WHY MERGE REALLY IS AUTONOMOUS AND PARSIMONIOUS

James M. McQueen, Anne Cutler
Max-Planck-Institute for Psycholinguistics, Nijmegen, The Netherlands
and **Dennis Norris**
MRC Cognition and Brain Sciences Unit, Cambridge, U.K.

ABSTRACT

We briefly describe the Merge model of phonemic decision-making, and, in the light of general arguments about the possible role of feedback in spoken-word recognition, defend Merge's feedforward structure. Merge not only accounts adequately for the data, without invoking feedback connections, but does so in a parsimonious manner.

1. MERGE

Norris et al. [12] reviewed studies of phonemic decisions, and concluded that neither standard interactive nor standard autonomous models were now tenable. The former are challenged by findings showing variability in lexical effects [2] and no inhibitory effects in nonwords [5], as well as by the latest data on compensation for coarticulation [14] and subcategorical mismatch [6,9]. The latter are challenged by demonstrations of lexical involvement in phonemic decisions on nonwords [1,10]. Norris et al. proposed the Merge model (see Figure 1) to account for the data about phonetic processing in spoken-word recognition, while remaining faithful to the basic principles of autonomy.

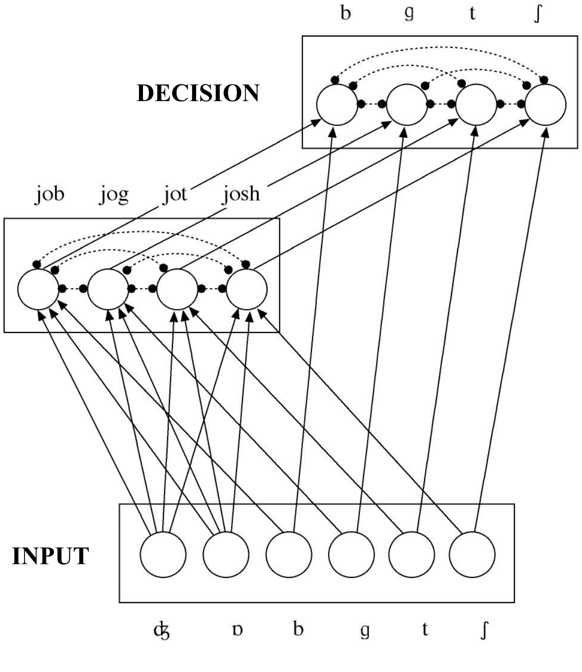

Figure 1. The Merge model

As Figure 1 shows, activation in Merge spreads from the input nodes to the lexical and the phoneme decision nodes, and from the lexical nodes to phoneme decision nodes; inhibitory competition operates at the lexical and phoneme decision levels. Excitatory connections (solid lines, arrows) are unidirectional; inhibitory connections (dotted lines, closed circles) are bidirectional.

In Merge, sublexical 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 merges information from the two sources. 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.

There are no inhibitory links between phoneme nodes at the prelexical level. Inhibition here would lead to categorical decisions which would be difficult for other levels to overturn, so that 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. 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.

The necessity of between-unit inhibition at the decision level, but not at the level of perceptual processing itself, is in itself an important motivation for Merge's architecture. Perceptual processing and decision making have different requirements and therefore cannot be performed effectively by the same units. Any account of phonemic decision making should separate decision processes from prelexical processes.

The units at the phoneme decision level in Merge are necessarily, phonemic. The prelexical representations are also phonemic. Note, however, that nothing hinges on this assumption. The prelexical units could be replaced with an other type of representation (features, gestural units, etc.). If these representations were connected in the appropriate way to the units at the lexical and decision levels, the model would work in the same way and its basic architecture would be unchanged.

