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

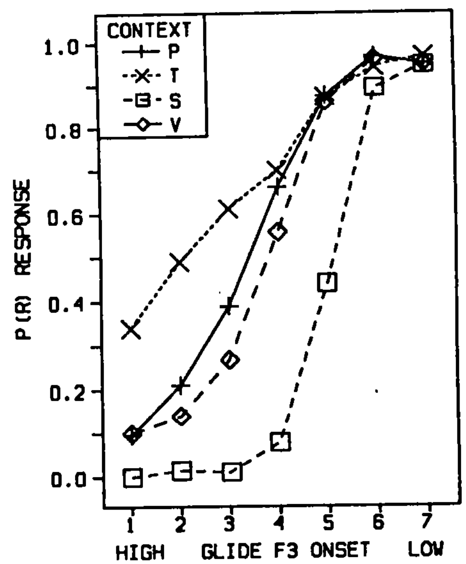


FIG. 1. Observed probability of an *r* response as a function of the glide $F_3$ onset level and context (after Massaro & Cohen, Experiment 2, 1983).

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

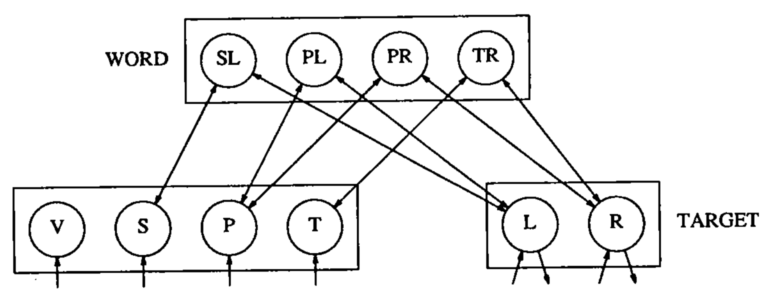


FIG. 2. Network used in the simulation of the IAC model applied to the phonological constraints experiment of Massaro and Cohen (1983). The inhibitory connections between units within the word, context, and target levels are not shown in the network.



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 $\alpha$ is the weight on excitatory connections, $\gamma$ 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, $\Delta act_i$, 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 $\Delta act_i$:

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

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-

parameter SIAC model (assuming 40 trials) to each subject. On a SUN-3/50 with 17.1-MHz-68881 floating point, this model took 269.8 cpu seconds per subject to run compared to 3.7 cpu seconds per subject for the regular FLMP—74 times slower. The best fitting parameter values for the simple SIAC model are given in Table 1. These parameter values require some explanation. The parameter values for the glide reflect the amount of activation of the $R$ target unit. The activation of the $L$ target unit is one minus this value. The parameter value for the context reflects the amount of activation of the specified context unit when that context was presented; all other context units had activation zero. In addition, the amount of activation of a context unit must be considered with respect to the network in Fig. 2. The same activation has different consequences depending on which context unit is activated. Thus, a .5 activation of the context unit $P$ has a very different outcome than a .5 activation of $S$. The target units $L$ and $R$ are activated equally given activation of $P$, whereas only the target unit $L$ received activation given activation of $S$.

Figure 3 shows the observed and predicted results for three typical subjects. The root mean squared deviation (RMSD) values between the observed and predicted data for individual subject fits and the average of the individual subject fits are given in the first line of Table 2. Given the relatively poor fit of the model (.1113 average RMSD), two additional models were fit in order to determine if increasing the number of free parameters would bring the model into line with the results. In a 12-parameter model, an additional parameter was estimated for the standard deviation of the normal noise added to the inputs. Given the large number of iterations required by the model, the parameters from the 11-parameter model were used as starting values for these parameters in the search. The standard deviation of the noise was permitted to vary between 0 and 1. As can be seen in Table 2, this model yielded a small improvement in the fit (average RMSD of .0987), with an average standard deviation

TABLE 1

Best-Fitting Parameters from Simple SIAC Model with 40 Simulated Trials for the Results of the Massaro and Cohen (1983) Experiment 2

| | Glide | | | | | | | Context | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Subject | L | | | | | | R | P | T | S | V |
| 1 | .2600 | .2900 | .3200 | .4100 | .5300 | .9500 | .9500 | .5000 | .7400 | .5000 | .5000 |
| 2 | .3323 | .4008 | .4894 | .5217 | .6533 | .8123 | .9500 | .5000 | .5104 | .9500 | .5000 |
| 3 | .4070 | .5300 | .5900 | .6122 | .7466 | .9200 | .9500 | .5000 | .5000 | .9800 | .5000 |
| 4 | .0500 | .0350 | .2600 | .4211 | .6082 | .6572 | .6823 | .5000 | .9800 | .8568 | .5000 |
| 5 | .0500 | .0800 | .0500 | .4670 | .7190 | .9500 | .9500 | .5000 | .5000 | .5582 | .5000 |
| 6 | .3149 | .3500 | .4421 | .5488 | .7170 | .9500 | .9500 | .5000 | .5000 | .5300 | .5000 |
| 7 | .3800 | .3762 | .4363 | .5508 | .6489 | .9500 | .9500 | .5000 | .9800 | .9800 | .5000 |
| Average | .2563 | .2946 | .3697 | .5045 | .6604 | .8842 | .9118 | .5000 | .6729 | .7650 | .5000 |



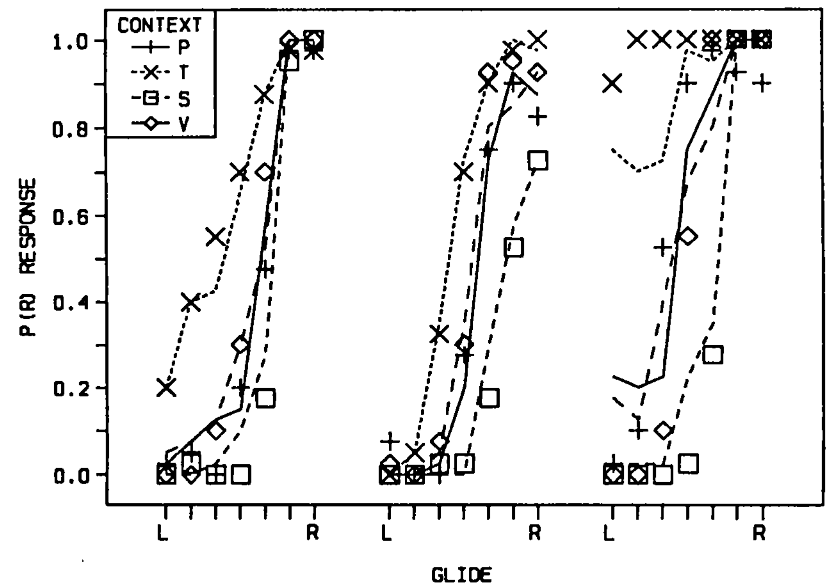


FIG. 3. Observed (points) and predicted (lines) probability of an $r$ response for three typical subjects as a function of the glide $F_3$ onset level and context (after Massaro and Cohen, Experiment 2, 1983). Predictions are for the SIAC model.

parameter at .1263. A full-blown SIAC model was fit to the results by allowing the four control parameters of the IAC process (*estr*, *alpha*, *gamma*, and *decay*) to be additional free parameters for the minimization search. Because these parameters varied, the SIAC results were directly computed rather than using the table lookup method. This 16-parameter model yielded essentially no improvement, with an average RMSD of .0939 (see Table 2).

Given the possibility that the model's optimal parameter values might change when more parameters are added and that the values found for the 11-parameter model forced the predictions of the SIAC models with more parameters to a local minimum, we tested the 16-parameter model with a new set of starting values and a different random seed. Although the parameter values and predictions differed slightly, an ANOVA on the RMSD values indicated that there was no improvement in the goodness of fit (average RMSD of .0999) for the seven subjects, $F(1,6) = 1.83$, $p = .224$.

The number of simulated trials was equated with the number in the actual experiment, and a small number of trials might not give a truly representative sampling from the noise distribution. This might be especially true given the necessarily repetitive use of the same random se-

TABLE 2
The RMSD Values for the Fit of the Individual Subjects and the Average of the Individual Subject Fits for the FLMP and SIAC Models Fit to the Results of Massaro and Cohen (1983) Experiment 2

| Model | NP | NT | S1 | S2 | S3 | S4 | S5 | S6 | S7 | AVE |
|---|---|---|---|---|---|---|---|---|---|---|
| FLMP | 11 | | .0307 | .0484 | .0896 | .0419 | .0306 | .1162 | .0251 | .0546 |
| FLMP | 11 | 40 | .0206 | .0537 | .0896 | .0409 | .0259 | .1062 | .0259 | .0518 |
| FLMP | 11 | 300 | .0312 | .0484 | .0902 | .0451 | .0344 | .1132 | .0250 | .0554 |
| FLMP | 11 | 1000 | .0309 | .0505 | .0871 | .0410 | .0273 | .1170 | .0254 | .0542 |
| SIAC | 11 | 40 | .0634 | .1177 | .1327 | .0489 | .1381 | .1293 | .1487 | .1113 |
| SIAC | 11 | 300 | .0477 | .1083 | .1228 | .0572 | .0912 | .1167 | .1546 | .0998 |
| SIAC | 11 | 1000 | .0462 | .1055 | .1218 | .0527 | .0900 | .1218 | .1543 | .0989 |
| SIAC | 11 | 10000 | .0465 | .1050 | .1261 | .0562 | .0954 | .1205 | .1554 | .1007 |
| SIAC | 12 | 40 | .0528 | .1173 | .1191 | .0489 | .1381 | .1293 | .0852 | .0987 |
| SIAC | 12 | 300 | .0433 | .1064 | .1087 | .0533 | .0922 | .1152 | .0817 | .0858 |
| SIAC | 12 | 1000 | .0429 | .0985 | .1036 | .0489 | .0912 | .1207 | .0765 | .0832 |
| SIAC | 12 | 10000 | .0433 | .0959 | .1051 | .0492 | .0961 | .1173 | .0795 | .0838 |
| SIAC | 16 | 40 | .0520 | .1117 | .1191 | .0480 | .1121 | .1293 | .0852 | .0939 |
| SIAC-INT | 16 | 40 | .0502 | .0813 | .0976 | .0561 | .1013 | .1220 | .0417 | .0786 |

quence in the simulation during the parameter estimation process. Perhaps the luck of the draw might be particularly favorable or unfavorable to a model? The use of a larger number of trials should reveal whether the poor fit of the SIAC model could be due to this potential problem. On the other hand, there could be substantive limitations in the model independent of the number of simulated trials used during the model test.

Accordingly, we also tested the 11- and 12-parameter models with 300, 1000, and 10,000 simulated trials. The simulations were also started with the same random seed in each iteration of the STEPIT fit. Even with this constraint the 10,000 trial fit took about 35 cpu hours per subject on a SUN-3/60 with 20.5 MHz 68881 floating point. Some small improvement was seen in the fits; the RMSDs are given in Table 2. Overall, the RMSD of fit for the 11-parameter model changed from .1113 to .0998 for 300 trials to .0989 for 1000 trials, and .1007 for 10,000 trials. For the 12-parameter model, the overall RMSD decreased from .0987 to .0858 for 300 trials to .0832 for 1000 trials, and slightly increased to .0838 for 10,000 trials. To summarize, we do see a small improvement in the fit with a larger number of trials, but note a severely diminished return for our effort above 300 trials or so. Most importantly, the rather poor fit of the SIAC model cannot be blamed on sampling variability given a small number of simulated trials.

