COGNITIVE PSYCHOLOGY 23, 558–614 (1991)

Integration versus Interactive Activation: The Joint Influence of Stimulus and Context in Perception

DOMINIC W. MASSARO AND MICHAEL M. COHEN

University of California, Santa Cruz

Empirical results from both reading and speech perception indicate that stimulus and context information have independent influences on perceptual recognition. Massaro (1989) argued that these data are inconsistent with an interactive activation and competition (IAC) model (McClelland & Rumelhart, 1981), and consistent with the fuzzy logical model of perception (FLMP) (Massaro, 1979; 1989). McClelland (1991) then modified the interactive activation model to be stochastic rather than deterministic and to use a best one wins (BOW) decision rule, allowing it to predict independent influences of stimulus and context. When tested against real data, however, the network proposed by McClelland and extended by us gives a poorer description of actual empirical results than the FLMP. To account for the dynamics of information processing, the SIAC model, an interactive model based on the Boltzmann machine, and the FLMP are formulated to make quantitative predictions of performance as a function of processing time. It is shown that the dynamic FLMP provides a better description of the time course of perceptual processing than does interactive activation. The SIAC and Boltzmann models have difficulty predicting 1) context effects given little processing time and 2) a strong stimulus influence given substantial processing time. Finally, we demonstrate that the FLMP predicts that context can improve the accuracy of performance, in addition to providing a bias to respond with the alternative supported by context. In summary, there is now both empirical and theoretical evidence in favor of the FLMP over SIAC models of pattern recognition. We therefore argue that interactive activation is both less consistent with empirical results and not necessary to describe the joint influence of stimulus and context in language perception.

INTRODUCTION

Psychologists have long been intrigued with the finding that context appears to influence perception. The same stimulus information in different contexts can produce different perceptual events. In reading, Cattell (1886) demonstrated that readers could recognize more letters when they formed words than when they were randomly sequenced. In speech perception, studies showed that a sentence context facilitated recognition of a spoken word (Bagley, 1900). A recent example of a context effect in psycholinguistic research is the influence of phonological constraints in speech perception (Massaro, 1989). Each test stimulus was a consonant cluster syllable beginning with one of the three consonants /p/, /t/, or /s/ followed by a glide consonant ranging (in five levels) from /l/ to /r/, followed by the vowel /i/. There were 15 test stimuli created from the factorial combination of the three initial-consonant contexts times the five levels of the glide consonant. Subjects who were instructed to listen to each test syllable and to respond whether they heard /l/ or /r/, were influenced by both the glide consonant and the context.

Two models of these phonological context effects are the fuzzy logical model of perception (FLMP) (Massaro, 1989) and the TRACE model (Elman & McClelland, 1986; McClelland & Elman, 1986). Both models provide a detailed description of the integration of top-down and bottom-up sources of information in speech perception. These two models share a variety of processing assumptions and make highly similar predictions. They are information-processing models and assume some perceptual processing followed by decision. Continuous, not just categorical, information is available during perceptual processing and at the decision stage. Both the original interactive activation and competition (IAC) models and the FLMP assumed decision rule based on the relative goodness of match. These similarities and others (Massaro, 1987, 1988; Massaro & Cohen, 1987; McClelland, 1991) are responsible for similar predictions in most situations. Thus, differentiating between the models requires a fine-grained analysis of experiments specifically aimed at testing between the models.

It is important to analyze both performance and models of performance in terms of stages—sequential algorithms or equations specifying processing between stimulus input and response output (Massaro & Friedman, 1990). Even if they are only implicit, models of pattern recognition necessarily distinguish between evaluation of the available sources of information and integration of these sources. In the TRACE model, bottom-up stimulus information is evaluated at the feature level, whereas the phoneme level allows for the integration of information from the feature level and the word level. Consider recognition of the glide in a stop-consonant glide syllable, such as /pli/. The acoustic information about the stop is evaluated at the featural level and activates phonemes and words in memory. The same is true for the glide. These two activation processes overlap in time and interact with one another. Most importantly, the featural

The research reported in this paper and the writing of the paper were supported, in part, by grants from the Public Health Service (PHS R01 NS 20314), the National Science Foundation (BNS 8812728), a James McKeen Cattell Fellowship, and the graduate division of the University of California, Santa Cruz. The authors would like to thank Jay McClelland for making his modified version of the SIAC program available (Rumelhart & McClelland, 1988), Steve Kitzis and Dan Friedman for valuable discussions, and Steve Kitzis, Jay McClelland, two anonymous reviewers, and Steve Palmer for comments on earlier versions of this paper. Requests for reprints should be sent to Dr. Dominic Massaro, Department of Psychology, University of California, Santa Cruz, CA 95064.