2. MERGE AND THE DATA

Simulations with Merge conducted by Norris et al. [12] demonstrated 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 data which had proved problematic for previously available models. Merge could account for inhibitory effects of competition in nonwords with subcategorical mismatches [6,9], facilitatory effects in nonwords which are more like real words relative to those which are less like real words [1], and the lack of inhibitory effects in nonwords which diverge from real words near their end [5]. It also explained how both facilitatory and inhibitory effects come and go according to task demands [2,9].

Further, Merge is also able to explain other basic lexical effects observed in the literature, for example in phonetic categorization and phoneme restoration. The explanation for lexical effects in phonetic categorization parallels that of effects in phoneme monitoring: lexical node activation can bias phoneme decision-node activation such that an ambiguous phoneme in a word-nonword continuum will tend to be labelled in a lexically-consistent manner (e.g. as /d/ in a *tice-dice* continuum). Likewise, lexical activation boosts phoneme decision-node activation so that there will tend to be more phonemic restorations in words than in nonwords.

Similarly, Merge can account for the results of Pitt and McQueen [14], in particular the fact that these authors found that lexical effects and effects of transitional probability were dissociated. Models such as TRACE [8], in which transitional probability effects arise from lexical information, cannot account for such a dissociation. In Merge, the influence of transitional probabilities on compensation for coarticulation can be modelled by adding a process sensitive to these probabilities at the prelexical level. The lexicon can influence decisions to ambiguous phonemes 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 processes at the prelexical level, so that the two loci are separate and dissociation of the two sorts of effect is to be expected.

3. DEFINING FEEDBACK

Merge was designed as an existence proof of the ability of feedforward models to account for certain crucial empirical findings involving the relationship of prelexical and lexical processing in spoken-word recognition. Reactions to the target article [12] from some researchers, however, suggested that the crucial terminology involved in the debate ('feedback; top-down; interaction') is at times used inconsistently, even by the same authors in different publications.

Norris et al.'s target article focussed on the issue of feedback, because this is the crux of the debate in the literature. Norris et al. 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 the term `interaction' is indeed most commonly used in this way, that is, to characterise information flow between processes, it is true that the term is sometimes also used instead to indicate that two different kinds of information are somehow combined. These two senses of interaction have been termed 'process interaction' versus 'information interaction' [11]. It is important to note that information interaction does not imply process interaction. For example, one might draw 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 interact with one another. Merge has no process interaction and no feedback.

The sense of 'top-down' which predominates in the psychological literature concerns 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. Although this sense of top-down gets close to the concept of feedback, and is often used synonymously with feedback in the literature, it is not identical. Nonspecific top-down flow of information, in generalised attentional activation for example, is not the same as specific feedback from particular lexical items which alters the processing of specific phonemes.

But again, 'top-down' is also sometimes used in a less well-defined sense, which appears to resemble 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. If, for instance, lexical knowledge were used in any way to influence decisions about phonemes, this would be described as lexical and phonemic information being combined in a top-down fashion. This is quite independent of the issue of feedback, or even direction of information flow. In this sense, strictly feedforward models like the Race model [2,3], Merge [12], and FLMP [7] would be top-down.

4. WHY FEEDBACK IS UNNECESSARY

4.1. Feedback in word recognition

In models like TRACE [8], interaction alters the tendency of the model to emit particular responses but does not help the model to perform lexical processing more accurately. The best performance that can be expected from a word recognition system is that it reliably identifies the word with the lexical representation which best matches the input. Thus a recognition system that simply matched the input against each lexical entry, and then selected the entry with the best fit, would provide optimal isolated-word recognition performance, limited only by the accuracy of the representations. Adding activation feedback from lexicon to input would not improve recognition accuracy. In order to make word recognition more accurate, any feedback would have to enable the system to improve the initial perceptual representation of the input, so that the match to the lexicon could become more exact. This could only be done by actually altering the sensitivity of phoneme recognition, as described below. Feedback could have the effect of speeding word identification by reducing the recognition threshold. But recognition would then be based on less perceptual information, which of course allows for the possibility that recognition would become less rather than more accurate. Note that the argument that feedback cannot improve recognition does not depend on any assumptions about the quality or reliability of the input. The same logic applies whether the speech is clearly articulated laboratory speech or natural conversational speech in a noisy background.

