Chapter 15

On the Similarity of Categorization Models

Michael M. Cohen & Dominic W. Massaro
Program in Experimental Psychology
University of California
Santa Cruz, CA 95064 U.S.A.

Send Correspondence to
Michael M. Cohen
Program in Experimental Psychology
University of California
Santa Cruz, CA 95064 U.S.A.

Chapter for F. G. Ashby (Ed.) *Multidimensional Models of Perception and Cognition*. We thank Greg Ashby, an anonymous reviewer, and Antoinette Gesi for their helpful comments. The authors of this collaborative effort are listed in alphabetical order. The research reported in this paper and the writing of the paper were supported, in part, by grants to Dominic W. Massaro 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.

15.1 Introduction

Most of the chapters in this book describe probabilistic multidimensional models of perceptual and cognitive tasks. This fact acknowledges several important properties of human information processing. First, performance is not deterministic, but is variable or probabilistic. A subject responds in one way to a stimulus on one trial and responds in another way on another trial. Thus, performance is often characterized by some probability value representing overall response probability to a repeated presentation of a stimulus. Probabilistic performance might result from probabilistic differences in processing, probabilistic differences in the physical stimulus information from trial to trial, or probabilistic representations of prototype items in memory. Second, the term multidimensional refers to the multiple sources of information that influence perception and cognition (Massaro & Cohen, submitted; Massaro & Friedman, 1990).

In a previous paper, Massaro & Friedman (1990) presented and compared various existing models of how multiple sources of information influence perception and decision. The question they addressed was how individuals process two or more sources of information that may reinforce or conflict to various degrees. The models were analyzed in terms of a prototypical pattern recognition task and the application of extant models to this task. The central concerns were the processes assumed by the models, the similarities and differences in predictions of the models, their optimality properties, and empirical validity.

Our goal in this chapter is to extend the analyses carried out by Massaro and Friedman (1990) by comparing several additional classes of models. The models analyzed and tested by Massaro and Friedman were the Fuzzy Logical Model of Perception (FLMP), a two-layer Connectionist Model (CMP), a model derived from the Theory of Signal Detection (TSD), a Linear Integration Model (LIM) derived from results of Functional Measurement, and a model based on multidimensional scaling (MDS). The goal of these analytical and descriptive exercises was to lay the groundwork for valid experimental tests of the models. Simulations of the models and predictions of the results by the same models were carried out to provide a measure of identifiability or the extent to which the models can be distinguished from one another. The models were also contrasted against empirical results from tasks with two and four response alternatives and with graded responses. The results indicated FLMP was optimal (mathematically equivalent to Bayes's theorem), and was the only psychological model to provide an adequate account of performance in these tasks.

The goal of the present paper is to attempt to formulate the models previously shown to be inadequate, in such a manner to bring them in line with the predictions of the FLMP and with empirical results. In contrast to our general research strategy of strong inference (Platt, 1964) and competition between models, we attempt here a reconciliation of current models. We will see that, in some cases, there are straightforward formulations that give equivalence among some models and the FLMP. In other cases, an additional free parameter or some other modification is necessary to bring a model in line with the desired predictions. The model classes that we will consider are: Fuzzy Logical Model of Perception (FLMP), Gaussian MDS Model (GMM), Exponential MDS Model (EMM), Theory of Signal Detection (TSD), Feed-forward Connectionist Model (FCM), and Interactive Activation and Competition (IAC). Some of these models are based on spatial representations, others are grounded in probability or truth value, and others are described in activation of neural-like units. Each of these different representational foundations is referred to as a model's currency (Massaro & Friedman, 1990). As in Massaro and Friedman (1990), we will first present each of these models theoretically in the context of a prototypical pattern recognition task combining two sources of stimulus information and allowing two possible responses. We will then consider how the models compare in a four-response task. For both the two- and four-response situations, we derive predictions of the models, compare these predictions to those of other models, and test the models against one another with both hypothetical FLMP data and experimental data, using model fitting techniques. We do not completely describe here the psychological theory underlying each

model because adequate summaries are available elsewhere (Massaro & Friedman, 1990, Shepard, 1987).

15.2 Prototypical Two-Response Categorization Task

We use a letter-processing task carried out by Massaro and Hary (1986) with the two categories, G and Q, as the response alternatives. Two sources of information were manipulated with a number of levels of each source of information. A range of letters between G and Q was created by independently varying the obliqueness of the straight line and the closedness of the gap in the letter (Figure 15.1). Seven levels of closedness were made by removing 0, 2, 3, 4, 7, 9 and 10 points from the right side of the oval of a capital letter Q. Similarly, the obliqueness of the line varied between the horizontal and 11, 21, 29, 38, 51, and 61 degrees of obliqueness measured from the horizontal. The resultant 49 test letters make up the factorial design.

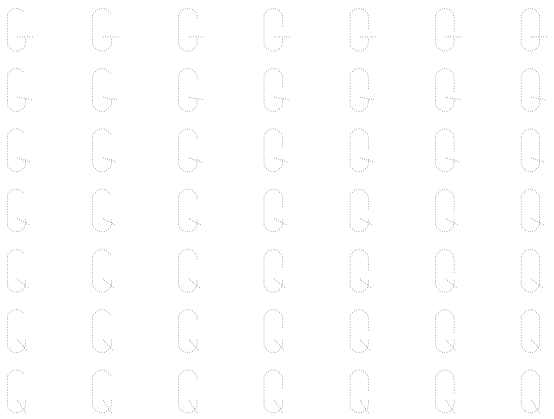

Figure 15.1. Forty-nine test letters, varying between G and Q, created by varying the obliqueness of the straight line (row factor) and the closedness of the gap in the oval (column factor) (after Massaro & Hary, 1986).

Nine subjects saw each of the test letters (shown in Figure 15.1) for 400 msec 12 times in random order (Massaro & Hary, 1986). On each trial they labeled the test letter as a Q or a G. Figure 15.2 gives the observed performance for two typical subjects. The probability of a Q response given each test letter is the dependent variable. Because the Q and G identifications sum to one, the probability of a Q response to each test letter, $P(Q)$, completely represents the identification judgments. Thus we have 49 independent observations to describe performance given the 49 test letters.

15.3 Fuzzy Logical Model of Perception (FLMP)

We initiate our derivation of specific models with the fuzzy logical model of perception (FLMP). We begin with this model because it has proven successful in accounting for a wide range of categorization results (Massaro, 1987a; Massaro & Friedman, 1990; McClelland, in press). Thus, the model's predictions serve as a standard for the predictions of other models.

Underlying the FLMP is the assumption that well-learned patterns are recognized in accordance with a general algorithm, regardless of the modality or particular nature of the patterns (Massaro, 1984, 1987b; Oden, 1981, 1984). The model consists of three operations in perceptual recognition: feature evaluation, feature integration, and pattern classification. Continuously-valued features are evaluated, integrated, and matched against prototype descriptions in memory, and an identification decision is made on the basis of the relative goodness of match of the stimulus information with the relevant prototype descriptions.

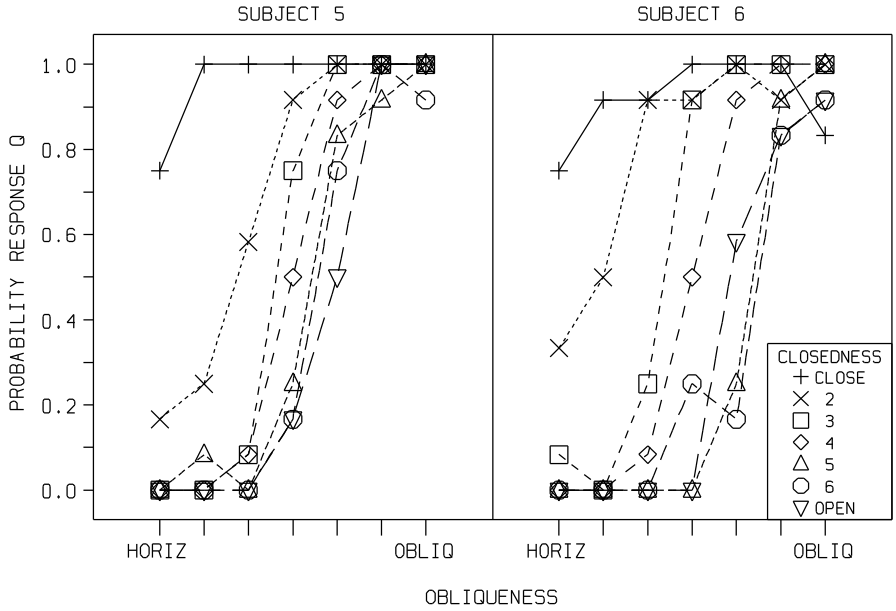

Figure 15.2. Observed probability of Q responses for the forty-nine test letters presented in Figure 15.1 (created by varying the obliqueness of the straight line and the closedness of the gap in the oval) The results are for two typical subjects (from Massaro & Hary, 1986).

Given multiple features, it is useful to have a common metric representing the degree of match of each feature. Two features which define a prototype can be related to one another more easily if they share a common currency. To serve this purpose, fuzzy truth values (Goguen, 1969; Zadeh, 1965) are used because they provide a natural representation of the degree of match. Fuzzy truth values lie between zero and one, corresponding to a proposition being completely false and completely true. The value .5 corresponds to a completely ambiguous situation whereas .7 would be more true than false and so on. Fuzzy truth values, therefore, not only can represent continuous rather than just categorical information, they also can represent different kinds of information.