### *SIAC Model with Intrinsic Noise*

Up to this point, the SIAC model tests have assumed variability added to the inputs. Processing itself is deterministic. A second type of SIAC model proposed by McClelland assumes variability added at each processing cycle. We shall call this model the stochastic IAC-Intrinsic Noise (SIAC-INT) model. Given the possibility that this type of model would give a better description of actual results, we tested this model against the results in the same manner. The computation of the SIAC-INT model is similar to that for the prior SIAC except that noise is not added to the external inputs directly but rather is added to the activation of each of the nodes at each step. Equation 9 represents this version of the model and replaces the prior Eq. 3.



$$net_i = \alpha \times exc_i + \gamma \times inh_i + estr \times ext_i + gsd \times RN \tag{9}$$

where $RN$ is a normally distributed random noise value and $gsd$, the Gaussian standard deviation, controls the magnitude of the noise contribution. Rather than using the instantaneous target activations for a BOW decision, running average activations were computed, as in McClelland (in press), according to the formula

$$ract_t = \lambda\ act_t + (1 - \lambda)ract_{t-1} \tag{10}$$

where $ract_t$ is the running average activation at time $t$, $act_t$ is the activation at time $t$, $ract_{t-1}$ is the running average activation at time $t - 1$, and $\lambda$ (set to .05 in current simulations) controls the degree of averaging. A third feature of our SIAC-INT simulations was that the external inputs could cover a wider range (at the suggestion of McClelland, personal communication). We allowed these values to range from $-.5$ to 1.5. A 16-parameter SIAC-INT model (estimating the same parameters as the 16-parameter SIAC model) was fit, yielding a mean RMSD of .0786, a nonsignificant improvement [$F(1,6) = 5.251$, $p = .060$] over the 16-parameter SIAC model. Table 3 gives the best-fitting parameter values for the model.

We now describe and test an alternative to interactive activation, one we believe is more predictive and parsimonious.

### *Fuzzy Logical Model of Perception*

A critical assumption of the application of the FLMP in the phonological constraints study is that the featural information from the glide and the phonological context provide *independent* sources of information. Even with this constraint, however, it is important to demonstrate that adding context can improve performance—not just change bias—relative to the case in which only stimulus information is presented. Consider our identification task in which a set of seven syllables along a /li/–/ri/ continuum were factorially combined with four different initial consonant contexts /p/, /t/, /s/, or /v/. We assume that subjects adopt the prototypes $R$ and $L$ in the task, and evaluate and integrate the two sources of information with respect to these prototypes. The stimulus featural information in the glide $i$ supporting the $R$ prototype can be represented by the

TABLE 3
Best-Fitting Parameters from 16-Parameter SIAC-INT Model with 40 Simulated Trials for the Results of the Massaro and Cohen (1983) Experiment 2

| Subject | Glide | | | | | | | Context | | | | Control | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | L | | | | | | R | P | T | S | V | estr | α | γ | decay | gsd |
| 1 | .2300 | .2816 | .3502 | .4245 | .5576 | .9500 | .9500 | .4100 | .7040 | .7171 | .5000 | .0547 | .1633 | .1300 | .1000 | .0385 |
| 2 | .3923 | .4008 | .4986 | .5324 | .6628 | .8753 | .9200 | .4700 | .5254 | .9800 | .5000 | .1300 | .1900 | .1385 | .1000 | .1414 |
| 3 | .4356 | .5324 | .6110 | .6882 | .8534 | .9950 | .9500 | .4700 | .5900 | .9800 | .5000 | .1002 | .1822 | .1169 | .1000 | .0857 |
| 4 | .0500 | .0350 | .2600 | .4211 | .6082 | .6272 | .6823 | .5000 | .8900 | .8568 | .5000 | .1300 | .1576 | .1196 | .0993 | .1208 |
| 5 | .0500 | .0800 | .0500 | .4670 | .6350 | .9500 | .9500 | .5000 | .0500 | .7682 | .5000 | .1300 | .1300 | .0948 | .1000 | .1414 |
| 6 | .3149 | .3339 | .4421 | .5788 | .8070 | .9800 | .9800 | .5000 | .5600 | .7300 | .5000 | .1000 | .1831 | .1009 | .1000 | .1120 |
| 7 | .3816 | .3878 | .4693 | .5395 | .6416 | .9500 | .9500 | .4700 | .9800 | .9200 | .5000 | .0985 | .1330 | .0970 | .1088 | .0445 |
| Average | .2649 | .2931 | .3830 | .5216 | .6808 | .9039 | .9118 | .4743 | .6142 | .8503 | .5000 | .1062 | .1627 | .1140 | .1012 | .0978 |





truth value $f_i$, and $(1 - f_i)$ specifies the stimulus support for $L$. The value $c_j$ represents how much context $j$ supports the prototype $R$, and the degree to which the phonological context supports the prototype $L$ is indexed by $(1 - c_j)$. Truth values index the degree of support of each source of information for each alternative. These values range between 0 and 1 reflecting no support to complete support, with .5 corresponding to a neutral support in a two-alternative task.

Given two independent sources of information, the total degree of match with the prototypes $R$ and $L$ is determined by integrating these two sources. Following Goguen (1969), feature integration involves a multiplicative combination of two truth values (Oden & Massaro, 1978). Therefore, the degree of match to $R$ and $L$ for a given syllable can be represented by

$$R = f_i \times c_j \qquad (11)$$

$$L = (1 - f_i) \times (1 - c_j) \qquad (12)$$

The decision operation maps these outcomes of integration into responses by way of a relative goodness rule (RGR) (Luce, 1959, 1963; Massaro & Friedman, 1990; Shepard, 1957). The probability of an /r/ response given test stimulus $S_{ij}$ is predicted to be

$$P(r|S_{ij}) = \frac{f_i c_j}{f_i c_j + (1 - f_i)(1 - c_j)} \qquad (13)$$

In fitting our experimental data to the FLMP model there are—as with the simple SIAC model—11 parameters. These include four parameters giving the r-ness of the context $c_j$ and seven parameters giving the r-ness of the glide $f_i$. The best fitting parameter values for the FLMP model are given in Table 4. The parameter values $f_i$ and $c_j$ correspond to the amount of support for the alternative /r/ given the level of glide and context,

TABLE 4
Best-Fitting Parameters for the Ordinary FLMP for the Results of the Massaro and Cohen (1983) Experiment 2

| Subject | Glide | | | | | | | Context | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | L | | | | | | R | P | T | S | V |
| 1 | .0177 | .0437 | .0753 | .1794 | .5369 | .9905 | .9991 | .4502 | .9307 | .1471 | .6417 |
| 2 | .1430 | .2536 | .4064 | .5088 | .8523 | .9862 | .9909 | .7835 | .7529 | .0725 | .4730 |
| 3 | .0754 | .2227 | .5937 | .9558 | .9906 | .9990 | .9999 | .7395 | .9617 | .0192 | .9215 |
| 4 | .0031 | .0107 | .0896 | .3434 | .8476 | .9545 | .9780 | .3522 | .8199 | .0488 | .4932 |
| 5 | .0016 | .0200 | .0991 | .3453 | .8888 | .9877 | .9998 | .8140 | .2596 | .2084 | .5802 |
| 6 | .0606 | .1103 | .2372 | .6704 | .9677 | .9997 | .9997 | .6570 | .8617 | .0776 | .6019 |
| 7 | .0009 | .0065 | .0596 | .3806 | .9668 | .9999 | .9999 | .9449 | .9999 | .0129 | .6622 |
| Average | .0432 | .0954 | .2230 | .4834 | .8644 | .9882 | .9953 | .6773 | .7981 | .0838 | .6248 |

respectively. Figure 4 shows the close agreement of the observed and predicted results of the FLMP model for the same three subjects shown in Fig. 3. The average of the RMSDs of fits of the seven subjects was .0546. Table 2 gives the RMSDs for individual subjects and the average of the individual fits.

To be fair to the SIAC model, which is a simulation over a given number of trials, we evaluated a simulation version of the FLMP model which also used Eq. 13 as a basis for its predictions. This 11-parameter model, which started with the parameter estimates from the ordinary FLMP, ran a series of 40 simulated trials for each condition. On each trial, a uniformly distributed random number was generated. If it was less than or equal to $P(r|S_{ij})$, as computed by Eq. 13, then the simulation $P(r|S_{ij})$ was incremented by 1/40. The random sequence was restarted at the beginning of the prediction calculation of each set of results with the same seed as used in the SIAC model fit. On a SUN-3/50 with 17.1-MHz-68881 floating point, this model took 55.7 cpu seconds per subject to run compared to 3.7 cpu seconds per subject for the regular FLMP—21 times slower. Table 2 gives the individual and mean RMSDs for the model fit. The overall fit of this model was .0518, actually a small (but nonsignificant) improvement over the ordinary FLMP. Decreasing the number of

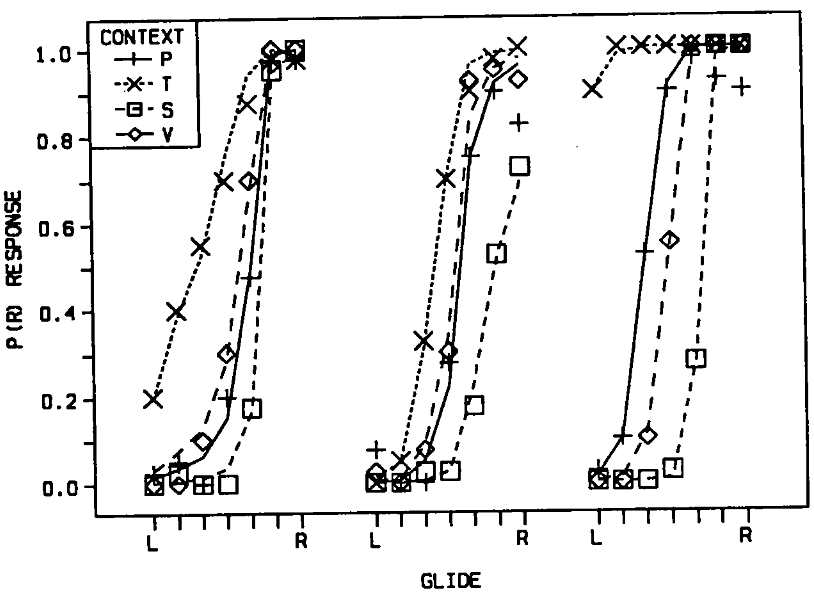


FIG. 4. Observed (points) and predicted (lines) probability of an *r* response for three typical subjects as a function of the glide $F_3$ onset level and context (after Massaro and Cohen, Experiment 2, 1983). Predictions are for the ordinary FLMP.



trials can in some cases increase rather than decrease the goodness-of-fit of a model. This can occur when the number of trials is equivalent to the number actually used in the experiment. The granularity of the prediction, at 1/40, now exactly corresponds to the possible observed values, allowing predictions which were only nearby to migrate to the actual observed values. As with the SIAC model, we also evaluated the simulation FLMP with greater numbers of trials. The fits for each subject are given in Table 2. As can be seen, the mean fits for 300 trials (RMSD = .0554), 1000 trials (RMSD = .0542), and 10,000 trials (RMSD = .0546) were essentially the same as for the ordinary FLMP.

Using ANOVAs, the RMSDs from the 11-parameter, 40-trial simulation FLMP model were compared with the RMSDs from the SIAC 40 trial models. All four SIAC models—the 11-parameter [$F(1,6) = 13.093$, $p = .011$], the 12-parameter [$F(1,6) = 12.633$, $p = .012$], the 16-parameter [$F(1,6) = 17.091$, $p = .007$], and the 16-parameter SIAC-INT [$F(1,6) = 9.709$, $p = .020$] models—provided an inferior fit relative to that of the simulation FLMP. As can be seen by the RMSD values and the mismatch between the observations and predictions in Fig. 3, the SIAC model cannot describe these results as accurately as the FLMP.

The parameter values in Tables 1 and 3 show one possible reason why the SIAC model fit the results more poorly than the FLMP. The parameter values for the /p/ and /v/ context did not move from .5 for the specific architecture of the SIAC model shown in Fig. 2. As pointed out by Stephen Kitzis (personal communication) and McClelland (personal communication), the network shown in Fig. 2 has two limitations. First, the /v/ context does not activate any units in the word layer and cannot support one target alternative more than the other, having only an inhibitory effect within the context layer. Second, the /p/ context must support /r/ as much as /l/ because the /p/ context unit is connected to a word unit containing /r/ and a word unit containing /l/. Thus, it seems likely that the poor fit of the SIAC models relative to the FLMP is due at least in part to the limitations of the network developed by McClelland (1991) and extended by us. An architecture is required that would allow free parameters for the /p/ and /v/ contexts. An obvious network to achieve this flexibility is to add word units /vl/, /vr/, /sr/, and /tl/ to the word level in Fig. 2, for a total of 14 nodes, using 16 parameters. For this model, the connection weights between a word unit and appropriate context and target units would be free parameters rather than fixed at one. However, this model requires an immense amount of cpu time to test against actual results. The test of a single subject required over 100 hours of computation. In another paper (Cohen & Massaro, in press), we develop a new simpler six-node network that brings the asymptotic predictions of a SIAC model very close to those of the FLMP. This network was used for modeling the integration of multiple sources of featural information in letter perception. In the network, complementary information supporting the letters "Q" or "G" is represented in three layers of two nodes each (angle features, gap features, and memory). The activation used to generate a response is taken from the outputs of the nodes at the memory layer.

