article of Norris et al. is based on psycholinguistic experiments with isolated words, and does not address the recognition of word sequences. The recognition of word sequences can present problems different from those for isolated words because when words are concatenated the segments can undergo modifications that are not evident in utterances of isolated words.

We begin by assuming that a listener has access to two kinds of language-specific knowledge. The language has a lexicon in which each item is represented in terms of a phoneme sequence, with each phoneme consisting of an array of distinctive features. The listener also has knowledge of a set of rules specifying certain optional modifications of the lexically-specified features that can occur in running speech. These modifications frequently occur at word boundaries, and are less evident in single-word utterances. (There are, of course, also obligatory morphophonemic rules.)

As acousticians with a linguistic orientation, we take the following view of the process of human speech recognition (Stevens 1995). There is an initial stage in which landmarks are located in the signal. These landmarks include acoustic prominences that identify the presence of syllabic nuclei, and acoustic discontinuities that mark consonantal closures and releases. The acoustic signal in the vicinity of these landmarks is processed by a set of modules, each of which identifies a phonetic feature that was implemented by the speaker. The input to a module is a set of acoustic parameters tailored specifically to the type of landmark and the feature to be identified. From these landmarks and features, and taking into account possible rule-generated feature modifications, the sequence of words generated by the speaker is determined. This process cannot, however, be carried out in a strictly bottom-up fashion, since application of the rules operates in a top-down manner. A typical rule specifies a lexical feature that potentially undergoes modification, it states the modified value of the feature, and it specifies the environment of features in which this modification can occur (cf Chomsky & Halle 1968). Thus it is necessary to make an initial hypothesis of a word sequence before rules can be applied. This initial hypothesis must be made based on a partial description of the pattern of features derived from the feature modules.

As an example, consider how the words can be extracted in the sentence "He won those shoes," as produced in a casual style. The /ð/ is probably produced as a nasal consonant, and the /z/ in "those" is usually produced as a palato-alveolar consonant, and may be devoiced. Acoustic processing in the vicinity of the consonantal landmarks for the word "those" will yield a pattern of features that does not match the lexically-specified features for this word. The feature pattern may, however, be sufficient to propose a cohort of word sequences, including the word "nose" as well as "those." Application of rules to the hypothesized sequence containing "those" will lead to a pattern of landmarks and features that matches the pattern derived from the acoustic signal. One such rule changes the nasal feature of the dental consonant from [−nasal] to [+nasal] when it is preceded by a [+nasal] consonant (Manuel 1995). (Close analysis will reject the word "nose," since the rule that creates a nasal consonant from /ð/ retains the dental place of articulation.) Another rule palatalizes the final /z/ when it precedes the palatoalveolar /š/ (Zue & Shattuck-Hufnagel 1979).

We conclude, then, that a model for word recognition in running speech must be interactive. That is, the process must require analysis by synthesis (Stevens & Halle 1967), in which a word sequence is hypothesized, a possible pattern of features from this sequence is internally synthesized, and this synthesized pattern is tested for a match against an acoustically derived pattern. When the utterance consists of isolated words, as in the experiments described in Norris et al.'s target article, there is minimal application of rules, and the acoustically based features match the lexically specified features. Consequently isolated word recognition can be largely based on bottom-up or autonomous analysis, as proposed by the authors.

No compelling evidence against feedback in spoken word recognition

Michael K. Tanenhaus, James S. Magnuson, Bob McMurray, and Richard N. Aslin

Department of Brain and Cognitive Sciences, University of Rochester, Rochester, NY 14627. **{mtan; magnuson; mcmurray}@bcs.rochester.edu aslin@cvs.rochester.edu www.bcs.rochester.edu**

Abstract: Norris et al.'s claim that feedback is unnecessary is compromised by (1) a questionable application of Occam's razor, given strong evidence for feedback in perception; (2) an idealization of the speech recognition problem that simplifies those aspects of the input that create conditions where feedback is useful; (3) Norris et al.'s use of decision nodes that incorporate feedback to model some important empirical results; and (4) problematic linking hypotheses between crucial simulations and behavioral data.

Norris et al. have provided a valuable service to the field by organizing and evaluating the literature concerning lexical influences on phonemic decisions in spoken word recognition. We believe their analysis will sharpen the discussion of issues in spoken word recognition and help shape the future research agenda in the field. Nonetheless, we find their case against feedback unconvincing for the following reasons.

1. Occam's razor has a double-edged blade. Norris et al. invoke Occam's razor to support their a priori claim that models without feedback should be preferred to models with feedback. Occam's razor, however, applies only when there is no empirical basis for preferring one model over another. In fact, there is considerable evidence for feedback connections in various cortical areas and for feedback in perceptual and cognitive processes. In visual perception, where the links between brain mechanisms and perception are best understood, there is evidence for feedback connections and processing interactions at both high and low levels (Churchland et al. 1994; Wandell 1995). There is also evidence for feedback at auditory levels presumably preceding phonemic processing (Yost & Nielsen 1977). Moreover, as Norris et al. acknowledge in section 3.5, feedback is likely at higher levels in language comprehension. Why, then, should sub-lexical processing be uniquely devoid of feedback? Given the ubiquitous nature of feedback in the brain, it is simpler to hypothesize feedback than to make sublexical processing a special case.

2. Feedback is surely helpful. Norris et al. argue that feedback cannot improve the efficiency of word recognition. This is only true given the sort of idealized input representation they use, consisting of noise-free discrete phonemes. Speech, however, is characterized by noise and variability (due to coarticulation, talker dif-

https://doi.org/10.1017/S0140525X00303247 Published online by Cambridge University Press

ferences, etc.). Given a more realistic characterization of the input, feedback would be helpful, as it is in higher level language processing.

3. Feedback is required by the data and is incorporated into Merge. Norris et al. admit that lexical effects on phonemic decisions in non-words provide evidence against autonomous models of spoken word recognition. Merge allows for this feedback and differs from other autonomous models by adding phonemic decision nodes where phonemic and lexical information can be integrated. Although lexical information influences phonemic decisions in Merge, the autonomy of phonemic processing is preserved, because information at the lexical units is unaffected by the phonemic decision units. Parsimony issues aside, distinguishing interaction at the decision level from interaction at the "perceptual" level is at worst impossible and at best requires clearer linking assumptions between the model and the data.

4. Simulations against TRACE and in support of Merge are problematic. Although we are not trying to defend TRACE as a fully plausible model, it is important to note that the simulations challenging TRACE and supporting Merge depend upon particular parameter settings and questionable linking assumptions between the data and the models. Consider the subcategorical mismatch simulations that play a central role in Norris et al.'s arguments. The relevant Merge activations are shown in Figure 2 in section 5.2.1.

Compare the target activations for W1W1 from Figure 2A with the activations for N3W1 and W2W1 (the correct target is W1 for all three conditions). Clearly, the activations follow different timecourses. W1W1 precedes N3W1, which precedes W2W1. The puzzle, however, is that mean lexical decisions are fastest to W1W1 and slower (but equivalent) to N3W1 and W2W1. Marslen-Wilson and Warren (1994) reported that TRACE does not predict this pattern, but rather predicts the W1W1 < N3W1 < W2W1 ordering that is present in the activation functions. Merge is able to capture the empirical lexical decision pattern, despite showing similar activation patterns as TRACE, but only when a particular decision threshold (.20) is assumed. Activations for W1W1 cross this threshold at Cycle 8, and activations for N3W1 and W2W1 cross together at Cycle 10. With a slightly lower threshold, say .19, N3W1 would be faster than W2W1.

Norris et al. would like to conclude that this is compelling evidence for Merge and against TRACE. Their argument is that feedback in TRACE prevents the model from getting the activations just right; in their simulations with a mock-up of TRACE, they could not find a set of parameters that would yield a threshold where N3W1 and W2W1 will be treated the same without introducing other deviations from the actual lexical decision data. Their simulations of the sub-categorical mismatch findings might be a powerful argument against TRACE, if we had strong independent reasons to believe that (1) a particular all-or-none decision threshold of precisely .20 is correct, and (2) the feedback parameter in TRACE creates a fatal flaw which makes it impossible to find a threshold that would correctly simulate the lexical decision data. We find both of these assertions implausible.

More crucially, we should ask why lexical decisions are not mirroring the highly similar activation patterns predicted by both TRACE and Merge. Why do activations for W2W1, which lag behind activations for N3W1, have similar mean lexical decision times? The answer lies in the activation patterns and the linking hypotheses between the activations and lexical decision times. Early on in W2W1, W2 becomes quite active, following the same trajectory as W1W1 through Cycle 8. If one assumes that faster lexical decisions tend to be affected by earlier states of the system than slower lexical decisions or that the system is affected by noise, the distribution of lexical decisions in the W2W1 condition will contain a small proportion of fast "yes" times, based on activation of W2, as well as some slow "yes" responses based on the activation of W1. Whereas the means might be similar for the N3W1 and W2W1 conditions, the distributions are likely to differ in ways that are clearly testable but not revealed by mean lexical decision times alone.

More generally, we believe that arguments about model architecture on the basis of simulations of the type appealed to by Norris et al. are extremely important. However, the arguments are only as strong as the linking hypotheses between the model and the data. Norris et al. have simply not made a compelling case that feedback is unnecessary in the architecture or in the simulations used to support their Merge model.

ACKNOWLEDGMENT

This work was supported by a grant NSF SBR-9729095.

Why not model spoken word recognition instead of phoneme monitoring?

Jean Vroomen and Beatrice de Gelder

Department of Psychology, University of Tilburg, 5000 LE Tilburg, The Netherlands. **j.vroomen@kub.nl**

cwis.kub.nl/~fsw_1/psychono/persons/jvroomen/index.htm

Abstract: Norris, McQueen & Cutler present a detailed account of the decision stage of the phoneme monitoring task. However, we question whether this contributes to our understanding of the speech recognition process itself, and we fail to see why phonotactic knowledge is playing a role in phoneme recognition.

Psycholinguistics is a strange research domain. Once, the noble aim was to understand human language processing, or, more in particular, to understand how humans recognize words when they hear sounds. There was no obvious way to tackle that question because spoken language processes themselves were not particularly designed for introspection or any other direct method. Psycholinguists therefore invented clever tasks like phoneme monitoring and lexical decision. These tasks, so was the idea, would allow one to tap the underlying processes and deliver the data on which models of speech recognition could be built. TRACE (McClelland & Elman 1986), and indeed Shortlist (Norris 1994b) are an example of that. With the present work of Norris et al. though, it seems that the focus has been shifted from trying to understand spoken word recognition toward trying to understand the ingenious methods that psycholinguists come up with. We wonder whether this move will lead towards a deeper understanding of the speech recognition process.

A decade ago, the relation between data and theory was straightforward. For example, in TRACE there was a bank of phoneme detectors that mediated between articulatory features and words. The (too) strong assumption was that the activation level of a particular phoneme was reflected in the time a subject needed to detect that specific phoneme. One could have anticipated that this assumption was a bit of an oversimplification. At that time, it was already well known that the phoneme was, at least to some extent, an invention, and not so much a natural concept. Different populations with little knowledge about the alphabet (young children, dyslexics, illiterates, Chinese, and other non-alphabetic readers) were unable to explicitly represent speech as a concatenation of phonemes, yet did not have any apparent difficulty recognizing spoken words (see, e.g., Bertelson 1986 for a review). A task like phoneme monitoring requiring an explicit decision about the presence of a phoneme could thus be expected to be related with alphabetic reading instruction, but not so for spoken word recognition.

Norris et al. now formalize this distinction in a model that segregates recognition of phonemes from decisions about phonemes. They make a strict distinction between phoneme recognition units and phoneme decision units. Decision units are very different from recognition units. Decision units are strategic, they are made on the fly, they receive information from the word level, and they have inhibitory connections. None of those properties is shared by phoneme recognition units. Phoneme recognition units are what

https://doi.org/10.1017/S0140525X00303247 Published online by Cambridge University Press

they always were: they are assumed to mediate between the speech signal and words. In fact, almost nothing is said in Norris et al. about recognition units that has not been said previously. In our view, this is disturbing if the ultimate goal is to understand speech recognition, and not phoneme monitoring, lexical decision, or whatever other task psycholinguists have invented or will invent in the future.

One can of course argue that it pays to understand the tools one is working with. In this particular case, it was the decision stage in the phoneme monitoring task that troubled our view. Basically Norris et al. argue that we have been misled and that many of the feedback phenomena occurred at a task-specific decision stage. This may well be correct, but it should be realized that this task-specific decision stage is also the least interesting part of the word recognition process. Indeed, the phoneme decision stage is in fact superfluous. One can recognize words without phoneme decision units: Decision units only exist because the experimenter told a subject to perform a task with phonemes. In our view, there is a distinction between being critical about a task and knowing its weaknesses versus modelling its weaknesses. Why should one model that aspect of a task which is ultimately the least informative? Would it not be better to try instead to model spoken word recognition?

The ultimate question, in our view, is what has been learned from Norris et al.'s model about speech recognition itself. The architecture they propose is a straightforward one: Phonemes activate words, and words compete. The main argument for the absence of feedback from word recognition units to phoneme recognition units is a logical one: Phonemes are already recognized fast and accurately, and sending information back from words to phonemes simply does not improve word recognition. So far, this may well be correct, but Norris et al. make a surprising exception to this strictly bottom-up process. They allow "lower"-order statistical knowledge about transitional phoneme probabilities to play a role in phoneme recognition. To us, this seems a strange move in a strictly bottom-up recognition process.

First, it seems to be a matter of arbitrary labels to call transitional phoneme probabilities "low," and lexical feedback "high." There is nothing inherently low or high in any of these kinds of information. Maybe one is precompiled, the other is computed online, but the principle is that in both cases information from a different source than the speech signal itself enters the recognition process. It is difficult, then, to understand on what principle the distinction is based: why is lexical information excluded, but not transitional probabilities?

Second, it seems at least debatable whether transitional phoneme probabilities will help phoneme recognition if, as argued before, phonemes are already recognized fast and accurately. Are phonemes recognized fast and accurately because the speech signal itself is processed efficiently, or because of the help of transitional probabilities? Third, how is the transitional knowledge about phonemes learned if not by some form of feedback to the phoneme recognition stage? Finally, instead of using phoneme-sized units, why not have higher-order recognition units like syllables that already incorporate the transitional phoneme information?

Phonemic organization does not occur: Hence no feedback

Richard M. Warren

Department of Psychology, University of Wisconsin-Milwaukee, Milwaukee, WI 53201. **rmwarren@uwm.edu**

Abstract: I agree with Norris et al. that feedback to a phonemic level is never necessary, but disagree strongly with their reason why this is true. I believe the available evidence indicates that there is no feedback because there is no phonemic level employed in the perceptual processing of speech.

I will explain how I came to the belief that there is no phonemic level of analysis, and how testing this concept led to: (1) confirmatory experiments based upon delays in identifying target speech sounds, (2) the discovery of the phonemic restoration illusion, and (3) the discovery of the vowel sequence illusion.

While attending a rather dull colloquium during a sabbatical in 1969, I decided to play a game, and find out how quickly I could detect the occurrence of a particular phoneme when produced by the speaker. After trying one phonemic target after another, it became apparent that several words would go by before I could identify a target speech sound. I surmised that identification of phonetic components in running speech could not be accomplished directly, but was inferred following linguistic organization. This led to a formal experiment carried out that year that demonstrated that phonetic targets required more time for identification than did the monosyllables containing them, whether the monosyllable was a word in a sentence, a word in a list, or an item in a list of nonsense syllables (Warren 1971). Further, I reported in this paper that when sentences were used, context that accelerated or delayed the identification time for a word produced a corresponding change in the time required for identification of its phonetic components. Encouraged by these observations, I reasoned that since phonemes appear to be inferred following syllabic or lexical organization, individual phonemes replaced by noise should be "restored" after the word is recognized. In addition, the listener should be unable to tell where in a word the noise had occurred. It was found that this "phonemic restoration" did indeed occur: When told that a speech sound in a sentence had been replaced by noise, listeners could identify neither which speech sound was absent nor where the noise had occurred even after listening to the sentence several times (Warren 1970; Warren & Obusek 1971; Warren & Warren 1970; see also Warren 1999).

Perhaps the most direct evidence we found indicating that phonemic segmentation by listeners is a consequence, not a contributing factor in linguistic organization, is furnished by the vowel sequence illusion (Warren et al. 1990; 1996). (A compact disk accompanying Warren [1999] includes a demonstration of this illusion.) When listeners heard repeated sequences consisting of from three to ten isochronous steady-state vowels having durations below the 100 msec threshold for identification of the vowels in their appropriate order, there was a loss not only of order identification, but the vowels themselves could not be identified (when heard in isolation, the individual vowels at these durations were readily recognizable). Between 30 and 100 msec/vowel, native speakers of English heard an obligatory organization into linguistic "temporal compounds" consisting of syllables that were either monosyllabic English words or syllables that occurred as part of polysyllabic English words. Listeners could segment these syllables into illusory consonants and vowels that seemed as real as those occurring in normal speech. Different arrangements of the same phonetic elements formed different compounds. A similar combining of acoustic elements into order-dependent temporal compounds (along with the loss of ability to identify the constituent sounds) had been reported earlier for sequences of nonlinguistic sounds presented at item durations below the threshold for order identification (Warren & Ackroff 1976; Warren & Bashford 1993; Warren et al. 1991).

Based upon this earlier work, Warren and Gardner (1995) reasoned that because recognition of constituent phonemes and their orders was not required for linguistic organization, then it might be inappropriate to attribute receptive aphasias to an inability to distinguish the order of brief components as had been suggested by several investigators (e.g., Brookshire 1972; Carmon & Nachshon 1971; Tallal & Piercy 1973). Warren and Gardner (1995) designed an experiment to determine whether aphasic listeners could (or could not) distinguish between the order of brief sounds. We tested adults with receptive aphasia who had suffered cerebral accidents at least one year earlier, and were judged to be neurologically stabilized. It was hypothesized that they would be able to discriminate different arrangements of sounds as long as they

https://doi.org/10.1017/S0140525X00303247 Published online by Cambridge University Press

were not required to use language skills to name the components in their proper order. We found that the aphasic subjects were indeed able to distinguish between the two possible arrangements of repeating sequences consisting of either three steady-state vowels or three tones with different frequencies when the items in the sequences had isochronous durations ranging from 10 msec through 1 sec. However, when the individual sounds (vowels or tones) were 3 sec in duration, they could not distinguish between different arrangements, presumably because an inability to linguistically encode and recall the item names in appropriate order (of course, normal controls could easily accomplish the task with the longer as well as the shorter item durations).

Finally, I have my doubts about two statements made in the last paragraph of the target article. (1) "The empirical challenge of testing theories can be met only in the context of specific models." (last para.). I would go along with the statement if the word "model" was changed to either "experiments" or "evidence." The ingenuity used in constructing, attacking, defending, and modifying the ever shifting complex array of interactive and autonomous models for speech perception could be put to better use by testing theories (and assumptions) directly. (2) The authors state that "Whether it [the Merge model] passes future empirical challenges remains to be seen." (last para.). I believe it is not necessary to wait – the evidence available at this time indicates that there is no phonemic level of analysis and hence no merging with the lexical level as assumed by the Merge model.