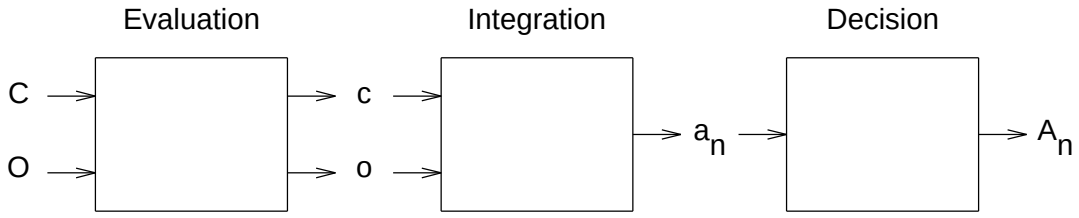

Figure 15.3. Schematic representation of the three stages of processing in the FLMP. The three stages are illustrated for the sources of information closedness C and obliqueness O in the G-Q task. The evaluation of the degree to which the oval is closed and the straight line is oblique produce values c and o that are made available to the integration process. Integration of the values gives an overall value a_n indicating the degree of support for alternative n. The decision process maps the information made available to it into a response A_n.

The three operations between presentation of a pattern and its categorization, as illustrated in Figure 15.3, can be formalized mathematically. Feature evaluation gives the degree to which a given dimension supports each test alternative. Given a test letter in the QG task, the featural evaluation stage determines the degree to which each of the alternatives A_n (in this case Q and G)

are supported by each feature of the visual information. The physical input (represented in uppercase) is transformed to a psychological value (represented in lowercase), e.g., Closedness C would be transformed to c, and analogously for dimension Obliqueness O. The notation used in the present article is described in the Appendix. Each dimension provides a feature value at feature evaluation. Using fuzzy truth values, a value between zero and one is assigned to the closedness and obliqueness dimensions, indicating the degree to which these features support the Q and G alternatives. To develop hypothetical predictions of the FLMP model the following feature values were used: for closedness, going from open to closed: .01, .10, .30, .50, .70, .90, .99, and for obliqueness, going from horizontal to oblique: .03, .20, .40, .60, .80, .92, .95.

These features values are then integrated within the Q and G prototypes. The prototypes are defined by:

Q: Closed Oval & Oblique Line
G: Open Oval & Horizontal Line

Given a prototype's *independent* specifications for the oval and straight-line features, the value of one of these features cannot change the value of the other feature at feature integration. Given a two-alternative forced-choice task and the opposing features being manipulated, it is reasonable to assume that closed and open are opposites (or negations) of one another, as are oblique and horizontal. Using the definition of fuzzy negation as 1 minus the feature value (Zadeh, 1965), we can represent the prototypes in terms of the degree to which the oval is closed and the line is oblique:

Q: Closed & Oblique
G: (1 - Closed) & (1 - Oblique)

The integration of the features defining each prototype can be represented by the product of the feature values. Feature integration consists of a multiplicative combination of feature values supporting a given alternative. These products are represented in Figure 15.3 by the lowercase a_n. If c_j and o_k are the hypothetical feature values from stimulus level j of closedness and k of obliqueness supporting alternative Q, then the total support is given by the product of c_j and o_k. Similarly, the support for G, the other alternative, is given by the product of $(1-c_j)$ and $(1-o_k)$. Figure 15.4 shows the total support for alternatives G and Q given the hypothetical feature values. The linear results reflect the multiplicative integration of the feature values.

The third operation is decision which uses a relative goodness rule (RGR) based on the support for each of the test alternatives to give the probability that a given alternative is selected. In this case, the probability of response Q given a specific stimulus $C_j\ O_k$ is

$$P(Q \mid C_j \text{ and } O_k) = \frac{c_j\ o_k}{c_j\ o_k + (1-c_j)(1-o_k)} \tag{15.1}$$

where the denominator is equal to the sum of the merit of all relevant alternatives. Figure 15.5 shows the probability of selecting alternative Q given the support for each alternative shown in Figure 15.4.

To summarize, evaluation in the FLMP involves the representation of each source of information in terms of a truth value, between zero and one, indicating the merit of a particular alternative. Integration consists of a multiplicative combination of truth values. The decision utilizes an RGR.

15.4 Bayesian Probability Model (BPM)

The Bayesian approach to combining multiple sources of information has at its core a theorem derived independently by Reverend Thomas Bayes (about 1701-1761) and Pierre Laplace (1749-1827) (Stigler, 1986). This theorem states that

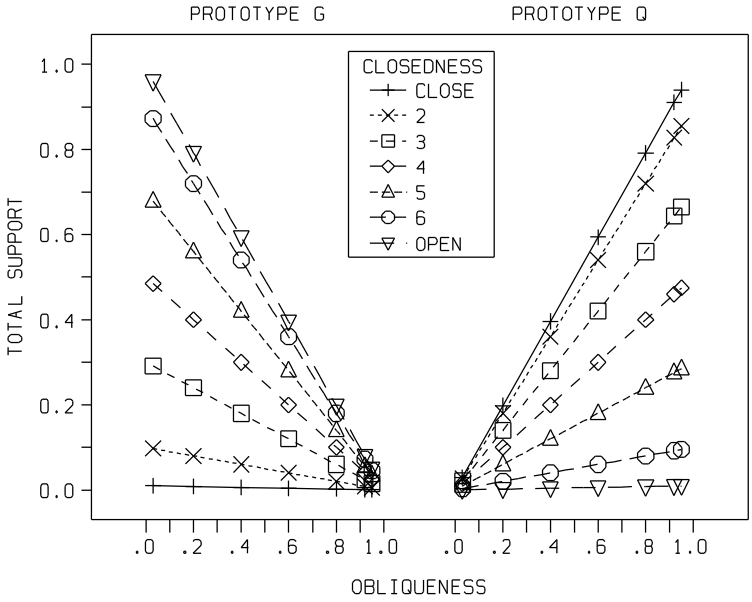

Figure 15.4. Total support for alternative Q and G prototypes based on hypothetical closedness and obliqueness features given in the text. The locations on the abscissa are scaled according to the value of the obliqueness feature.

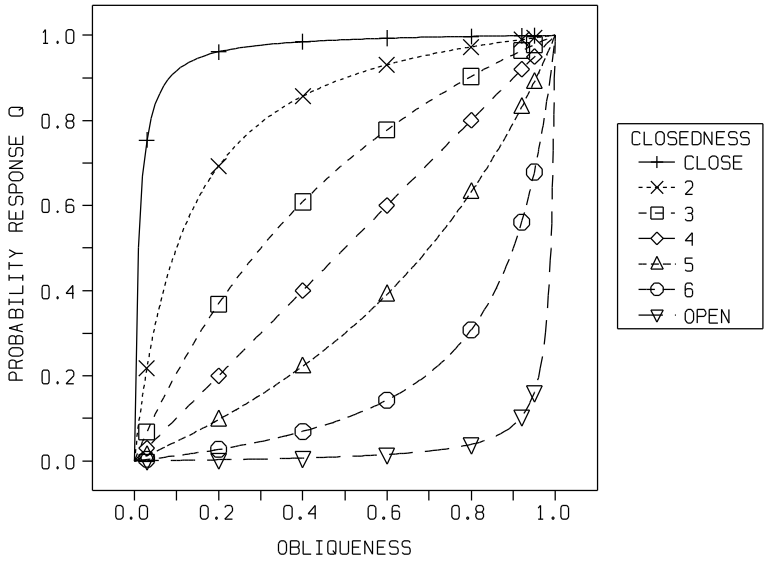

Figure 15.5. Probability for alternative Q based on the total support for the Q and G prototypes shown in Figure 15.4. The locations on the abscissa are scaled according to the value of the obliqueness feature.

$$P(h_1 \mid e) = \frac{P(e \mid h_1)\,P(h_1)}{\sum_i P(e \mid h_i)\,P(h_i)} \tag{15.2}$$

where $P(h_i \mid e)$ is the probability that some hypothesis h_i is true given that some evidence e is observed; $P(e \mid h_i)$ is the probability of the evidence e, given that the hypothesis h_i is true, and $P(h_i)$ is the *a priori* probability of the hypothesis h_i. The probability of hypothesis h_1 given some evidence e is equal to the probability of the evidence given the hypothesis times the *a priori* probability of the hypothesis, divided by the sum of analogous products for all possible hypotheses. If the *a*

priori probabilities of all possible hypotheses are equal, Bayes theorem reduces to

$$P(h_1 \mid e) = \frac{P(e \mid h_1)}{\sum_i P(e \mid h_i)} \tag{15.3}$$

Bayes theorem specifies how different sources of evidence are combined. Given two independent pieces of evidence e_1 and e_2 and equal *a priori* probabilities, the probability of a hypothesis h_1 is equal to

$$P(h_1 \mid e_1 \text{ and } e_2) = \frac{P(e_1 \text{ and } e_2 \mid h_1)}{\sum_i P(e_1 \text{ and } e_2 \mid h_i)} \tag{15.4}$$

$$= \frac{P(e_1 \mid h_1)\, P(e_2 \mid h_1)}{\sum_i P(e_1 \mid h_i)\, P(e_2 \mid h_i)} \tag{15.5}$$

Equation 15.5 has a direct correspondence to our evaluation and integration processes. This equation describes optimal information integration in the currency of probability under two assumptions. First, the prior probabilities of all relevant response alternatives are equal. Second, it is assumed that the sources of evidence are evaluated independently of one another, as was earlier assumed for the FLMP. Following our QG paradigm example above, and analogous to Equation 15.5, Equation 15.6

$$P(Q \mid C_j \text{ and } O_k) = \frac{P(C_j \mid Q)\, P(O_k \mid Q)}{P(C_j \mid Q)\, P(O_k \mid Q) + P(C_j \mid G)\, P(O_k \mid G)} \tag{15.6}$$

gives the probability of selecting Q based on the two sources of information C and O, with $P(C \mid Q)$ representing evaluation of the closedness source (in terms of the subjective probability "currency") and $P(O \mid Q)$ representing separate evaluation of the obliqueness source. Comparing Equations 15.1 and 15.6 we see a direct correspondence between the fuzzy feature values in the FLMP and the probabilities of evidence given hypotheses in the BPM, as well as the common use of an RGR.