In conclusion, we accept that SIAC models with a BOW decision rule can predict results consistent with the independence of stimulus information and context. What is important to remember, however, is that the independence prediction is *not* directly due to interactive activation. In fact, the independence prediction appears to occur in spite of it. In other work, we have demonstrated that the independence prediction always occurs regardless of the underlying activation functions if noise is added before or during perceptual processing and a BOW decision rule is used (Massaro & Cohen, 1989). McClelland's (1991) modification of the IAC model allows the predictions of the model to no longer be constrained by the underlying activation functions. The prediction of the SIAC models, therefore, cannot be taken as direct evidence for the interactive processing assumed by the models. This outcome is particularly disturbing because connectionist models are claimed to account for the microstructure of cognition. Even though we devised a network to account for the observed results, we argue for the FLMP on the basis of parsimony. Relative to the FLMP, the SIAC models require a larger number of parameters that must be set at some arbitrary values or estimated from the observed results. In the next section, the nature of the interactive activation in SIAC models is examined more closely in order to gain some insight into the underlying processing that is assumed and how it compares to other noninteractive models such as the FLMP.

### *Outcome of Evaluation and Integration Processes*

Following the logic of Massaro and Friedman (1990), it is helpful if not essential to analyze perceptual recognition in terms of three stages of information processing: evaluation, integration, and decision. Evaluation is defined as the analysis of each source of information. Integration combines the outputs made available by the evaluation process. Decision maps the outcome of either evaluation or integration into a response. This framework was used to describe a feed forward network model (Massaro & Friedman, 1990) and to describe a SIAC model (Cohen & Massaro, in press). In contrast to feedforward models, there is an apparent blurring between evaluation and integration in the SIAC model. In the SIAC model, there is activation of the Target (phoneme) and Context units by external inputs (see Fig. 2). These lower-level activations are fed forward to a layer of Word units, which are also connected in top-down fashion to some units at the lower layer. Processing continues through a sequence of time steps (cycles) in which each unit updates its activation value by summing the weighted activations of all units feeding into the unit (McClelland, in press, Equation 1). Given interactive activation, the representation of a unit at the Target layer might be influenced by context, because the unit also integrates information from the layer of Word units. This is what we mean by nonindependence.



In the three-stage framework used by Massaro and Friedman (1990), the activations of the target units get mapped into a response by way of a decision process. Decision consists of either a relative goodness rule (RGR) in the original model or a best one wins (BOW) in the revised model. The RGR predicts that the probability of a response is equal to the goodness of match of the alternative relative to the sum of the goodness of match values of all relevant alternatives in the task. The BOW decision rule always chooses the alternative corresponding to the target unit with the highest activation. Before analyzing the activations and the predictions of the SIAC models in this framework, however, it is worthwhile to analyze the FLMP because its predictions were viewed as a benchmark for McClelland's revision of the SIAC model.

In the FLMP, evaluation makes available a continuous truth value representing each source of information. Integration involves a multiplicative combination of these truth values with respect to prototype definitions of the response alternatives. The outcome of integration gives goodness-of-match values with each of the alternative prototypes. Figure 5 shows these values for the FLMP, applied to the phonological constraints study, as a function of feature values varying between 0 and 1 in steps of .05 for three different context values supporting /r/: .25, .5, and .75. As can be seen in the figure, the goodness-of-match values change linearly with a linear change in target feature values. The rate of change depends on the context. The context value can be seen directly when the input feature value is .5—completely ambiguous or neutral. With a completely neutral context /p/, the goodness-of-match to the alternative /r/ is equal to one-half of the input feature value. With the context /s/, the goodness-of-match with alternative /l/ is .75 times the input feature value whereas the goodness-of-match with alternative /r/ is $(1 - .75) = .25$ times the input feature value.

We stress that since activation is the currency of SIAC models, it is the changes in activation that are important, not simply the probability of a response. McClelland (1991) did not present any activation functions of the SIAC model—only the predicted response functions given an RGR or BOW decision were computed. This omission seems particularly paradoxical because network models putatively illuminate the microstructure of processing and do not simply predict stimulus-response functions. It

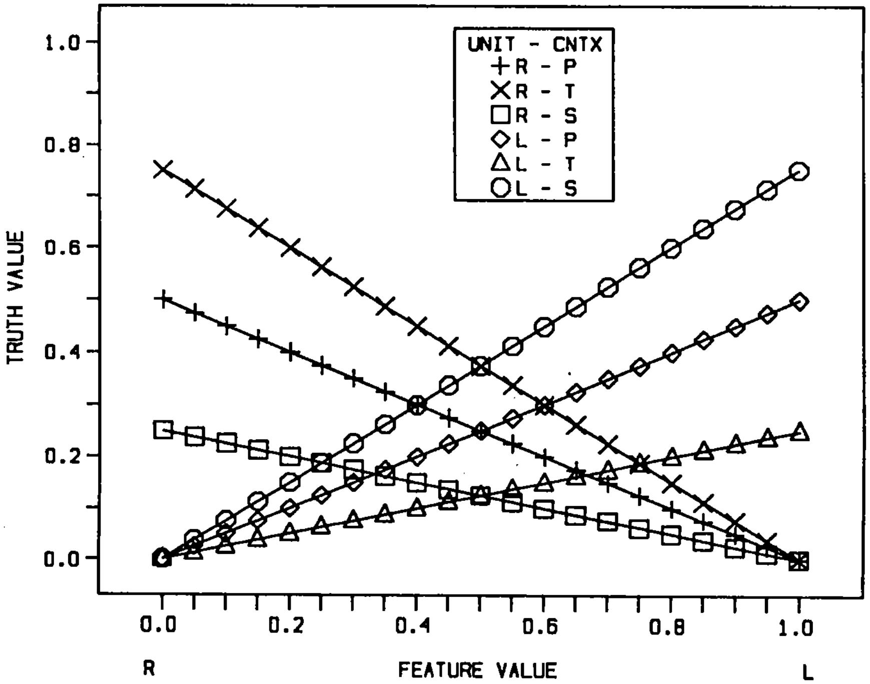


FIG. 5. Truth (Goodness-of-match) values for the R and L units, resulting from the integration of the two sources of information feature value and context (CNTX) according to the FLMP.

should be valuable to generate and analyze the asymptotic activation functions of SIAC models.

Figure 6 plots the activation values for the predictions of the simple SIAC model and network used by McClelland (1991). These activation values correspond to those of the /r/ and /l/ units after 60 cycles through the network. The context feature value was normally 0 and set to .5 when the corresponding context was present. The target feature value was varied between 0 and 1 in steps of .05. By looking at these activations and contrasting them with the analogous truth values of the FLMP, we can better observe the consequences of the interactive algorithm. Although the SIAC model is an interactive and potentially complex system, its predicted activation functions are fairly easy to understand. The figure shows that the activation is not a linear function of the feature value. The nonlinearity has two components. First, there is the sharpening of each activation curve at the crossover point of the /l/ and /r/ curves for a given context. Second, the rate of change in activation as a function of a constant change in feature value is greater when the context and feature value



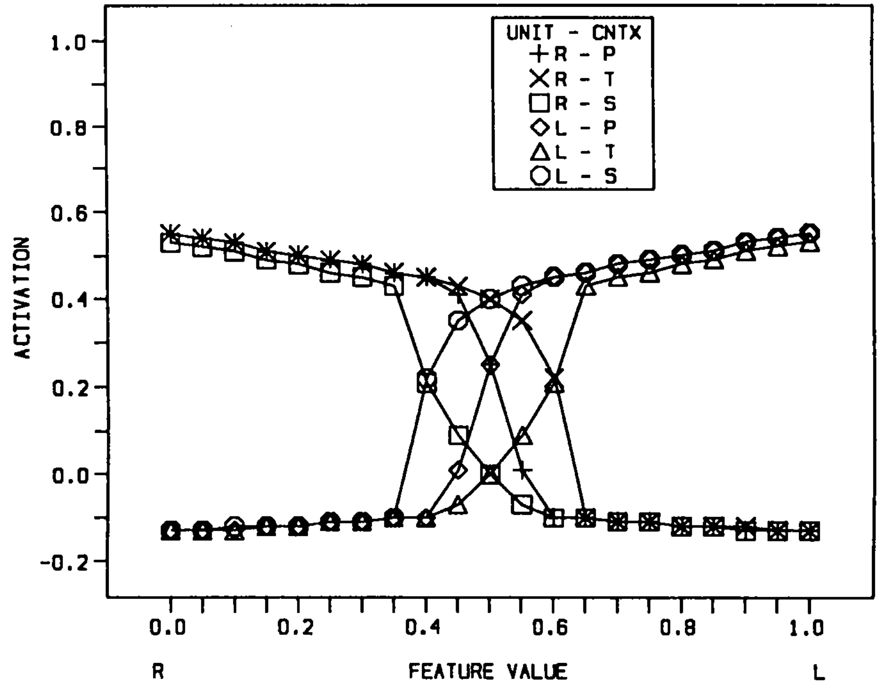


FIG. 6. Activation values for the R and L units, resulting from the integration of the two sources of information feature value and context (CNTX), according to the SIAC model.

are inconsistent with one another than when they are consistent with one another. That is, the rate of change is greater when the parameter values of the two sources of information favor different responses. These two forms of nonlinearity appear to be direct consequences of the combination of inhibitory connections within a layer of units and two-way connections between units embedded in different layers.

It might be valuable to compare the outcome of the SIAC model in Fig. 6 with that of the FLMP in Fig. 5. In the FLMP, the difference between adjacent levels of the feature value is constant for a given context. Given the context /s/, the difference between features values .3 and .4 is equal to the difference between .6 and .7. The activations of the SIAC model shown in Fig. 6, on the other hand, indicate that the difference between .3 and .4 is significantly larger than the difference between .6 and .7. Given the context /s/, the difference between .3 and .4 on the /r/ side of the feature continuum is larger than the difference between .6 and .7 on the /l/ side of the continuum. More generally for the IAC model, the difference between two adjacent feature values is larger on the side of the feature continuum that conflicts with the context information. The RGR preserves this difference in the response predictions, as demonstrated in Massaro (1989) and McClelland (1991). The BOW decision rule with noise added to the inputs or their processing, on the other hand, cancels out this difference. Thus the predicted responses no longer reflect the activation functions; they only transmit information about the location of the crossover of the activations of the /l/ and /r/ units (Massaro & Cohen, 1989).

Interactive activation networks differ from simple feedforward networks because they have two-way connections between units. If the units are taken to represent sources of information, then SIAC models do not maintain a single feed-forward flow through the processes of evaluation, integration, and decision. The reason is that integration feeds back to evaluation and changes the outcome of evaluation. Although noise added to the inputs or at each processing cycle and a BOW decision rule allow the classic result of independence to be simulated at asymptote, it is important to acknowledge the nonlinearity of the underlying activations. Therefore, the SIAC model will have difficulty in predicting the classic result when a BOW decision cannot be used. For example, the SIAC model would have difficulty predicting the classic result—independence of stimulus information and context—when subjects rate the degree to which the stimulus is /l/ or /r/. In the next sections, we extend the contrast between the SIAC models and FLMP to results on the dynamics of perceptual processing.

## PREDICTING DYNAMIC BEHAVIOR

One of the attractions of SIAC is that it putatively accounts for the dynamics of perception. Thus, it is important to contrast the SIAC and the FLMP in this domain. McClelland (1991) concluded that the SIAC is to be preferred because it and not the FLMP can account for the dynamics of perception. Our tests of these models, however, indicate otherwise.

### *The Dynamic FLMP*

Although the FLMP has not been tested directly against results on the time course of perceptual processing, there is a natural extension of the model when it is combined with another model of the time course of processing (Massaro, 1970a, 1975a). This model has received support in a backward recognition masking (BRM) task. In the backward masking paradigm, a brief target stimulus is presented followed after a variable interstimulus interval (ISI) by a second stimulus (the mask).