4.2. Feedback in phoneme recognition

In general, although interaction cannot improve the accuracy of word recognition, it can affect phoneme recognition if the input consists entirely of words. If a phoneme cannot be distinguished clearly by the phoneme level alone, biasing interaction from the lexical level can make it more likely that the phoneme will be identified as the appropriate phoneme for a particular word. Of course, if the input were a nonword or mispronunciation, such a biasing effect would harm phoneme recognition performance rather than help it, in that a deviant phoneme would be misidentified as the phoneme expected on the basis of the lexical feedback.

Given that feedback does not help word recognition, however, it is unclear what is gained by making prelexical representations concur with decisions already made at the lexical level. Once a word decision has been reached, there is no obvious reason why the representations which served as input to the word level should then be modified. Feedback might improve explicit phoneme identification, but this is not usually an objective -- the explicit recognition objective in normal spoken language processing is word identification. Only if feedback can improve sensitivity would it be of assistance to phonemic processing. This would, for instance, be manifested by improved discriminability along a phonetic continuum with at least one word endpoint (e.g. *dice* to *tice*) as opposed to the same continuum with two nonword endpoints (e.g. *dife* to *tife*). So far, there is no evidence that such effects occur.

5. WHY MERGE IS AUTONOMOUS

Some of the commentators on the Merge article [12] suggested that Merge is not autonomous. This is incorrect, for three reasons. First, as described above, the central issue in debates about model architecture has been whether feedback is necessary, that is, whether information flow should be unidirectional or bidirectional. In Merge, as is clear from Figure 1, the input phonemes feed into words, but words do not feed back to input phonemes. Input phonemes also feed to decision units, but decision units also do not feed back to input phonemes. Finally, words feed to decision units, but decision units again do not feed back to words. There is absolutely no bidirectional information flow in the model. Merge's accurate simulation of the results from many experiments shows that the data can be captured by a model with only unidirectional information flow, that is, an autonomous model.

Second, it is important to emphasize that any demonstration of lexical effects on phoneme identification (which must surely be based on phonemic codes) does not by itself constitute evidence of `top-down' processing. All researchers in the field have been in agreement about the existence of lexical effects on phoneme identification for more than 20 years (see [3] for a review). Furthermore, lexical codes have always influenced phonemic codes in autonomous models. In the (feedforward only) Race model [2,3], lexical access makes the lexically-based phonological code of the word available. The fact that lexical information can influence phonemic decisions in Merge does not make the model non-autonomous.

Third, note that Merge's decision units are flexible and configurable according to task demands, so they do not constitute a Fodorian [4] module. But, once configured for the task, they take input from two sources (lexical and prelexical) and then produce an output without any interference or feedback from subsequent processes. This again justifies the label 'autonomous'.

6. WHY MERGE IS PARSIMONIOUS

Merge incorporates decision nodes; but there are no explicit decision nodes in, for instance, TRACE [8]. Are these decision nodes thus an added extra which interactive models such as TRACE can do without? In fact, all models, TRACE included, need some form of decision mechanism. Merge simply makes that mechanism explicit.

Most psychological theories give a less than complete account of how a model might be configured to perform various experimental tasks. In many phoneme-monitoring studies, for instance, listeners have to monitor only for word-initial phonemes. By definition, this demands that positional information from the lexicon is combined with information about phoneme identity. Although the results from such tasks were modelled by the Race model [2,3] and by TRACE [8], neither model 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 specific phonemes in word-final position in nouns when a signal light turned red. Correct responding would require combining phonemic, syntactic, and crossmodal information. But this does not mean that we have hard-wired {final, /t/, noun, red} 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, syntactic 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 needs to be flexible enough to be able to cope with any task (from the very simple to the very complex) that an experimenter might devise.