One difference between the FLMP and the Bayesian predictions is that the two alternatives are defined as negations of their sources of information in the FLMP. Thus, closed and open are assumed to be negations as are oblique and horizontal. The probabilities $P(e \mid h_1)$ and $P(e \mid h_2)$ in Bayes's theorem need not sum to one as do the corresponding fuzzy features. When applied to a factorial design with two response alternatives, however, we should note that the predictions of the BPM, given in Equation 15.6, can be rearranged (by dividing top and bottom by $P(C_j \mid Q)\, P(O_k \mid Q)$) to give the form:

$$P(Q \mid C_j \text{ and } O_k) = \frac{1}{1 + \left(\frac{P(C_j \mid G)}{P(C_j \mid Q)}\right)\left(\frac{P(O_k \mid G)}{P(O_k \mid Q)}\right)} \quad . \tag{15.7}$$

or the equivalent:

$$= \frac{1}{1 + l_{GQ}(C_j)\, l_{GQ}(O_k)} \tag{15.8}$$

where

$$l_{GQ}(C_j) = \frac{P(C_j \mid G)}{P(C_j \mid Q)} \tag{15.9}$$

is the likelihood ratio of category G to Q given evidence C_j and

$$l_{GQ}(O_k) = \frac{P(O_k \mid G)}{P(O_k \mid Q)} \tag{15.10}$$

is the likelihood ratio of category G to Q given evidence O_k. Given that each fraction on the right side of the denominator of Equation 15.8, is indexed by a single subscript, a single value of likelihood ratio is sufficient for each source of evidence. We shall call equations like 15.8 the canonical *likelihood product form* and note that it is an instance of the more general form given in Green and Swets (1966) Appendix 1-A, Equation 1.A.7:

$$P(h_j \mid e_1, e_2, \cdots e_i \cdots e_n) = \frac{1}{1 + \frac{P(h_k)}{P(h_j)} \prod_{i=1}^{n} l_{kj}(e_i)} \tag{15.11}$$

where $e_1, e_2, ..., e_n$ are n events (sources of evidence) concerning hypotheses h_j and h_k and $l_{kj}(e_i)$ is the likelihood ratio of hypothesis h_k to h_j for evidence e_i. The *likelihood product form* for the FLMP can be obtained from Equation 15.1 by dividing top and bottom by (c_j o_k):

$$P(Q \mid C_j \text{and} O_k) = \frac{1}{1 + \left(\frac{1-c_j}{c_j}\right)\left(\frac{1-o_k}{o_k}\right)} \tag{15.12}$$

from which we obtain

$$l_{GQ}(C_j) = \frac{1- c_j}{c_j} \tag{15.13}$$

$$l_{GQ}(O_k) = \frac{1-o_k}{o_k} \tag{15.14}$$

for the likelihood ratios in the FLMP. Comparing Equations 15.8 and 15.12, the BPM and FLMP (and any other that can be put in this form) are mathematically equivalent. We will call any model which can be put in *likelihood product form* a member of the class of likelihood product models (LPM).

15.5 Gaussian MDS Model (GMM)

This model considers perceptual features as dimensions in a multidimensional psychological space. Figure 15.6 illustrates the model using our QG categorization example. In the figure, the center of each prototype distribution is denoted by a symbol (e.g. Q in the upper right quadrant), surrounded by concentric circles indicating 1, 2, and 3 standard deviations from the center. Feature analysis locates a stimulus event at a particular point in the space. In our notation, [*obliqueness*, *closedness*] indicates a specific location as a coordinate pair in the multidimensional space. For expositional simplicity rather than theoretical necessity, we assume that the prototypes are symmetrically centered at [P, P] and [$-P$, $-P$]. In the example shown in Figure 15.6, P=1. Given a pair of perceptual features associated with a stimulus (denoted by S in Figure 15.6), the perceiver evaluates the similarity of S to each of the prototypes by transforming the distance between S and each prototype according to the Gaussian e^{-d^2} (Nosofsky, 1986; Ennis, 1988). We will examine two forms of the GMM using different metrics: a Euclidean GMM (GMM-E) and a city-block GMM (GMM-CB). For the GMM-E, the Euclidean distances $d(S, Q)$ and $d(S, G)$ from stimulus S at [o_k, c_j] to the center of the Q and G prototype distributions are given by

$$d(S, Q) = \sqrt{(P-o_k)^2 + (P-c_j)^2} \tag{15.15}$$

and

$$d(S, G) = \sqrt{(-P-o_k)^2 + (-P-c_j)^2}. \tag{15.16}$$

Transforming these Euclidean distances by the Gaussian function, we arrive at the similarities:

$$s(S, Q) = e^{-[(P-o_k)^2 + (P-c_j)^2]} \tag{15.17}$$

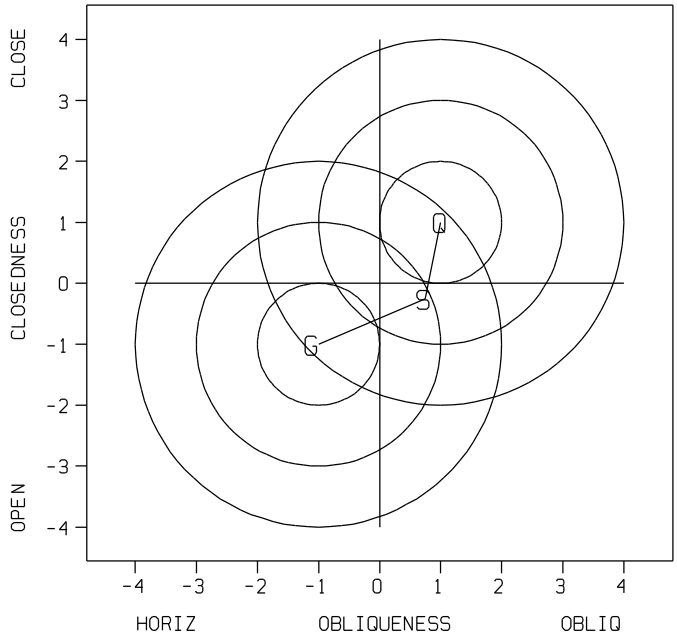

Figure 15.6. Gaussian multidimensional model of QG paradigm. The letters G and Q represent the center of the of the probabilistic prototype distributions corresponding to G and Q, respectively. The concentric circles indicate 1, 2, and 3 standard deviations from the center, as described by the Gaussian. The letter S represents a stimulus input with coordinates [o , c].

and

$$s(S, G) = e^{-[(-P-o_k)^2 + (-P-c_j)^2)} \tag{15.18}$$

Given these similarities, the perceiver uses an RGR (Shepard, 1957, Nosofsky, 1986) to decide among alternatives. Putting the similarities in an RGR, we have:

$$p(Q \mid c_j, o_k) = \frac{e^{-[(P-o_k)^2 + (P-c_j)^2]}}{e^{-[(P-o_k)^2 + (P-c_j)^2]} + e^{-[(-P-o_k)^2 + (-P-c_j)^2]}} \quad . \tag{15.19}$$

Because $e^{a+b} = e^a e^b$ we can factor Equation 15.19 as follows:

$$p(Q \mid c_j, o_k) = \frac{e^{-(P-o_k)^2} e^{-(P-c_j)^2}}{e^{-(P-o_k)^2} e^{-(P-c_j)^2} + e^{-(-P-o_k)^2} e^{-(-P-c_j)^2}} \quad . \tag{15.20}$$

With the factoring of the exponential terms into products, we see clearly the multiplicative combination of the two sources of information in the model. Given this factoring we can draw the equivalences:

$$P(c \mid Q) = e^{-(P-c_j)^2} \tag{15.21}$$

$$P(c \mid G) = e^{-(-P-c_j)^2} \tag{15.22}$$

$$P(o \mid Q) = e^{-(P-o_k)^2} \tag{15.23}$$

$$P(o \mid G) = e^{-(-P-o_k)^2} \tag{15.24}$$

For the *likelihood product form* of the model (viz Equation 15.8) we have:

$$l_{GQ}(C_j) = e^{[(P-c_j)^2 - (-P-c_j)^2]} = e^{-4\,P\,c_j} \tag{15.25}$$

$$l_{GQ}(O_k) = e^{[(P-o_k)^2 - (-P-o_k)^2]} = e^{-4\,P\,o_k} \tag{15.26}$$

Figure 15.7 shows how the hypothetical QG stimuli whose response probabilities are shown in Figure 15.5 are arranged in the two-dimensional space.

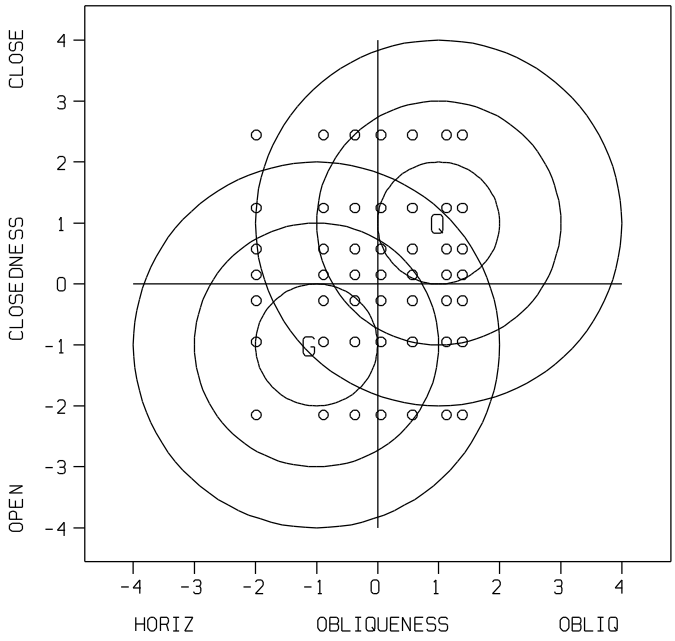

Figure 15.7. Gaussian multidimensional model of QG paradigm showing location of hypothetical stimuli assuming Euclidean distance metric.