Consider the three masking functions shown in Fig. 7, representing the performance of three different young adults in the first study of auditory backward recognition masking (Massaro, 1970b). The two test alternatives were brief tones (20 ms) differing in frequency and the subjects identified the test tone as high or low in pitch. The curves are plotted in $d'$ values to eliminate any large contribution of decision bias and to provide a measure of performance that can be justified in terms of a formal description of the processes involved. The $d'$ measure can be conceptualized as the distance between means of two distributions of events corresponding to the two different types of test trials. The masking function reflects changes in this distance as a function of processing time before the onset of the masking tone.



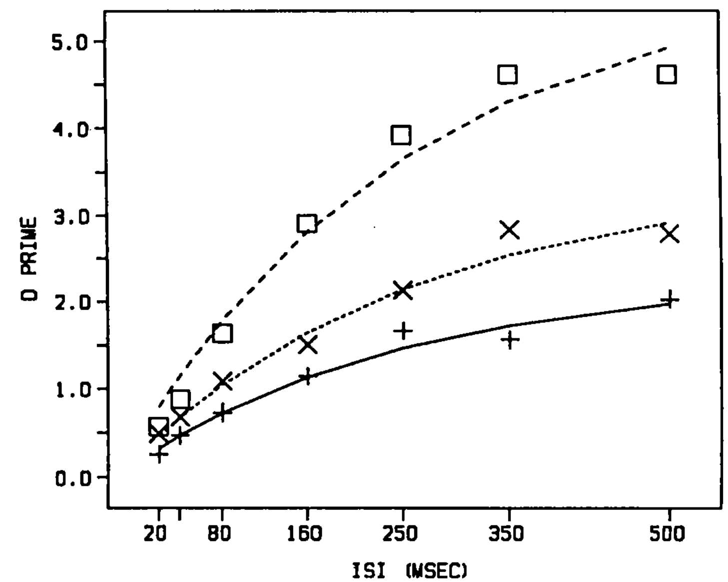


FIG. 7. Observed (points) and predicted (lines) recognition accuracy (measured in $d'$) as a function of the interstimulus interval (ISI) between the test and masking tones (after Massaro, 1970a).

The backward masking results have been explained within the framework of a model of auditory information processing in which the target sound is transduced by the listener's sensory system and stored in a preperceptual auditory store which briefly holds a single auditory event (Massaro, 1972, 1975a). Perceptual processing of the sound involves resolving the features of the sound to produce a synthesized representation of some segment. A second sound replaces the first in the preperceptual auditory store and terminates any further reliable processing of the first sound. Because of the transient nature of preperceptual memory, and backward masking when a second sound occurs before resolution of the first sound is complete, the duration of preperceptual memory as well as the rate of processing places a limit on how much can be perceived.

The amount of time that the target information is available in preperceptual memory can be carefully controlled by manipulating the duration of the ISI. The accuracy of target identification increases as ISI lengthens (Massaro, 1970b). The rate of the improvement reflects the rate at which the stimulus information is processed. Performance typically asymptotes at an ISI of roughly 250 ms and this interval is believed to reflect the duration of the preperceptual auditory store (Cowan, 1984; Kallman & Massaro, 1983). That is, the mask no longer affects performance because the preperceptual trace is no longer available for processing.

According to the theory, subjects have continuous information about the test stimulus and this information accumulates gradually with the processing time available before the onset of the mask. Perceptual recognition cannot be considered to be all-or-none or accurate-inaccurate at any time during the processing interval. Given noise in the system, however, identification accuracy is probabilistic and increases systematically with increases in processing time. Analogous to the RGR, responses can be considered to be probabilistic because subjects match their response probabilities to the relative goodness of match of the response alternatives. The masking stimulus serves to terminate any additional processing of the test, but it does not work retroactively. Masking does not reduce the amount of processing that has occurred before the occurrence of the mask; it can only preclude further processing. It is important to distinguish between the potential information given unlimited processing time and the rate of reaching that level of information. Increasing the discriminability between the test alternatives should increase the asymptote but not necessarily the rate of the masking function.

The functions of $d'$ in Fig. 7 can be described accurately by a negatively accelerated exponential growth function of processing time,

$$d' = \alpha(1 - e^{-\theta t}) \tag{14}$$

The parameter $\alpha$ is the asymptote of the function and $\theta$ is the rate of growth to the asymptote. The function can be conceptualized as representing a process that resolves some fixed proportion of the potential information that remains to be resolved per unit of time. The same increment in processing time results in a larger absolute improvement in performance early relative to late in the processing interval.

A reasonable assumption is that feature evaluation would follow the same negatively accelerating growth function found in backward recognition masking tasks. Early in featural evaluation, the perceiver would have some information about each feature (dimension), but the information would not be sufficient to inform the perceiver about the identity of the stimulus. Integration of the separate features (dimensions) would be updated continuously as the featural information is being evaluated. Similarly, decision (and thus response selection) could occur at any time after the stimulus presentation. For example, a response could be initiated before sufficient information is accumulated—as might occur in speed-accuracy experiments.



Following the theoretical analysis of backward masking, a masking stimulus would terminate any additional processing of the test stimulus. While the masking function given in Eq. 14 represents the time course of only a single feature's evaluation, this dynamic model can be combined with the FLMP to describe how multiple sources of information are evaluated and integrated over time. There is evidence for the parallel evaluation of multiple sources of information in the BRM task. Moore and Massaro (1973), for example, asked subjects to identify both loudness and timbre in the BRM task. On each trial, the subject was cued to identify either the loudness (loud or soft), the timbre (dull or sharp), or both of these dimensions of the test tone. Typical backward masking functions were found under each of the attention conditions. In addition, the subjects were able to identify the two dimensions of the test tone (loudness and timbre) about as accurately as one.

Although we illustrate the three stages—evaluation, integration, and decision—as discrete, in reality they would operate continuously. Evaluation of a source of information would follow Eq. 14. The truth value $T_{xA}$ of a source of information $x$ supporting a given alternative $A$ changes over time $t$ toward some asymptotic value $\alpha$ from an ambiguous initial value (.5). This change can be described as the sum of a negatively accelerated transition from 0 to the asymptotic stimulus value $\alpha$ and a negatively accelerated transition from the initial value .5 going to 0.

$$T_{xA}(t) = \alpha(1 - e^{-\theta t}) + .5(e^{-\theta t}). \tag{15}$$

The output from evaluation would be fed continuously to the integration process—which would operate in the same manner as assumed in the FLMP. Integration would be fed forward to decision which would compute the relative goodness of match of the alternatives. Given the concern with the dynamics of processing, an additional process must be implemented. This process would determine when a subject would actually initiate a response in the task. In most identification tasks with unlimited response time, it seems reasonable to assume that the subject waits until evaluation of the sources of information is near asymptote. With limited processing time, the decision system would initiate a response when no additional information is being accumulated. Thus, the decision system could maintain some running memory of the change in relative goodness values, and if this change is less than some minimum in a given time period, then a response could be initiated. In tasks with speeded responses, the decision system would simply initiate a response at the required time based on the relative goodness values at that time. We now test this dynamic FLMP against empirical results.

*Context Effects and Backward Masking*

In an experiment reported by Massaro (1979), a reader was asked to read lowercase letter strings with an ambiguous test letter between *c* and *e*. It is possible to gradually transform the *c* into an *e* by extending the horizontal bar. The interpretation of the ambiguous letter differs in the different letter strings. To the extent the bar is long, there is good visual information for an *e* and poor visual information for a *c*. Now consider the letter presented as the first letter in the context *-oin* and the context *-dit*. Only *c* is orthographically admissible in the first context since the three consecutive vowels *eoi* violate English orthography. Only *e* is admissible in the second context since the initial cluster *cd* is an inadmissible English pattern. In this case, the context *-oin* favors *c*, whereas the context *-dit* favors *e*. The context *-tsa* and *-ast* can be considered to favor neither *e* nor *c*. The first remains an inadmissible context whether *e* or *c* is present, and the second is orthographically admissible for both *e* and *c*.

The experiment factorially combined six levels of visual information with these four levels of orthographic context, giving a total of 24 experimental conditions. The bar length of the letter took on six values going from a prototypical *c* to a prototypical *e*. The test letter was presented at each of the four letter positions in each of the four contexts. The test string was presented for a short duration followed after some short interval by a masking stimulus composed of random letter features. Subjects were instructed to identify the test letter on the basis of what they saw. The results of the experimental test are shown in Fig. 8. As can be seen in the figure, both the test letter and the context influenced performance in the expected direction. Further, the effect of context is larger for the more ambiguous test letters along the stimulus continuum.

This study also evaluated context effects as a function of processing time controlled by backward masking. The test stimulus was presented for 30 ms. Four masking interstimulus intervals (5, 40, 95, or 210 ms) were tested in that task at each of the other experimental conditions. The points in Fig. 10 show the probability of an *e* response as a function of the bar length of the test letter and the four contexts at each of the four masking intervals. Each point represents data from 176 trials (16 observations from each of 11 subjects). As can be seen in the figure, performance was more chaotic at the short masking intervals. That is, less processing time leads to less orderly behavior—as expected from research on the time course of perceptual processing. Even for unambiguous test letters, subjects did not make consistent identification judgments at short masking intervals. According to perceptual processing theory, there was not sufficient time for feature evaluation and integration to take place before the onset of the masking stimulus.



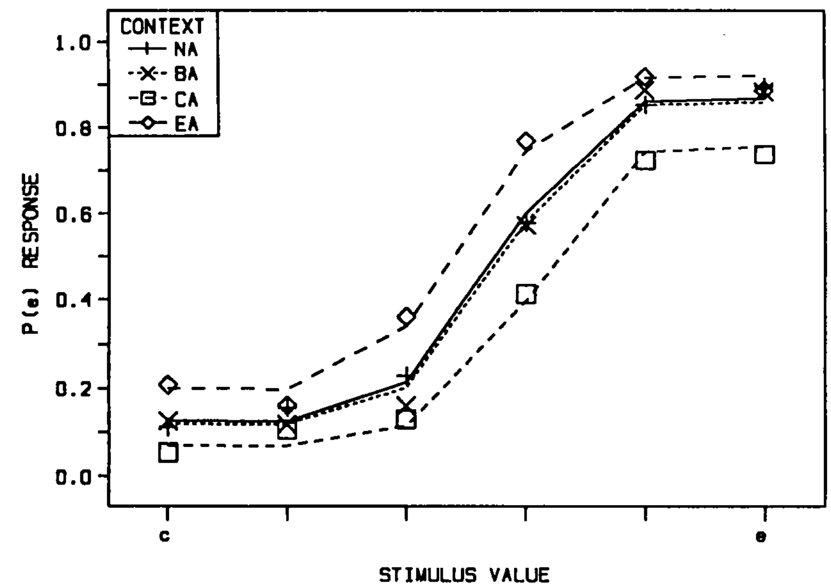


FIG. 8. Probability of *e* identifications as a function of the stimulus value (bar length) of the test letter and the orthographic context (after Massaro, 1979).

Both the test letter and the context influenced performance at all masking intervals. The effect of test letter was attenuated at the short relative to the long processing time. That is, the identification functions covered a larger range across the *e–c* continuum with increases in processing time. Context has a significant effect at all masking intervals. In fact, the context effect was larger for the unambiguous test letters at the short than at the longer masking intervals. This result follows naturally from the tradeoff between stimulus information and context in the FLMP. Context has a smaller influence to the extent the stimulus information is unambiguous.

Given the four masking intervals in the task it is possible to describe performance in terms of the change in featural information $F$ and orthographic context $C$ across the four masking intervals. Implementing Eq. 15 for both context and featural information, the change in featural information can be described by

$$F = \alpha_{Fi}(1 - e^{-\theta t}) + .5(e^{-\theta t}) \quad (16)$$

and the change in contextual information $C$ by

$$C = \alpha_{C_j}(1 - e^{-\theta t}) + .5(e^{-\theta t}) \quad (17)$$

for the processing time $t$ (the sum of the masking interval and the 30-ms stimulus duration).