ACKNOWLEDGMENT

Preparation of this commentary was supported by the National Institute on Deafness and Other Communication Disorders, Grant DC00208.

Occam's razor is a double-edged sword: Reduced interaction is not necessarily reduced power

D. H. Whalen

Haskins Laboratories, 270 Crown St., New Haven, CT 06511.
whalen@haskins.yale.edu
macserver.haskins.yale.edu/haskins/STAFF/whalen.html

Abstract: Although Norris, McQueen & Cutler have provided convincing evidence that there is no need for contributions from the lexicon to phonetic processing, their simplification of the communication between levels comes at a cost to the processes themselves. Although their arrangement may ultimately prove correct, its validity is not due to a successful application of Occam's razor.

The evidence for modularity in phonetic processing is extensive and not convincingly refuted, though the issue is far from settled. Norris, McQueen & Cutler lay out several sound arguments against the need for interaction, especially the fact that lexical feedback cannot, in principle, improve phonetic processing. They have accounted for an impressive array of facts, though of course there are always more to be dealt with. In my view, the theory makes its gains by complicating the total system rather than simplifying it. Allowing decisions at the phonemic level is an addition, though a clearly necessary one. The elimination of lexical feedback is a simplification, but it comes at the cost of recreating much of the information that feedback provides in the speech process itself. The results of Pitt and McQueen (1998) suggest that this is necessary, but it is a complication nonetheless. While there are no generally recognized criteria for determining which of two theories is more parsimonious, it is clear that the degree of complication in the speech process due to this recreation of information is as great if not greater than simply allowing feedback from the lexicon, where that information will still need to be represented in some fashion.

The target article touches on many areas, none of which can be definitively covered in one article; I will limit my comments to the section dealing with subcategorical mismatches (sect. 4.6). Although my own subcategorical mismatch design (Whalen 1991) did not address exactly the questions at hand, it suggests that the results that are modelled are not, in fact, treating subcategorical mismatches. The point of studying such mismatches is that they do not impinge on consciousness (Whalen 1984; Whalen & Samuel 1985) and thus should not allow for cognitive processing strategies that are irrelevant to normal speech perception. The results of Marslen-Wilson and Warren (1994) are most likely not to be such mismatches. Two-thirds of their data is ignored because the fricatives and voiceless stops did not give significant results. Only the voiced stops gave rise to a mismatch effect, but it was of such a large magnitude (122 msec) that it could only reasonably have been due to overt ambiguity. The replication of that result (McQueen et al., 1999a) apparently has an equally large effect (134 msec). My own results were on the order of 10–20 msec. If there is uncertainty about the identity of the final stop, then surely nonphonetic means of resolving the ambiguity will be brought in. I suspect that the absence of an effect in the other two-thirds of the cases was a lack of power. The modeling, then, is of an effect much different from the one supposedly under study, and one that would be expected to bring in other kinds of processes (which would require a more powerful system to simulate, not a less powerful one).

Norris et al. also adopt the computationally simple scheme of allowing phonetic information to accumulate from time-slice to time-slice, but there is solid evidence that this is not the way humans perceive speech. There are effects of consonant decisions on vowels and vowel decisions on consonants that do not seem to proceed in purely early-to-late temporal order (Whalen 1989). Similarly, later-occurring information affects the distinction between /b/ and /w/ (Miller & Liberman 1979). Whether or not this entails segment-sized entities, it suggests that phonetic processes are still active after a time-slice has passed. Simply sending the raw acoustic signal to the lexicon is unlikely to be the way that word recognition occurs (though it has certainly been proposed before, e.g., Klatt 1980). The model proposed in the target article seems, in this regard, likely to need changing. A model that integrates information across spectral slices is more complex than one that does not but, again to Occam's chagrin, the more complicated model is called for.

ACKNOWLEDGMENTS

The writing of this commentary was supported by grants DC-02717, HD-01994 and DC-00403 to Haskins Laboratories.

Feedback consistency effects

Johannes C. Ziegler[a,b] and Guy C. Van Orden[c]

[a]*Macquarie Centre for Cognitive Science, Macquarie University, Sydney, NSW 2109, Australia;* [b]*LPC-CNRS, Aix-en-Provence, France;* [c]*Cognitive Systems Group, Department of Psychology, Arizona State University, Tempe, AZ 85287-1104.* **ziegler@newsup.univ-mrs.fr** **guy.van.orden@asu.edu**

Abstract: Models are not adequately evaluated simply by whether they capture the data, after the fact. Other criteria are needed. One criterion is parsimony; but utility and generality are at least as important. Even with respect to parsimony, however, the case against feedback is not as straightforward as Norris et al. present it. We use feedback consistency effects to illustrate these points.

According to Norris et al., the case against feedback has a parallel in visual word recognition (sect. 3.5.1). Strict feedforward models mimic effects that previously motivated feedback, such as the word superiority effect. According to Norris et al., Occam's razor dictates a preference for the exclusively feedforward explanations. With regard to the word superiority effect, Grainger and Jacobs (1994) described a feedforward version of their interactive-

https://doi.org/10.1017/S0140525X00303247 Published online by Cambridge University Press

activation, dual-read-out model that produced a word superiority effect. To do so, however, required a sophisticated, post-access, recognition process that could infer a target letter from a word representation.

It is not straightforward whether a post-access process of symbolic inference is simpler or more plausible than feedback. Occam's razor may cut in more than one direction. For example, Bradley and Forster (1987) warned against couching explanations in post-access, recognition processes, because post-access mechanisms will protect indefinitely any theory of lexical access. "Any unwelcome facts about language performance can be attributed to [post-access processes of] recognition" (p. 110). According to them, this renders the pre- versus post-lexical distinction vacuous and justifies our abandoning the distinction, altogether. Feedforward models could be constructed to mimic any data pattern, allowing variegated intermediate levels, response criteria, weighted representations, read-out strategies, and so on. But how are such models simpler than a single, general, feedback principle?

It appears to us that feedforward models accumulate *ad hoc* explanations each time they confront a new feedback phenomenon. Also, counter-intuitive feedback effects exist that would never have been predicted by exclusively feedforward models. Feedback consistency effects are a good case in point. Resonance models predict that symmetrical, consistent relations between spelling and phonology imply faster and more accurate word recognition (e.g., Van Orden & Goldinger 1994). Inconsistent relations, including relations that feed back from phonology to spelling, add time and variability to word recognition performance. The predicted feedback consistency effect is highly nonintuitive. Activation should always flow forward, from spelling to phonology. Why should it matter in *visual* word recognition that a *pronunciation* may have more than one *spelling*?

Feedback consistency effects were first corroborated in performance of English lexical decision (Stone et al. 1997). Words such as *hurt,* with phonological rimes (/_ûrt/) that could be spelled in multiple ways (*_urt, _ert, _irt*) yielded slower lexical decision times and more errors than words with rimes spelled in only one way. Subsequently, Ziegler et al. (1997a) replicated the basic finding in French, and Ziegler and Ferrand (1998) extended it to performance of auditory lexical decision (also in French). What is feedforward for visual presentation is feedback for auditory presentation, and vice versa, a parsimonious qualitative symmetry. Frost et al. (submitted) also observed feedback consistency effects in both visual and auditory lexical decision tasks, in visual and auditory identification accuracy, and in visual word-familiarity judgments.

Feedback consistency effects were not discussed in the target article, possibly because a recent study in French failed to replicate the effect and attributed previous feedback consistency effects to a confound with familiarity (Peereman et al. 1998). We followed up on this failure with new studies, however, which yielded reliable feedback consistency effects using their materials. Feedback consistency also predicted rated familiarity, in new studies that properly treated rated familiarity as a dependent variable (Ziegler & Van Orden, submitted). Incidentally, Jacobs et al. (1998) report that the feedforward version of their interactive-activation, dual-read-out model could not simulate the original feedback consistency effect, while one with feedback could.

Feedback consistency effects satisfy a strong test of the feedback hypothesis. Nevertheless, one could surely construct exclusively feedforward models, after the fact, to mimic the effects – one model for visual word recognition (cf. Taft 1982) and another model for spoken word recognition. The point remains, however, that feedback consistency effects were not anticipated by feedforward models. Data are often indeterminate with respect to modeling frameworks; the same behavioral patterns can be mimicked in more than one framework. For example, key results in physics, sometimes described as falsification of the Newtonian (mechanical) view, were (or could have been) accommodated by the traditional view, after the fact (Einstein & Infeld 1966/1938). According to Einstein and Infeld, the mechanical view was not naively falsified, but collapsed finally under the weight of *ad hoc* assumptions (given an alternative theory with demonstrated utility and generality, cf. Lakatos 1970). It is possible that feedforward models will suffer the same fate. They are conspicuous for their expendable *ad hoc* mechanisms, which multiply to accommodate existing feedback effects, and other classes of findings, such as ubiquitous strategy effects. We may expect such models to become increasingly arbitrary (Gibbs & Van Orden 1998; Stone & Van Orden 1993; Van Orden et al. 1999).

As feedback consistency effects illustrate, resonance models demonstrate their utility by predicting nonintuitive findings (see also Gottlob et al. 1999; Kawamoto & Zemblidge 1992; Lukatela & Turvey 1998; Van Orden et al. 1999; Ziegler & Jacobs 1995; Ziegler et al. 1997b, etc.). They also provide general links to other areas of science. Resonance models instantiate nonlinear dynamical systems theory (Carello et al. 1992; Farmer 1990). Resonance models are attractor networks simulated as nonlinear iterative maps. Nonlinear iterative maps approximate solutions to systems of nonlinear differential equations (Peitgen et al. 1992). Thus resonance models, as dynamical systems, link cognitive performance to contemporary mathematical frameworks that concern the behavior of complex systems. This link suggests additional rigorous tests of feedback models (e.g., tests that focus on stability, cf. Kelso 1995) and reveals truly generic behavioral phenomena (e.g., Case et al. 1995; Raczaszek et al. 1999; Tuller et al. 1994). Most important, this link suggests general hypotheses that may eventually bridge disciplines (cf Kugler & Turvey 1987).

ACKNOWLEDGMENTS

Guy Van Orden is supported by an Independent Scientist Award, National Institute of Neurological Disorders and Stroke (1 K02 NS 01905).

Authors' Response

Feedback on feedback on feedback: It's feedforward

Dennis Norris,[a] James M. McQueen,[b] and Anne Cutler[b]

[a]*Medical Research Council Cognition and Brain Sciences Unit, Cambridge, CB2 2EF, United Kingdom;* [b]*Max-Planck-Institute for Psycholinguistics, 6525 XD Nijmegen, The Netherlands;* **dennis.norris@mrc-cbu.cam.ac.uk www.mrc-cbu.cam.ac.uk james.mcqueen; anne.cutler@mpi.nl www.mpi.nl**

Abstract: The central thesis of our target article is that feedback is never necessary in spoken word recognition. In this response we begin by clarifying some terminological issues that have led to a number of misunderstandings. We provide some new arguments that the feedforward model Merge is indeed more parsimonious than the interactive alternatives, and that it provides a more convincing account of the data than alternative models. Finally, we extend the arguments to deal with new issues raised by the commentators such as infant speech perception and neural architecture.

R1. Definitions

Many commentators' points rest on particular interpretations of the terms "top-down" and "interaction." In several cases, commentators have used these terms quite differently from the way we used them in the target article and, in some cases, quite differently from the way they have used

https://doi.org/10.1017/S0140525X00303247 Published online by Cambridge University Press

them in their own previous writings. When we introduced these concepts we made it clear that we were using them in the senses which most closely correspond to the notion of feedback. Remember, feedback is the central issue, not only because it is the focus of the target article, but because it is the focus of the debate in the literature.

R1.1. Interaction. In the target article, we used the term interaction as synonymous with feedback. Two stages which interact are linked by feedback as well as feedforward connections, that is, each can influence the other.

Although "interaction" is most commonly used to characterise the information flow between processes, interaction is sometimes used instead to make statements about how different kinds of information are used or combined. So, if lexical and phonemic knowledge are combined in making phonemic decisions, one might want to say that lexical and phonemic knowledge interact. We can call these two senses of interaction "process interaction" and "information interaction" (Norris 1980). Information interaction does not imply process interaction. For example, one might (like **Luce et al.**) make no distinction between lexical and phonemic processes, but still characterise lexical and phonemic information as different kinds of knowledge. In Merge, lexical and phonemic knowledge are combined in the decision nodes, but no processes in the model interact with one another. Merge has no process interaction and no feedback. We have therefore not adopted a narrow definition of interaction as **Pitt** and **Slowiaczek** suggest, but we have tried not to conflate the two quite distinct senses of interaction.

R1.2. Top-down. The sense of top-down which predominates in the psychological literature refers to the direction of information flow within the system. In this architectural sense, flow of information from one process back to previous processes in the chain is referred to as top-down. Merge is not top-down. Lexical units give output only to decision units which are themselves output units and are not part of the processing chain delivering input to lexical units. Note that top-down does not refer to the direction of lines on the page. If it did, classification of models would depend how one drew the lines, and possibly on how one held the page! We leave it as an exercise for the reader to redraw Figure 1 so that all lines from the input phoneme and lexical levels point upwards.

Note that although this sense of top-down gets close to the concept of feedback, and is generally used synonymously with feedback in the literature, it is not identical. Nonspecific top-down flow of information, such as might be involved in generalised attentional activation, would not in any way be the same as specific feedback from particular lexical items which altered the processing of specific phonemes. The target article concerns itself with specific feedback, and not with nonspecific top-down effects, such as attention, which are not part of a lexicon-phoneme feedback loop.

Top-down is also sometimes used in a less well-defined sense that appears to be a close synonym of information interaction. In this second sense, top-down is used to mean that information at one level of analysis is brought to bear in processing information specifiable at a more fine-grained level of description. So if lexical knowledge is used in any way to influence decisions about phonemes, this is evidence that lexical and phonemic information are combined in a top-down fashion. As the target article demonstrates, this is quite independent of the issue of feedback, or even direction of information flow. Strictly feedforward models like the Race model, Merge, and FLMP are, in this sense, top-down. (Thus although both Shortlist and Merge are feedforward and bottom-up in terms of information flow, they do use the lexical constraints that **Benkí** wants them to use.)

In concluding this section we emphasise that our choice of terminology and definitions is not arbitrary. It reflects the core issues in the "interaction" debate that has been pursued in the literature for over 20 years. This has been part of a more general debate about the architecture of the language processing system, perhaps most clearly set out by Forster (1979). During this time there has been no disagreement about whether lexical information can influence phoneme identification. The debate has been about process interaction and feedback (Cutler et al. 1987; Elman & McClelland 1988; Frauenfelder et al. 1990; Massaro & Cohen 1991; McClelland 1991; Pitt 1995; Pitt & McQueen 1998; Samuel 1997).

R2. Theory

None of the commentaries has explained why feedback might be necessary. **Tanenhaus et al., Montant, Shillcock,** and **Stevens** all express their conviction that it really should (under certain circumstances) be helpful. But without specific reasons why our arguments might not hold under such circumstances, pleas like "feedback is surely helpful" (Tanenhaus et al.) remain wishful thinking. In the following sections we discuss the general points of theory that were raised. The majority of the commentaries have concentrated on issues concerning the Merge model itself, raising three main concerns: that Merge might not be "ecologically valid"; that, contrary to our characterisation, Merge might really be a top-down or interactive model after all; and that Merge might not really be simpler than interactive models.

R2.1. Ecological validity. In order to make the case for a feedforward model of speech perception we must be able to explain data from laboratory tasks that have been presented as evidence for feedback. Merge was designed to explain these data in a manner consistent with the principles of Shortlist, which is concerned with modelling word recognition in continuous speech. Some commentators question the ecological validity of Merge (**Appelbaum, Benkí, Vroomen & de Gelder**). After all, Merge has been used to explain behaviour in laboratory tasks involving metalinguistic judgements. In part this is true. None of us is primarily concerned with explaining laboratory data rather than naturalistic processing. Psycholinguists have to live with the fact that the experimental tasks they use do not directly reveal the inner workings of the speech perception system. These tasks do, however, give us some very good clues, whereas naturalistic observation of speech perception tells us nothing at all about processing architecture. To make the best use of these clues, models like Merge must attempt to explain both normal processing and performance in laboratory tasks. The data that Merge explains have on occasion been taken as evidence for feedback, so we cannot ignore

https://doi.org/10.1017/S0140525X00303247 Published online by Cambridge University Press

these data. The commentators who criticise the ecological validity of Merge present no alternative.

R2.2. Terminological confusion. Some commentators seem to be in a state of terminological confusion. This worries us because it indicates that there is confusion over the use of some fundamental terms in the literature. More worrying still is the fact that some commentators (**Pitt, Samuel**) who have published papers using terms like "top-down" in the standard sense of direction of information flow used in the target article, use the terms in a quite different sense in their commentaries.

Appelbaum and **Samuel** suggest that the interaction debate has not been about processing interaction and information flow (see sect. R1, Definitions) and that we should now call feedforward models like Merge and FLMP interactive. **Pitt** believes that we have narrowed the meaning of "interactivity" by restricting it to cover only top-down feedback. Interestingly enough, if we look at recent papers on interaction written by Pitt and Samuel, we see that the opening paragraph of each of these papers defines both the terms and the issues very clearly (Pitt 1995; Pitt & McQueen 1998; Pitt & Samuel 1993; Samuel 1997). We quote here from Samuel (1997, p. 97) (although the clearest definition is to be found in Pitt & McQueen), "Some models hypothesize strictly bottom-up connections between the lower level (phonemic) and higher (lexical), while others posit bidirectional information flow." The fact that bidirectional information flow really is the issue is confirmed in an earlier paper by McClelland (1991, p. 3), which makes it clear that the debate is whether "perception involves a bidirectional flow of information," a point endorsed by Massaro and Cohen (1991) who cite the same quotation from McClelland. It is not surprising that our own papers contain many similar quotations (e.g., Cutler et al. 1987; McQueen 1991; McQueen et al. 1999a; Norris 1992).

Why does **Samuel** now think that Merge is interactive and nonautonomous? Given that he has adopted the standard conventions in the past, it is hard to know why he adopts different interpretations here. Part of Samuel's problem may be attributable to the fact that he wrongly equates phoneme and decision nodes in Merge with the feature and phoneme nodes of TRACE. In TRACE, features feed into phonemes, which in turn feed into words. In Merge only the input phonemes feed into words. Decision units cannot be equated with phoneme nodes in TRACE as they do not feed into lexical units. But Samuel has chosen to call the connections from lexical to decision nodes "top-down." He then states that "Norris et al. offer several reasons for including such top-down connections, and they are exactly correct: Top-down lexical influences are necessary." It is important that decision nodes are influenced by the lexicon, but this influence does not involve top-down flow of information. Information in these connections passes from input to output.