To summarize, the GMM-E conceptualizes the representation of stimuli and prototypes as occurring in a multidimensional space, with the perceiver using the Gaussian transformed Euclidean distances as similarities which are then translated to response probabilities with an RGR. Theoretical analysis shows that the model is a member of the LPM class.

An interesting note regarding the use of a Euclidean distance measure here is that it is what makes the GMM-E consistent with the FLMP which is based on the assumption of *independent* fuzzy feature values. This runs contrary to the common belief that the Euclidean distance is more appropriate for integral dimensions, while the alternate city-block distance measure,

$$d(S, Q) = |P - o_k| + |P - c_j| \tag{15.27}$$

is more appropriate for separable dimensions (Shepard, 1984). For the GMM-CB in which the latter measure is used, the similarity (e.g. to Q) using a Gaussian function is

$$s(S, Q) = e^{-((P-o_k)^2 + 2|P-c_j||P-o_k| + (P-c_j)^2)}, \tag{15.28}$$

which is not factorable into two parts containing c and o independently, and thus not a member of the LPM class. Another way to look at this situation would be to say that the the critical factor here for independence is that the exponent α in the similarity function $e^{-d^{\alpha}}$ equals the exponent r in the Minkowski distance equation.

Although the original derivation of MDS stressed similarities of stimuli to response alternatives (and functions relating distances to similarities) in the MDS space, the model is mathematically equivalent to one based on probabilities of stimuli belonging to response categories. In this view, each alternative would be represented by a bivariate normal probability distribution (reflecting the variability of category instances in the world) centered at a particular location in the space, rather than by a single point. The MDS models discussed here can be considered relatively simple models, subsumed by the more general formalization called *General Recognition Theory* (Ashby & Gott, 1988; Ashby & Perrin 1988; Ashby & Townsend, 1986). Ashby and Perrin (1988) offer a general

Gaussian recognition model as an alternative to traditional MDS models. In that model, similarity is a function of the overlap of perceptual distributions. They argue that similarity judgments between two stimuli result from a judgment of the degree to which a pair of perceptual distributions overlap. The decision process divides the similarity space into response regions, one associated with each response. Identification judgments are not constrained by any distance axioms. Ashby and Gott (1988) found evidence for integration and for an optimal noise-free decision process in that subjects did not make independent decisions on each component or use the distance to each prototype. See chapters 6 and 16 for a more thorough discussion of *General Recognition Theory*.

15.6 Exponential MDS Model (EMM)

The Exponential MDS model (Shepard, 1957) is essentially equivalent to the Gaussian MDS model discussed in section 5, except that distance is translated into similarity by the negative exponential e^{-d} rather than the Gaussian e^{-d^2}. As with the GMM, we examine two different metrics: a Euclidean EMM (EMM-E) and a city-block EMM (EMM-CB). For the EMM-E, distances are computed in the same way as the GMM-E, given in Equations 15.15 and 15.16:

$$s(S, Q) = e^{-\sqrt{(P-o_k)^2 + (P-c_j)^2}} \tag{15.29}$$

and

$$s(S, G) = e^{-\sqrt{(-P-o_k)^2 + (-P-c_j)^2}} \quad . \tag{15.30}$$

Putting these similarities in an RGR, we have

$$p(Q \mid c_j, o_k) = \frac{e^{-\sqrt{(P-o_k)^2 + (P-c_j)^2}}}{e^{-\sqrt{(P-o_k)^2 + (P-c_j)^2}} + e^{-\sqrt{(-P-o_k)^2 + (-P-c_j)^2}}} \quad . \tag{15.31}$$

Because of the $\sqrt{\ }$ term in the exponentials, the support (similarities) for the two alternatives cannot be factored into two parts containing c and o independently. Thus, this model makes different mathematical predictions from those of the LPM class.

If a city-block rather than Euclidean distance is used, however, then the EMM model falls within the LPM. For the EMM-CB, the distances to each prototype are:

$$d(S, Q) = |P-o_k| + |P-c_j| \tag{15.32}$$

and

$$d(S, G) = |-P-o_k| + |-P-c_j| \tag{15.33}$$

and corresponding similarities, instead of Equations 15.29 and 15.30, will be:

$$s(S, Q) = e^{-(|P-o_k| + |P-c_j|)} \tag{15.34}$$

and

$$s(S, G) = e^{-(|-P-o_k| + |-P-c_j|)} \tag{15.35}$$

Putting these similarities in an RGR, we have

$$p(Q \mid c_j, o_k) = \frac{e^{-(|P-o_k| + |P-c_j|)}}{e^{-(|P-o_k| + |P-c_j|)} + e^{-(|-P-o_k| + |-P-c_j|)}} \tag{15.36}$$

which can be factored into

$$p(Q \mid c_j, o_k) = \frac{e^{-|P-o_k|}\, e^{-|P-c_j|}}{e^{-|P-o_k|}\, e^{-|P-c_j|} + e^{-|-P-o_k|}\, e^{-|-P-c_j|}} \quad . \tag{15.37}$$

We can then draw the equivalences:

$$P(c \mid Q) = e^{-|P - c_{EMM}|} \tag{15.38}$$

$$P(c \mid G) = e^{-|-P - c_{EMM}|} \tag{15.39}$$

$$P(o \mid Q) = e^{-|P - o_{EMM}|} \tag{15.40}$$

$$P(o \mid G) = e^{-|-P - o_{EMM}|}. \tag{15.41}$$

and construct the likelihood ratios:

$$l_{GQ}(C_j) = e^{(|P - c_j| - |-P - c_j|)} \tag{15.42}$$

$$l_{GQ}(O_k) = e^{(|P - o_k| - |-P - o_k|)} \tag{15.43}$$

for the *likelihood product form*. Figure 15.8 shows how the QG stimuli whose response probabilities are shown in Figure 15.5 are arranged in the two-dimensional space given the city-block metric. For this solution P was set to 6.

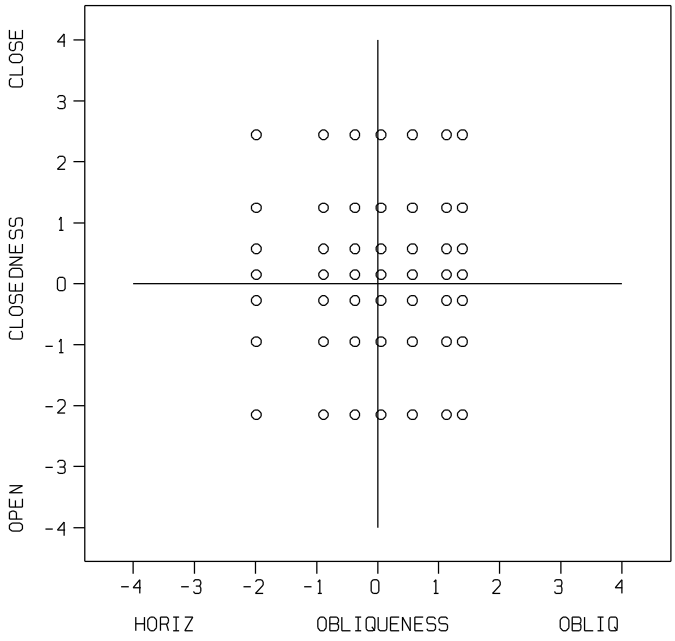

Figure 15.8. EMM multidimensional space for QG task showing location of hypothetical stimuli assuming a city block metric.

Thus, this EMM-CB is an LPM class member, and, at least for identification performance with two prototypes, the EMM-CB is equivalent to the GMM-E, both of which are LPM class members. This equivalence exists even though the two distance metrics make different predictions regarding ordering of distances between pairs of stimuli. When there are more than two prototype locations on a given dimension, however, the constraints imposed by the different similarity and distance functions between prototypes cause the EMM-CB and the GMM-E to make different predictions. Nosofsky (1985, 1987) proved this result for two dimensional data sets in which there are 16 prototypes, with 4 prototype locations equally spaced on each dimension. Given only two prototypes locations on each dimension, however, the locations of stimuli on each dimension are not constrained in this fashion.

15.7 Theory of Signal Detection (TSD)

The original TSD model (Green & Swets, 1966; Peterson, Birdsall, & Fox, 1954) offers a framework for a categorization model. This model also serves as a foundation for the multivariate TSD model *General Recognition Theory* (Ashby & Townsend, 1986). The multivariate TSD model is illustrated in Figure 15.9, as applied to our QG two response categorization task. As in the GMM model we have Q and G prototypes defined as distributions located symmetrically at $[P, P]$ and $[-P, -P]$ (with P=1) in a multidimensional feature space. We assume equal uncorrelated variance for each prototype (and stimulus) distribution. In terms of the notation of this volume we would say that the covariance matrix of the model is the identity matrix (ones on the main diagonal and zero elsewhere). We assume that the perceiver establishes a decision rule based on the use of a criterion at the line of equal likelihood between the two alternatives Q and G. This decision rule is equivalent to the minimum distance bound described by Ashby and Gott (1988). Given the symmetric locations of the alternative distributions, the criterion lies on the main anti-diagonal ($c + o = 0$).