0010-0285/91 $7.50

information (degree of activation) passed on to integration at the phoneme level changes with the consonant context. Similarly, the top-down activation due to context depends on the featural information from the glide. That is, in the TRACE model, the evaluation (representation) of each source of information is influenced by the processing of the other source of information. The acoustic information about the glide is evaluated differently at the featural level as a function of the nature of the initial consonant and its processing and vice versa.

Using a signal detection framework, Massaro (1989) demonstrated that the TRACE model predicts sensitivity differences in the phonological constraints experiment—rather than just bias differences. In TRACE, context influences the discriminability of the stimulus information specifying or representing the glide consonant. The discrimination of two adjacent levels along the /li/-/ri/ continuum differs for different contexts. In Massaro's experiment, the effect of phonological context turned out to be only a biasing effect rather than an effect on sensitivity, thus contradicting the predictions of the TRACE model. On the other hand, the results were well-described by a fuzzy logical model of perception (FLMP)—whose distinguishing feature (relative to TRACE) is independence of stimulus information and context at the evaluation stage of processing. When analyzed in the signal detection framework, the FLMP correctly predicts that context in the phonological constraints experiment should influence only bias and not sensitivity.

Note on Bias versus Sensitivity Effects

Although the signal detection framework is valuable, it can be somewhat misleading to describe the possible outcomes of an identification task as sensitivity and bias. Strictly speaking, sensitivity is used here to refer to the representation of the stimulus featural information, not to any arbitrary measure of performance. Bias is used to describe any influence of context that does not result from a change in the representation of the featural information about the glide. Interactive activation predicts that the initial-consonant context influences the representation of the featural information about the glide whereas the FLMP does not (Massaro, 1989). Both models can predict that an additional source of information can influence performance, such as making it more orderly and accurate.

According to the FLMP, the effects of stimulus information and context are symmetrical. Context can bias the response to stimulus information or stimulus information can bias the response to context. This mutual influence or bias is more apparent in McClelland's (1991, Fig. 4) plots of the z-score transformations of the percentage judgments than in Massaro's (1989) plots of z-score differences along the stimulus continuum. Thus, it is just as accurate to describe the influence of stimulus information on the effect of context as a bias effect as it is to use bias to describe the influence of context on the effect of stimulus information.

In the FLMP, stimulus information and context function as two independent sources of information at evaluation. Each biases the response given in the presence of the other source of information. However, as will be illustrated in the derivation of the FLMP, two sources of information can be more informative than just one. In this manner, the FLMP also predicts sensitivity effects at the outcome of the integration of the two sources of information. That is, the combination of stimulus information and context can produce more accurate performance than produced by either source presented alone.

Revised Interactive Activation Models

McClelland (1991) placed the blame for TRACE's failure to predict Massaro's results on the decision stage of the model rather than on interactive activation during the evaluation stage. By adding noise to the input or to its processing, and by assuming a decision rule of choosing the response alternative corresponding to the most active phoneme unit, the predictions of a new stochastic IAC (SIAC) model and TRACE were brought into line with a biasing effect of context. Thus, the new TRACE appeared to be consistent with the empirical observations (and the predictions of the FLMP). According to McClelland, while TRACE and the FLMP are equally able to capture the observed data, TRACE and interactive activation models in general are to be preferred because they account for the increase in accuracy given context, the mutual influence of the multiple parts (source of information) of a pattern, and the dynamics of information processing, whereas the FLMP does not. We dispute these claims in the present paper.

At this point, we should emphasize our agreement with McClelland's acknowledgment that the critical point is the falsification of interactivity itself—bidirectional propagation of information—rather than just some specific model implementing it. If the assumption of interactivity is falsified, "the whole idea that perception involves a bidirectional flow of information would be ruled out" (McClelland, 1991, p. 3). Even so, the investigator is limited to testing various implementations of interactive activation models—a daunting task given the intensive computation that is required. When several implementations are shown to be inadequate, however, doubt begins to be cast on the underlying theory, at least until its proponents uncover an adequate version.