Possibly **Samuel** believes that any information flow from lexical to phonemic representations is "top-down": "if lexical nodes influence the activation of phonemic codes, a model is *not* autonomous." Note that the effect of this would be to redefine "top-down" so that any demonstration of lexical effects on phoneme identification (which must surely be based on phonemic codes) is "top-down." All researchers in the field have been in agreement about the existence of lexical effects on phoneme identification for more than 20 years (see Cutler & Norris 1979 for review). Furthermore, lexical nodes have always influenced phonemic codes in bottom-up models. In the Race mode, lexical access makes the lexically based phonological code of the word available. In Samuel's terms, lexical nodes activate phonemic codes. If we were to adopt the terminology of Samuel's commentary everybody would accept that the data argue for "top-down" processing and all of the models would be "top-down" too. Have all of us who have worked on this question (including Samuel) been wasting our time? No. We have all been addressing the much more interesting question of whether there is top-down feedback. Furthermore, Samuel himself has made some rather ingenious contributions to this debate (e.g., Samuel 1997). We only hope that his terminological *volte face* is just a temporary aberration and not an attempt to rewrite history and pretend that he believed in what we are now proposing all along. It is not that we do not want him to agree with us. But we think he should agree on our terms.

Appelbaum suggests that we have reinterpreted the interactive/autonomy distinction. But in fact it is Appelbaum who seems to have interpreted the distinction incorrectly. In an earlier paper, Appelbaum (1998) assumed that lexical effects on phonemic processing (e.g., Ganong 1980) were evidence of "top-down information flow" (Appelbaum 1998, p. 321) and hence evidence against a modular stage of phonetic perception. The Race model (Cutler & Norris 1979) had long ago shown that lexical effects are entirely consistent with a modular stage of phonetic perception, and Merge maintains modular prelexical processes. Remember, decision nodes are not prelexical. Appelbaum's attempts to use lexical effects on phoneme decisions as evidence against modularity are therefore flawed; and her criticisms of our terminology may stem from a misreading of the literature.

Appelbaum, Pitt, and **Samuel** also seem confused by our application of the term "autonomous" to Merge. As we pointed out, autonomy is properly applied to stages rather than models, and Merge "preserves the essential feature of autonomous models – independence of prelexical processing from direct higher-level influence" (sect. 5.1, para. 7). Prelexical processing, and that is what the debate is about, is autonomous in Merge. The appropriate classification of the decision units is less straightforward. The decision units are flexible and configurable according to task demands, so they certainly do not constitute a Fodorian (Fodor 1983) module. Once configured for the task, however, they take input from two sources (lexical and prelexical) and then produce an output without interference or feedback from subsequent processes. This justifies the label "autonomous."

Finally in this section we should respond to the claim of **Tanenhaus et al.** that there is feedback from lexical to decision nodes. Where there is no feedforward (decision to lexical) there can be no feedback. The lexical-to-decision connections are feedforward.

R2.3. Parsimony. The question of parsimony rests largely on the issue of whether the decision nodes in Merge are an added extra that interactive models can do without (see **Doeleman et al., Gow, Murray, Pitt, Slowiaczek,** and **Whalen**). For example, there are no explicit decision nodes in TRACE so, although TRACE has interaction, it has no counterpart of Merge's decision nodes. How then can we claim that Merge is simpler than TRACE? There are two

https://doi.org/10.1017/S0140525X00303247 Published online by Cambridge University Press

parts to our answer. As we explained in the target article, one is that even if TRACE is as simple as Merge, it cannot account for the data (e.g., Pitt & McQueen 1998). We will remind readers of the details of this argument when discussing comparisons between Merge and TRACE in a later section. The second is that all models need some form of decision mechanism. Merge only appears more complex because it makes that mechanism explicit.

R2.3.1. Decision processes in Merge. Most psychological theories give a less than complete account of how a model might be configured to perform various experimental tasks. For example, TRACE and Merge must be able to perform either lexical decision or phoneme identification depending on the requirements of the task. In early phoneme-monitoring studies, listeners were typically required to monitor only for word-initial phonemes (Foss 1969). By definition, this demands that positional information from the lexicon is combined with information about phoneme identity. Neither Race nor TRACE ever specified a mechanism for performing this part of the task. This is unsurprising because there is practically no limit to the complexity of the experimental tasks we might ask our subjects to perform. Listeners could no doubt be trained to monitor for word-initial phonemes in animal words when a signal light turned green. Correct responding would require combining phonemic, lexical, semantic, and cross-modal information. But this does not mean that we have hard-wired {initial /p/, animal, green} nodes just sitting there in case someone dreams up precisely such an experiment. It certainly does not mean that we should conclude that the processes of colour perception, semantic processing, and phoneme perception all interact in normal speech recognition. A far more likely explanation is that a number of simple non-interacting processes deliver output to a system that can monitor and merge those outputs to produce a response. This system has to have enough flexibility to cope with all manner of bizarre tasks that experimenters, and the world in general, can throw at it. In Merge we have finessed the issue of how this system configures itself, and assumed that we can represent the process of combining different sources of information by a set of decision nodes. Merge does one extra thing. Although we can devise phoneme identification tasks that necessarily take account of lexical information, in the simplest phoneme identification tasks listeners could, in principle, ignore the output of the lexicon (and in fact often appear to do so; Cutler et al. 1987). In Merge we assume that listeners sometimes monitor the phonemic and lexical levels even when this is not explicitly required by the task, and that this is the source of lexical effects in phoneme identification.

Additional evidence that we need something more than just the phoneme nodes of TRACE to perform phoneme identification was reviewed in section 7 of the target article. The ability to perform phoneme identification is not an automatic consequence of being able to recognise spoken words. For instance, it is greatly facilitated by having learned to read an alphabetic script (Read et al. 1986). Furthermore, neuroimaging work reveal different patterns of brain activity in tasks involving explicit phonological decisions from those involving passive listening to speech (Demonet et al. 1994; Zatorre et al. 1992; see Norris & Wise, 1999, for review).

In conclusion then, the decision nodes in Merge do not undermine its parsimony compared to other models. All models must make allowance for the facts that there is a flexible and configurable decision mechanism, that listeners have to learn to interpret the workings of prelexical processes, and that explicit phonological decisions appear to activate parts of the brain not activated during normal speech recognition. The important point is that the decision process is not an optional extra. Without some such process listeners could not perform the experimental tasks we give them. The decision process is not something Merge has but other models can do without. All models need a decision process. Our claim is that when that decision process is taken into account we see that it is probably responsible for lexical effects in phoneme identification, leaving normal speech perception as a feedforward process.

R2.3.2. Rewiring decision nodes. The decision process has to be very flexible. Our suggestion that the connections in Merge might be rewired on the fly is the subject of criticism by both **Grainger** and **Grossberg.** Grossberg's worry about the plausibility of "rewiring" seems to apply to the very literal rewiring that might be done by a neural electrician. Our intention is to capture the process of reconfiguring network connectivity as in the Programmable Blackboard model of McClelland (1986). As we have argued above, all models must be able to configure themselves according to task demands. Grossberg's ART model must find a way of doing this too.

Grainger suggests that rewiring is implausible and unimplementable. The original suggestion for wiring on the fly, as proposed for Merge and Shortlist, rests on the assumption that it is worth adding an extra mechanism in order to save the need to have vast (possibly astronomical) numbers of permanent connections. The issue of rewiring is quite orthogonal to the question of feedback. However, it should be clear that if two different representations (say lexical and decision) are to be wired dynamically, then there must be some way to identify pairs of representations that are to be wired together. Lexical representations should therefore not be considered to be single unstructured nodes. They must contain the form-based lexical representation which can be dynamically associated with an appropriate decision node. It has always been part of Shortlist and the Race model that the lexicon explicitly represents phonological form. Grainger's assumption that a dynamically rewirable version of Merge would have no lexical representation of phonological form is bizarre.

Note that if we set aside the issue of rewiring on the fly, Merge simply does not have the problems **Grainger** supposes. In the simulations we presented, the decision nodes are effectively the output representations. Activating a word activates its phonological form on the decision nodes.

For some reason, **Grainger** believes that the problem of merging lexical and phonemic information presents a problem for Merge which is not faced by his own DROM model (Grainger & Jacobs 1994) simply because the DROM can combine letter and spelling information "at leisure." The speed of the process does not alter the logic of the connectivity. It is fortunate that Merge does not have this problem as DROM would have exactly the same problem.

R2.3.3. Feedback consistency. A further issue of parsimony is raised by **Ziegler & Van Orden** who believe that models with feedback have been able to generate important

https://doi.org/10.1017/S0140525X00303247 Published online by Cambridge University Press

theoretical predictions such as the feedback consistency effect in reading which "would never have been predicted by exclusively feedforward models." Interesting to note, Norris (submitted) demonstrates that the reported data on feedback consistency effects in reading can be well explained by the feedforward multiple-levels model (Norris 1994a) without any modification whatsoever. The reason that a feedforward model can simulate a "feedback consistency" effect is that the effect is not actually due to feedback at all, but to the type frequency of body-rime correspondences. Other things being equal we might expect most rimes to appear in a roughly equal number of words. If those rimes are always spelled in the same way, then the type frequency of each body-rime correspondence will be roughly equal. But, for rimes that are feedback inconsistent (i.e., spelled in more than one way), the major body-rime correspondence will tend to have a lower type frequency than in feedforward consistent words. Feedback consistency has an effect on naming because it tends to alter the type frequency of the correspondence. Feedforward models like the multiple-levels model are sensitive to type frequency. Feedforward models predict the data and correctly explain it as an effect of type frequency which has nothing to do with feedback from phonological to orthographic processing.

R3. Comparison of Merge with other models

R3.1. Merge versus TRACE. Throughout the target article, we claim that Merge is more parsimonious than interactive models. It is quite possible that Merge could be theoretically sound, but actually less parsimonious than interactive models. If models with and without feedback were otherwise equal, and the trade-off were simply between having the phoneme decision units units required by Merge and having feedback, it is hard to see which would be more parsimonious. This is essentially the point raised by **Murray, Pitt,** and **Tanenhaus et al.** How do we set about choosing between similar models? As we pointed out in section R2.3.1 above, all models need some form of decision process. Merge incorporates that explicitly, other models do not. So, comparing Merge and TRACE for parsimony is not actually comparing like with like. TRACE has extra hidden complexity, even though it may have fewer free parameters (Pitt). But most importantly, Merge still satisfies Occam's precept better than TRACE does. Occam's razor favours the most parsimonious theory consistent with the data; TRACE (the original or our modified version) is inconsistent with the data from Pitt and McQueen (1998). TRACE is also inconsistent with the development of phonological awareness with literacy without adding something akin to decision units; and finally TRACE is unable to account for detection of mispronunciations. We did our best to show that TRACE could be modified to account for the subcategorical mismatch data, but that is not enough.

In discussing the Merge simulations, **Tanenhaus et al.** state that we would like to conclude that the superior performance of Merge over the interactive model simulation is "compelling evidence for Merge and against TRACE." This is incorrect. As we point out (sect. 6.1), "With Merge-like dynamics, an interactive model could approximate the correct data pattern." The importance of the simulations is to demonstrate that a feedforward model can account for the subcategorical mismatch data and to show how a model like TRACE might be modified to simulate that data too. The compelling evidence against TRACE comes from the data from Pitt and McQueen and the fact that TRACE fails to account for the bigger picture.

Tanenhaus et al. believe that we have made "questionable linking assumptions between the data and the models" (without saying why they believe this), and they seem to take exception to our assumption that positive lexical decision responses should be made when any lexical node exceeds a threshold. Note that we make exactly the same assumptions about response thresholds for both Merge and the interactive model. There is convincing neurophysiological evidence that reaction times are determined by neural activation thresholds in eye movement control (Hanes & Schall 1996). Both **Tanenhaus et al.** and **Gaskell** remark that the threshold in Merge needs to be precisely set to simulate the correct pattern of data. This is true, but follows from the need to match the model's performance to that of subjects. Subjects in these experiments make few errors. To respond correctly, they must place their decision criterion high enough not to be exceeded by nonword activation and low enough to always be exceeded by word activation. In Merge this requirement ties the criterion down to a range of 0.06 activation units and in the interactive model about 0.1 units. In both models a high criterion within this range leads to equally fast responses to N3W1 and W2W1, whereas the lowest possible criterion would lead to slightly faster N3W1 responses. With the lowest criterion, the N3W1 advantage is twice as large for the interactive model as for Merge. Contrary to what Tanenhaus et al. claim, we would not expect to see evidence of fast lexical decision responses based on early activation of W2 if subjects are responding accurately. Also, contrary to their claims, the RT distributions of our data are clearly unimodal and not bimodal. Because W2W1 and W2N1 follow the same trajectory until cycle 8 there is no way that subjects could possibly make fast "Yes" responses to W2W1 based on early W2 activation without also making erroneous "Yes" responses to W2N1. This is not surprising because the final phoneme is not fully presented until cycle 9. Note that the error rate to W2N1 items is only 3%.

R3.2. Merge versus FLMP: FLMP is running a different race. In terms of the central argument about feedforward processing there is no fundamental conflict between Merge and FLMP. But **Massaro**'s and **Oden**'s commentaries now make us think that, in processing terms, FLMP must be much more different from Merge than we had originally thought.

Both **Oden** and **Massaro** criticise us for having misrepresented FLMP when discussing their account of the Ganong effect, where we say that "the support for a word has nothing to do with the perceptual evidence for that word" (sect. 6.3, para. 6). Oden points out that when they say "support for the voiced alternative given by the following context" (Massaro & Oden 1995, p. 1054) they are not saying that *gift* supports /g/, but that *ift* supports /g/. The evidence for *ift* is independent of the evidence for /g/ whereas the evidence for *gift* would not be. But why is the probability of responding /g/ dependent on the evidence for *ift*? The sequence *ift* does not support /g/ any more than it supports any other phoneme. The word *gift* might support /g/, but there is simply no reason why the sequence *ift* should support any onset phoneme in the absence of infor-

https://doi.org/10.1017/S0140525X00303247 Published online by Cambridge University Press

mation about lexical forms. Oden's claim that *ift* supports /g/ only makes sense if the relevant information is derived from the lexicon. So, when making a phonetic decision, the listener must decompose the input into the phoneme of interest and the residue. Then the residue *-ift* can be fed into the lexicon to determine that /g/ is a possible word onset in this context and /k/ is not. This way the context is independent of the support that /g/ provides for the word *gift*. Of course, at the same time *gift* is also being fed into the lexicon so that the word can be recognised. All of this is to avoid violating independence by feeding only *gift* into the lexicon and allowing lexical information to bias interpretation of /g/. Perhaps Massaro and Oden will think that our attempt to discover the processes behind the FLMP equations has led us to misrepresent them again. But in fact this is the heart of our criticism of FLMP. Although the FLMP equations are simple, they do not specify a process model, and it is far from clear what the underlying process model should be (for similar criticisms see Grossberg et al. 1997). Also, within the broad scope of the FLMP equations, there seems to be just too much room for manoeuver in how they are used to explain any particular piece of data.

This flexibility is apparent in **Oden**'s commentary. In response to our criticism of the FLMP account of compensation for coarticulation, Oden offers a new explanation of sequential effects in FLMP terms of decisions about the "candidate identity of the sequence of words." The implication of this statement is that compensation for coarticulation takes place not at a prelexical, or even a lexical level, but at a new level representing sequences of words. The one straightforward thing we can say about this explanation is that it is wrong. As we will show later in section R4.3, there is abundant evidence that compensation for coarticulation is prelexical. Compensation applies even to nonword stimuli. Therefore, as we originally argued, FLMP still has no plausible account of the Pitt and McQueen data.

Oden suggests that the inhibition in Merge might produce all-or-none decisions. This tends not to be true given the levels of inhibition employed at the decision stage. As we pointed out, adding noise would also stop the model being deterministic. However, there is no doubt that there is work to be done in developing the model to account for both response probability with ambiguous input and speed of responding with unambiguous input (see Carpenter 1999; Ratcliff et al. 1999; Usher & McClelland 1995).

Both **Massaro** and **Meyer & Levelt** criticise us for concentrating too much on modeling activation levels. Massaro assumes we believe that activations are somehow better than response probabilities; Meyer & Levelt suggest that it is preferable to use the Luce choice rule than to allow inhibitory effects on activation. However, the Luce rule is not simply an alternative to inhibition, because models still need a mechanism whereby the rule can be implemented. Any complete model needs an account of how response probabilities are computed. Network models have the ability to suggest mechanisms which show how differences in activation can be translated into differences in response probabilities and latencies (Carpenter 1999; Page 2000).

Massaro criticises models with hidden units as being untestable. Contrary to his claim, however, there is no connection between the fact that networks with hidden units can approximate any measurable function (Hornik et al. 1989) and their testability. Nothing in this work implies that a network trained on a given data set (such as speech input) will then behave as people do (in some experimental task on which it was not explicitly trained). A clear example of this comes from Seidenberg and McClelland's (1989) model of reading aloud. The model is trained to translate orthography to phonology. The model succeeds or fails (it actually does a bit of both) depending on its ability to simulate human reading behaviour, something it was never explicitly trained to do. A model trained directly to reproduce the correct RTs and error rates might not be testable, but then it would just be a redescription of the data. Massaro's criticism of models with hidden units is fallacious, and so, in consequence, is his attempt to extend such criticism to network models in general.

R3.3. Merge versus the distributed cohort model. The commentary by **Gaskell** shows that the Distributed Cohort Model (DCM, referred to in the target article as the post-lexical model) can be modified to overcome the technical criticisms we made in the target article and to simulate the subcategorical mismatch data. This suggests that the subcategorical mismatch data might not be as diagnostic as we originally thought, and that at least one other bottom-up model can account for the data. Although the model still cannot explain the variability in the effect, for exactly the reasons we originally suggested, **Marslen-Wilson** suggests that the DCM probably needs a decision mechanism that can shift attention from the normal phonological output to a lower-level auditory output less influenced by lexical factors. This is essentially the same explanation as in Merge where attention can be shifted between levels. However, the recurrent net architecture still fails as a model of continuous speech recognition for the reasons pointed out by Norris (1994b). Other shortcomings of recurrent networks in speech recognition are highlighted by **Nearey.** Page (2000) presents a more general critique of models relying on distributed representations. One problem that DCM faces is that it is not clear how lexical decisions could be made. Presentation of an input word leads to a pattern of activation across semantic units. Without some independent lexical representation that specifies exactly what pattern of semantic unit activation is to be expected for each word, there is no way to determine whether a given activation pattern actually corresponds to a word or not.