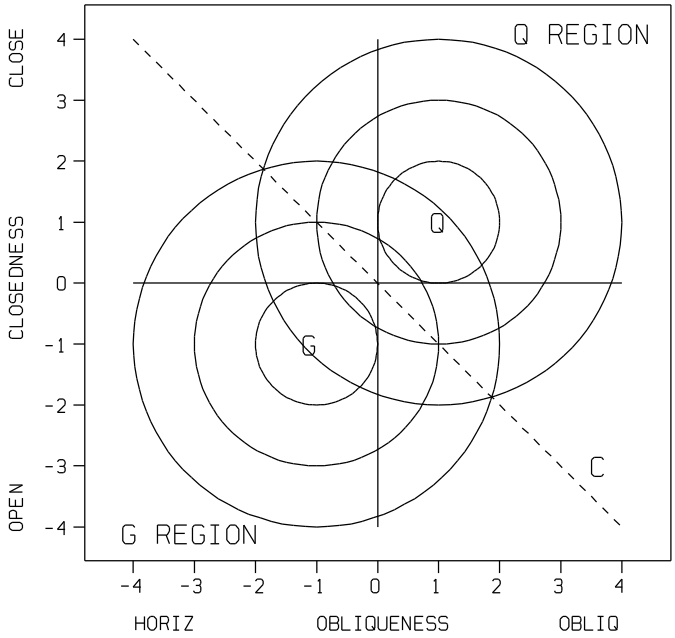

Figure 15.9. Multidimensional TSD model of QG task. The Q and G prototypes are defined as distributions located symmetrically in a multidimensional feature space, indicated by G and Q corresponding to the centers. In addition, a criterion (C) at the line of equal likelihood between the two prototype distributions is shown.

In this model it is assumed that stimuli are noisy, varying in location from trial to trial, with bivariate normal distributions. The distributions for three stimuli are shown in the center of the two-dimensional space in Figure 15.10. On a given trial, if the stimulus falls in the Q region, above and to the right of the criterion line, then a Q response is made. Similarly, for the G region below and to the left of the criterion line, a G response is made. Given that the stimuli are noisy, however, the two-dimensional location of a particular stimulus will sometimes fall in one region and sometimes in the other. To predict response probabilities, it is necessary to determine what volume of each stimulus distribution (what proportion) falls into each region. These proportions are related to the distance of the center of the distribution from the criterion line. Taking a hypothetical stimulus on the main diagonal centered at $[c, o]$ (e.g. the stimulus in Figure 15.10 at [.33, .33]), we can see that the Euclidean distance d back along the main diagonal to the criterion line is given by:

$$d = \sqrt{x^2 + x^2} = \sqrt{2\,x^2} = \sqrt{2}\,x \tag{15.44}$$

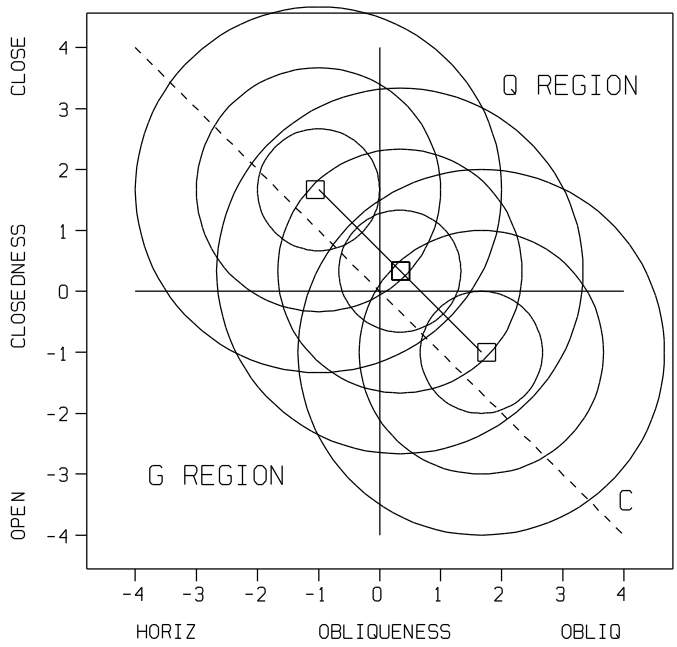

Figure 15.10. Multidimensional TSD model of QG task. Three stimulus distributions, eqidistant from the criterion line, are shown in the multidimensional space.

where $x = c = o$. Given that $x = (c + o)/2$, we arrive at

$$d = \sqrt{2}\,\frac{(c+o)}{2} = \frac{1}{\sqrt{2}}(c+o) \tag{15.45}$$

when the stimulus falls on the main diagonal. We can also see that the other two hypothetical stimuli which lie on the same anti-diagonal have the same distance to the criterion line, since an anti-diagonal line has an equation of the form $c + o = k$, with the result that the distance must always be $k/\sqrt{2}$. If we assume equal variance, uncorrelated normal distributions, then the probability of a Q response is given by:

$$P(Q \mid C_j \text{ and } O_k) = \Phi(\frac{1}{\sqrt{2}}(c_j + o_k)). \tag{15.46}$$

where $\Phi(x)$ is the cumulative standard normal distribution (see Chapter 1). We call this model the normal TSD (TSD-N). The predictions of this model are mathematically different from the LPM (to what degree, we will evaluate later).

Rather than the cumulative normal area Φ, suppose we substitute the similar corrected logistic cumulative function

$$L(x) = \frac{1}{1 + e^{-k\,x}} \tag{15.47}$$

(with correction factor $k = \frac{\pi}{\sqrt{3}}$ to equate for the variance $\frac{\pi^2}{3}$ of the logistic) in Equation 15.46. We then obtain:

$$P(Q \mid C_j \text{ and } O_k) = \frac{1}{1 + e^{-\frac{k}{\sqrt{2}}(c_j + o_k)}} = \frac{1}{1 + e^{-\frac{\pi}{\sqrt{6}}(c_j + o_k)}} \tag{15.48}$$

$$= \frac{1}{1 + e^{-\frac{\pi}{\sqrt{6}}c_j}\, e^{-\frac{\pi}{\sqrt{6}}o_k}} \quad . \tag{15.49}$$

We call this model the logistic TSD (TSD-L). This model is equivalent to the *likelihood product*

form with the assignments:

$$l_{GQ}(C_j) = e^{-\frac{\pi}{\sqrt{6}}c_j} \tag{15.50}$$

and

$$l_{GQ}(O_k) = e^{-\frac{\pi}{\sqrt{6}}o_k} \tag{15.51}$$

and we can also draw the equivalences:

$$e^{-\frac{\pi}{\sqrt{6}}c_{GMM}} = \frac{1-c_{FLMP}}{c_{FLMP}} \tag{15.52}$$

$$e^{-\frac{\pi}{\sqrt{6}}o_{GMM}} = \frac{1-o_{FLMP}}{o_{FLMP}} \tag{15.53}$$

or going in the other direction,

$$c_{GMM} = \frac{-\sqrt{6}}{\pi}\ln\left(\frac{1-c_{FLMP}}{c_{FLMP}}\right) \tag{15.54}$$

$$o_{GMM} = \frac{-\sqrt{6}}{\pi}\ln\left(\frac{1-o_{FLMP}}{o_{FLMP}}\right) \tag{15.55}$$

Given these relationships, Figure 15.11 shows how the QG stimuli whose hypothetical response probabilities are shown in Figure 15.5 are arranged in the two-dimensional space given the logistic TSD model. To summarize, we find a mathematical equivalence between the TSD-L and the LPM class of models, and a strong similarity (due to the similarity of the cumulative normal and the corrected logistic) to those theories for the TSD-N.

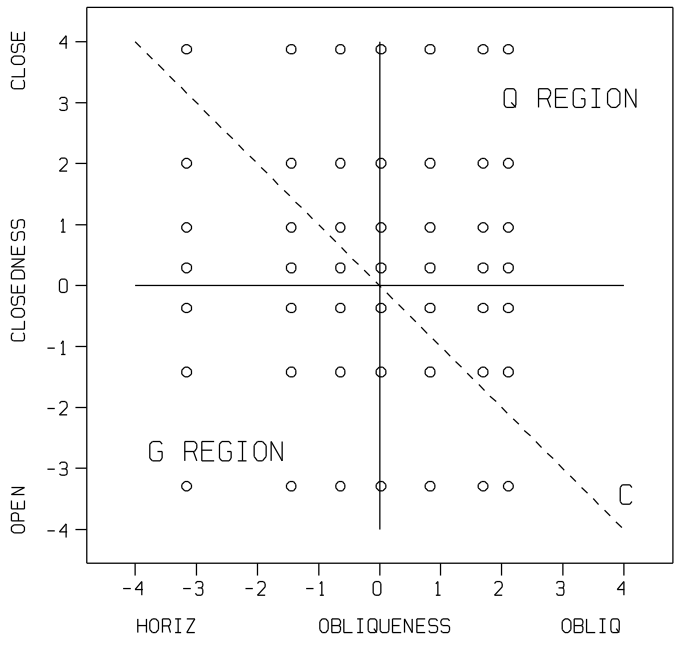

Figure 15.11. Multidimensional TSD-L model of QG task showing location of hypothetical stimuli assuming a logistic cumulative function.

Massaro and Friedman (1990) derived a somewhat different formulation of TSD. Their starting point was the derivation of Green and Swets (1966, Appendix 9-A) for the optimal combination

of two observations in a detection task. With both observations in the detection task taken into account, the d' value for the two-observation task was proven to be

$$d' = \sqrt{(d'_1)^2 + (d'_2)^2} \tag{15.56}$$

where d'_1 and d'_2 are the d's of the single observations. This solution was extended by Massaro and Friedman to take into account the directions of the component distances by adjusting the signs in Equation 15.56 to preserve direction. The current implementation of the TSD, however, places the observations in a multidimensional space, and derives the predictions on the basis of their location with respect to a criterion line. Thus, the model departs from the TSD framework as described by Green and Swets (1966), and has many of the same properties of the MDS models.