In fitting the FLMP to the group data, 11 parameters were used: six $\alpha_{Fi}$ values for the six levels going from $c$ to $e$, four $\alpha_{C_j}$ values for the four contexts, and one additional parameter for $\theta$. Figure 9 shows how the predicted feature values (left panel) and context values (right panel) changed with processing time. The points in Fig. 9 give the predicted values that are inserted in the integration and decision algorithm for each of the 96 conditions. The predicted probability of an $e$ response is therefore computed in an analogous fashion to Eq. 13 in the phonological constraints experiment. Figure 10 gives the predictions of the FLMP (lines) along with the observed (points) data. The RMSD between the observed and predicted points was .0501. Table 5 gives the parameter values for the fit of the model.

As in the fit of the phonological constraints data, we evaluated a simulation version of the FLMP model. This model, also using 11 parameters, started with the parameter estimates from the ordinary FLMP, and ran a series of 176 simulated trials for each condition. On each trial, a uniformly distributed random number was generated. If it was less than or equal to the probability of an $e$ response as predicted by the FLMP, then

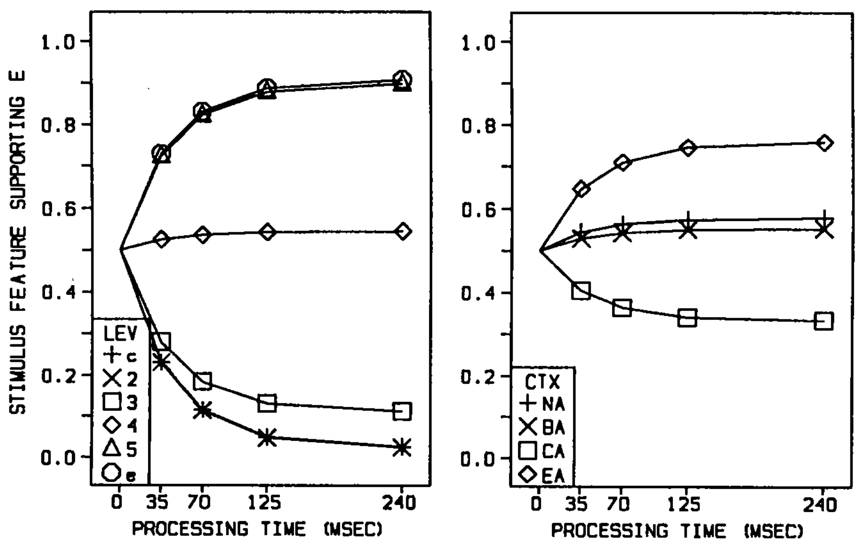


FIG. 9. The development with processing time of feature values supporting alternative $e$ for the dynamic FLMP model. The left panel shows the stimulus feature value as a function of processing time with e–c level (LEV) as the curve parameter. The right panel shows the context feature value as a function of processing time with orthographic context (CTX) as the curve parameter.



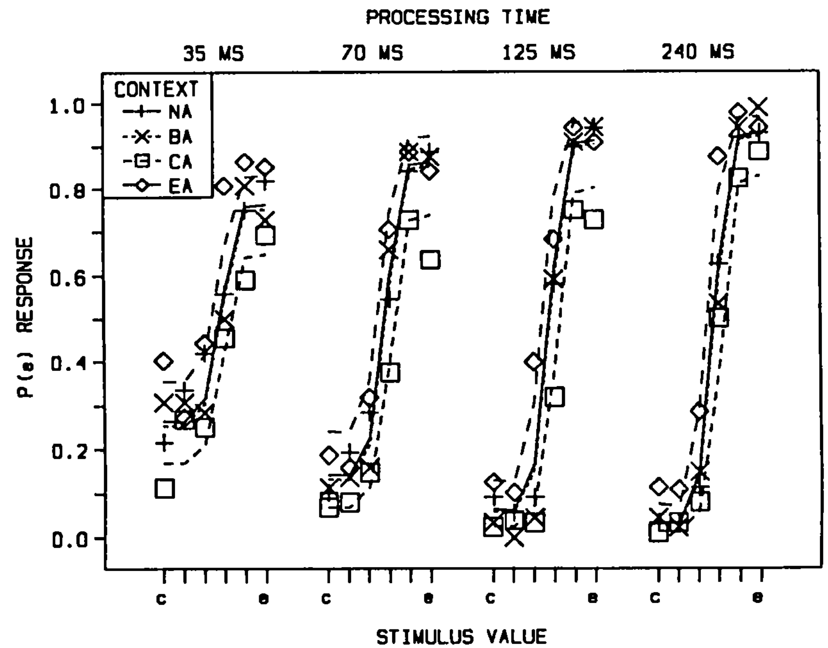


FIG. 10. Observed (points) and predicted (lines) probability of $e$ identifications as a function of the bar length of the test letter, the orthographic context, and the processing interval between the onset of the test stimulus and the onset of the masking stimulus for the dynamic FLMP model (results after Massaro, 1979).

the simulation predicted $P(e/S_{ij})$ was incremented by 1/176. The overall fit of this model was .0561, a small decrement in the goodness of fit relative to the ordinary FLMP.

We also fit a variety of SIAC models to the data. The topology of the SIAC network used was the same as that for the phonological constraints experiment discussed above and shown in Fig. 2. The new network, shown in Fig. 11, differs only in the labels attached to the units. In this case, the word units labelled *EAB* (*E admissible both*) and *CAB* (*C admissible both*) correspond to /pr/ and /pl/. The word units labelled *EAO* (*E admissible only*) and *CAO* (*C admissible only*) correspond to /tr/ and /sl/. The simplest model requires 11 parameters: six $e$ target value inputs (with the $c$ target receiving the additive complement), four context inputs (with the nonselected contexts set to 0), and a parameter $\theta$ which multiplicatively translates processing time (in ms) into the number of cycles used in the SIAC process. As assumed in the fit of the dynamic FLMP, a reasonable assumption is that the mask would terminate further processing of the test—an assumption compatible with Rumelhart and McClelland's (1982) implementation of the original IAC model. As with the initial phonological constraints SIAC model, the SIAC control parameters *estr*,

TABLE 5

Best-Fitting Parameters for FLMP, SIAC, and Boltzmann Models for the Results of Massaro (1979)

| Model | SIAC | SIAC | SIAC-INT | FLMP | SFLMP | FLMP-$\theta$ | FLMP-D | FLMP-C | BOLTZ |
|---|---|---|---|---|---|---|---|---|---|
| Parameters | 16 | 11 | 16 | 11 | 11 | 12 | 12 | 9 | 15 |
| RMSD | .0953 | .1135 | .0688 | .0501 | .0561 | .0500 | .0501 | .0508 | .0748 |
| LL | −839 | −1005 | −500 | −385 | −436 | −386 | −386 | −388 | −583 |
| AIC | 1710 | 2032 | 1032 | 792 | 894 | 795 | 796 | 794 | 1181 |
| Stimulus | | | | | | | | | |
| c | .3222 | .3348 | .1506 | .0234 | .0234 | .0245 | .0234 | −.0334 | −.7701 |
| 2 | .3337 | .3246 | .1288 | .0220 | .0164 | .0231 | .0221 | .0319 | −.7130 |
| 3 | .3908 | .3947 | .2500 | .1094 | .1094 | .1125 | .1094 | .1385 | −.4824 |
| 4 | .5207 | .5211 | .5556 | .5441 | .5138 | .5468 | .5440 | .6071 | .0877 |
| 5 | .6583 | .6661 | .8402 | .8982 | .9004 | .8983 | .8981 | .9291 | .6837 |
| e | .6417 | .6541 | .8349 | .9078 | .9111 | .9077 | .9078 | .9379 | .5813 |
| Context | | | | | | | | | |
| NA | .5000 | .5000 | .4400 | .5781 | .5715 | .5730 | .5781 | .5000* | .1554 |
| BA | .9699 | .9999 | .0999 | .5525 | .5585 | .5484 | .5525 | .5000* | 2.9174 |
| CA | .9865 | .9017 | 1.4425 | .3329 | .3292 | .3379 | .3329 | .2767 | 2.9939 |
| EA | .4451 | .4901 | 1.3396 | .7599 | .7599 | .7451 | .7600 | .7156 | 2.9490 |
| $\theta$ | .5500 | .0965 | .2252 | .0237 | .0237 | .0238 | .0237 | .0225 | — |
| $\theta_c$ | — | — | — | — | .0282 | — | — | — | — |
| gsd | .1398 | .1414* | .1854 | — | — | — | — | — | — |
| estr | .5752 | .1000* | .1180 | — | — | — | — | — | −.4199 |
| istr | — | — | — | — | — | — | — | — | .3099 |
| $\alpha$ | .1464 | .1000* | .1601 | — | — | — | — | — | — |
| $\gamma$ | .0894 | .1000* | .0954 | — | — | — | — | — | — |
| decay | .0447 | .1000* | .1087 | — | — | — | — | — | — |
| $T_0$ | — | — | — | — | — | — | — | — | .5767 |
| $\theta_T$ | — | — | — | — | — | — | — | — | .1080 |
| bias | — | — | — | — | — | — | — | — | −1.1786 |
| delay (msec) | — | — | — | — | — | — | 0.0050 | — | — |

*Note.* Parameters denoted by * for the 11-parameter SIAC model and 9-parameter FLMP-C were fixed values. Also given are the number of parameters for each model, the RMSD, the log-likelihood (LL) of the model and the Akaike Information Criterion (AIC) statistic.





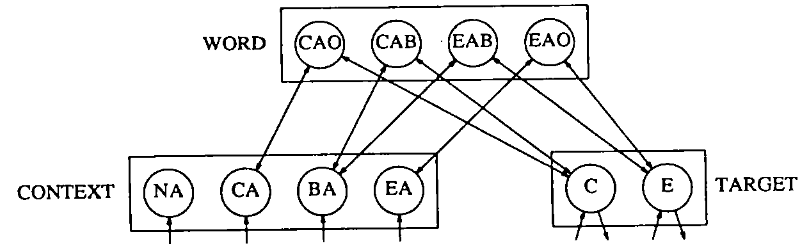


FIG. 11. Network used in the simulation of the IAC model applied to the experiment of Massaro (1979). The inhibitory connections between units within the word, context, and target levels are not shown in the network.

*alpha, gamma,* and *decay* were set to .1 and the standard deviation of the input noise used was set to .14142, all following McClelland (in press).

For each of the 96 experimental conditions, 176 simulation trials were run by adding normal noise samples to both the Target and Context input values. For each of the simulated trials, a BOW decision was made on the final target activations; if the activation of the *e* Target node was greater than that for *c* Target node, then the probability of an *e* response was incremented by 1/176. The RMSD obtained for this model was .1135, about twice that found for the dynamic simulation FLMP. Figure 12 shows the fit of the 11-parameter SIAC model to the masking data. A second, 12-parameter SIAC model was run (starting with the parameters of the 11-parameter model) which included the standard deviation of the noise as a parameter. No improvement was seen in the RMSD with a final RMSD value of .1132. As is apparent in the figure, this SIAC model does not predict a context effect at short processing times. This result follows from the fact that interactive activation, as implemented with the four control parameters set at .1, requires more than two or three cycles before the effects of top-down information on bottom-up representations are apparent.