R3.4. Merge and ART. The following line of reasoning is pursued by **Montant:** ART uses feedback, ART is good, therefore this is evidence in favour of feedback. Remember that our central claim is that "Feedback is never necessary." We also pointed out that the best a recognition system can do is to select the stored representation that best matches its input. This holds for ART as much as anything else. In ART, the feedback is part of the mechanism for assessing the degree of match between bottom-up input and stored representations. The same result could be achieved without feedback. Indeed, although most versions of ART use feedback, the feedback is not necessary and is not needed in ART2-A (Carpenter et al. 1991). Grossberg et al. (1997a) demonstrate the similarity between ART and the FLMP equations which do not require feedback to be implemented. Feedback is also absent from the related learning mechanisms proposed by Page (2000). So the fact that ART has such an impressive track record, and normally uses feedback, in no way counters our thesis about feedback not being necessary.

https://doi.org/10.1017/S0140525X00303247 Published online by Cambridge University Press

A potentially more interesting criticism based on ART comes from **Luce et al.** They also think that feedback is needed for ART and is therefore a good idea. But they argue that in ART phonemes and words are just lists of different lengths, so the whole issue of feedback between words and phonemes simply does not arise. Although it is true that phonemes are represented at the list level in ART, they are also represented at a lower level as the elements from which lists are composed. We can see this clear distinction between levels in action in Grossberg's (e.g., Cohen & Grossberg 1986) account of the word superiority effect in reading, which relies on feedback from the list (i.e., letter and word) level to the previous letter level. We presume that the account of lexical effects on phoneme identification would have a similar explanation in ART, in that phonemes could be identified directly from a phoneme level. The alternative that Luce et al. suggest is that attention can be shifted between words and phonemes by attending to different sizes of list. Normally longer list units like words would mask (inhibit) shorter units like phonemes. Such masking would be stronger for phonemes in words than in nonwords. So, while attention to lists of length 1 might overcome the problems faced by phonemes in words, there is no reason why it should lead to facilitation of responses to words. If the crucial representations for phonemes and words are both at the list level, then the model cannot explain the effects of lexical context on phoneme identification.

Overall, our view of ART is that it may well be able to provide the basic building blocks for implementing a model of speech perception. It has addressed many issues which have been ignored by other models. However, the basic principles of ART place few constraints on how the components might be put together to form a complete model, and it is not clear that feedback would necessarily be central to such a model. ART is now being used to simulate real data on speech perception, and we look forward to an ART-based testable psychological model of speech perception.

Grossberg himself argues that feedforward models fail to explain phenomena such as phoneme restoration and backward effects in time (Repp 1980; Repp et al. 1978). First, the argument concerning phoneme restoration is flawed because it depends on the assumption that the source of the restored phonemic percept is in the input representation rather than being derived from the lexical representation. Second, the existence of backward effects in time has nothing to do with the feedforward/feedback distinction. Shortlist, a strictly feedforward model, simulates a range of backward-in-time phenomena (e.g., Norris et al. 1995; 1997).

R4. Data

Thirty years ago Foss (1969) introduced the phoneme-monitoring task to psycholinguistics (presumably precipitating **Warren**'s musings over phoneme detection during a colloquium that year). We, like many users of the task, would not want to claim that it taps directly into necessary stages of speech processing. Indeed, this is one of the motivating factors for our development of Merge. However, we do believe that in the past three decades spoken-word perception has been an enormously active research field in which real theoretical progress has been made, and that this is in part due to Foss and the other pioneers who equipped psycholinguistics with the necessary empirical tasks. That these tasks often involved metalinguistic decisions is a consequence of our inability to measure perception directly; **Doeleman et al., Marslen-Wilson, Meyer & Levelt,** and **Murray** all remark on the undesirability of this situation. Meyer & Levelt further claim that the study of speech production is less bedevilled by the indirect measurement problem than the study of perception, because in their work on production they are able to measure (and model) onset of articulation. We suspect that this claim should be taken with a pinch of salt; articulation can be seen as the bottleneck of an otherwise far more rapid speech production process (Levinson 2000), and this allows for the possibility that production processes such as lexical access are not directly reflected in articulation speed at all. For perception, however, both Murray and Meyer & Levelt point to the usefulness of recently developed eye movement paradigms (Tanenhaus et al. 1995). So far these tasks can only be used with a restricted set of specified response options (a display of which the subject knows all the members in advance), which means that many issues are as yet outside their range of usefulness; we certainly hope that this type of approach will be incorporated in further tasks of greater refinement in the near future. Even better, of course, would be appropriate refinement of brain imaging techniques; these are still laughably far from being able to provide insight into the sort of questions dealt with in the experiments we have discussed (such as the processing difference involved in hearing two versions of *job* in which the *jo-* portion comes respectively from *jod* or from *jog*).

At the moment, however, the data on phonemic decision making provide the only insight into such questions, and none of the commentaries lead us to revise our conclusion that the Merge model currently provides the best available account of these data. In this section, we discuss the comments addressed to specific questions about the decision data. The presentation follows the order we adopted for our review of the data in the target article (sect. 4), but ends with a new subsection on acoustic-phonetic processing.

R4.1. Variability of lexical effects. The variability of lexical effects on phonetic categorization and phoneme monitoring tasks is a challenge to models with feedback. No commentator contests this claim. **Pitt,** however, draws attention to a specific kind of variability of lexical involvement in phonetic categorization which we did not discuss in the target article. This is that lexical effects in categorization change over time (Fox 1984; McQueen 1991; Pitt & Samuel 1993). Pitt questions whether Merge could deal with this variability. It can. Lexical involvement in the model tends to build up, and then decay over time (although the situation that is being modelled is somewhat different, the lexical effects in Merge's subcategorical mismatch simulations [see Fig. 3b] increase as lexical activation builds up, and then decrease as phoneme node activation reaches asymptote). It is possible that experiments on word-initial categorization have tended to tap into the incrementing phase of lexical involvement (the standard finding is that there are larger lexical effects in slower responses), while those on word-final categorization (here there are smaller lexical effects in slower responses) have tended to tap into the decrementing phase. We have already begun to address this issue experimentally (McQueen et al. 1999b).

https://doi.org/10.1017/S0140525X00303247 Published online by Cambridge University Press

It is important to note that the pattern of lexical involvement in word-final categorization, though not problematic for Merge, is in fact problematic for models with feedback, like TRACE (as McQueen 1991 argued). TRACE simulations in McClelland (1987, Fig. 1.2, p. 12) show that, as processing continues, lexical feedback acts to increase the difference in activation between the phoneme nodes for the lexically-consistent and lexically-inconsistent phonemes (/t/ and /d/ given *dar*ʔ in McClelland's example). TRACE therefore wrongly predicts that lexical involvement in word-final categorization should build up over time. We thank **Pitt** for reminding us about another of TRACE's frailties.

R4.2. Facilitation versus inhibition in phoneme monitoring. The results presented by **Connine & LoCasto** show that listeners were faster to detect the target /m/ in the nonword *chorum* (which is close to the word *chorus*) than in the control nonword *golum* (which is not close to any real word). This finding replicates Wurm and Samuel (1997) and supports Wurm and Samuel's argument that more word-like nonwords are easier to process than less word-like nonwords. This is one reason why Frauenfelder et al. (1990) may have failed to find inhibitory lexical effects in nonwords like *vocabutaire.* Despite **Samuel**'s protestations, however, it remains the case that there was no *direct* evidence for this kind of inhibitory effect when the target article was written. Why Connine and LoCasto's commentary is important is that it now provides us with direct evidence of lexical inhibition. Listeners were slower to detect the /ʃ/, for example, in the nonword *chorush* than in the control nonword *golush.* It would appear that when the target phoneme is close enough to the sound it replaces in the base word (/ʃ/ is only one feature different from the /s/ in *chorus*) there is sufficient support for the lexically-consistent sound (/s/) to overcome the benefit due to *chorush* being more word-like than *golush,* resulting in a small net inhibitory effect.

Connine & LoCasto claim that their results are inconsistent with the Merge model. Specifically, they suggest that the bottom-up priority rule in Merge might have to be abandoned. They are right to suspect that we would be loath to remove this rule; it serves the important function of preventing hallucinations. These new results are, however, consistent with Merge and in fact provide support for the rule. The inhibitory effect appears to occur only when the target phoneme is phonetically close to the phoneme it replaces, that is, when the target itself provides some bottom-up support for the replaced sound. Since /ʃ/ activates the /s/ decision node, the word node for *chorus,* following the bottom-up priority rule, can also activate the /s/ decision node. Due to the resulting competition between the /s/ and /ʃ/ nodes, /ʃ/ decisions will be delayed. When the target provides no support for the sound it replaces (/m/ in *chorum* differs in place, manner and voicing from /s/) the bottom-up priority rule will prevent lexical activation from supporting the lexically-consistent phoneme, and no inhibition will be observed.

We agree with **Connine & LoCasto** that attentional processes have an important role to play in language processing. No model of phonemic decision making has a satisfactory attentional component. Merge, like any other model, would be strengthened if it included a fuller account of attentional factors.

R4.3. Compensation for coarticulation. The results of Pitt and McQueen (1998) are particularly important in the feedback debate. They found a dissociation between a lexical effect on the labeling of word-final fricatives and no lexical effect on the labeling of following word-initial stops (e.g., categorization of the ambiguous fricative in "jui? ?apes" as /s/, but no increased tendency to label the following stop as /k/, consistent with compensation for coarticulation following /s/). This dissociation is very problematic for models with feedback, like TRACE. If feedback modified the activation of the /s/ node at the phoneme level in TRACE, the compensation for coarticulation mechanism at that level of processing ought to have been triggered. The results are however consistent with the Merge model, in which the lexicon can influence fricative decision nodes, but cannot influence the prelexical compensation mechanism.

Some commentators question this argument. **Pitt** points out that the compensation process may not be purely prelexical, while, as we have already discussed, **Massaro** and **Oden** wish to maintain their view that the process operates at a high-level integration stage in FLMP. The evidence, however, suggests strongly that compensation for coarticulation has a prelexical locus. Pitt and McQueen's (1998) data in fact suggest this: It would be hard to explain the dissociation in lexical involvement between the fricative and stop decisions if compensation did take place at the decision stage. Mann and Repp's (1981) original demonstration of fricative-stop compensation was based on nonsense strings (like /skɑ/ and /uʃtɑ/), suggesting that the process does not depend on lexical access (contrary to Oden's suggestion). Most authors therefore agree that fricative-stop compensation is prelexical (Elman & McClelland 1988; Pitt & McQueen 1998). **Brancazio & Fowler** argue that liquid-stop compensation is also owing to a prelexical process.

A particularly striking demonstration that liquid-stop compensation does not operate at the phoneme decision stage is provided by Mann (1986b). Japanese listeners who could not identify English /l/ and /r/ correctly showed appropriate compensation in their labeling of stops following /l/ and /r/ (i.e., more /gɑ/ responses after /ɑl/ than after /ɑr/). These subjects showed the same amount of compensation as both native English speakers and Japanese listeners who were able to identify /l/ and /r/. The process responsible for compensation for coarticulation between liquids and stops (and, by extension, probably the mechanism for fricative-stop coarticulation) therefore appears to operate at the prelexical stage, that is, at a level of processing below that at which explicit phoneme decisions are made.

Pitt and McQueen (1998) also showed that compensation for coarticulation following ambiguous fricatives could be triggered by Transitional Probability (TP) biases in the nonword contexts in which the ambiguous fricatives were placed. Previous demonstrations of lexical involvement in compensation for coarticulation in which the word contexts had TP biases (Elman & McClelland 1988) could thus be due to a prelexical process sensitive to TPs (and not to lexical feedback). The remarks of **Brancazio & Fowler, Doeleman et al.,** and **Massaro** suggest that they may have misunderstood Pitt and McQueen's results. It is therefore important to emphasize that Pitt and McQueen did not show that the compensatory effect was owing to a TP bias rather than to a bias based on sensitivity to coarticulation.

https://doi.org/10.1017/S0140525X00303247 Published online by Cambridge University Press

That is, they did not show that there was a TP mechanism instead of a compensation for coarticulation mechanism. Rather, they showed that the process which compensates for coarticulation could be triggered by a TP bias in the context. In other words, the compensation process can be activated either by unambiguous fricatives or by an ambiguous fricative in a context where, for example, TPs favor /s/ over /ʃ/.

Vroomen & de Gelder question the distinction between TPs and lexical information. They ask why, in the Merge account, statistical regularities can play a role in prelexical processing while lexical knowledge cannot. Our answer is that the data, in particular the dissociation between the effects of lexical biases and TP biases in Pitt and McQueen (1998), but also other similar dissociations (Vitevitch & Luce 1998; 1999), suggest that the effects result from two distinct sources of information, and that TP information is stored prelexically. Vroomen and de Gelder also question whether TP information can assist prelexical processing which is already very robust. As we discuss in section R4.7, we agree that prelexical processing is very efficient. Pitt and McQueen (1998) therefore suggest that TPs will be of most value when the speech signal is ambiguous (even if that is a relatively rare phenomenon outside the psycholinguistic laboratory). Vroomen & de Gelder's suggestion that TPs could only be learned using some form of feedback is incorrect. TPs are statistical regularities in the speech signal, and thus can be learned from the signal by a feedforward system.

R4.4. Phonemic restoration and selective adaptation. In his commentary, **Samuel** expresses concern that we are unwilling to accept the data in Samuel (1997) as evidence of feedback. He attributes two misconstruals to us. First, he tries to undermine our point that the adaptation produced by restored phoneme looks different from that obtained with real phonemes. There is no need to tally up just how many adaptation effects reported in the literature are limited to the category boundary and how many are spread over the entire continuum; the point remains that the effects with real and restored phonemes in Samuel (1997; see Figs. 1 and 2, pp. 102 and 104) do not look the same. This worried us, and still does.

Second, **Samuel** suggests that we have distorted the results of Samuel (1997) and of Samuel and Kat (1996), claiming that we suggested that the adaptation occurs at the lexical level. We did not. We agree with Samuel and Kat (and others) that adaptation may well operate at several different levels of processing, but we did not propose that the lexical level is one of them. The crucial issue is the locus of the adaptation effect with restored (noise-replaced) phonemes. We suggested that Merge could account for Samuel's (1997) data if the locus of the adaptation effect with restored phonemes is found to be at the decision stage (which can indeed be equated with Samuel and Kat's "categorical" level; in both accounts, these levels are responsible for categorical decisions). We also argued for a type of bottom-up priority in selective adaptation, that is, that adaptation effects are driven primarily by the information in the speech signal, rather than by phonemic precepts. The failure to find lexical effects with intact adaptors (Samuel 1997, Experiment 3) is thus consistent with the proposed account in the Merge model. Lexical context may bias processing at the decision level with noise-replaced adaptors but not with intact adaptors for two reasons: because there is no ambiguity at the decision level which lexical context can act upon when the adaptors are intact; and because intact adaptors will produce adaptation primarily at lower levels, which (at least in the Merge model) can not be modulated by the lexicon.

In short, there is nothing in **Samuel**'s commentary to change our view of the Samuel (1997) data. We agree that these are potentially crucial data in the feedback debate. However, the locus of the adaptation effect with noise-replaced adaptors remains to be established. Given the importance of these findings, we have attempted to replicate them, but have been unable to do so (McQueen et al. 1999c).

R4.5. Lexical effects on phonemic decisions in nonwords. It is argued by **Newman** and **Brancazio & Fowler** that a prelexical mechanism sensitive to simple TPs cannot be responsible for the effects on phonetic categorization in nonwords reported by Newman et al. (1997). We agree that since simple (diphone) probabilities were controlled by Newman et al. they cannot be the source of the effect. Higher-order (longer range) probabilities may have played a role, however. Newman in fact suggests that the probabilities between the initial and final consonants in her materials may have had an effect (though Brancazio & Fowler argue that these probabilities were also controlled). But what about the probabilities of the complete strings? Brancazio and Fowler assume that these were all zero. The CVCs only have zero probability on a strictly syllabic account, however. Though all of Newman et al.'s items were nonwords, and none appear as syllables in English words, some of the sequences do nevertheless occur in English words (e.g., *beysh* appears in *probation, kice* in *skyscraper*). In a count of the CELEX database (Baayen et al. 1993), we found that in one of Newman et al.'s sets (*beysh-peysh/beyth-peyth*) the triphone probability biases made the same (correct) predictions as the lexical neighborhood biases, while in another set (*gice-kice/gipe-kipe*) the triphone probabilities made the opposite (i.e., incorrect) predictions to the lexical neighborhoods. In the other four sets in Newman et al. (two which showed neighborhood effects, and two which showed no effects), we found no matching triphones (except for *toish* in *toyshop*).

Note that a dictionary-based count is a blunt instrument that can only approximate the frequencies of triphones (or diphones, or whatever) in continuous speech. The CELEX count misses strings across word boundaries in running speech, like *gice* in "big icecream," which at least theoretically might modulate prelexical TP sensitivities. The CELEX analyses nevertheless suggest that although some of the effects reported in Newman et al. are almost certainly due to the effects of lexical neighborhoods, some may be due to a prelexical mechanism sensitive to higher-order sequential probabilities. As we suggested in the target article, more work needs to be done to tie down the locus or loci of these effects. We also need to know more about the nature of the TP mechanism. As Pitt and McQueen (1998) pointed out, we do not yet know what the limits of listeners' TP sensitivity are (whether diphone, triphone, or even longer sequences are involved; whether syllable structure constrains sensitivity or not; and so on).

Whether these effects prove to be entirely prelexical, entirely lexical, or a mixture of the two, they do not challenge

https://doi.org/10.1017/S0140525X00303247 Published online by Cambridge University Press

the Merge model. Though **Newman** agrees that Merge could explain effects driven by a prelexical mechanism, she questions whether the model could explain effects at the lexical level, arising from the joint influence of gangs of words. She is right that there are strict limits on the number of words activated at the lexical level in Merge (as in the Shortlist model, whose name in fact reflects this property). In Shortlist, the default maximum number of candidate words considered to begin at any particular segmental position is 30. As we pointed out in the target article, the number of words can be reduced considerably without impairing Shortlist's performance; that is, correct recognition is still achieved even when only two words are allowed in the shortlist (Norris 1994b). We did not mean to imply however that the maximum (in Merge or Shortlist) should be as small as two. Indeed, other effects of competitor neighborhood size on word recognition (Norris et al. 1995; Vroomen & de Gelder 1995) have suggested that the shortlist maximum should be larger than two. Shortlist is able to simulate such data successfully with the maximum remaining at the default of 30 (Norris et al. 1995; 1997). Although Newman et al.'s data have not yet been simulated, we think that it is reasonable to assume that Merge, operating with a similar shortlist size, would capture effects due to gangs of lexical neighbours (the largest gang in Newman et al. 1997 had 14 members).

R4.6. Subcategorical mismatch. The possibility is raised by **Whalen** that the mismatches in the cross-spliced items in McQueen et al. (1999a) and in Marslen-Wilson and Warren (1994) had overt ambiguities, and thus that listeners used nonphonetic means to resolve these ambiguities. The data speak against this possibility. Although it is true that trained phoneticians could, with careful listening, possibly detect the mismatches in the materials, we do not believe that the naive subjects used in McQueen et al.'s experiments were able to detect the mismatches, at least when they were presented with the full items. Once the listeners had heard each final stop, they were able to identify it rapidly and accurately. If the materials had been overtly ambiguous, one would not expect mean phonetic decision latencies and error rates on the order of 650 msec and 5% and mean lexical decision latencies and error rates of about 470 msec and 8% (McQueen et al. 1999a, Experiments 1 and 3, cross-spliced conditions). The gating experiment in McQueen et al. shows that listeners were sensitive to the information in the pre-splice portions of the words (as does the forced-choice vowel identification task). But only in the earlier gates did listeners tend to respond with words consistent with the pre-splice portions of the cross-spliced items (e.g., *sloot* responses to the W2W1 word *sloop,* made from the [slo] from *sloot* and the [p] from *sloop*). Once listeners had heard the final stop, over 85% of their responses reflected the identity of the release burst (e.g., *sloop* responses). We therefore believe that the effects in these experiments reflect the operation of bottom-up speech processing, as modeled in Merge, rather than conscious ambiguity-resolution processes.