15.8 Feed-Forward Connectionist Model (FCM)

Network models are to some extent grounded in the metaphor of neural information processing. These models are usually referred to as connectionist, because information is represented in terms of the connections among the neural-like units (Minsky & Papert, 1968, 1988; Rosenblatt, 1958; Rumelhart & McClelland, 1986). These units are assumed to exist at different layers; for example, NETtalk (Sejnowski & Rosenberg, 1987) consists of units at the input, hidden, and output layers. The units interact with one another via connections with positive or negative weights that are either specified in advance or learned through feedback. Because of the extreme power of models with hidden units, we will consider here only a specific two-layer connectionist model that is most comparable to the alternative models that we have considered.

The FCM is assumed to have input and output layers of neural units, with all input units connected to all output units. It is assumed that each source of information and each response alternative is represented by a single unit at the input layer. Figure 15.12 gives a schematic representation of two input units connected to two output units for our QG task.

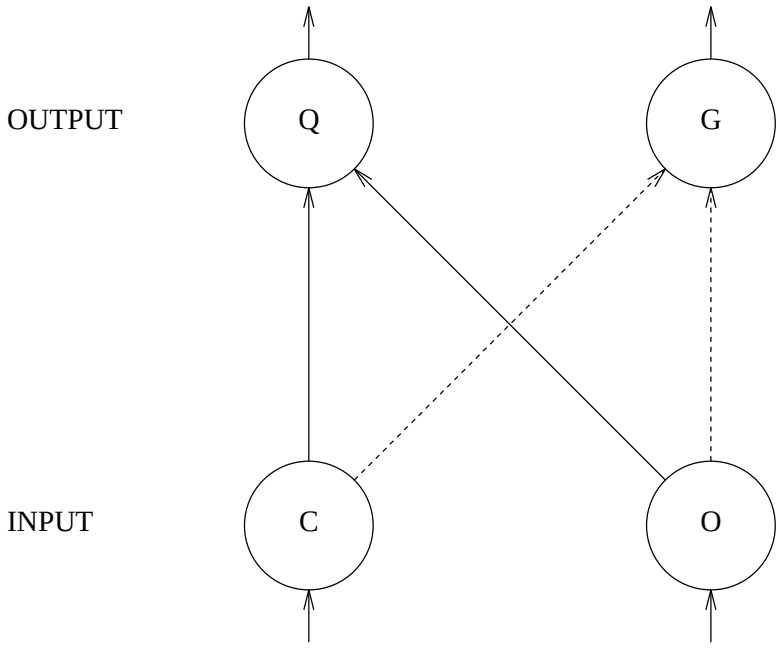

Figure 15.12. An FCM model for the QG task. The two layers of units contain input units corresponding to the features Closedness (C) and Obliqueness (O), and output units for the alternatives Q and G. Solid arrows indicate connections with weight 1, and dashed arrows indicate connections with weight -1.

The output units accumulate the sums of input activations. Each of these is given by the multiplicative combination of an input activation and a weight w which we assume is either 1 or -1. With two inputs c_j and o_k, the activation entering output unit Q is $(c_j + o_k)$, Analogous to the use of negation in the FLMP, it can be assumed that the weight on the activation entering output unit G is the negative of the weight entering Q (Massaro & Cohen, 1987). In this case, the activation entering output unit G is $(-c_j + -o_k)$. Although an additional "threshold" unit connected to each output unit is sometimes assumed in connectionist models, its use with just two response alternatives is not necessary to bring the model into line with the FLMP. The total activation leaving an output unit is given by the sum of the input activations, passed through a sigmoid squashing function (Rumelhart, Hinton, & Williams, 1986) (which is, incidentally, equivalent to the logistic function):

$$act_Q = \frac{1}{1 + e^{-(c_j + o_k)}} \tag{15.57}$$

$$act_G = \frac{1}{1 + e^{-(-c_j + -o_k)}}. \tag{15.58}$$

The "neural processing" of a connectionist model does not specify completely the stimulus-response function. The activations at the output layer have to be mapped into a response, and a RGR has been assumed to describe this mapping (Rumelhart et al., 1986):

$$P(Q \mid C_j \text{ and } O_k) = \frac{\frac{1}{1 + e^{-(c_j + o_k)}}}{\frac{1}{1 + e^{-(c_j + o_k)}} + \frac{1}{1 + e^{-(-c_j + -o_k)}}} \quad . \tag{15.59}$$

Noting the fact that

$$\frac{1}{1 + e^{-x}} + \frac{1}{1 + e^{x}} = 1 \quad , \tag{15.60}$$

the denominator of Equation 15.59 is equal to one, and we can simplify Equation 15.59 to:

$$P(Q \mid C_j \text{ and } O_k) = \frac{1}{1 + e^{-(c_j + o_k)}} = \frac{1}{1 + e^{-c_j}\, e^{-o_k}} \quad . \tag{15.61}$$

This is equivalent to the *likelihood product form* with the assignments:

$$l_{GQ}(C_j) = e^{-c_j} \tag{15.62}$$

and

$$l_{GQ}(O_k) = e^{-o_k} \tag{15.63}$$

and we can also draw the equivalences:

$$e^{-c_{FCM}} = \frac{1 - c_{FLMP}}{c_{FLMP}} \tag{15.64}$$

$$e^{-o_{FCM}} = \frac{1 - o_{FLMP}}{o_{FLMP}} \tag{15.65}$$

or going in the other direction,

$$c_{FCM} = -\ln\left(\frac{1 - c_{FLMP}}{c_{FLMP}}\right) \tag{15.66}$$

$$o_{FCM} = -\ln\left(\frac{1 - o_{FLMP}}{o_{FLMP}}\right) \quad . \tag{15.67}$$

To summarize, evaluation in the FCM consists of the activation and inhibition of neural-like units. Integration involves the summation of the separate activations which are then passed through the sigmoid squashing function. Decision follows the RGR. Mathematical analysis of this FCM reveals that it is also an LPM (for the case of two response alternatives).

15.9 Interactive Activation and Competition (IAC) Models

Interactive activation models are similar in spirit and design to feed forward models. However, IAC models assume that processing occurs over many processing cycles rather than just one. Furthermore, there are inhibitory lateral connections among units within a layer and two-way connections between units in different layers as illustrated in Figure 15.13.

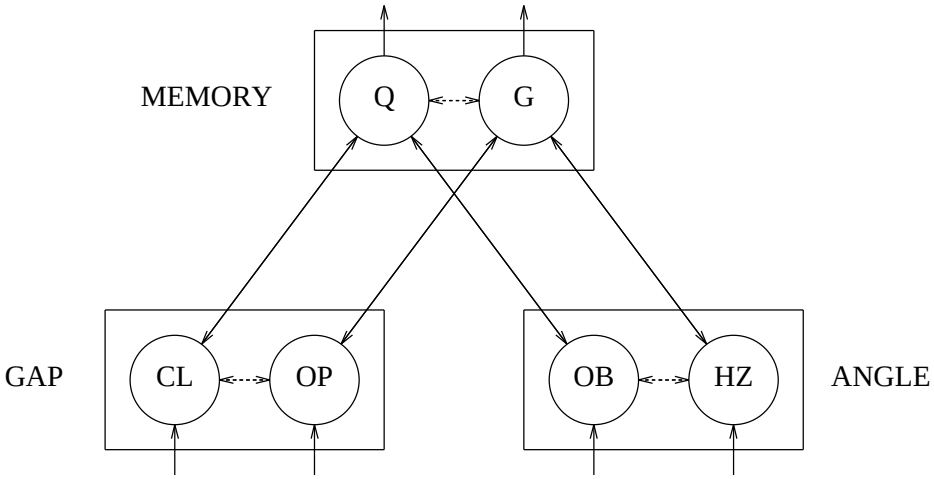

Figure 15.13. An IAC model for the QG task. The three layers of units contain input units corresponding to the gap and angle features, respectively, and to the "prototypes" or memory of G and Q. Units between layers have bidirectional excitatory connections indicated by solid arrows. Units within a layer have bidirectional inhibitory connections indicated by dashed arrows.

According to interactive activation, the information from one source can modify the processing (and representation) of other information sources (McClelland & Rumelhart, 1981; McClelland & Elman, 1986). Presentation of stimulus information from one source activates (and laterally inhibits) input units associated with that source, as illustrated in Figure 15.13. In this IAC model network, the two units for a given input layer receive complementary inputs. For example, if the CL unit receives .8 then the OP unit receives .2. These input units in turn activate the memory units of the top layer, which in turn feed backward and also activate the input units associated with the other input source, and so on during several cycles of processing. It should be noted that, contrary to most uses of the IAC, we are taking the activations of the highest level units as determinate of the response.