McClelland (1991) appears to take the following tack in his modification of the interactive activation model. Given the empirical results showing the independence of bottom-up and top-down information in z-score transformations of the percentage of identification judgments, the question is whether IAC models predict similar functions. He observes how the nonlinear activation process and interactive activation violate this prediction when a relative goodness rule (RGR) is used at the decision stage. Since independence is the correct result, he developed a new algorithm to produce it. Making the interactive processing stochastic and using a best one wins (BOW) decision rule was sufficient for the new model to simulate the pattern of data predicted by independence of top down and bottom-up information. These properties of the new model, however, cancel any unique effects produced by the interactive activation algorithm (Massaro & Cohen, 1989). Thus, the new SIAC model is able to make independence predictions even though the processing produced by the interactive activation algorithm is fundamentally nonindependent at the evaluation stage of processing. Moreover, although McClelland argues that the interactive processing is valuable for predicting the time course of processing, he has not demonstrated this for the new interactive activation model by actually fitting it to experimental results.

We accept McClelland's demonstration that SIAC models can now produce the asymptotic pattern of data predicted by independence. However, we question whether the proposed SIAC network can easily describe actual empirical results. In the present paper, we test the new SIAC model using several different data sets. In the first section, we compare its asymptotic predictions with those of the FLMP, using the results of a phonological experiment by Massaro and Cohen (1983). We find that the empirical tests of the SIAC models require immense computer resources and time and, in several instances, provide less adequate descriptions of the results than does the more parsimonious FLMP. The FLMP consistently produces a better fit than a variety of SIAC models. We also compare the asymptotic activations predicted by the SIAC model to the corresponding truth values predicted by the FLMP. We contrast the nonlinearity of the SIAC activations with the linear truth values of the FLMP, and explain why the nonlinear activations are problematic. In the second section, we extend the SIAC model and the FLMP to describe the dynamics of perceptual processing and contrast their predictions for data from a backward masking experiment by Massaro (1979). The FLMP describes the time course of processing more accurately than a corresponding SIAC model. A model based on the Boltzmann machine was also tested because McClelland (1991) proved that this model predicts independence at equilibrium. We have found, however, that the Boltzmann machine fails to predict the dynamics of information processing, in much the same way as SIAC models. In the final section, we show that the FLMP also predicts the word superiority effect and the dynamics of context effects in a Reicher-Wheeler task.

PREDICTING ASYMPTOTIC BEHAVIOR

Different traditions have emerged in computer simulations and mathematical models. Given a closed mathematical expression for a model, it is straightforward to test it against actual results by deriving quantitative predictions using parameter estimation. In simulations of models without closed expressions, as in the SIAC model, fitting the model to actual results is not carried out because it would be difficult and tedious. Consistent with the simulation approach, McClelland demonstrated that a SIAC model could predict response functions with similar shapes to those observed in experimentation. At the level of simulation, therefore, both the SIAC and the FLMP are qualitatively consistent with the independent influences of stimulus information and context. Nevertheless, the two models might *not* be equally descriptive of actual empirical results; this can only be determined by quantitative comparisons against real data. In this section, we use model fitting techniques to test the two models against results showing top-down effects of phonological constraints.

The "new" data that we will now consider came from a study investigating the role of phonological constraints in speech perception (Massaro & Cohen, 1983). Relative to the Massaro (1989) experiment, Massaro and Cohen (1983) tested a larger number of experimental conditions and recorded more observations per condition. The results should, therefore, provide more sensitive and reliable tests among the models. In Experiment 2 of that article we presented subjects with CCV syllables with the first consonant being /p/, /t/, /s/, or /v/, the second consonant being one of seven glides equally spaced on a continuum between /l/ and /r/, and with the vowel /i/. The glide was changed from /l/ to /r/ by changing its initial third formant (F_3) frequency from high to low. Seven subjects from an introductory psychology class were each presented with each of the 28 possible experimental conditions (4 context times 7 glide) 40 times in four sessions run over a two-day period. Subjects made their responses by pressing one of eight buttons combining context and glide identifications, but we will concern ourselves mainly with the data pertaining to glide identification, except to note that subjects were 95% correct in context identification. Readers are referred to the original paper for further details of the stimuli and procedures. Figure 1 shows the proportion of observed /r/ identifications for the 28 experimental conditions averaged over the seven subjects. The effects of context and glide level were highly significant, with each independent variable having its largest effect when the other was most ambiguous.