Finally, a full-blown SIAC model was fit to the results by allowing the four control parameters of the SIAC process (*estr, alpha, gamma,* and *decay*) to be additional free parameters for the minimization search. This model took a very long time to run, given the large number of SIAC computation cycles needed during most of the fit: four context conditions, times 258.5 cycles (35 + 70 + 125 + 240 ms times $\theta$ = .55) for the four processing durations, times six stimulus values, times 176 simulated trials, gives about 1,000,000 SIAC cycles for each complete set of predictions given a set of 16-parameter values. Each of the 674-parameter estimating iterations of the model took about 90 cpu minutes on a SUN-3/180 with 16.2 MHz 68881 floating point for a total of over a thousand hours of computation—about 6 weeks time. Figure 13 shows the final fit of the

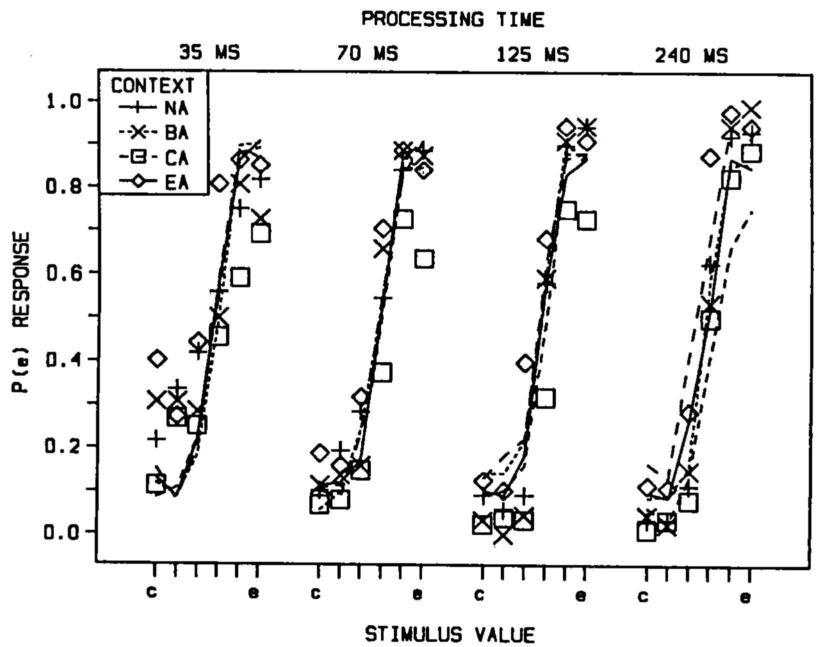


FIG. 12. Observed (points) and predicted (lines) probability of *e* identifications as a function of the bar length of the test letter, the orthographic context, and the processing interval between the onset of the test stimulus and the onset of the masking stimulus for the 11-parameter SIAC model (results after Massaro, 1979).

16-parameter SIAC model to the masking data, which had an RMSD of .0953, a slight improvement over the 11- and 12-parameter models, but still inferior to the FLMP. A comparable 16-parameter SIAC-INT model was also fit to the data and provided a fit with an RMSD of .0688. Figure 14 shows the final fit of this model. Table 5 gives the parameter values for each of the SIAC models of the masking experiment.

Given that the FLMP, SIAC, and SIAC-INT models are being fit to a single set of group data, it is not possible to use replications over subjects to give a statistical test between the models. An alternate statistical test between models is discussed at the end of the next section.

### *Boltzmann Machine Model*

McClelland (1991) has shown that at equilibrium or "asymptotically," as he sometimes terms it, the Boltzmann machine network may be characterized mathematically as the product of terms representing bias, context, and stimulus information, and is thus equivalent to the classic FLMP or Bayesian model. McClelland goes further by arguing that the Boltzmann Machine and SIAC models in general are to be preferred to the



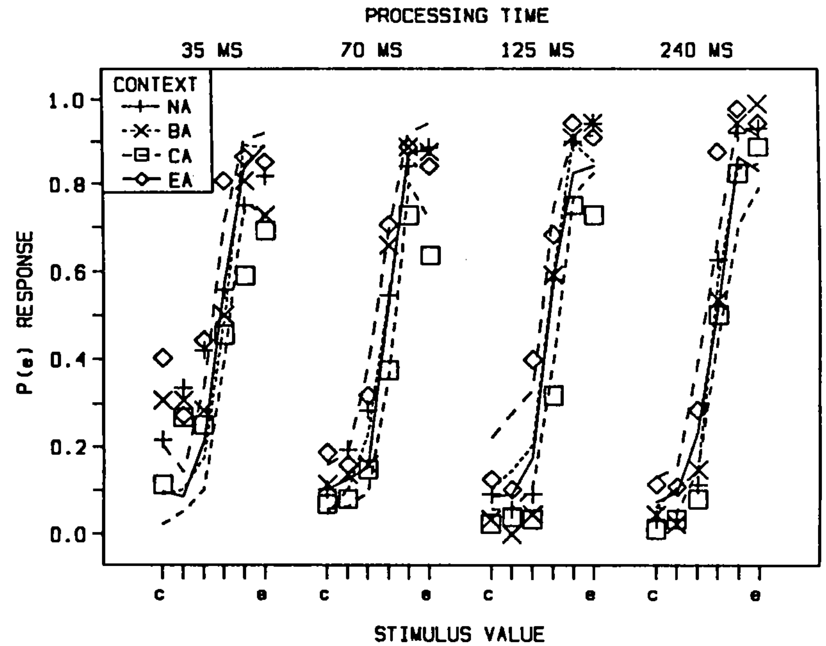


FIG. 13. Observed (points) and predicted (lines) probability of *e* identifications as a function of the bar length of the test letter, the orthographic context, and the processing interval between the onset of the test stimulus and the onset of the masking stimulus for the 16-parameter SIAC model (results after Massaro, 1979).

FLMP because they might be able to explain the dynamics of perceptual processing. However, the fact that the Boltzmann machine can be shown to be decomposable to a product form does not generalize to the larger class of mathematically intractable stochastic interactive activation models. Moreover, even the Boltzmann Machine has not yet been shown to be well-behaved over time and accurate in predicting data. We now consider the dynamics of this model and compare it to a dynamic extension of the FLMP.

Consider how a Boltzmann Machine network might be used to model processing in a backward masking task. Because the mask probably terminates processing of the stimulus, possibly before the network has reached equilibrium, we need to know whether its output can be characterized as "classic" prior to that point. We carried out a number of simulations to explore these possibilities using the standard network shown in Fig. 11.

In the first simulation, we were interested in the case in which both the stimulus inputs and the context supported the alternative *e*. The stimulus input for *e* was set to .41 and the input for *c* was set to −.41. The input

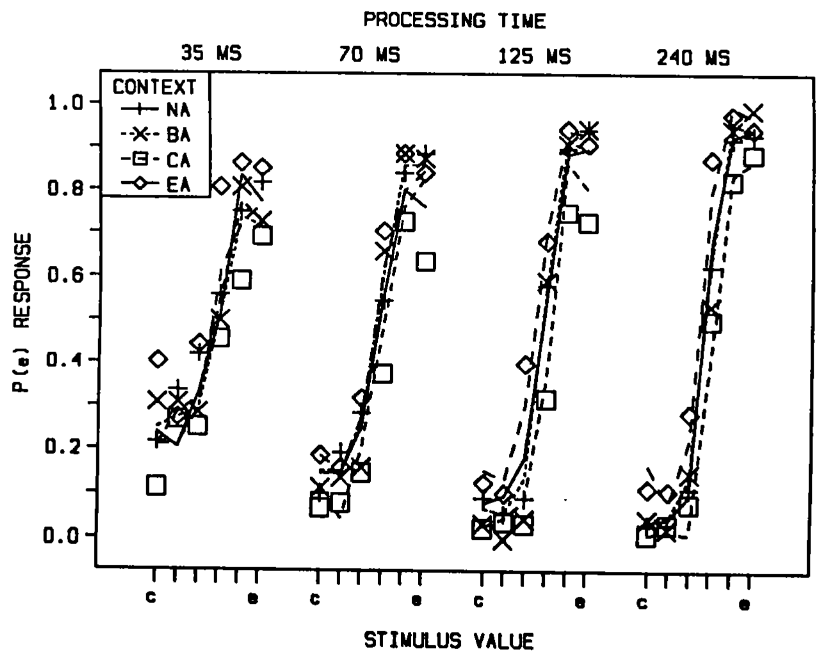


FIG. 14. Observed (points) and predicted (lines) probability of $e$ identifications as a function of the bar length of the test letter, the orthographic context, and the processing interval between the onset of the test stimulus and the onset of the masking stimulus for the 16-parameter SIAC-INT model (results after Massaro, 1979).

to the $e$ context unit was set to 3 and inputs to the other context units were set to 0. In contrast to the situation of consistent context and feature inputs, during a simulation, the initial activation $act_i$ of each unit was set to 0. During each cycle, the temperature was first set according to a simulated annealing schedule which gradually decreased the temperature of the units according to the negatively decelerating function given by the equation

$$Temp(c) = T_0 e^{-\theta_T(c-1)} \tag{18}$$

where $c$ is the number of computation cycles, $T_0$ is the starting temperature (set at .5), and $\theta_T$ is the cooling rate (set at .12). Asynchronous updates of randomly picked units were carried out: 1760 updates were done during each cycle, resulting in about 176 updates for each of the 10 units (comparable to the 176 synchronous updates of the 10 units for the SIAC models). For each update, the input net activation $net_i$ for unit $i$ was calculated to be:



$$net_i = istrength \times \left(bias_i + \sum_{j=1}^{nunits} w_{ij} \times act_j\right) + estrength \times ext_i, \tag{19}$$

where *istrength* controls the contribution of the internal units, $bias_i$ is the bias level of unit $i$, $w_{ij}$ weights the inputs to unit $i$ from unit $j$, and *estrength* controls the contribution of the external input $ext_i$. The activation $act_i$ was computed according to the stochastic rule:

$$\text{if } R < logistic\ (net_i,\ Temp),\ act_i = 1, \text{ else } act_i = 0 \tag{20}$$

where

$$logistic(x, Temp) = \frac{1}{1 + e^{-\frac{x}{Temp}}} \tag{21}$$

and $R$ is a uniformly distributed random number between 0 and 1. At the end of each update, the network is analyzed to see whether: (1) neither target $e$ nor target $c$ is on, (2) both target $e$ and target $c$ are on, (3) only target $c$ is on, or (4) only target $e$ is on. A tally is kept of how many times out of the 1760 updates per cycle each of these four cases occurred. Figure 15 shows the number of times out of 1760 each of these cases occurred over 101 cycles of processing, leaving the context and stimulus information on the entire time and with *istrength* = .6, *estrength* = .6, and *bias* = 0. Following the simulated run, we computed the expected proportion of $e$ responses, $p(e)$ as:

$$p(e) = \frac{f(e\ only)}{f(e\ only) + f(c\ only)} \tag{22}$$

where $f(x)$ is the frequency of $x$, disregarding the cases in which neither or both targets were on (as McClelland suggests), and setting $p(e)$ to zero if the denominator was zero. The function is shown by the solid line in Fig. 16 shows the $p(e)$ value calculated from the number of cases shown in Fig. 15.

In contrast to the situation of consistent context and feature inputs, a second simulation was carried out with the same context information supporting $e$ (value 3), but with feature information supporting $c$ by reversing the values for $e$ and $c$ ($e = -.41$; $c = .41$). This is shown by the dashed line in Fig. 16. We see that in this case the predicted proportion of $e$ responses first decreases and then suddenly increases dramatically. This prediction seems to be due to propagation delay in the network. The feature information arrives earlier at the target nodes than the context information which is delayed in effect because it must be mediated by the

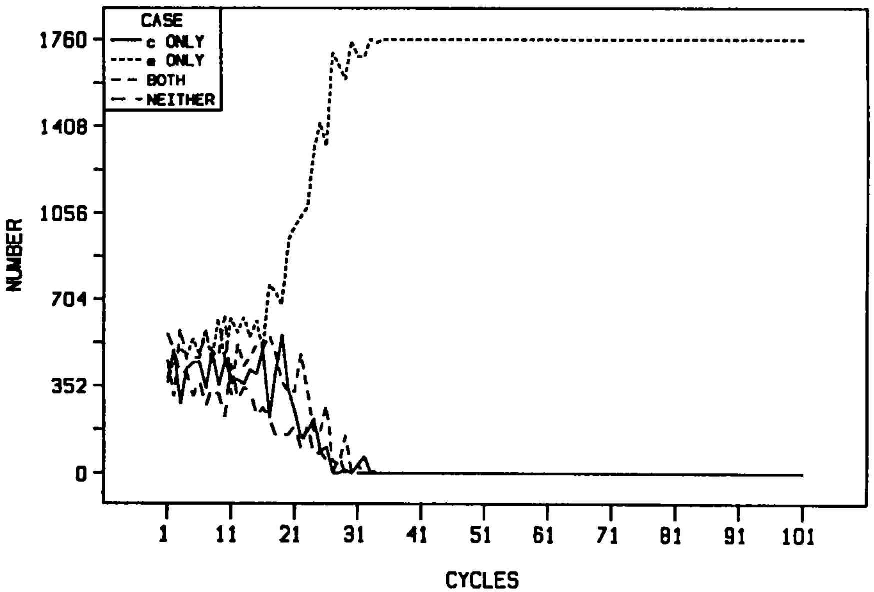


FIG. 15. Number of cases of four target activation outcomes for the Boltzmann Machine model over 101 cycles of activity with context supporting *e* (*EA* = 3) and feature information (*e* = .41; *c* = −.41) supporting *e*.

words in memory. The degree to which this propagation-delay result occurs will depend on the particular network topology, the network weights, the cooling schedule, the input values, and the control parameters of the Boltzmann process. However, the early activation of the presented letter and later activation of the contextually appropriate alternative seems endemic in models which assume top down effects on target nodes. Although the reversal in Fig. 16 is probably counterfactual, a careful study of dynamic processing results of individual subjects should show these reversals if they occur. A relatively fine-grained study would be required, because, as seen in Fig. 16, the reversal occurs for only a limited period of time during processing and too widely spaced temporal conditions might miss the effect.