The formal algorithm of the IAC model is as follows (McClelland & Rumelhart, 1988). Initially, for each unit i, the activation act_i is set to the resting level $rest$. Then, on each computation 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 activation of the sending units act_i, and path weights, w_{ij} as follows:

$$exc_i = \sum_j \max(0, w_{ij}) \quad \max(0, act_j) \tag{15.68}$$

$$inh_i = \sum_j \min(0, w_{ij}) \quad \max(0, act_j) \tag{15.69}$$

All weights w_{ij} are either +1 or -1, so Equation 15.68 adds up all the activations on positive pathways and Equation 15.69 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 the exc_i, inh_i, and external inputs ext_i:

$$net_i = \alpha\, exc_i + \gamma\, inh_i + estr\ \ ext_i \tag{15.70}$$

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

$$\text{if } net_i > 0,\ \Delta act_i = net_i\,(M - act_i) - decay\,(act_i - rest) \tag{15.71}$$

$$\text{if } net_i < 0,\ \Delta act_i = net_i\,(act_i - m) - decay\,(act_i - rest) \tag{15.72}$$

where M is the maximum allowed activation, m is the minimum allowed activation, and $decay$ is the rate of activation decay. Then each act_i is adjusted by adding the Δact_i:

$$act_i = act_i + \Delta act_i \tag{15.73}$$

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

$$\text{if } act_i > M,\ act_i = M \tag{15.74}$$

$$\text{if } act_i < m,\ act_i = m \tag{15.75}$$

In most simulations, the IAC model has been used with the following standard set of control parameters: $estr$=.1, α=.1, γ=.1, $decay$=.1, M=1., m=-.2, and $rest$=-.1 (McClelland, 1990). In some simulations of the IAC model, running averages across cycles have been used rather than those directly computed at the cycle of interest (McClelland, 1990). We utilize the latter. Typically, 60 cycles of processing are carried out.

In McClelland and Rumelhart's (1981) original formulation of the model which we will call the IAC-RGR, the activations were translated into strengths by the exponential function:

$$S_i = e^{k\ act_i} \tag{15.76}$$

(with k commonly equal to 5) and these strengths were evaluated by an RGR to obtain the probability of a particular response. For the QG task, the form would be:

$$P(Q) = \frac{S_Q}{S_Q + S_G} \quad . \tag{15.77}$$

Due to certain incorrect predictions of the IAC model pointed out by Massaro (1989) (which will be further discussed in Section 15.10), the IAC model was revised to instead use a best-one-wins (BOW) decision rule applied directly to the activations, with normal noise being added to the input feature values. We call this model the IAC-BOW. We will examine both the original and revised IAC model.

The IAC model is basically additive with nonlinearities introduced by the change in additivity as activation boundaries are approached (Equations 15.71 and 15.72), by nonlinear decay (Equations 15.71 and 15.72), and by the activation and competition processes over time. Given the nonlinear characteristics of the model, it is unfortunately impossible to exactly examine the mathematical behavior of this model by a general closed form analysis, and we will have to be content with our simulations and empirical analysis given in Sections 15.10 and 15.11.

We note, however, an important demonstration by McClelland (1990) that a model closely related to the IAC-BOW — the *Boltzmann machine* (Hinton and Sejnowski, 1983, 1986) — at its equilibrium state can be reduced to multiplicative terms each representing a single source of information and is thus equivalent to the FLMP.

15.10 Model tests with hypothetical two-response data

The predictions of each model were fit to the predictions of the FLMP (the hypothetical 49 data points shown in Figure 15.5) using the program STEPIT (Chandler, 1969). A model is represented to the analysis program STEPIT as an algorithm for computing the sum of squared deviations between the observed and predicted data as a function of a set of parameters. By iteratively adjusting the parameters of the model, STEPIT minimizes the deviations between the observed and predicted points. Thus, STEPIT finds a set of parameter values which when put in the model, come closest to predicting the observed data. The metric of goodness of fit used is the root mean squared deviation (RMSD) between the observed and predicted data. Although many of these models are expected to fit the hypothetical data perfectly, it is of interest to examine the the parameter values from the fits. For some of these models we could directly compute the parameters from those found for the FLMP, but not all models have exact transforms possible, and it is worthwhile to carry out the fitting procedures for all models. It also may be, for example, that some of the parameter estimates are illogical or not meaningful, and so even though mathematically equivalent, models could be distinguished on the basis of the reasonableness of the parameter estimates. For the most part, the estimated parameters are presented in figure form.

For the fit of the FLMP, 14 parameters were used, 7 for the c feature values and 7 for the o features. There was a perfect fit to the data. The fit of the FLMP model to the data is shown in Figure 15.5.

For the fit of the BPM, 28 parameters were used, 7 each for the $P(c \mid Q)$, $P(c \mid G)$, $P(o \mid Q)$, and $P(o \mid G)$ probabilities. As expected, there was a perfect fit to the FLMP predictions.

For the fit of the GMM models, 14 parameters were used, 7 for the spatial c coordinates and 7 for the o coordinates. Figure 15.7 gives the locations of the best fitting stimulus locations for the GMM-E (a perfect fit). Although theoretically different, the GMM-CB provided an extremely close fit (RMSD=.0060) with a slightly different arrangement of the stimuli.

The EMM models also used 14 parameters for spatial coordinates. Figure 15.8 gives the locations of the best fitting stimulus locations for the EMM-CB (a perfect fit). Interestingly, the EMM-E gave an almost perfect fit of (RMSD=.0002). This similarity of models has been previously noted by Indow (1974) and Ennis (1988), who reported a number of cases in which city-block data is well fit by Euclidean distance.

The TSD models also used 14 parameters for spatial coordinates. Figure 15.11 gives the locations of the best fitting stimulus locations for the TSD-L (a perfect fit). Although theoretically different from the FLMP, the TSD-N did a respectable (RMSD=.0064) job at mimicking the FLMP predictions.

The fit of the IAC models is somewhat more complex. We begin with a fit of the IAC-RGR with the typical k value of 5. This model used 14 parameters representing the 7 c and 7 o input feature values. To fit this model, the IAC algorithm (McClelland & Rumelhart, 1988) (recoded in FORTRAN (f77) for use with STEPIT) was run to compute the target activations after 60 time cycles. The fit of this model was relatively poor (RMSD=.0457). To understand the nature of the problem the IAC-RGR has in mimicking the FLMP predictions, we plot in the left panel of Figure 15.14 the fit transformed by the inverse logistic. In the figure, we can see that while the inverse logistic of the FLMP hypothetical data is additive, five of the seven IAC-RGR predicted lines tend to converge at both ends. This nonadditivity of the effects of the two variables was originally pointed out by Massaro (1989) in an analysis of the combination of phonological context and stimulus information in the TRACE model (McClelland & Elman, 1986), which is closely related to the IAC. This nonadditivity disagrees with the classically observed additivity

To improve the fit of the IAC-RGR, we set the k value to 15. The right panel of Figure 15.14 gives the inverse logistic of the predicted probability of a Q response for the FLMP and the IAC-

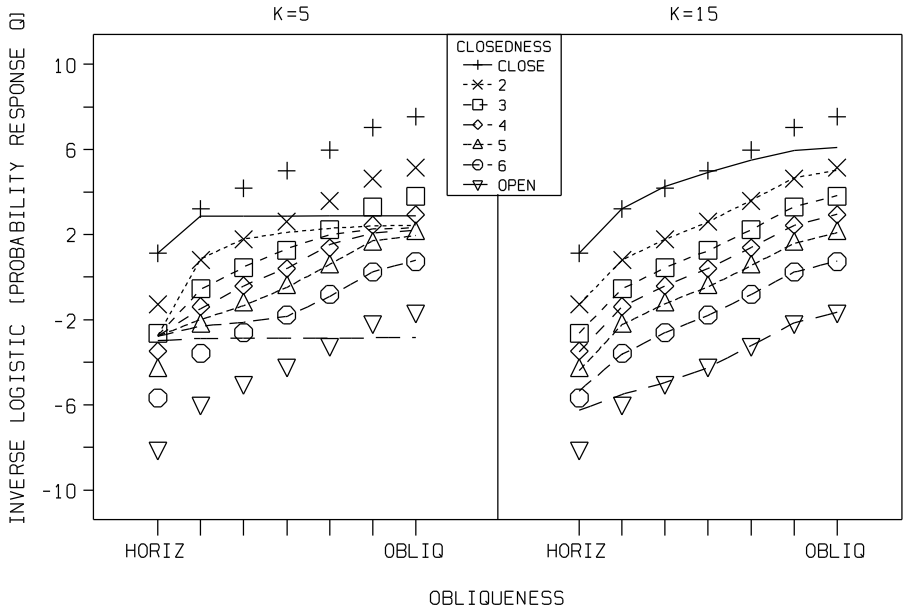

Figure 15.14. Inverse logistic transform of the predicted probability of a Q response according to the FLMP (points) and IAC-RGR model (lines) for k=5 (left panel) and k=15 (right panel).

RGR with k=15. As can be seen in this figure, the IAC-RGR with k=15 gave a good much better fit (RMSD=.0007), although there was still a slight nonadditivity. Table 15.1 gives the best fitting feature values found by STEPIT for the k=5 and k=15 cases. Although using larger k values allows better mimicking of the FLMP, this is an unattractive solution because the input feature values take on unrealistic values.

Table 15.1. The estimated input feature values for the two IAC-RGR models (with k=5 and k=15) for the best fit of the hypothetical data.

