

**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](mailto:massaro@fuzzy.ucsc.edu)  
<http://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, Mc-Queen & Cutler are to be applauded in articulating the bald hypothesis that feedback is never necessary in speech recognition.

### **338 BEHAVIORAL AND BRAIN SCIENCES (2000) 23:3**

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 /ift/ or /is/. 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).