A similar result can also be seen in the dynamics of the SIAC-INT model. The two functions in Fig. 17 show the predicted proportion of *e* responses over 176 simulated trials for the model with consistent and conflicting context and feature information is presented. In the simulation the following parameters were used: $estr = .3$, $\alpha = .3$, $\gamma = .1$, $decay = .1$, $gsd = .14142$, $\lambda = .05$, EA context = 3, and other contexts = 0. In one function, the *e* input was .7 and the *c* input was .3. As can be seen in this function, the consistent context and feature information quickly drive



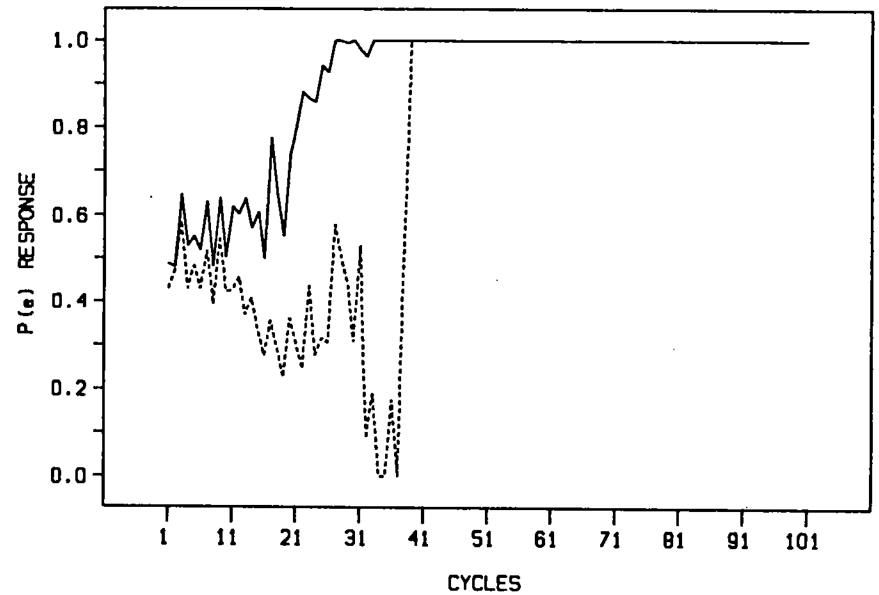


FIG. 16. Proportion of e responses for the Boltzmann Machine model over 101 cycles of activity with context supporting *e* (*EA* = 3.) and feature information (*e* = .41; *c* = −.41) supporting *e* (solid line). Proportion of e responses for the for the Boltzmann Machine model over 101 cycles of activity with context supporting *e* (*EA* = 3) and feature information (*e* = −.41; *c* = .41) supporting *c* (dashed line).

the responses toward *e*. For the other function, the *e* input was .3 and the *c* input was .7. In this case, the responses initially follow the feature information which favors *c*, but after about 50 cycles, responses mostly favor *e*.

We now consider how predictions by the Boltzmann model can be tested against the data of the masking experiment. The extreme noisiness of the predictions of the pure Boltzmann model might preclude a good fit to the data and disallow evaluation of the degree to which the results prior to equilibrium achieve the classic results (even in simulations with 10,000 asynchronous updates per cycle). To allow the model a better chance to fit the data we made the additional assumption that the number of the four possible outcomes given above are time averaged according to the formula:

$$af_{i,t} = \lambda f_{i,t} + (1 - \lambda)af_{i,t-1} \tag{23}$$

where $af_{i,t}$ is the averaged frequency of event type $i$ at cycle $t$, derived from the weighted average of the actual frequency $f_{i,t}$ and the prior running average frequency $af_{i,t-1}$, controlled by the weight $\lambda$. For the model fits, $\lambda$ was set at .05. There were six parameters for the feature values for

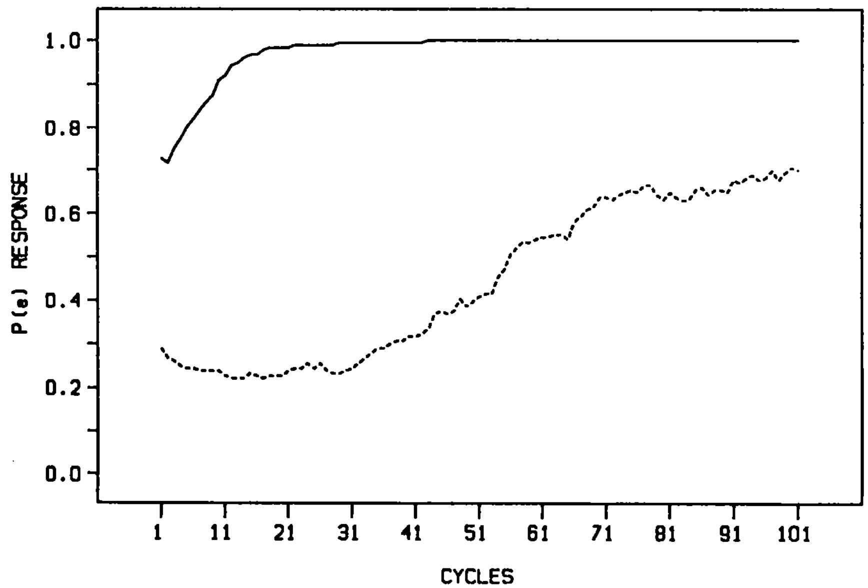


FIG. 17. Proportion of e responses for the SIAC-INT model over 101 cycles of activity with context supporting $e$ ($EA$ = 3) and feature information ($e$ = .7; $c$ = .3) supporting $e$ (solid line). Proportion of e responses for the SIAC-INT model over 101 cycles of activity context supporting $e$ ($EA$ = 3) and feature information ($e$ = .3; $c$ = .7) supporting $c$ (dashed line).

$e$ in the possible range $-3$ to $+3$, with the $c$ feature values receiving the negations of the $e$ inputs, four parameters for context in the possible range 0 to $+3$ with the values for the nonselected contexts set at 0. Other free parameters of the model were *bias* ($-3$ to $+3$), *estr* (0-1), and *istr* (0-1). In addition, two parameters $T_0$ and $\theta_T$ (both 0-1) controlled the cooling schedule as given in Eq. 18. In the model fit, it was assumed that each cycle represented 4 ms of processing time. This yielded 60 cycles for the longest processing time, equivalent to what was used in the phonological constraints SIAC fits. The model fits was carried out 1760 updates per cycle, resulting in about 176 updates for each of the 10 nodes in the network. This corresponds to the number of trials actually carried out in the experiment and the number of simulated trials in the SIAC fits.

Figure 18 shows the predicted probability of $e$ identifications as a function of the levels of context and feature for the four processing durations for the Boltzmann model. The RMSD for the predictions of the model was .0748. Table 5 gives the best fitting parameters for the model.

As mentioned previously, fitting-group data precludes carrying out the usual ANOVA on RMSD with subjects as the random variable. Instead, we use the Akaike Information Criterion (AIC) statistic to compare mod-



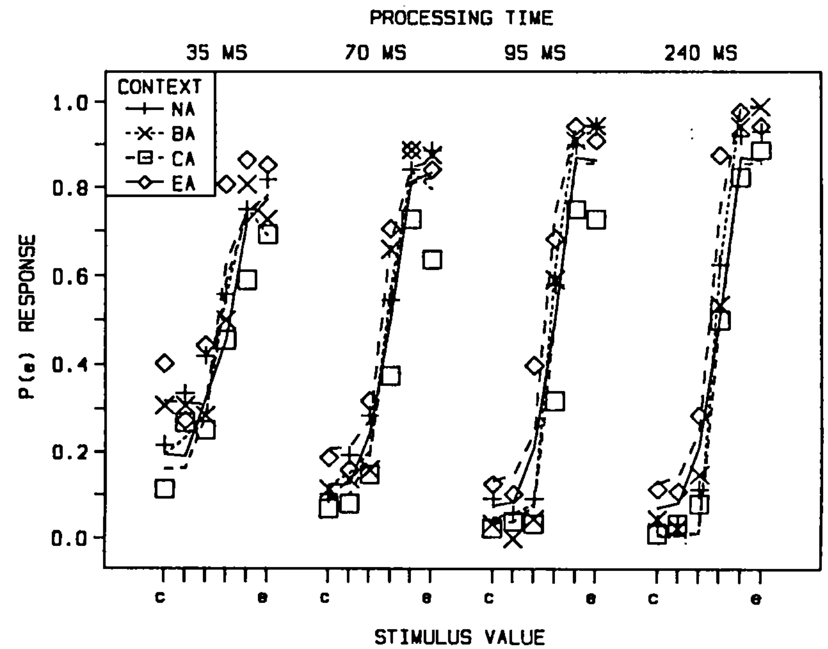


FIG. 18. Predicted probability of $e$ identifications as a function of the bar length of the test letter, the orthographic context, and the processing interval between the onset of the test stimulus and the onset of the masking stimulus for the Boltzmann model.

els (Akaike, 1974; Sakamoto, Ishiguro, & Kitagawa, 1986). This formal theory takes into account the likelihood of a model fit and also the number of parameters used by the model. When several models give an approximately equally likely fit of the observed data, the AIC statistic would say that we should choose the model with the fewest parameters. In this sense, the inclusion of the number of parameters in computing the AIC allows us to contrast different models with a varying number of free parameters.

We note first of all that our model fits using STEPIT minimized the squared deviations of the observed and predicted data which yields a maximum likelihood fit; thus the likelihood of the obtained fit is the maximum likelihood. The general form for the exact likelihood ($L$) of this obtained fit is given by the product of the multinomial distributions for each stimulus condition:

$$L = \prod_s \left[ \frac{\left(\sum_r f_{sr}\right)!}{\prod_r f_{sr}!} \times \prod_r p_{sr}^{f_{sr}} \right] \tag{24}$$

where $f_{sr}$ is the observed frequency of response $r$ to stimulus $s$ and $p_{sr}$ is the predicted proportion of response $r$ to stimulus $s$. The log-likelihood ($LL$) of the fit is given by:

$$LL = \sum_s \left[ ln\left(\left(\sum_r f_{sr}\right)!\right) - \sum_r ln(f_{sr}!) + \sum_r f_{sr} ln(p_{sr}) \right]. \qquad (25)$$

The AIC statistic is computed as:

$$AIC = -2(maximum\ LL) + 2(number\ of\ parameters). \qquad (26)$$

Smaller AIC values are preferred. From the relationship between the AIC quantity and entropy, if the difference in AICs between models is at least 1 or 2, then the difference is considered to be significant. If the difference is much less than 1, then the models are equally good in describing the data. Table 5 gives the *LL* and AIC values for each of the models. As can be seen in the table, the AIC of the FLMP was 240 less than the AIC of the SIAC-INT and 389 less than the AIC of the Boltzmann model and therefore the latter models can be rejected in favor of the FLMP. The dynamic predictions of the SIAC-INT and Boltzmann Machine were significantly poorer than the dynamic FLMP.