SIAC Model with Input Noise

We begin with the SIAC model with noise added to the stimulus inputs. The network we used, shown in Fig. 2, assumes three layers of units: Target, Context, and Word. All units within the Context layer are bidirectionally connected to all units within the Word layer. Similarly, all

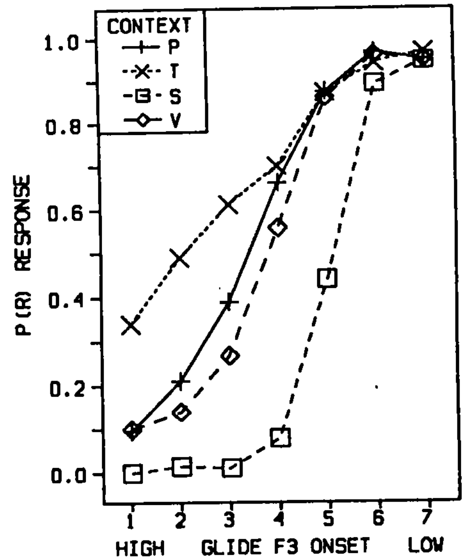

FIG. 1. Observed probability of an r response as a function of the glide F_3 onset level and context (after Massaro & Cohen, Experiment 2, 1983).

units within the Target layer are bidirectionally connected to all units within the Word layer. This network is identical to that used by McClelland (1991, Fig. 5), except that an additional context unit for /v/ is added. In Fig. 2, only the excitatory connections (which are bidirectional between each pair of units) are shown. Within each layer, each unit sends inhibitory connections to all other units. Given a stimulus presentation, external inputs, ext_i, are applied to the Context and Target units. External inputs are values representing the stimulus input to the respective units. These units pass on activation to units in the Word layer, which in turn

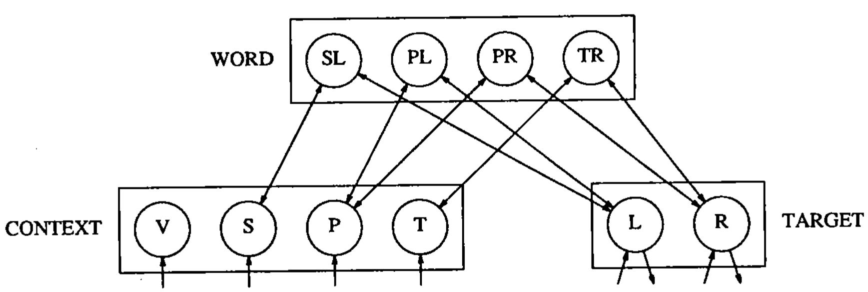

FIG. 2. Network used in the simulation of the IAC model applied to the phonological constraints experiment of Massaro and Cohen (1983). The inhibitory connections between units within the word, context, and target levels are not shown in the network.

pass on activation back to the Context and Target units. Processing continues in this manner for a number of cycles. Note that there are no word units for /vl/ or /vr/, because no words with these consonant clusters in initial position occur in English.

The formal algorithm of the SIAC model is as follows (McClelland & Rumelhart, 1988). Initially, for each unit, i, its activation, act_i, is set to the resting level, $rest$. Then, on each computational cycle of the model for each unit, i, the excitatory input, exc_i, and inhibitory input, inh_i, are computed from the product of the sending units and path weights as follows:

$$exc_i = \sum_j \max(0,w_{ij}) \times \max(0,act_j) \tag{1}$$

$$inh_i = \sum_j \min(0,w_{ij}) \times \max(0,act_j) \tag{2}$$

where w_{ij} is the weight from unit j to unit i. All weights w_{ij} are either $+1$ or -1, so that Eq. 1 adds up all the activations on positive pathways and Eq. 2 adds up all the activations on negative pathways. Activations less than 0 are ignored in these summations. Next, for each unit, i, the summed net input, net_i is computed from the weighted sum of exc_i, inh_i, and external inputs, ext_i:

$$net_i = \alpha \times exc_i + \gamma \times inh_i + estr \times ext_i \tag{3}$$

where α is the weight on excitatory connections, γ is the weight on inhibitory connections, and $estr$ is the weight on external inputs. Next, the change of activation for each unit for the upcoming cycle, Δact_i, is computed as:

$$\text{if } net_i > 0,\ \Delta act_i = net_i(M - act_i) - decay(act_i - rest) \tag{4}$$

$$\text{if } net_i < 0,\ \Delta act_i = net_i(act_i - m) - decay(act_i - rest) \tag{5}$$

where M is the maximum allowed activation, m is the minimum allowed activation, and $decay$ is the rate at which each unit returns to resting state. Then each act_i is adjusted by adding Δact_i:

$$act_i = act_i + \Delta act_i \tag{6}$$

Finally, each act_i is adjusted, if necessary, to remain in the interval m to M:

$$\text{if } act_i > M,\ act_i = M \tag{7}$$

$$\text{if } act_i < m,\ act_i = m \tag{8}$$

McClelland (1991) used the SIAC model with the following set of control parameters: $estr = .1$, $\alpha = .1$, $\gamma = .1$, $decay = .1$, $M = 1.$, $m = -.2$, and $rest = .1$. In the network, the effects of stimulus and context are combined via the units in the word layer. The activations of Word units are fed back to the Target and Context units, changing their activations in a manner that reflects the activations of both Target and Context units. In this manner, the joint effect of Target and Context are represented in the activations of units in both the Target and Context layers. The activations of the *R* and *L* Target units after 60 cycles of processing were used as inputs to the BOW decision rule.