Merge does not explicitly model how this system configures itself, but the decision nodes do explicitly represent the process of combining different sources of information. Merge does one further thing. Although we can devise phoneme identification tasks that necessarily take account of lexical information, in the simplest phoneme identification tasks listeners can and do sometimes ignore the output of the lexicon [2]. Merge allows for the possibility that listeners sometimes monitor the phonemic and lexical levels even when this is not explicitly required by the task. This is the source of lexical effects in phoneme identification, which can then come and go as a function of task demands.

Note also that 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 [15]. Furthermore, neuroimaging work reveals different patterns of brain activity in tasks involving explicit phonological decisions from those involving passive listening (see [13] for a review).

Thus the decision nodes in Merge do not undermine its parsimony compared to other models. All models must allow for flexible and configurable decision mechanisms, for listeners' strategic choices in task performance, and for differing brain activation patterns consequent upon explicit phonological decisions as opposed to normal speech recognition. The important point is that the decision process is not an optional extra. Without some such process listeners could not ever perform the experimental tasks required of them in psycholinguistic laboratories. The decision process is not something Merge has but other models can forgo; all models need a decision process. When that decision process is taken into account, as in the Merge model, it can be seen that this process is probably responsible for lexical effects in phoneme identification, leaving normal speech perception as a feedforward process.

7. CONCLUSIONS

Feedback in models of speech recognition does not actually help the process of recognising words, and it is also not necessary to explain the process of recognising phonemes. The Merge model, which incorporates no feedback, accounts for data from phonemic decision-making in a parsimonious and ecologically valid manner.

8. ACKNOWLEDGEMENTS

Corresponding author: J.M. McQueen, MPI for Psycholinguistics, Wundtlaan 1, 6525 XD Nijmegen, The Netherlands. Email: james.mcqueen@mpi.nl.

9. REFERENCES

[1] Connine, C.M., Titone, D., Deelman, T. & Blasko, D. (1997) Similarity mapping in spoken word recognition. *Journal of Memory and Language*, **37**, 463-480.

[2] Cutler, A., Mehler, J., Norris, D.G. & Segui, J. (1987). Phoneme identification and the lexicon. *Cognitive Psychology*, **19**, 141-177.

[3] Cutler, A. & Norris, D. (1979). Monitoring sentence comprehension. In Cooper, W.E. & Walker, E.C.T. (Eds.), *Sentence processing: Psycholinguistic studies presented to Merrill Garrett* (pp. 113-134). Erlbaum.

[4] Fodor, J.A. (1983). *The modularity of mind*. MIT Press.

[5] Frauenfelder, U.H., Segui, J. & Dijkstra, T. (1990). Lexical effects in phonemic processing: Facilitatory or inhibitory? *Journal of Experimental Psychology: Human Perception and Performance*, **16**, 77-91.

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

[7] Massaro, D.W. (1987*). Speech perception by ear and eye: A paradigm for psychological inquiry*. Erlbaum.

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

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

[10] 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**, 873-889.

[11] Norris, D.G. (1980*). Serial and parallel models of comprehension*. PhD Thesis: University of Sussex.

[12] Norris, D.G., McQueen, J.M. & Cutler, A. (2000). Merging information in speech recognition: Feedback is never necessary. *Behavioral and Brain Sciences*, **23**.

[13] Norris, D.G. & Wise, R. (1999). The study of prelexical and lexical processes in comprehension: Psycholinguistics and functional neuroimaging. In Gazzaniga, M. (Ed.), *The Cognitive Neurosciences* (pp. 867-880). MIT Press.

[14] Pitt, M.A. & McQueen, J.M. (1998). Is compensation for coarticulation mediated by the lexicon? *Journal of Memory and Language*, **39**, 347-370.

[15] 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.