R4.7. Speech processing. Speech recognition is a difficult and complex process. Several of the commentators seem to have based assumptions of top-down feedback solely on intuitions that a complex process must necessarily be error-prone, and hence incapable of succeeding on its own without reference to other processing levels. Thus we read that speech "is characterized by noise and variability" (**Tanenhaus et al.**) and that ambiguity in speech signals "makes it very unlikely that a pure bottom-up analysis can be efficient" (**Montant**) so that "feedback would be helpful" (**Tanenhaus et al.**); the system should not be designed to be error-free and optimal because it is not actually error-free and optimal (**Connine & LoCasto**). These commentators are psychologists, and their intuitions do not appear to be shared by those commentators who are phonetic scientists. The assumption of error-prone front-end processing can be contrasted with the explicit detail of speech processing laid out in the commentary by **Kingston,** in which we see a picture of massive redundancy producing a bottom-up information flow of such richness that there is room for portions of it to be lost without damage to the end result. Other commentators who are phonetic scientists (**Benkí, Nearey, Stevens, Whalen**) likewise display no such intuition-based assumptions about defective front-end processing warranting feedback: for Whalen, the claim that lexical feedback cannot in principle improve speech processing is "sound"; for Benkí our arguments against feedback are "convincing" and lexical effects should better be viewed in terms of bias; Nearey points out that the human system even in high noise with arbitrary and unpredictable input does a remarkably good job, far better than any existing ASR system; Stevens accepts bottom-up processing alone for the same situation of words in isolation.

A comparable contrast between views can be seen in the remarks of **Warren, Nearey,** and **Slowiaczek** on the issue of phonemic representations in speech processing. As we pointed out in the target article (sect. 7), the framework we have proposed is compatible with a range of possible front-end implementations. Certainly the experimental evidence (including of course that of our own work on subcategorical mismatch) indicates that listeners process speech input continuously and not as a sequence of independent phonemes. The evidence from his own laboratory which Warren so amply cites is fully consistent with the consensus position. Warren interprets such evidence as indicating that phonemes have no role to play in human speech processing. Nearey, however, on the basis of his own work, argues for "phoneme-like units," and Slowiaczek makes a strong case for phonemic representations on the basis of evidence from phonological priming. Our own position is closer to that of the latter two commentators, but the crucial point here is that nothing in the Merge/Shortlist framework depends on whether or not phonemic representations intervene in speech recognition. Phonemic decisions are based on output from the decision nodes, which are separate from the direct input-to-lexicon processing path.

Stevens, taking up the issue of the input-to-lexicon path, describes a casual-speech multi-word utterance the recognition of which, he maintains, involves the kind of top-down processes which the target article argues against. However, the processes he describes do not involve feedback. He proposes acoustic processing that produces a pattern of features; these features in turn generate a cohort of potential word sequences. This is exactly the process of multiple activation of candidate word sequences embodied in Shortlist and indeed most current spoken-word recognition models. Stevens then proposes application of rule-based transformations of the activated word forms. Rules are, of course, by definition not lexically stored information. Application

https://doi.org/10.1017/S0140525X00303247 Published online by Cambridge University Press

of the rules will then "lead to a pattern that matches the pattern derived from the acoustic signal." This is exactly the bottom-up priority embodied in Merge and Shortlist. Feedback, in contrast, would allow the reverse – transformation of the pattern derived from the acoustic signal to match the lexical form. That is, where Stevens's rules allow the system to accept, for instance, a nasal as a possible instantiation of a voiced fricative, top-down feedback would result in the system altering its analysis of the input, and deciding that what had been heard was a voiced fricative and not a nasal at all. Stevens does not think that this happens, and nor do we: there is no feedback in speech recognition.

Finally, the speech scientist commentators point to some levels of complexity which we had not considered explicitly in the target article: **Whalen** describes non-sequential context effects requiring integration of acoustic information across time, **Nearey** discusses the need for temporal sensitivity in the front-end processor, and **Stevens** (as also the commentary by **Gow**) highlights the fact that phonological processes can transform featural representations of phonetic information. There are many further aspects still to the complexity of speech processing. But complexity is not ipso facto a warrant for feedback.

R5. The wider context of language processing

Several commentators relate our arguments to aspects of human language processing beyond the circumscribed domain of the evidence we reviewed. We used research on phonemic decision-making in speech recognition as a clear case study in which to examine the need for feedback in modeling the research evidence. But speech recognition is just one function of the human language processing system. This system not only recognises speech but also produces it; the relationship between our model and models of speech production has been raised by **Meyer & Levelt.** The system processes auditory information for speech recognition; but it is also capable of drawing on visual information to the same end, as noted by **Brancazio & Fowler.** The system recognises words; but it also recognises sentence structure, raised in the commentaries by **Isel** and **Shillcock.** Furthermore, the adult listener's recognition system has developed from an initial state via a process of language acquisition in the child, as **Jusczyk & Johnson** discuss; and it is implemented, as a number of commentators stress, in the neural architecture of the human brain. All these comments provide welcome views of the place of Merge in the wider context of language processing.

R5.1. Production and perception. It is proposed by **Meyer & Levelt** that certain representational levels in the language processing system are shared between production and perception, and that feedback must therefore necessarily occur between those levels. This speculation prompts two obvious remarks. One is that sharing of resources at these levels is as yet unsupported by empirical evidence. Experiments summarised by Levelt et al. (1999) support tripartite lexical processing in production (lexical concepts, syntactic words, phonological forms), but to our knowledge such a division is not indicated by empirical evidence for perception (although note that **Gaskell** proposes a division between lexical content and form, implemented without feedback in the DCM). The second remark is that bidirectional connectivity is the prerequisite for feedback, but is not itself feedback; feedback occurs when the connections are used in both directions during the same processing operation. If connections between two levels are used in only one direction during language production, and only in the other direction during language recognition, there is no feedback. Certainly there is room for further investigation of such issues.

R5.2. Syntactic processing. As **Shillcock** points out, the recognition of function words is dependent upon syntactic context and hence might be more likely to involve feedback. Studies in Dutch (Haveman 1997) have in fact shown comparable priming effects for function and content words, and no evidence for the prediction of function words from syntactic context. **Isel,** responding to our remarks in section 3.5.2 about the relationship between syntactic and semantic processing during comprehension, describes ERP studies which indicate early independence of the processing of lexical gender and of sentence semantics. Modulation of the effects of gender by semantic factors occurs only at a later processing stage. Gender is a property of words which does not alter its type (masculine, feminine, and neuter, in the case of the study cited by Isel) as a function of syntactic structure, but can alter its expression; for example, to mark case relations. Gender type can thus hardly serve as the prototypical measure of syntactic processing; indeed, disambiguation via gender type has been shown not to block concurrent availability of alternate parses of an ambiguous syntactic structure (Brown et al., in press; Van Berkum et al. 1999). There is however separate electrophysiological evidence that syntactic analysis of verb agreement is independent of semantic processing (Hagoort & Brown, in press). Similar studies have shown separate processing effects of content and function words, at least in visual processing (Brown et al. 1999; Ter Keurs et al. 1999). As we pointed out in section 3.5.2, current models of syntactic/semantic processing differ significantly with respect to feedback; we welcome the growing attention paid to sorting out these differences via neurophysiological investigations.

R5.3. Audio-visual processing. We note that **Brancazio & Fowler** observe that Merge, like other models of speech processing, fails to incorporate any obvious mechanism for exploiting visual information. Visual information is, of course, not *necessary* for speech perception. Indeed the McGurk effect is evidence that speech perception can be adversely affected by visual information – it is only when looking at the face producing /ga/ that we decide we are hearing /da/; close the eyes and the speaker's production of /ba/ is veridically available to the listener's consciousness. Although it is tempting to relegate this effect to domains external to the speech perception model, the phenomenon is nonetheless robust and poses an intriguing set of questions (which, it should be remarked, **Massaro** and his colleagues have not shied from addressing in FLMP simulations). Moreover, as Brancazio & Fowler point out, the range of data currently available suggest a prelexical locus for the McGurk effect, which could make it a useful experimental tool. We are therefore very interested to hear of Brancazio & Fowler's planned exploitation of audio-visual effects to test predictions from autonomous models such as Merge

https://doi.org/10.1017/S0140525X00303247 Published online by Cambridge University Press

versus feedback models, and we look forward to the results of their study. (Although space constraints prevented them from describing their planned materials in detail, we hope that as well as controlling transition probability of the consonant-consonant sequences they also, for the reasons discussed above in sect. R4.3, will be able to control the probability of the vowel-to-consonant transitions.)

R5.4. The development of phonemic processing. As **Jusczyk & Johnson** point out, any speech recognition system in place in an adult listener has its beginnings in a system developed by an infant. And an infant begins by knowing no words, so the system must be capable of developing without the use of information flowing from word representations to prelexical representations. This is of course not in itself an argument that the adult system must also make no use of top-down information flow. As Jusczyk & Johnson observe, a reorganisation of the system to allow feedback in the stable state is conceivable. They also observe that the decision nodes of Merge may imply reorganisation or elaboration of the system beyond what is available in the initial state, for phonemic decision is not, as we argued in section 7, a necessary operation in infant development. Neuro-imaging evidence certainly exists, which suggests that such a reorganisation distinguishes phonological processing by literate versus illiterate language users (Castro-Caldas et al. 1998), and evidence from aphasic listeners also suggests a dissociation of phonemic decision-making and speech comprehension (Basso et al. 1977; Riedel & Studdert-Kennedy 1985).

Note that **Jusczyk & Johnson**'s assumption that phonemic decision plays no role in language development stands in marked contrast to **Doeleman et al.**'s claim that phonemic decision-making *is* part of infant perception. Here Doeleman et al. confuse infants' ability to discriminate with adults' ability to identify. Years of speech perception research have been based on the difference between identification tasks and discrimination tasks; discriminating a difference between two inputs is not at all the same thing as identifying the nature of the difference. In fact the studies with very young infants to which Doeleman et al. refer have inter alia shown infants to be capable of discriminations that adults cannot make; thus infants in an English-language environment discriminate a change from dental to retroflex stops, both of which English-speaking adults unhesitatingly categorise as a /t/ (neither dental nor retroflex, but alveolar place of articulation in their dialect; Werker & Tees 1984). That the discrimination performance is actually not based on phonemic identification was shown by Moon et al. (1992): in their study, infants could tell the difference between *pat* and *tap* but not between *pst* and *tsp*. The phonemic changes in Moon et al.'s two pairs were identical; in the first pair, however, the medial vowel resulted in a possible syllable, while in the second pair the medial fricative resulted in non-syllabic input which the infants clearly could not decompose as adult listeners would have done.

Not all infant perception research involves simple discrimination; researchers can now also establish whether infants prefer one of two types of input which they can discriminate. **Jusczyk & Johnson** list an impressive array of evidence gleaned from such preference tasks concerning the speech perception capacities of very young infants, and the list could be much longer. But Jusczyk & Johnson hold that these discrimination and preference capacities do not constitute phonemic decision, and we agree. Phonemic decision is knowing, for instance, that *cup* and *cat* begin in the same way, and it is not observed, even in societies which encourage such awareness, till age three or four (Bradley & Bryant 1983; Liberman 1973). Phonemic decision-making is, as we argue in section 7 of the target article, separate from the normal route from input to lexicon, which by that age is fully in place.

R5.5. Neural implementation. A number of commentators (**Doeleman et al., Grossberg, Luce et al., Montant, Tanenhaus et al.**) raise the question of whether the existence of widespread neural backprojections in the brain might undermine our case against feedback. The answer here is that it depends on what those backprojections actually do. For example, backprojections might be involved in non-specific attentional control over the entire prelexical system. The presence of such backprojections would be entirely consistent with our case against feedback (see definitions). More generally, we have very little understanding of how information processing algorithms are implemented in neural hardware. Backprojections might well be part of the neural mechanisms, such as gain control, required to implement an informationally feedforward system with neural hardware. That is, the existence of backprojections may not be manifest at all at the psychological or information processing level. Alternatively, backprojections might be involved in learning but play no role in processing learned material (see Norris 1993).

The relation between processing models and their neural implementation is surely one of the most exciting areas for future research. But we should remember that the gulf between psychological models and their possible neural implementation is currently enormous.

R6. Conclusion

The feedback from the commentaries leaves us convinced that feedback in spoken word recognition is never necessary. There is still no good theoretical case for assuming that there should be feedback from lexical to prelexical processing in speech recognition. The data are consistent with a feedforward model like Merge, but inconsistent with a feedback model like TRACE. A model based on ART might possibly be able to explain some of the data, but it is far from clear that the feedback in ART is necessary. Advances in neurobiology might well illuminate this debate but, as we have cautioned, the mapping between neurobiological data and psychological theory is not straightforward. In the meantime progress will come from the development of Merge and other models to give a better computational account of human speech recognition, one that can be subjected to rigorous empirical test.

References

Letters "a" and "r" appearing before authors' initials refer to target article and response, respectively

Aderman, D. & Smith, E. E. (1971) Expectancy as a determinant of functional units in perceptual recognition. *Cognitive Psychology* 2:117–29. [aDN]

Allen, J. (1994) How do humans process and recognize speech? *IEEE Transactions on Speech and Audio Processing* 2(4):567–77. [TMN]

https://doi.org/10.1017/S0140525X00303247 Published online by Cambridge University Press

Alonso, J. M., Cudeiro, J., Perez, R., Gonzalez, F. & Acuna, C. (1993) Influence of layer 5 of area 18 of the cat visual cortex on responses of cells in layer 5 of area 17 to stimuli of high velocity. *Experimental Brain Research* 93:363–66. [MM]

Altmann, G. T. M. & Steedman, M. (1988) Interaction with context in human sentence processing. *Cognition* 30:191–238. [aDN]

Appelbaum, I. (1998) Fodor, modularity and speech perception. *Philosophical Psychology* 11:317–30. [rDN]

Azuma, T. & van Orden, G. C. (1997) Why safe is better than fast: The relatedness of a word's meaning affects lexical decision times. *Journal of Memory and Language* 36:484–504. [MGG]

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

Bard, E. & Shillcock, R. (1993) Competitor effects during lexical access: Chasing Zipf's tail. In: *Cognitive models of speech processing: The second Sperlonga meeting,* ed. G. T. M. Altmann & R. Shillcock. Erlbaum. [RSN]

Basso, A., Casati, G. & Vignolo, L. A. (1977) Phonemic identification defect in aphasia. *Cortex* 13:85–95. [rDN]

Becker, C. A. (1980) Semantic context effects in visual word recognition: An analysis of semantic strategies. *Memory and Cognition* 8:493–512. [aDN]

Bertelson, P. (1986) The onset of literacy: Liminal remarks. *Cognition* 24:1–30. [JV]

Boland, J. E. (1997) The relationship between syntactic and semantic processes in sentence comprehension. *Language and Cognitive Processes* 12:423–84. [aDN]

Boland, J. E. & Cutler, A. (1996) Interaction with autonomy: Multiple output models and the inadequacy of the Great Divide. *Cognition* 58:309–20. [aDN]

Boothroyd, A. & Nittrouer, S. (1988) Mathematical treatment of context effects in phoneme and word recognition. *Journal of the Acoustical Society of America* 84(1):101–14. [JRB, TMN]

Bourlard, H. & Morgan, N. (1994) *Connectionist speech recognition: A hybrid approach.* Kluwer Academic. [TMN]

Bradley, D. C. & Forster, K. I. (1987) A reader's view of listening. *Cognition* 25:103–34. [JCZ]

Bradley, L. & Bryant, P. E. (1983) Categorising sounds and learning to read – a causal connection. *Nature* 301:419–21. [arDN]

Bradlow, A. R. & Kingston, J. (1990) Cognitive processing in the perception of speech. *Journal of the Acoustical Society of America* 88:S56. (Abstract). [JK]

Brancazio, L. (1998) Contributions of the lexicon to audiovisual speech perception. Unpublished doctoral dissertation, University of Connecticut. [LB]

(1999) Lexical influences on the McGurk effect. *Proceedings of the International Conference on Auditory-Visual Speech Processing '99,* 67–73. [LB]

Brookshire, R. H. (1972) Visual and auditory sequencing by aphasic subjects. *Journal of Communication Disorders* 5:259–69. [RMW]

Brown, C. M., van Berkum, J. J. A. & Hagoort, P. (in press) Discourse before gender: An event-related potential study on the interplay of semantic and syntactic information during spoken language understanding. *Journal of Psycholinguistic Research.* [rDN]

Brown, C. M., Hagoort, P. & Ter Keurs, M. (1999) Electrophysiological signatures of visual lexical processing: Open- and closed-class words. *Journal of Cognitive Neuroscience* 11:261–81. [rDN]

Bullier, J., Munk, M. H. J. & Nowak, L. G. (1993) Corticocortical connections sustain interarea synchronization. *Concepts in Neuroscience* 4:101–17. [MM]

Burton, M. W. (1992) Syllable priming in auditory word recognition. Paper presented at the 33rd annual meeting of the Psychonomic Society, St. Louis, MO. [LMS]

Burton, M. W., Baum, S. R. & Blumstein, S. E. (1989) Lexical effects on the phonetic categorization of speech: The role of acoustic structure. *Journal of Experimental Psychology: Human Perception and Performance* 15:567–75. [aDN]

Cairns, P., Shillock, R., Chater, N. & Levy, J. (1995) Bottom-up connectionist modelling of speech. In: *Connectionist models of memory and language,* ed. J. P. Levy, D. Bairaktaris, J. A. Bullinaria & P. Cairns. University College London Press. [JRB, aDN]

(1997) Bootstrapping word boundaries: A bottom-up corpus based approach to speech segmentation. *Cognitive Psychology* 33:111–53. [MGG]

Carello, C., Turvey, M. T. & Lukatela, G. (1992) Can theories of word recognition remain stubbornly nonphonological? In: *Orthography, phonology, morphology, and meaning,* ed. R. Frost & L. Katz. North-Holland. [JCZ]

Carmon, A. & Nachshon, I. (1971) Effect of unilateral brain damage on perception of temporal order. *Cortex* 7:410–18. [RMW]

Carpenter, G. A. & Grossberg, S. (1987) A massively parallel architecture for a self-organizing neural pattern recognition machine. *Computer Vision, Graphics and Image Processing* 7:54–115. [aDN]

Carpenter, G. A., Grossberg, S. & Rosen, D. (1991) ART2-A: An adaptive resonance algorithm for rapid category learning and recognition. *Neural Networks* 4:493–504. [rDN]

Carpenter, R. H. S. (1999) Visual selection: Neurons that make up their minds. *Current Biology* 9:R595–98. [rDN]

Case, P., Tuller, B., Ding, M. Z. & Kelso, J. A. S. (1995) Evaluation of a dynamical model of speech perception. *Perception and Psychophysics* 57:977–88. [JCZ]

Castro-Caldas, A., Petersson, K. M., Reis, A., Stone-Elander, S. & Ingvar, M. (1998) The illiterate brain: Learning to read and write during childhood influences the functional organization of the brain. *Brain* 121:1053–63. [rDN]