The SIAC model was fit to the observed data using the program STEPIT (Chandler, 1969). Fits were obtained for the seven individual subjects. A model is represented to the analysis program STEPIT as a set of prediction equations or an algorithm for generating the model's predictions. In both cases, the model has a set of unknown parameters. These free parameters are first set at some starting value and a set of predictions is made with these values. A measure of goodness of fit is computed. Then the parameters are changed and another set of predictions and another measure of goodness of fit are computed. By comparing the measures of goodness of fit, iteratively adjusting the parameters of the model, and using a modified direct search technique, STEPIT minimizes the sum of squared deviations between the observed and predicted values. Thus, STEPIT finds the set of parameter values which allow the model to predict the observed data most closely.

For the SIAC model, the estimated parameters were four external activation values for the four possible contexts and seven external activations for the /r/ target. Following McClelland (1991), only the external activation for the node corresponding to the actual context was made nonzero, and the *L* Target unit received an activation value which was the additive complement of that received by the *R* Target unit. This gives a total of 11 parameters. In our preliminary fits of the model, several other parameters of the model (*estr, alpha, gamma,* and *decay*) were set to .1 and the standard deviation of the input noise used was set to .14142, since these were the values assumed by McClelland (1991).

McClelland (1991) used 10,000 simulated trials in his simulation of the SIAC model under a given set of input conditions (i.e., a given level along the /r/–/l/ continuum and a given context). Although this assumption is reasonable for determining asymptotic predictions, it is difficult to implement and somewhat unrealistic when the model is applied to experimental results. It is difficult to implement because 10,000 simulated trials require an immense amount of computer time and this number of simulated trials must be carried out for each set of parameters during the parameter estimation process. A large number of simulated trials is also somewhat unrealistic because subjects in Massaro and Cohen's (1983) Experiment 2 were tested for just 40 trials per condition. In terms of the SIAC model, each subject had just 40 opportunities to sample the activations of the units associated with each response alternative under each experimental condition. Because we might expect the goodness of the fit of the predicted results to differ depending on the number of observations, it is reasonable to allow the SIAC model the same number of trials to compare with the observed results. We will, however, also explore how the models behave as the number of simulated trials is varied.

The computation of the model predictions began by setting the 28 predicted probabilities of /r/ response given the 7 glide times 4 context conditions to 0 and resetting the random number generator. Then for each of the 28 experimental conditions, 40 simulated trials occurred. On each simulated trial, random deviates from a normal (Gaussian) distribution computed by the *Box-Muller* method (Press, Flannery, Teukolsky, & Vetterling, 1988) with a standard deviation of .14142 were added to each of the current pair of external input parameters (for context and target). Then the IAC algorithm (McClelland & Rumelhart, 1988) was run to compute the target activations after 60 time cycles. If the final activation of the /r/ target node was greater than or equal to the final activation of the /l/ target node, then 1/40 was added to the predicted probability of an /r/ response. In order for the parameter estimation routine to operate properly and converge on an optimal fit of the results, it was necessary to employ the same sequence of random numbers on each overall computation run. (If this had not been done, the parameter values would not have a reliable effect on the predictions, and STEPIT would spuriously accept or reject parameter value modifications.) This allowed STEPIT to make reliable adjustments in the parameter values, even though noise was being added to the input.

The central IAC subroutine from McClelland and Rumelhart (1988) was recoded in FORTRAN (f77) for use with STEPIT. For the tests of models with fixed SIAC parameters, we were able to speed up the lengthy parameter search process by precomputing the IAC activations and then using a table lookup technique with two-dimensional interpolation. To construct a lookup table for each of the four contexts, we computed the activation after 60 cycles for 101 possible context activation values going from $-.5$ to 1.5 in steps of .02 each combined with 101 possible /r/ target activation values going from $-.5$ to 1.5 in steps of .02. Given the activations of the target and the context, the nearest activation values could be found in the table and interpolation would give a very close approximation of the directly computed activations predicted by the IAC model at 60 cycles.

Several hundred adjustments to the set of parameter values (calls to STEPIT) were needed to maximize the goodness-of-fit of the 11-