The dynamic FLMP gave a good description of the results with the assumption that the build-up of context occurs simultaneously with and at the same rate as the build-up of stimulus information. To test these assumptions directly, two expanded forms of the FLMP were tested: one (FLMP-D) assumed that there was a delay in the build-up of context information (delay time subtracted from context processing time), and the other (FLMP-θ) assumed that context occurred at a slower rate than stimulus information (with different θ values for feature and context information). These fits are presented in Table 5. Even though these two models used an additional free parameter, their fits were not better than the simple FLMP. Thus, we have strong evidence that, contrary to the predictions of IAC models, the build-up of context and stimulus information occur simultaneously and at the same rate. Table 5 also gives the fit of a constrained FLMP model (FLMP-C) which was tested to see whether the poor fit of the IAC models might be due to the fact that the NA and BA contexts cannot differentially bias target activations. In the FLMP-C, the context values for these two contexts were fixed at .5. As can be seen in the table, the fit of this model was essentially equivalent to the ordinary FLMP, and far superior to any of the IAC models.

To summarize this section, backward masking was used to provide performance measures at different points during perceptual processing, making it possible to observe how stimulus information and context interact over time. The empirical results revealed that context can have a



substantial influence even at very short processing intervals. In addition, the relative contribution of stimulus information increased over processing time. Both of these empirical results cannot be simultaneously predicted by the SIAC and Boltzmann models. An additional analysis was carried out on the results of the *e–c* experiment to illustrate more directly how the contribution of stimulus and context varies with processing time. A stimulus effect was computed across the four levels of context by subtracting the probability of an *e* response given the endpoint *c* letter from the probability of an *e* response given the endpoint *e* letter. Figure 19 shows that the observed size of the stimulus effect increased with processing time. In the similar manner, a context effect was computed at each processing interval by subtracting the probability of an *e* response given the *c* admissible context from the probability of an *e* response given the *e* admissible context. As can be seen in Fig. 20, there was a substantial context effect at short processing times and it remained relatively constant with increases in processing time.

In the SIAC models, contextual information can become available only after feedback from the evaluation of stimulus information. In order to maximize the fit of these results, the 16-parameter SIAC model required a very short time (1.82 ms) for each processing cycle. This allows the model to predict a substantial context at short processing intervals (see

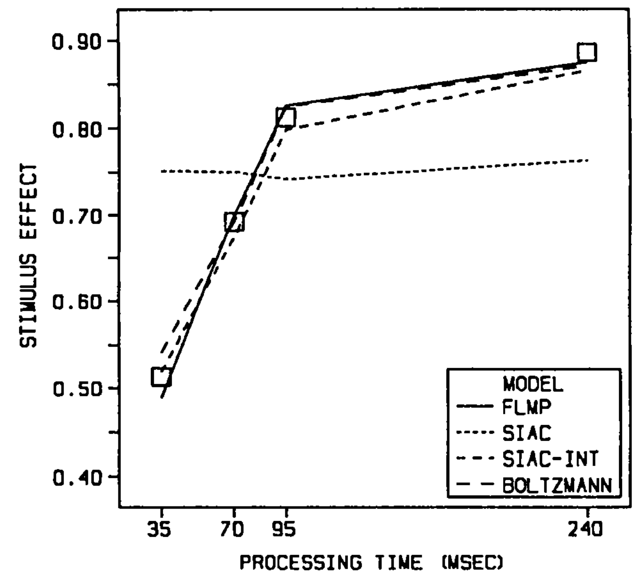


FIG. 19. Observed (points) and predicted (lines) size of the stimulus effect as a function of processing time for the FLMP, 16-parameter SIAC, SIAC-INT, and Boltzmann models of e–c masking data.

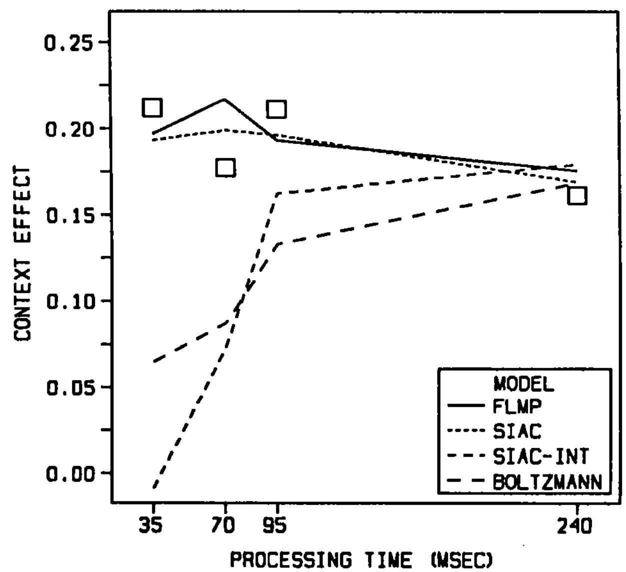


FIG. 20. Observed (points) and predicted (lines) size of the context effect as a function of processing time for the FLMP, 16-parameter SIAC, SIAC-INT, and Boltzmann models of e–c masking data.

Fig. 20). Given this short time per cycle, however, the target unit activations become asymptotic too quickly and the observed increase in the stimulus effect (shown in Fig. 19) with additional processing time cannot be predicted. The SIAC-INT model maximized the goodness of fit with a longer cycle time (4.44 ms), and is able to predict the increase in stimulus effect with increases in processing time (see Fig. 19). However, Fig. 20 shows that context effects are not predicted at short processing intervals, contrary to the observed results. The Boltzmann model with a cycle time of 4 ms behaves like the SIAC-INT model. In the FLMP, contextual information is evaluated simultaneously with stimulus information and the model accurately describes the temporal course of these contributions to performance (see Figs. 19 and 20).

The interpretation of the results in Figs. 19 and 20 requires some understanding of the relative nature of the stimulus effect and context effect. Figures 19 and 20 give the observed stimulus and context effects, respectively. However, the stimulus effect depends on the size of the context effect and vice versa. The stimulus effect was 3 or 4 times larger than the context effect, reflecting the fact that the stimulus letter was more influential than context letters. In addition, although the contribution of context increased at the same rate as the contribution of stimulus, the latter increased to a larger asymptote (see Fig. 9). These two results are responsible for fact that the effect of context shown in Fig. 20 (and also Fig. 10) did not change much with increases in processing time. Although the absolute contribution of context increases with increases in processing time, as illustrated in Fig. 9, so does the larger absolute contribution of stimulus information. In the FLMP, the contribution of one source of information is small to the extent the contribution of another source of information is large. Thus, the increase in the context effect is washed out by the larger increase in stimulus effect.



## WORD SUPERIORITY EFFECT (WSE)

Another prototypical example of a context effect in psychological research has been dubbed the Word Superiority Effect or the WSE. A letter in a word is better recognized than a letter in a nonword or even better than a letter presented alone. Many theories have been proposed to account for this influence of context on word recognition (the latest by Richman & Simon, 1989). In previous research, Massaro (1979) attempted to classify all previous accounts of the WSE and to test among them. The previous accounts could be classified as independence and nonindependence theories. This use of an independence property appears to be equivalent to Ashby and Townsend's (1986) later definition of perceptual independence. In independence theories, context does not influence sensory processing at the letter level. In nonindependence models, on the other hand, context modifies sensory processing at the letter level. A categorization of some existing theories helps clarify this taxonomic distinction. Nonindependence theories have taken a variety of forms. In one class of theories, orthographic context (for familiarity) has been assumed to influence a feature extraction stage. According to this view, the sensory resolution of a letter should be better in familiar than in unfamiliar letter strings. In another case of nonindependence theories, it is assumed that higher-order units intervene in the processing sequence to change the perceptual analyses which are employed. These models also predict that orthographic context influences the sensory featural processing of the letter string. In hypothesis-testing models, for example, nonindependence arises because higher-order information actually directs the nature of the featural analyses. Thus, hypothesis-testing models also predict that orthographic structure modifies featural analyses.

The SIAC model is also an example of a nonindependence theory, because top-down connections from the word level to the letter level allow context to modify the representation at the letter level. Although the model can account for many of the existing results on the WSE, it is important to stress that interactive activation is not *necessary* to account for these results. The FLMP, for example, does so without interactive activation. In the FLMP, context operates independently of featural analysis, simply by providing an additional source of information at the level of primary recognition, which synthesizes a percept and passes it on to synthesized visual memory (Massaro, 1984). McClelland (1991) claims that the FLMP cannot predict an accuracy advantage for words in the Reicher-Wheeler task and that some form of IA is necessary to predict such an effect. We prove below that the FLMP (without interactive activation) does predict an accuracy advantage given two sources of information relative to just one, and then demonstrate that the dynamic extension of the FLMP provides a good description of the time course for the WSE, as measured by Massaro and Klitzke (1979).

*Backward Masking and the WSE*

Before applying the FLMP to Massaro and Klitzke's (1979) results, it is important to mention how performance in the Reicher-Wheeler task is viewed (Thompson & Massaro, 1973; Massaro, 1975a). Visual information and orthographic context are integrated during the perceptual processing of a letter string, and the resulting percept is influenced by both these sources of information. The Reicher-Wheeler control does not eliminate a possible influence of orthographic context during perception; the control only precludes a postperceptual guessing advantage for words. As an example, suppose we are interested in comparing a word *WORD* with a test nonword *ORWD*. In both cases, the two test alternatives are *D* and *K* for the fourth position. If the subject has no visual information about the test string, he or she would have no advantage in the word relative to the nonword condition. However, if a curved feature from the fourth letter was derived from the visual information, then the candidates for this position might be *D, O,* or *Q*. If the first three letters *WOR* were also recognized in the test word, then orthographic context would eliminate the candidates *O* and *Q,* leaving *D* as the only perceptual alternative. Recognizing *ORW* in the nonword condition would not constrain the alternatives for the fourth position, thus making perception of any of the three alternatives equally likely. The advantage of words over nonwords in the Reicher-Wheeler task results from this contextual difference.

McClelland (personal communication) questioned why orthographic context is assumed to support the correct alternative in the implementation of the FLMP for the Reicher-Wheeler task. He is correct in observing that the orthographic context also supports the incorrect test alternative. However, the orthographic context does provide information against other incorrect letter alternatives—as illustrated in the example in the previous paragraph. Thus, the orthographic context can have a significant influence on perceptual performance. Two sources of information can lead to better performance than just one. If only the context is presented or if only the visual information about the test letter is presented, no advantage of words over nonwords is predicted with a Reicher-Wheeler control. Given both orthographic context and visual information, however, a word superiority effect is predicted in the Reicher-Wheeler task.



We now show that the dynamic FLMP not only predicts a WSE, it predicts a subtle interaction among the WSE, backward masking, and lateral masking. Johnston and McClelland (1973) found a WSE over letters when the test display was followed by a mask, but not when no mask was presented, and offered three possible explanations. Massaro (1975b) explained this effect in terms of the tradeoff between the positive contribution of orthographic context and the negative effect of lateral masking. To test this explanation, Massaro and Klitzke (1979) employed four types of display in the Reicher-Wheeler task: words, nonwords, letters, and letters flanked by dollar signs. On each trial, one of these test displays was presented followed by a masking display after one of seven stimulus onset asynchronies (SOAs) or no mask was presented. Two choice alternatives were presented ¼ s after the test display as in the standard Reicher-Wheeler control. The masking stimulus varied from trial to trial and was composed of nonsense letters created by selecting random feature strokes from the letters of the alphabet.

The intensity of the test letters was varied throughout the experiment to keep overall performance at 75% correct (with the constraint that each condition was tested with the same intensity level). Given that subjects get much better in the task over the course of the experiment, it was necessary to continually lower the intensity of the test stimulus throughout the experiment. Thus, although the intensity of the test and mask were initially equated, the masking stimulus was significantly more intense than the test stimulus during the major portion of the experiment. (This difference in intensity becomes important in the application of the models to the results.)

Six subjects were tested for 5 days, with the first day treated as practice. Half the subjects received seven experimental sessions and the other half received eight. There were 56 or 64 observations per subject at each of the 32 conditions (four contexts times eight masking intervals). For group data analysis, the percentage correct judgments were pooled across subjects giving a total of 360 observations per condition. The points in Fig. 21 present the probability of a correct identification for the group as a function of the test display and the SOA.

Massaro and Klitzke (1979) described the WSE and its change across masking conditions using the negatively accelerating function, given by Eq. 14. They showed how three variables interact: orthographic context, processing time, and lateral masking. Given the lateral interference of adjacent letters on each other, the sensory information in the word, nonword, and letter-in-dollar-signs conditions must necessarily be less than