Cheesman, M. F. & Greenwood, K. G. (1995) Selective adaptation by context conditioned fricatives. *Journal of the Acoustical Society of America* 97:531–38. [aDN]

Chomsky, N. & Halle, M. (1968) *The sound pattern of English.* Harper and Row. [KNS]

Churchland, P. S., Ramachandran, V. S. & Sejnowski, T. J. (1994) A critique of pure vision. In: *Large-scale neuronal theories of the brain,* ed. C. Koch & J. L. Davis. MIT Press. [MKT]

Cluff, M. S. & Luce, P. A. (1990) Similarity neighborhoods of spoken two-syllable words: Retroactive effects on multiple activation. *Journal of Experimental Psychology: Human Perception and Performance* 16:551–63. [aDN]

Cohen, M. & Grossberg, S. (1986) Neural dynamics of speech and language coding: Developmental programs, perceptual grouping, and competition for short-term memory. *Human Neurobiology* 5:1–22. [SG, rDN]

Colombo, L. (1986) Activation and inhibition with orthographically similar words. *Journal of Experimental Psychology: Human Perception and Performance* 12:226–34. [LMS]

Connine, C. M. (1994) Vertical and horizontal similarity in spoken word recognition. In: *Perspectives on sentence processing,* ed. C. Clifton, Jr., L. Frazier & K. Rayner. Erlbaum. [CMC]

Connine, C. M. & Clifton, C. (1987) Interactive use of lexical information in speech perception. *Journal of Experimental Psychology: Human Perception and Performance* 13:291–99. [CMC, aDN]

Connine, C. M. & Titone, D. (1996) Phoneme monitoring. *Language and Cognitive Processes* 11:635–64. [aDN]

Connine, C. M., Titone, D., Deelman, T. & Blasko, D. (1997) Similarity mapping in spoken word recognition. *Journal of Memory and Language* 37:463–80. [PAL, RSN, AGS]

Coulson, S., King, J. & Kutas, M. (1998) Expect the unexpected: Event-related brain potentials to morphosyntactic violations. *Language and Cognitive Processes* 13:21–58. [FI]

Crain, S. & Steedman, M. (1985) On not being led up the garden path: The use of context by the psychological syntax processor. In: *Natural language processing,* ed. D. R. Dowty, L. Kartunen & A. M. Zwicky. Cambridge University Press. [aDN]

Cutler, A. & Chen, H.-C. (1995) Phonological similarity effects in Cantonese word recognition. *Proceedings of the International Congress on Phonetic Sciences, Stockholm, Sweden, vol. 1,* 106–109. [LMS]

Cutler, A., Mehler, J., Norris, D. & Seguí, J. (1987) Phoneme identification and the lexicon. *Cognitive Psychology* 19:141–77. [JRB, arDN]

Cutler, A. & Norris, D. (1979) Monitoring sentence comprehension. In: *Sentence processing: Psycholinguistic studies presented to Merrill Garrett,* ed. W. E. Cooper & E. C. T. Walker. Erlbaum. [arDN]

(1999) Sharpening Ockham's razor. *Behavioral and Brain Sciences* 22:40–41. [ASM]

Cutler, A. & Otake, T. (1994) Mora or phoneme? Further evidence for language-specific listening. *Journal of Memory and Language* 33:824–44. [aDN]

David, A. S., Woodruff, P. W. R., Howard, R., Mellers, J. D. C., Brammer, M., Bullmore, E., Wright, I., Andrew, C. & Williams, S. C. R. (1996) Auditory hallucinations inhibit exogenous activation of auditory association cortex. *NeuroReport* 7:932–36. [RS]

Davis, M., Marslen-Wilson, W. D. & Gaskell, M. G. (1997a) Ambiguity and competition in lexical segmentation. In: *Proceedings of the Nineteenth Annual Conference of the Cognitive Science Society, Stanford, CA.* [WDM-W]

Davis, M., Gaskell, M. G. & Marslen-Wilson, W. D. (1997b) Recognising words in connected speech: Context and competition. In: *Proceedings of the Fourth Neural Computation and Psychology Workshop: Connectionist representations,* ed. J. A. Bullinaria. Springer-Verlag. [WDM-W]

Demonet, J. F., Price, C., Wise, R. & Frackowiak, R. (1994) A PET study of cognitive strategies in normal subjects during language tasks. Influence of phonetic ambiguity and sequence processing on phoneme monitoring. *Brain* 117:671–82. [rDN]

Deschenes, M. & Hu, B. (1990) Electrophysiology and pharmacology of the corticothalamic input to lateral thalamic nuclei: An intracellular study in the cat. *European Journal of Neuroscience* 2:140–52. [MM]

Donchin, E. & Coles, M.G.H. (1988) Is the P300 component a manifestation of context updating? *Behavioral and Brain Sciences* 11:357–74. [rDN]

Dumay, N. & Radeau, M. (1997) Rime and syllabic effects in phonological priming between French spoken words. *Proceedings of Eurospeech '97, Rhodes, Greece, vol. 4,* 2191–94. [LMS]

Eimas, P. D., Marcovitz Hornstein, S. B. & Payton, P. (1990) Attention and the role

https://doi.org/10.1017/S0140525X00303247 Published online by Cambridge University Press

of dual codes in phoneme monitoring. *Journal of Memory and Language* 29:160–80. [aDN]
Eimas, P. D. & Nygaard, L. C. (1992) Contextual coherence and attention in phoneme monitoring. *Journal of Memory and Language* 31:375–95. [aDN]
Einstein, A. & Infeld, L. (1966) *The evolution of physics.* Simon and Schuster. (Originally published in 1938). [JCZ]
Elman, J. L. & McClelland, J. L. (1988) Cognitive penetration of the mechanisms of perception: Compensation for coarticulation of lexically restored phonemes. *Journal of Memory and Language* 27:143–65. [DWM, arDN, AGS]
Emmorey, K. D. (1989) Auditory morphological priming in the lexicon. *Language and Cognitive Processes* 4:73–92. [LMS]
Farah, M. (1989) Semantic and perceptual priming: How similar are the underlying mechanisms? *Journal of Experimental Psychology: Human Perception and Performance* 15:188–94. [aDN]
Farmer, J. D. (1990) A Rosetta Stone for connectionism. *Physica D* 42:153–87. [JCZ]
Felleman, D. J. & van Essen, C. D. (1991) Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex* 1:1–47. [SG]
Finkel, L. H. & Edelman, G. M. (1989) The integration of distributed cortical systems by reentry: A computer simulation of interactive functionally segregated visual areas. *The Journal of Neuroscience* 9:3188–208. [MM]
Fodor, J. A. (1983) *The modularity of mind.* MIT Press. [WSM, arDN]
(1985) Précis of *The modularity of mind. Behavioral and Brain Sciences* 8:1–42. [aDN]
Forster, K. I. (1979) Levels of processing and the structure of the language processor. In: *Sentence processing: Psycholinguistic studies presented to Merrill Garrett,* ed. W. E. Cooper & E. C. T. Walker. Erlbaum. [WSM, arDN]
Foss, D. J. & Blank, M. A. (1980) Identifying the speech codes. *Cognitive Psychology* 12:1–31. [aDN]
Foss, D. J. (1969) Decision processes during sentence comprehension, effects of lexical item difficulty and position upon decision times. *Journal of Verbal Learning and Verbal Behavior* 8:457–62. [rDN]
Foss, D. J. & Gernsbacher, M. A. (1983) Cracking the dual code: Toward a unitary model of phoneme identification. *Journal of Verbal Learning and Verbal Behavior* 22:609–32. [aDN]
Fowler, C. A., Brown, J. M. & Mann, V. A. (1999) Compensation for coarticulation in audiovisual speech perception. *Proceedings of the XIVth International Congress of Phonetic Sciences, San Francisco,* 639–42. [JK]
(in press) Contrast effects do not underlie effects of preceding liquids on stop consonant identification by humans. *Journal of Experimental Psychology: Human Perception and Performance.* [LB]
Fox, R. A. (1984) Effect of lexical status on phonetic categorization. *Journal of Experimental Psychology: Human Perception and Performance* 10:526–40. [rDN]
Frauenfelder, U. H. & Peeters, G. (1998) Simulating the time-course of spoken word recognition: An analysis of lexical competition in TRACE. In: *Localist connectionist approaches to human cognition,* ed. J. Grainger & A. M. Jacobs. Erlbaum. [aDN]
Frauenfelder, U. H. & Seguí, J. (1989) Phoneme monitoring and lexical processing: Evidence for associative context effects. *Memory and Cognition* 17:134–40. [aDN]
Frauenfelder, U. H., Seguí, J.& Dijkstra, T. (1990) Lexical effects in phonemic processing: Facilitatory or inhibitory? *Journal of Experimental Psychology: Human Perception and Performance* 16:77–91. [CMC, PAL, arDN, AGS]
Frazier, L. (1979) *On comprehending sentences: Syntactic parsing strategies.* Indiana Linguistics Club. [aDN]
(1987) Sentence processing: A tutorial review. In: *Attention and performance XII: The psychology of reading,* ed. M. Coltheart. Erlbaum. [aDN]
Freeman, W. (1991) The physiology of perception. *Scientific American* 264:78–85. [PAL, MM]
Friederici, A. D. (1995) The time course of syntactic activation during language processing: A model based on neuropsychological and neurophysiological data. *Brain and Language* 50:259–81. [FI]
Friederici, A. D., Pfeifer, E. & Hahne, A. (1993) Event-related brain potentials during natural speech processing: Effects in semantic, morphological, and syntactic violations. *Cognitive Brain Research* 1:183–92. [FI]
Frith, C. & Dolan, R. J. (1997) Brain mechanism associated with top-down processes in perception. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences* 352:1221–30. [MM]
Frost, S. J., Fowler, C. A. & Rueckl, J. G. (1998) Bidirectional consistency: Effects of a phonology common to speech and reading. (submitted). [JCZ]
Ganong, W. F. (1980) Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance* 6:110–25. [LB, arDN, MP]
Gaskell, M. G. & Marslen-Wilson, W. D. (1995) Modeling the perception of spoken words. In: *Proceedings of the 17th Annual Conference of the Cognitive Science Society,* ed. J. D. Moore & J. F. Lehman. Erlbaum. [WDM-W, aDN]
(1997) Integrating form and meaning: A distributed model of speech perception. *Language and Cognitive Processes* 12:613–56. [MGG, DWG, JG, WDM-W, aDN]
(1998) Mechanisms of phonological inference in speech perception. *Journal of Experimental Psychology: Human Perception and Performance* 24:380–96. [aDN]
(1999) Ambiguity, competition and blending in speech perception. *Cognitive Science.* (in press). [WDM-W]
Gibbs, P. & Van Orden, G. C. (1998) Pathway selection's utility for control of word recognition. *Journal of Experimental Psychology: Human Perception and Performance* 24:1162–87. [JCZ]
Goldinger, S. D. (1998) Echoes of echoes? An episodic theory of lexical access. *Psychological Review* 105:251–79. [PAL]
(1999) Only the shadower knows: Comment on Hamburger and Slowiaczek (1996). *Psychonomic Bulletin and Review* 6:347–51. [LMS]
Goldinger, S. D., Luce, P. A., Pisoni, D. B. & Marcario, J. K. (1992) Form-based priming in spoken word recognition: The roles of competition and bias. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 18:1211–38. [aDN]
Gottlob, L. R., Goldinger, S. D., Stone, G. O. & Van Orden, G. C. (1999) Reading homographs: Orthographic, phonologic, and semantic dynamics. *Journal of Experimental Psychology: Human Perception and Performance* 25:561–74. [JCZ]
Gow, D. W. (submitted a) Assimilation and anticipation: Phonological modification and word recognition in connected speech. [DWG]
(submitted b) Does assimilation produce lexical ambiguity? [DWG]
Grainger, J. & Jacobs, A. M. (1994) A dual read-out model of word context effects in letter perception: Further investigations of the word-superiority effect. *Journal of Experimental Psychology: Human Perception and Performance* 20:1158–76. [JG, arDN, JCZ]
(1996) Orthographic processing in visual word recognition: A multiple read-out model. *Psychological Review* 103:674–91. [JG, aDN]
(1998) On localist connectionism and psychological science. In: *Localist connectionist approaches to human cognition,* ed. J. Grainger & A. M. Jacobs. Erlbaum. [JG]
Green, K. P. & Miller, J. L. (1985) On the role of visual rate information in phonetic perception. *Perception and Psychophysics* 38:269–76. [LB]
Grossberg, S. (1978) A theory of human memory: Self-organization and performance of sensory-motor codes, maps, and plans. In: *Progress in theoretical biology,* ed. R. Rosen & F. Snell. Academic Press. [SG]
(1980) How does a brain build a cognitive code? *Psychological Review* 87:1–51. [SG, MM]
(1986) The adaptive self-organization of serial order in behavior: Speech, language, and motor control. In: *Pattern recognition by humans and machines, vol 1: Speech perception,* ed. E. C. Schwab & H. C. Nusbaum. Academic Press. [SG, PAL]
(1995) The attentive brain. *American Scientist* 83:438–49. [SG]
(1999a) How does the cerebral cortex work? Learning, attention, and grouping by the laminar circuits of visual cortex. *Spatial Vision* 12:163–85. [SG]
(1999b) The link between brain learning, attention, and consciousness. *Consciousness and Cognition* 8:1–44. [SG, MM]
Grossberg, S., Boardman, I. & Cohen, M. (1997a) Neural dynamics of variable-rate speech categorization. *Journal of Experimental Psychology: Human Perception and Performance* 23:481–503. [SG, PAL, MM, rDN]
Grossberg, S., Mingolla, E. & Ross, W. (1997b) Visual brain and visual perception: How does the cortex do perceptual grouping? *Trends in Neurosciences* 20:106–11. [SG]
Grossberg, S. & Myers, C. (1999) The resonant dynamics of conscious speech: Interword integration and duration-dependent backward effects. *Psychological Review.* (in press). [SG]
Grossberg, S. & Stone, G. O. (1986) Neural dynamics of word recognition and recall: Attentional priming, learning, and resonance. *Psychological Review* 93:46–74. [SG, PAL]
Gunter, T. C., Friederici, A. D. & Schriefers, H. (1998) Gender violations, semantic expectancy and ERPs. Poster presented at the XII International Conference on Event-Related Potentials, Cambridge, USA. [FI]
Gunter, T. C., Stowe, L. A. & Mulder, G. (1997) When syntax meets semantics. *Psychophysiology* 36:126–37. [FI]
Hagoort, P. & Brown, C. M. (in press) Semantic and syntactic effects of listening to speech compared to reading. *Neuropsychologia.* [rDN]
Hahne, A. & Friederici, A. D. (1999) Electrophysiological evidence for two steps in syntactic analysis: Early automatic and late controlled processes. *Journal of Cognitive Neuroscience* 11(2):194–205. [FI]
Hamburger, M. B. & Slowiaczek, L. M. (1996) Phonological priming reflects lexical competition in auditory word recognition. *Psychonomic Bulletin and Review* 3:520–25. [LMS]

https://doi.org/10.1017/S0140525X00303247 Published online by Cambridge University Press

(1999) On the role of bias in dissociated phonological priming effects: A reply to Goldinger (1999). *Psychonomic Bulletin and Review* 6:352–55. [LMS]
Hanes, D. P. & Schall, J. D. (1996) Neural control of voluntary movement initiation. *Science* 274:427–30. [rDN]
Haveman, A. P. (1997) *The open-/closed-class distinction in spoken-word recognition.* Ph. D. thesis, University of Nijmegen. [rDN]
Hines, T. (1999) A demonstration of auditory top-down processing. *Behavior, Research, Methods, Instruments and Computers* 31:55–56. [MM]
Hintzman, D. (1986) "Schema abstraction" in a multiple-trace memory model. *Psychological Review* 93:411–28. [PAL]
Hornik, K., Stinchcombe, M. & White, H. (1989) Multilayer feedforward networks are universal approximators. *Neural Networks* 2:359–66. [DWM, rDN]
Hupé, J. M., James, A. C., Paynes, B. R., Lomber, S. G., Girad, P. & Bullier, J. (1998) Cortical feedback improves discrimination between figure and background by V1, V2 and V3 neurons. *Nature* 394:784–87. [MM]
Iverson, P., Bernstein, L. E. & Auer, E. T. (1998) Modeling the interaction of phonemic intelligibility and lexical structure in audiovisual word recognition. *Speech Communication* 26:45–63. [LB]
Jacobs, A. M., Rey, A., Ziegler, J. C. & Grainger, J. (1998) MROM-P: An interactive activation, multiple read-out model of orthographic and phonological processes in visual word recognition. In: *Localist connectionist approaches to human cognition,* ed. J. Grainger & A. M. Jacobs. Erlbaum. [JCZ]
Jusczyk, P. W. (1997) *The discovery of spoken language.* MIT Press. [aDN]
Jusczyk, P. W. & Aslin, R. N. (1995) Infants' detection of sound patterns of words in fluent speech. *Cognitive Psychology* 29:1–23. [PWJ]
Jusczyk, P. W., Hohne, E. A. & Bauman, A. (1999) Infants' sensitivity to allophonic cues for word segmentation. *Perception and Psychophysics* 61:1465–76. [PWJ]
Jusczyk, P. W., Houston, D. & Newsome, M. (1999) The beginnings of word segmentation in English-learning infants. *Cognitive Psychology* 39: [PWJ]
Kawamoto, A. H. & Zemblidge, J. (1992) Pronunciation of homographs. *Journal of Memory and Language* 31:349–74. [JCZ]
Kelso, J. A. S. (1995) *Dynamic patterns: The self-organization of brain and behavior.* MIT Press. [JCZ]
Kingston, J. (1991) Integrating articulations in the perception of vowel height. *Phonetica* 48:149–79. [JK]
Kingston, J. & Diehl, R. L. (1994) Phonetic knowledge. *Language* 70:419–54. [JK]
(1995) Intermediate properties in the perception of distinctive feature values. In: *Papers in laboratory phonology IV,* ed. B. Connell & A. Arvaniti. Cambridge University Press. [JK]
Kingston, J. & Macmillan, N. A. (1995) Integrality of nasalization and F_1 in vowels in isolation and before oral and nasal consonants: A detection-theoretic application of the Garner paradigm. *Journal of the Acoustical Society of America* 97:1261–85. [JK]
Kingston, J., Macmillan, N. A., Walsh Dickey, L., Thorburn, R. & Bartels, C. (1997) Integrality in the perception of tongue root position and voice quality in vowels. *Journal of the Acoustical Society of America* 101:1696–709. [JK]
Klatt, D. H. (1980) Speech perception: A model of acoustic-phonetic analysis and lexical access. In: *Perception and production of fluent speech,* ed. R. A. Cole. Erlbaum. [DHW]
Koster, C. J. (1987) *Word recognition in foreign and native language.* Foris. [aDN]
Kugler, P. N. & Turvey, M. T. (1987) *Information, natural law, and the self-assembly of rhythmic movement.* Erlbaum. [JCZ]
Kutas, M. & Hillyard, S. A. (1983) Event-related brain potentials to grammatical errors and semantic anomalies. *Memory and Cognition* 11:539–50. [FI]
Kutas, M. & Van Petten, C. (1988) Event-related potential studies of language. In: *Advances in psychophysiology, vol. 3,* ed. P. K. Ackles, J. R. Jennings & M. G. H. Coles. JAI Press. [FI]
Lakatos, I. (1970) Falsification and the methodology of scientific research programmes. In: *Criticism and the growth of knowledge,* ed. I. Lakatos & A. Musgrave. Cambridge University Press. [JCZ]
Levelt, W. J. M., Roelofs, A. & Meyer, A. S. (1999) A theory of lexical access in speech production. *Behavioral and Brain Sciences* 22(1):1–75. [ASM, arDN]
Levinson, S. C. (2000) *Presumptive meanings: The theory of generalized conversational implicature.* MIT Press. [rDN]
Liberman, I. Y. (1973) Segmentation of the spoken word and reading acquisition. *Bulletin of the Orton Society* 23:65–77. [arDN]
Lotto, A. J. & Kluender, K. R. (1998) General contrast effects in speech perception: Effect of preceding liquid on stop consonant identification. *Perception and Psychophysics* 60:602–19. [JK]
Luce, P. A. & Pisoni, D. B. (1998) Recognizing spoken words: The neighborhood activation model. *Ear and Hearing* 19:1–36. [MM]
Luce, R. D. (1959) *Individual choice behavior.* Wiley. [aDN]
(1963) Detection and recognition. In: *Handbook of mathematical psychology, vol. I,* ed. R. D. Luce, R. T. Bush & E. Galanter. Wiley. [aDN]
Luck, S. J., Chelazzi, L., Hillyard, S. A. & Desimone, R. (1997) Neural mechanisms of spatial selective attention in areas V1, V2, and V4 of macaque visual cortex. *Journal of Neurophysiology* 77:24–42. [MM]
Lukatela, G. & Turvey, M. T. (1998) Reading in two alphabets. *American Psychologist* 53:1057–72. [JCZ]
Lupker, S. J. & Colombo, L. (1994) Inhibitory effects in form priming: Evaluating a phonological competition explanation. *Journal of Experimental Psychology: Human Perception and Performance* 20:437–51. [LMS]
Luria, A. (1973) *The working brain.* Penguin Books. [PAL]
MacDonald, M. C., Pearlmutter, N. J. & Seidenberg, M. S. (1994) Lexical nature of syntactic ambiguity resolution. *Psychological Review* 101:676–703. [aDN]
MacKay, W. A. (1997) Synchronized neuronal oscillations and their role in motor processes. *Trends in Cognitive Sciences* 1:176–83. [MM]
Macmillan, N. A., Kingston, J., Thorburn, R., Walsh Dickey, L. & Bartels, C. (1999) Integrality of nasalization and F_1. II. Basic sensitivity and phonetic labeling. *Journal of the Acoustical Society of America* 106:2913–32. [JK]
Mann, V. A. (1980) Influence of preceding liquid on stop-consonant perception. *Perception and Psychophysics* 28:407–12. [JK]
(1986a) Phonological awareness: The role of reading experience. *Cognition* 24:65–92. [aDN]
(1986b) Distinguishing universal and language-dependent levels of speech perception: Evidence from Japanese listeners' perception of "l" and "r". *Cognition* 24:169–96. [rDN]
Mann, V. A. & Repp, B. H. (1981) Influence of preceding fricative on stop consonant perception. *Journal of the Acoustical Society of America* 69:546–58. [JK, arDN]
Manuel, S. Y. (1992) Vocal reduction and perceptual recovery in casual speech. *Journal of the Acoustical Society of America* 91:4(2):2388. [DWG]
(1995) Speakers nasalize /ð/ after /n/, but listerners still hear /ð/. *Journal of Phonetics* 23:453–76. [KNS]
Marrocco, R. T., McClurkin, J. W. & Young, R. A. (1982) Modulation of lateral geniculate nucleus cell responsiveness by visual activation of the corticogeniculate pathway. *The Journal of Neuroscience* 2:256–63. [MM]
Marslen-Wilson, W. D. (1993) Issues of process and representation in lexical access. In: *Cognitive models of speech processing: The second Sperlonga meeting,* ed. G. T. M. Altmann & R. Shillock. Erlbaum. [aDN]
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* 22:1376–92. [aDN]
Marslen-Wilson, W. D. & Tyler, L. K. (1980) The temporal structure of spoken language understanding. *Cognition* 8:1–71. [WDM-W]
Marslen-Wilson, W. D. & Warren, P. (1994) Levels of perceptual representation and process in lexical access: Words, phonemes, and features. *Psychological Review* 101:653–75. [LB, MGG, PAL, WDM-W, arDN, AGS, LMS, MKT, DHW]
Marslen-Wilson, W. D. & Welsh, A. (1978) Processing interactions during word recognition in continuous speech. *Cognition* 10:29–63. [DWG]
Massaro, D. W. (1973) Perception of letters, words, and nonwords. *Journal of Experimental Psychology* 100:49–53. [DWM]
(1975) *Understanding language: An information-processing analysis of speech perception, reading, and psycholinguistics.* Academic Press. [DWM]
(1978) A stage model of reading and listening. *Visible Language* 12:3–26. [aDN]
(1979) Letter information and orthographic context in word perception. *Journal of Experimental Psychology: Human Perception and Performance* 5:595–609. [DWM]
(1987) *Speech perception by ear and eye: A paradigm for psychological inquiry.* Erlbaum. [LB, aDN]
(1988) Some criticisms of connectionist models of human performance. *Journal of Memory and Language* 27:213–34. [DWM]
(1989a) Testing between the TRACE model and the Fuzzy Logical Model of Speech Perception. *Cognitive Psychology* 21:398–421. [aDN]
(1989b) Multiple book review of *Speech perception by ear and eye: A paradigm for psychological inquiry. Behavioral and Brain Sciences* 12:741–94. [aDN]
(1996) Integration of multiple sources of information in language processing. In: *Attention and performance XVI: Information integration in perception and communication,* ed. T. Inui & J. L. McClelland. MIT Press. [DWM, aDN]
(1998) *Perceiving talking faces: From speech perception to a behavioral principle.* MIT Press. [DWM, aDN]
Massaro, D. W. & Cohen, M. M. (1983a) Categorical or continuous speech perception: A new test. *Speech Communication* 2:15–35. [GCO]
(1983b) Phonological context in speech perception. *Perception and Psychophysics* 34:338–48. [DWM, aDN]
(1991) Integration versus interactive activation: The joint influence of stimulus and context in perception. *Cognitive Psychology* 23:558–614. [DWM, arDN, GCO]
Massaro, D. W., Cohen, M. M., Campbell, C. S. & Rodriguez, T. (submitted) Bayesian method of model selection validates FLMP. *Psychonomic Bulletin and Review.* [DWM]

https://doi.org/10.1017/S0140525X00303247 Published online by Cambridge University Press

Massaro, D. W. & Cowan, N. (1993) Information processing models: Microscopes of the mind. *Annual Review of Psychology* 44:383–425. [aDN]

Massaro, D. W. & Oden, G. C. (1980a) Evaluation and integration of acoustic features in speech perception. *Journal of the Acoustical Society of America* 67:996–1013. [GCO]

(1980b) Speech perception: A framework for research and theory. In: *Speech and language: Advances in basic research and practice, vol. 3,* ed. N. J. Lass. Academic Press. [GCO]

(1995) Independence of lexical context and phonological information in speech perception. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 21:1053–64. [aDN, GCO]

Mattys, S. L. & Jusczyk, P. W. (submitted) Phonotactic cues for segmentation of fluent speech by infants. [PWJ]

Mattys, S. L., Jusczyk, P. W., Luce, P. A. & Morgan, J. L. (1999) Word segmentation in infants: How phonotactics and prosody combine. *Cognitive Psychology* 38:465–94. [PWJ]

McClelland, J. L. (1986) The programmable blackboard model of reading. In: *Parallel distributed processing, vol. 2: Psychological and biological models,* ed. J. L. McClelland, D. E. Rumelhart & the PDP Research Group. MIT Press. [rDN]

(1987) The case for interactionism in language processing. In: *Attention and performance XII: The psychology of reading,* ed. M. Coltheart. Erlbaum. [rDN]

(1991) Stochastic interactive processes and the effect of context on perception. *Cognitive Psychology* 23:1–44. [arDN]

McClelland, J. L. & Elman, J. L. (1986) The TRACE model of speech perception. *Cognitive Psychology* 18:1–86. [TMN, aDN, JV]

McClelland, J. L. & Johnston, J. (1977) The role of familiar units in perception of words and nonwords. *Perception and Psychophysics* 22:249–61. [aDN]

McClelland, J. L. & Rumelhart, D. E. (1981) An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychological Review* 88:375–407. [aDN]

McClelland, J. L., St. John, M. & Taraban, R. (1989) Sentence comprehension: A parallel distributed processing approach. *Language and Cognitive Processes* 4:287–336. [aDN]

McGurk, H. & MacDonald, J. (1976) Hearing lips and seeing voices. *Nature* 264:746–48. [LB, aDN]

McNeill, D. & Lindig, K. (1973) The perceptual reality of phonemes, syllables, words, and sentences. *Journal of Verbal Learning and Verbal Behavior* 12:419–30. [PAL]

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* 17:433–43. [arDN, MP]

(1996) Phonetic categorisation. *Language and Cognitive Processes* 11:655–64. [aDN]

McQueen, J. M., Cutler, A., Briscoe, T. & Norris, D. (1995) Models of continuous speech recognition and the contents of the vocabulary. *Language and Cognitive Processes* 10:309–31. [aDN]

McQueen, J. M., Norris, D. & Cutler, A. (1994) Competition in spoken word recognition: Spotting words in other words. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 20:621–38. [aDN]

(1999a) Lexical influence in phonetic decision-making: Evidence from subcategorical mismatches. *Journal of Experimental Psychology: Human Perception and Performance* 25:1363–89. [MGG, PAL, WDM-W, arDN, AGS, DHW]

(1999b) The time course of lexical involvement in phonetic categorisation. Presented at the 137th meeting of the Acoustical Society of America, Berlin, March 1999. (Abstract in *Journal of the Acoustical Society of America* 105:1398). [rDN]

(1999c) Lexical activation produces impotent phonemic percepts. Presented at the 138th meeting of the Acoustical Society of America, Columbus, OH, November 1999. (Abstract in *Journal of the Acoustical Society of America* 106:2296). [rDN]

Mecklinger, A., Schriefers, H., Steinhauer, K. & Friederici, A. D. (1995) Processing relative clauses varying on syntactic and semantic dimensions: An analysis with event-related potentials. *Memory and Cognition* 23:477–94. [FI]

Mehler, J. & Segui, J. (1987) English and French speech processing. In: *The psychophysics of speech perception,* ed. M. E. H. Schouten. Martinus Nijhoff. [aDN]

Mergner, T., Huber, W. & Becker, W. (1997) Vestibular-neck interaction and transformation of sensory coordinates. *Journal of Vestibular Research* 7:347–67. [MM]

Miller, G. A., Heise, G. A. & Lichten, W. (1951) The intelligibility of speech as a function of the context of the test materials. *Journal of Experimental Psychology* 41:329–35. [JRB]

Miller, J. (1994) On the internal structure of phonetic categories: A progress report. *Cognition* 50:271–85. [DWG]

Miller, J. L. & Liberman, A. M. (1979) Some effects of later-occurring information on the perception of stop consonant and semivowel. *Perception and Psychophysics* 25:457–65. [DHW]

Monsell, S. & Hirsh, K. W. (1998) Competitor priming in spoken word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 24:1495–520. [LMS]

Moon, C., Bever, T. G. & Fifer, W. P. (1992) Canonical and non-canonical syllable discrimination by two-day-old infants. *Journal of Child Language* 19:1–17. [rDN]

Morais, J. (1985) Literacy and awareness of the units of speech: Implications for research on the units of perception. *Linguistics* 23:707–21. [aDN]

Morais, J., Bartelson, P., Cary, L. & Alegria, J. (1986) Literacy training and speech segmentation. *Cognition* 24:45–64. [aDN]

Morais, J., Cary, L., Alegria, J. & Bartelson, P. (1979) Does awareness of speech as a sequence of phones arise spontaneously? *Cognition* 7:323–31. [aDN]

Morton, J. (1969) Interaction of information in word recognition. *Psychological Review* 76:165–78. [aDN]

Morton, J. & Long, J. (1976) Effect of word transitional probability on phoneme identification. *Journal of Verbal Learning and Verbal Behavior* 15:43–51. [aDN]

Münte, T. F., Heinze, H. & Mangun, G. R. (1993) Dissociation of brain activity related to syntactic and semantic aspects of language. *Journal of Cognitive Neuroscience* 5:335–44. [FI]

Murray, W. S. & Rowan, M. (1998) Early, mandatory, pragmatic processing. *Journal of Psycholinguistic Research (CUNY Special Issue)* 27:1–22. [WSM]

Myung, I. J. (in press) The importance of complexity in model selection. *Journal of Mathematical Psychology.* [MP]

Myung, I. J & Pitt, M. A. (1997) Applying Occam's razor in modeling cognition: A Bayesian approach. *Psychonomic Bulletin and Review* 4:79–95. [DWM]

(1998) Issues in selecting mathematical models of cognition. In: *Localist connectionist approaches to human cognition,* ed. J. Grainger & A. Jacobs. Erlbaum. [MP]

Nearey, T. (1997) Speech perception as pattern recognition. *Journal of the Acoustical Society of America* 101(6):3241–54. [TMN]

(forthcoming) The factorability of phonological units in speech perception: Simulating results on speech reception in noise. In: *Festschrift for Bruce L. Derwing,* ed. R. Smyth. [JRB, TMN]

Neville, H. J., Nicol, J., Barss, A., Forster, K. & Garrett, M. (1991) Syntactically based sentence processing classes: Evidence from event-related brain potentials. *Journal of Cognitive Neuroscience* 3:155–70. [FI]

Newell, A. (1973) You can't play 20 questions with nature and win: Projective comments on the papers of this symposium. In: *Visual information processing,* ed. W. G. Chase. Academic Press. [aDN]

Newman, R. S., Sawusch, J. R. & Luce, P. A. (1997) Lexical neighborhood effects in phonetic processing. *Journal of Experimental Psychology: Human Perception and Performance* 23(3):873–89. [LB, RSN, arDN]

Nobre, A. C., Allison, T. & McCarthy, G. (1998) Modulation of human extrastriate visual processing by selective attention to colours and words. *Brain* 121:1357–68. [MM]

Norris, D. G. (1980) *Serial and parallel models of comprehension.* Ph. D. thesis, University of Sussex. [rDN]

(1986) Word recognition: Context effects without priming. *Cognition* 22:93–136. [JRB, aDN]

(1987) Syntax, semantics and garden paths. In: *Progress in the psychology of language, vol. 3,* ed. A. W. Ellis. Erlbaum. [aDN]

(1992) Connectionism: A new breed of bottom-up model? In: *Connectionist approaches to language processing,* ed. N. Sharkey & R. Reiley. Erlbaum. [rDN]

(1993) Bottom-up connectionist models of "interaction." In: *Cognitive models of speech processing: The second Sperlonga meeting,* ed. G. T. M. Altmann & R. Shillcock. Erlbaum. [JRB, WDM-W, aDN]

(1994a) A quantitative multiple-levels model of reading aloud. *Journal of Experimental Psychology: Human Perception and Performance* 20:1212–32. [arDN]

(1994b) Shortlist: A connectionist model of continuous speech recognition. *Cognition* 52:189–234. [MGG, JG, TMN, arDN, JV]

(1995) Signal detection theory and modularity: On being sensitive to the power of bias models of semantic priming. *Journal of Experimental Psychology: Human Perception and Performance* 21:935–39. [aDN]

(submitted) Feedback consistency in visual word recognition: Don't try going the wrong way up a one-way street. [rDN]

Norris, D. G., McQueen, J. M. & Cutler, A. (1995) Competition and segmentation in spoken word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 21:1209–28. [arDN]

Norris, D. G., McQueen, J. M., Cutler, A. & Butterfield, S. (1997) The possible-word constraint in the segmentation of continuous speech. *Cognitive Psychology* 34:193–243. [aDN]

Norris, D. G. & Wise, R. (1999) The study of prelexical and lexical processes in

https://doi.org/10.1017/S0140525X00303247 Published online by Cambridge University Press

comprehension: Psycholinguistics and functional neuroimaging. In: *The cognitive neurosciences,* ed. M. Gazzaniga. MIT Press. [rDN]
Oden, G. C. (1978) Semantic constraints and judged preference for interpretations of ambiguous sentences. *Memory and Cognition* 6:26–37. [GCO]
(1983) On the use of semantic constraints in guiding syntactic analysis. *International Journal of Man-Machine Studies* 19:335–57. [GCO]
Oden, G. C. & Massaro, D. W. (1978) Integration of featural information in speech perception. *Psychological Review* 85:172–91. [aDN, GCO]
Oden, G. C. & McDowell, B. D. (in preparation) The gradedness of perceptual experience in identifying handwritten words. [GCO]
Oden, G. C., Rueckl, J. G. & Sanocki, T. (1991) Making sentences make sense, or words to that effect. In: *Understanding word and sentence,* ed. G. B. Simpson. North-Holland. [GCO]
Ohala, J. J. & Feder, D. (1994) Listeners' normalization of vowel quality is influenced by "restored" consonantal context. *Phonetica* 51:111–18. [JK]
Osterhout, L. & Holcomb, P. J. (1992) Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language* 31:785–804. [FI]
(1993) Event-related potentials and syntactic anomaly: Evidence of anomaly detection during the perception of continuous speech. *Language and Cognitive Processes* 8:413–37. [FI]
Osterhout, L. & Mobley, L. A. (1995) Event-related brain potentials elicited by failure to agree. *Journal of Memory and Language* 34:739–73. [FI]
Otake, T., Yoneyama, K., Cutler, A. & van der Lugt, A. (1996) The representation of Japanese moraic nasals. *Journal of the Acoustical Society of America* 100:3831–42. [aDN]
Paap, K., Newsome, S. L., McDonald, J. E. & Schvaneveldt, R. W. (1982) An activation- verification model for letter and word recognition: The word superiority effect. *Psychological Review* 89:573–94. [aDN]
Page, M. (2000) Connectionist modelling in psychology: A localist manifesto. *Behavioral and Brain Sciences* 23(3). [JG, arDN]
Peereman, R., Content, A. & Bonin, P. (1998) Is perception a two-way street? The case of feedback consistency in visual word recognition. *Journal of Memory and Language* 39:151–74. [JCZ]
Peeters, G., Frauenfelder, U. & Wittenburg, P. (1989) Psychological constraints upon connectionist models of word recognition: Exploring TRACE and alternatives. In: *Connectionism in perspective,* ed. R. Pfeifer, Z. Schreter, F. Fogelman-Soulié & L. Steels. Elsevier. [aDN]
Peitgen, H.-O., Jürgens, H. & Saupe, D. (1992) *Chaos and fractals: New frontiers of science.* Springer-Verlag. [JCZ]
Pitt, M. A. (1995) The locus of the lexical shift in phoneme identification. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 21:1037–52. [JRB, arDN]
Pitt, M. A. & McQueen, J. M. (1998) Is compensation for coarticulation mediated by the lexicon? *Journal of Memory and Language* 39:347–70. [LB, TLD, JK, DWM, RSN, arDN, MP, DHW]
Pitt, M. A. & Samuel, A. G. (1993) An empirical and meta-analytic evaluation of the phoneme identification task. *Journal of Experimental Psychology: Human Perception and Performance* 19:699–725. [JRB, arDN]
Press, W. H., Flannery, B. P., Teukolsky, S. A. & Vetterling, W. T. (1986) *Numerical recipes: The art of scientific computing.* Cambridge University Press. [aDN]
Raczaszek, J., Tuller, B., Shapiro, L. P., Case, P. & Kelso, S. (1999) Categorization of ambiguous sentences as a function of a changing prosodic parameter: A dynamical approach. *Journal of Psycholinguistic Research* 28:367–93. [JCZ]
Radeau, M. (1995) Facilitation and inhibition in phonological priming. Paper presented at the 36th Annual Meeting of the Psychonomics Society, Los Angeles. [LMS]
Radeau, M. & Colin, C. (1996) Task effects in phonological priming between spoken words. Paper presented at the 37th Annual Meeting of the Psychonomics Society, Chicago. [LMS]
Radeau, M., Morais, J. & Dewier, A. (1989) Phonological priming in spoken word recognition: Task effects. *Memory and Cognition* 17:525–35. [LMS]
Radeau, M., Morais, J. & Segui, J. (1995) Phonological priming between monosyllabic spoken words. *Journal of Experimental Psychology: Human Perception and Performance* 21:1297–311. [LMS]
Radeau, M., Segui, J. & Morais, J. (1994) The effect of overlap position in phonological priming between spoken words. *Proceedings of ICSLP "94, Yokohama, Japan, vol. 3,* 1419–22. [LMS]
Ratcliff, R., Van Zandt, T. & McKoon, G. (1999) Connectionist and diffusion models of reaction time. *Psychological Review.* (in press). [rDN]
Read, C. A., Zhang, Y., Nie, H. & Ding, B. (1986) The ability to manipulate speech sounds depends on knowing alphabetic reading. *Cognition* 24:31–44. [arDN]
Reicher, G. M. (1969) Perceptual recognition as a function of meaningfulness of stimulus material. *Journal of Experimental Psychology* 81:274–80. [aDN]
Repp, B. H. (1980) A range-frequency effect on perception of silence in speech. *Haskins Laboratories Status Report on Speech Research* 61:151–65. [SG, rDN]
(1982) Phonetic trading relations and context effects: New experimental evidence for a speech mode of perception. *Psychological Bulletin* 92:81–110. [JK]
Repp, B. H., Liberman, A. M., Eccardt, T. & Pesetsky, D. (1978) Perceptual integration of acoustic cues for stop, fricative, and affricate manner. *Journal of Experimental Psychology: Human Perception and Performance* 4:621–37. [SG, rDN]
Rhodes, G., Parkin, A. J. & Tremewan, T. (1993) Semantic priming and sensitivity in lexical decision. *Journal of Experimental Psychology: Human Perception and Performance* 19:154–65. [aDN]
Riedel, K. & Studdert-Kennedy, M. (1985) Extending formant transitions may not improve aphasics' perception of stop consonant place of articulation. *Brain and Language* 24:223–32. [rDN]
Roberts, M. & Summerfield, Q. (1981) Audiovisual presentation demonstrates that selective adaptation in speech perception is purely auditory. *Perception and Psychophysics* 30:309–14. [aDN]
Rodd, J., Gaskell, M. G. & Marslen-Wilson, W. D. (1999) Semantic competition and the ambiguity disadvantage. In: *Proceedings of the Twenty-First Annual Conference of the Cognitive Science Society, Vancouver, Canada, August 1999.* Erlbaum. [MGG]
Roelfsema, P. R., Engel, A. K., König, P. & Singer, W. (1996) The role of neuronal synchronization in response selection: A biologically plausible theory of structured representations in the visual cortex. *Journal of Cognitive Neuroscience* 8:603–25. [MM]
Rösler, F., Friederici, A. D., Pütz, P. & Hahne, A. (1993) Event-related brain potentials while encountering semantic and syntactic constraint violation. *Journal of Cognitive Neuroscience* 5:345–62. [FI]
Rubin, P., Turvey, M. T. & van Gelder, P. (1976) Initial phonemes are detected faster in spoken words than in nonwords. *Perception and Psychophysics* 19:394–98. [aDN]
Rueckl, J. G. & Oden, G. C. (1986) The integration of contextual and featural information during word identification. *Journal of Memory and Language* 25:445–60. [GCO]
Rumelhart, D. & McClelland, J. (1982) An interactive activation model of context effects in letter perception: Part 2. The contextual enhancement effect and some tests and extensions of the model. *Psychological Review* 89:60–94. [aDN]
Saffran, J. R., Aslin, R. N. & Newport, E. L. (1996) Statistical learning by 8-month-old infants. *Science* 274:1926–28. [PWJ]
Saldaña, H. M. & Rosenblum, L. D. (1994) Selective adaptation in speech perception using a compelling audiovisual adaptor. *Journal of the Acoustical Society of America* 95:3658–61. [aDN]
Salin, P.-A. & Bullier, J. (1995) Cortico-cortical connections in the visual system: Structure and function. *Physiological Reviews* 75:107–54. [MM]
Samuel, A. G. (1981) Phonemic restoration: Insights from a new methodology. *Journal of Experimental Psychology: General* 110:474–94. [SG, aDN, AGS]
(1981b) The rule of bottom-up confirmation in the phonemic restoration illusion. *Journal of Experimental Psychology: Human Perception and Performance* 7:1124–31. [SG]
(1986) Red herring detectors and speech perception: In defense of selective adaptation. *Cognitive Psychology* 18:452–99. [aDN]
(1987) Lexical uniqueness effects on phonemic restoration. *Journal of Memory and Language* 26:36–56. [aDN]
(1996a) Does lexical information influence the perceptual restoration of phonemes? *Journal of Experimental Psychology: General* 125:28–51. [aDN]
(1996b) Phoneme restoration. *Language and Cognitive Processes* 11:647–54. [aDN]
(1997) Lexical activation produces potent phonemic percepts. *Cognitive Psychology* 32:97–127. [arDN, AGS]
Samuel, A. G. & Kat, D. (1996) Early levels of analysis of speech. *Journal of Experimental Psychology: Human Perception and Performance* 22:676–94. [arDN, AGS]
Sawusch, J. R. & Jusczyk, P. (1981) Adaptation and contrast in the perception of voicing. *Journal of Experimental Psychology: Human Perception and Performance* 7:408–21. [aDN]
Segui, J. & Frauenfelder, U. (1986) The effect of lexical constraints upon speech perception. In: *Human memory and cognitive capabilities: Mechanisms and performances,* ed. F. Klix & H. Hagendorf. North-Holland. [aDN]
Seidenberg, M. S. & McClelland, J. L. (1989) A distributed, developmental model of visual word-recognition and naming. *Psychological Review* 96:523–68. [WDM-W, rDN]
Shepard, R. (1984) Ecological constraints in internal representation: Resonant kinematics of perceiving, imagining, thinking, and dreaming. *Psychological Review* 91:417–47. [PAL]
Shillcock, R. C. & Bard, E. G. (1993) Modularity and the processing of closed class words. In: *Cognitive models of speech processing: The second Sperlonga meeting,* ed. G. T. M. Altmann & R. C. Shillcock. Erlbaum. [RS]
Sillito, A. M., Cudeiro, J. & Murphy, P. C. (1993) Orientation sensitive elements in

https://doi.org/10.1017/S0140525X00303247 Published online by Cambridge University Press

the corticofugal influence on centre-surround interactions in the dorsal lateral geniculate nucleus. *Experimental Brain Research* 93:6–16. [MM]
Silverstein, S. M., Matteson, S. & Knight, R. A. (1996) Reduced top-down influence in auditory perceptual organization in schizophrenia. *Journal of Abnormal Psychology* 105:663–67. [MM]
Singer, W. (1995) Development and plasticity of cortical processing architectures. *Science* 270:758–64. [MM]
Slaney, M. (1998) A critique of pure audition. In: *Computational auditory scene analysis,* ed. D. Rosenthal & H. G. Okuno. Erlbaum. [MM]
Slowiaczek, L. M. & Hamburger, M. B. (1992) Prelexical facilitation and lexical interference in auditory word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 18:1239–50. [aDN, LMS]
Slowiaczek, L. M., McQueen, J. M., Soltano, E. G. & Lynch, M. (2000) Phonological representations in prelexical speech processing: Evidence from form-based priming. *Journal of Memory and Language.* (forthcoming). [LMS]
Slowiaczek, L. M., Soltano, E. G. & McQueen, J. M. (1997) Facilitation of spoken word processing: Only a rime can prime. Paper presented at the Psychonomic Society Meeting, Philadelphia. [LMS]
Steedman, M. J. & Altmann, G. T. M. (1989) Ambiguity in context: A reply. *Language and Cognitive Processes* 4:SI105–22. [aDN]
Stevens, K. N. (1995) Applying phonetic knowledge to lexical access. In: *Proceedings of the 4th European Conference on Speech Communication and Technology, Madrid,* 1:3–11. [KNS]
Stevens, K. N. & Halle, M. (1967) Remarks on analysis by synthesis and distinctive features. In: *Models for the perception of speech and visual form,* ed. W. Wathen-Dunn. MIT Press. [KNS]
Stone, G. O., Vanhoy, M. D. & Van Orden, G. C. (1997) Perception is a two-way street: Feedforward and feedback phonology in visual word recognition. *Journal of Memory and Language* 36:337–59. [JCZ]
Stone, G. O. & Van Orden, G. C. (1993) Strategic processes in printed word recognition. *Journal of Experimental Psychology: Human Perception and Performance* 19:744–74. [JCZ]
Streeter, L. A. & Nigro, G. N. (1979) The role of medial consonant transitions in word perception. *Journal of the Acoustical Society of America* 65:1533–41. [aDN]
Taft, M. (1982) An alternative to grapheme-phoneme conversion rules? *Memory and Cognition* 10:465–74. [JCZ]
Tallal, P. & Piercy, M. (1973) Defects of non-verbal auditory perception in children with developmental aphasia. *Nature* 241:468–69. [RMW]
Tallon-Baudry, C., Bertrand, O., Delpuech, C. & Pernier, J. (1997) Oscillatory γ-band (30–70 Hz) activity induced by a visual search task in humans. *The Journal of Neuroscience* 17:722–34. [MM]
Tanenhaus, M. K. & Donnenwerth-Nolan, S. (1984) Syntactic context and lexical access. *The Quarterly Journal of Experimental Psychology* 36A:649–61. [RS]
Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M. & Sedivy, J. (1995) Integration of visual and linguistic information in spoken language comprehension. *Science* 268:1632–34. [ASM, rDN]
Taraban, R. & McClelland, J. L. (1988) Constituent attachment and thematic role assignment in sentence processing: Influences of content-based expectations. *Journal of Memory and Language* 27:1–36. [aDN]
Ter Keurs, M., Brown, C. M., Hagoort, P. & Stegemann, D. F. (1999) Electrophysiological manifestations of open- and closed-class words in patients with Broca's aphasia with agrammatic comprehension. *Brain* 122:839–54. [rDN]
Tincoff, R. & Jusczyk, P. W. (1996) Are word-final sounds perceptually salient for infants? Poster presented at LabPhon V, Northwestern University, Evanston, IL. [PWJ]
Tononi, G., Sporns, O. & Edelman, G. M. (1992) Reentry and the problem of integrating multiple cortical areas: Simulation of dynamics integration in the visual system. *Cerebral Cortex* 2:310–35. [MM]
Trueswell, J. C. & Tanenhaus, M. K. (1994) Towards a lexicalist framework of constraint-based syntactic ambiguity resolution. In: *Perspectives on sentence processing,* ed. C. Clifton, L. Frazier & K. Rayner. Erlbaum. [aDN]
Tuller, B., Case, P., Ding, M. & Kelso, J. A. S. (1994) The nonlinear dynamics of speech categorization. *Journal of Experimental Psychology: Human Perception and Performance* 20:3–16. [JCZ]
Turner, C. W. & Doherty, K. A. (1997) Temporal masking and the "active process" in normal and hearing-impaired listeners. In: *Modeling sensorimotor hearing loss,* ed. W. Jesteadt. Erlbaum. [MM]
Ullman, S. (1995) Sequence seeking and counter streams: A computational model for bidirectional information flow in the visual cortex. *Cerebral Cortex* 1:1–11. [MM]
Usher, M. & McClelland, J. (1995) On the time course of perceptual choice: A model based on principles of neural computation. Technical Report PDP.CNS.95.5, Department of Psychology, Carnegie-Mellon University. [rDN]
Van Berkum, J. J. A., Brown, C. M. & Hagoort, P. (1999) When does gender constrain parsing? Evidence from ERPs. *Journal of Psycholinguistic Research* 28:555–71. [rDN]
Van Orden, G. & Goldinger, S. (1994) Interdependence of form and function in cognitive systems explains perception of printed words. *Journal of Experimental Psychology: Human Perception and Performance* 20:1269–91. [PAL, JCZ]
Van Orden, G. C., Holden, J. G., Podgornik, M. N. & Aitchison, C. S. (1999) What swimming says about reading: Coordination, context, and homophone errors. *Ecological Psychology* 11:45–79. [JCZ]
Van Petten, C. & Kutas, M. (1991) Electrophysiological evidence for the flexibility of lexical processing. In: *Understanding word and sentence,* ed. G. B. Simpson. Elsevier. [FI]
Vitevitch, M. S. & Luce, P. A. (1998) When words compete: Levels of processing in spoken word recognition. *Psychological Science* 9:325–29. [JRB, RSN, arDN]
(1999) Probabilistic phonotactics and neighborhood activation in spoken word recognition. *Journal of Memory and Language* 40:374–408. [PAL, rDN]
Vroomen, J. & de Gelder, B. (1995) Metrical segmentation and lexical inhibition in spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance* 21:98–108. [arDN]
Wandell, B. A. (1995) *Foundations of vision.* Sinauer. [MKT]
Warren, R. M. (1970) Perceptual restoration of missing speech sounds. *Science* 167:392–93. [RMW]
(1971) Identification times for phonemic components of graded complexity and for spelling of speech. *Perception and Psychophysics* 9:345–49. [RMW]
(1984) Perceptual restoration of obliterated sounds. *Psychological Bulletin* 96:371–83. [SG, MM]
(1999) *Auditory perception: A new analysis and synthesis.* Cambridge University Press. [RMW]
Warren, R. M. & Ackroff, J. M. (1976) Two types of auditory sequence perception. *Perception and Psychophysics* 20:387–94. [RMW]
Warren, R. M. & Bashford, J. A., Jr. (1993) When acoustic sequences are not perceptual sequences: The global perception of auditory patterns. *Perception and Psychophysics* 54:121–26. [RMW]
Warren, R. M., Bashford, J. A., Jr. & Gardner, D. A. (1990) Tweaking the lexicon: Organization of vowel sequences into words. *Perception and Psychophysics* 47:423–32. [RMW]
Warren, R. M. & Gardner, D. A. (1995) Aphasics can distinguish permuted order of phonemes - but only if presented rapidly. *Journal of Speech and Hearing Research* 38:473–76. [RMW]
Warren, R. M., Gardner, D. A., Brubaker, B. S. & Bashford, J. A., Jr. (1991) Melodic and nonmelodic sequence of tones: Effects of duration on perception. *Music Perception* 8:277–90. [RMW]
Warren, R. M., Healy, E. W. & Chalikia, M. H. (1996) The vowel-sequence illusion: Intrasubject stability and intersubject agreement of syllabic forms. *Journal of the Acoustical Society of America* 100:2452–61. [RMW]
Warren, R. M. & Obusek, C. J. (1971) Speech perception and phonemic restorations. *Perception and Psychophysics* 9:358–62. [RMW]
Warren, R. M. & Sherman, G. L. (1974) Phonemic restorations based on subsequent context. *Perception and Psychophysics* 16:150–56. [SG]
Warren, R. M. & Warren, R. P. (1970) Auditory illusions and confusions. *Scientific American* 223:30–36. [RMW]
Watt, S. M. & Murray, W. S. (1996) Prosodic form and parsing commitments. *Journal of Psycholinguistic Research (CUNY Special Issue)* 25:91–318. [WSM]
Werker, J. & Tees, R. (1984) Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development* 7:49–63. [rDN]
Whalen, D. H. (1984) Subcategorical phonetic mismatches slow phonetic judgments. *Perception and Psychophysics* 35:49–64. [PAL, aDN, DHW]
(1989) Vowel and consonant judgments are not independent when cued by the same information. *Perception and Psychophysics* 46:284–92. [DHW]
(1991) Subcategorical phonetic mismatches and lexical access. *Perception and Psychophysics* 50:351–60. [aDN, DHW]
Whalen, D. H. & Samuel, A. G. (1985) Phonetic information is integrated across intervening nonlinguistic sounds. *Perception and Psychophysics* 37:579–87. [DHW]
Wurm, L. H. & Samuel, A. G. (1997) Lexical inhibition and attentional allocation during speech perception: Evidence from phoneme monitoring. *Journal of Memory and Language* 36:165–87. [CMC, arDN, AGS]
Yost, W. A. & Nielsen, D. W. (1977) *Fundamentals of hearing: An introduction.* Holt, Rinehart and Winston. [MKT]
Zatorre, R. J., Evans, A. C., Meyer, E. & Gjedde, A. (1992) PET studies of phonetic processing of speech: Review, replication and reanalysis. *Cerebral Cortex* 6:21–30. [rDN]
Ziegler, J. C. & Ferrand, L. (1998) Orthography shapes the perception of speech:

https://doi.org/10.1017/S0140525X00303247 Published online by Cambridge University Press

The consistency effect in auditory word recognition. *Psychonomic Bulletin and Review* 5:683–89. [JCZ]

Ziegler, J. C. & Jacobs, A. M. (1995) Phonological information provides early sources of constraint in the processing of letter strings. *Journal of Memory and Language* 34:567–93. [JCZ]

Ziegler, J. C., Montant, M. & Jacobs, A. M. (1997a) The feedback consistency effect in lexical decision and naming. *Journal of Memory and Language* 37:533–54. [JCZ]

Ziegler, J. C. & Van Orden, G. C. (1999) Feedback consistency and subjective familiarity. (submitted). [JCZ]

Ziegler, J. C., Van Orden, G. C. & Jacobs, A. M. (1997b) Phonology can help or hurt the perception of print. *Journal of Experimental Psychology: Human Perception and Performance* 23:845–60. [JCZ]

Zue, V. & Shattuck-Hufnagel, S. (1979) The palatalization of voiceless alveolar fricative /s/ in American English. In: *Proceedings of the 9th International Congress of Phonetic Sciences*, ed E. Fischer-Jorgensen, 1:215. [KNS]

https://doi.org/10.1017/S0140525X00303247 Published online by Cambridge University Press