		LEVEL							
FEATURE	k	1	2	3	4	5	6	7	RANGE
CLOSEDNESS	5	.0001	.4869	.4946	.4993	.5038	.5112	.9999	.9998
	15	.4966	.5007	.5030	.5044	.5059	.5082	.5123	.0157
OBLIQUENESS	5	.2636	.4932	.4986	.5029	.5085	.5145	.5173	.2537
	15	.4896	.4932	.4949	.4963	.4980	.4998	.5006	.0110

For the IAC-RGR to mimick the FLMP predictions, the input features must be very close to neutral values, with a very small range of feature values. In fact, a small range of feature values also occurred for the interior five levels of the k=5 case. Thus, the IAC-RGR model can only mimick the FLMP predictions by having the activations of the response units fall in a relatively neutral, non-asymptotic range. The left and right panels of Figure 15.15 give the response activations for the Q and G units for the k=5 and k=15 cases respectively. Within the relatively neutral range, the activation values are less susceptible to the nonlinearities of the IAC and therefore can better approximate the hypothetical data. Thus the model involves something closer to an addition of individual effects, transformed at the end by an exponential — that is, more like the multiplication of the two effects (which is what happens in the FLMP and related models). Figure 15.16 illustrates how the activations from the two cases are shaped by transforming them into strength values. The difference in the strength scales on the two graphs should be noted; the strengths resulting from k=15 are about 15

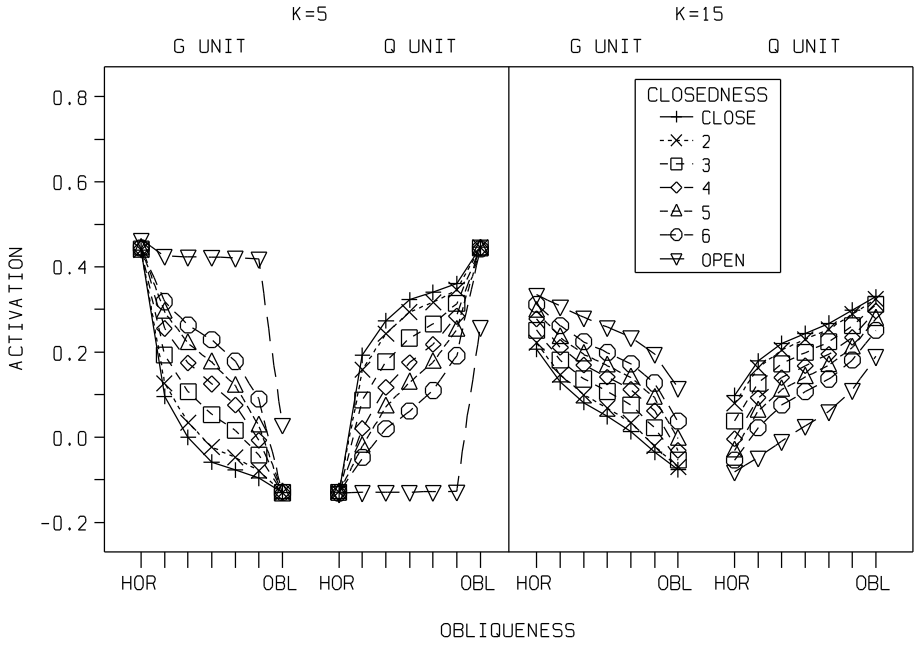

Figure 15.15. Activation of the G and Q units for the IAC-RGR model with k=5 (left panel) and k=15 (right panel) given the hypothetical parameter values in Table 15.1.

times larger. Comparing the right panel of Figure 15.16 with Figure 15.4, we see that the IAC-RGR with k=15 is able to replicate to some extent the multiplicative fan predicted by the FLMP. In summary, the IAC-RGR can apparently mimic the noninteractive nature of the LPM, but only by staying in a restricted linear parameter range. The small range of feature values is particularly damaging to a neural network interpretation because neurons are unlikely to represent activations to this fine level of resolution.

We now evaluate the IAC-BOW model with its alternative decision process of choosing the response with the highest activation. To compute the model's predictions, we first set the 49 probabilities of a Q response given the 7 closedness times 7 obliqueness conditions to 0 and reset the random number generator. Then for each of the 49 conditions 1000 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; see also Chapter 1) with a standard deviation (which was allowed to vary) initially set at .14142, were added to each of the current pair of parameters (for closedness and obliqueness). Then the IAC algorithm (McClelland & Rumelhart, 1988) was run to compute the target activations after 60 time cycles. If the final activation of the Q unit was greater than or equal to the final activation of the G target, then 1/1000 was added to the probability of a Q response. In order for the parameter estimation routine to operate properly it was necessary to employ the same sequence of random numbers on each overall computation run. This allowed STEPIT to make reliable adjustments in the parameter values, even though noise was being added to the input.

The IAC with the BOW decision rule gave a good fit to the FLMP predictions (RMSD=.0090) with the noise standard deviation at .1251. Figure 15.17 shows the fit of the model in terms of the inverse logistic of the proportion of Q responses. We note some deviations from additivity, but those are mostly due to the problem of computing the inverse logistic for 0 and 1. In those cases, the value was set at minus or plus 10.

Table 15.2 summarizes the fits of the various models, giving the root mean squared deviations (RMSDs) between the FLMP predictions and the predicted data for each model. As can be seen in

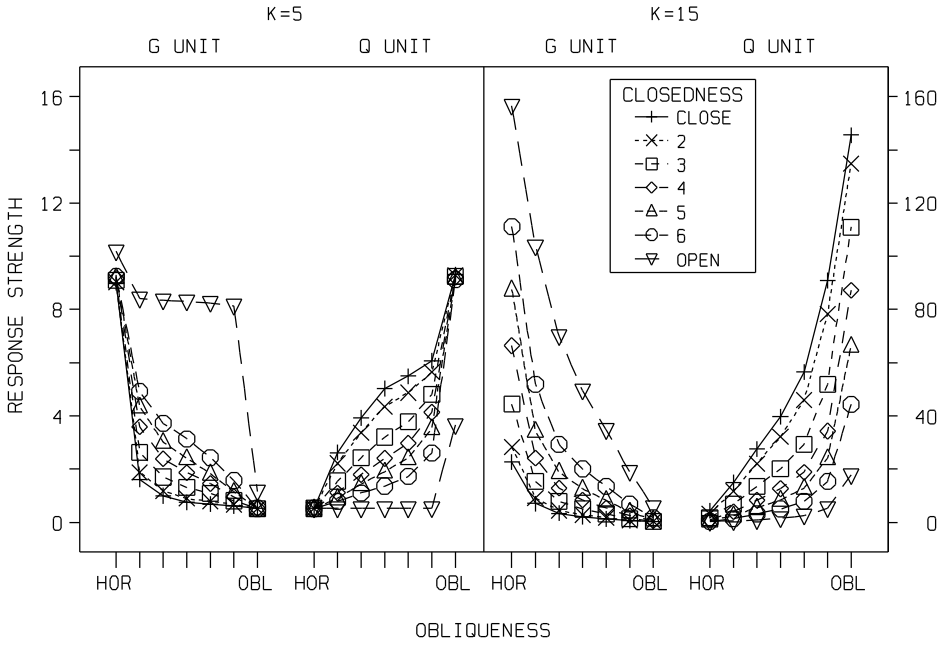

Figure 15.16. The strengths of the G and Q response alternatives for the IAC-RGR model with k=5 (left panel) and k=15 (right panel) given the activations in Figure 15.15. Note difference in response strength scales for the two panels.

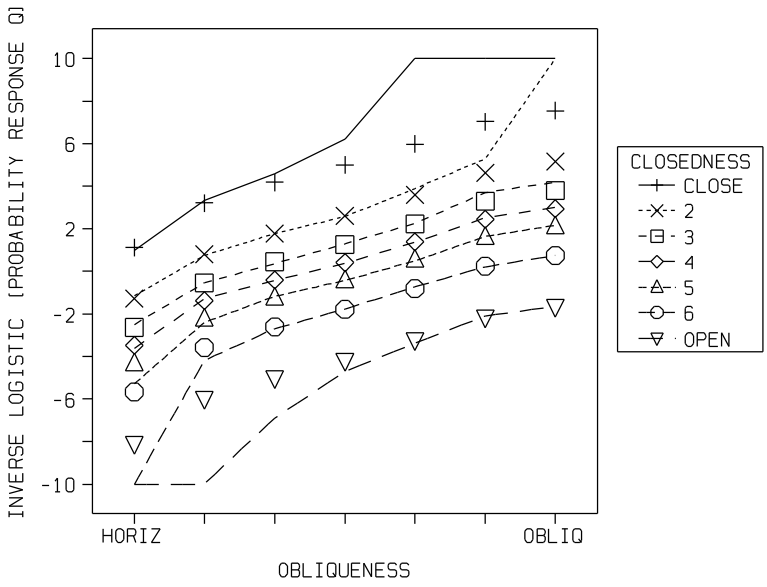

Figure 15.17. Inverse logistic transform of the predicted probability of a Q response according to the FLMP (points) and IAC-BOW model (lines).

the table, the FLMP, BPM, GMM-E, EMM-CB, TSD-L, and FCM models all had perfect fits to the FLMP hypothetical data, as predicted by the theoretical analysis. Of the models not exactly equivalent to the FLMP, the GMM-CB, EMM-E, TSD-N, IAC-RGR15 did a good job at mimicking the FLMP, the latter model doing so with unrealistic parameter values. Only the IAC-RGR5 made different predictions.

Table 15.2. Model fits of hypothetical FLMP data.

MODEL	RMSD
FLMP	.0000
BPM	.0000
GMM-E	.0000
GMM-CB	.0060
EMM-E	.0002
EMM-CB	.0000
TSD-L	.0000
TSD-N	.0064
FCM	.0000
IAC-RGR5	.0457
IAC-RGR15	.0007
IAC-BOW	.0090

15.11 Model tests with experimental QG data

Given the subtle differences between the LPM class and most of the others (and the large difference for the IAC-RGR5) observed for the pure hypothetical data, it is of interest to see whether we can discriminate between these models with real experimental data. Of course, we expect all the LPM class to give the same goodness of fit and predictions, but it is worthwhile to present the fits so that the parameters of the models can be examined. Table 15.3 gives the RMSD for each model for each of the nine subjects in the QG task, a fit of a mean subject, and the mean of the subject fits. We will graph the observed and predicted data for the two typical subjects shown in Figure 15.2.

Table 15.3. RMSDs for model fits of QG experimental data for 9 individual subjects, the fit of the mean subject, and the average of the 9 subject fits.

MODEL	S1	S2	S3	S4	S5	S6	S7	S8	S9	SM	AVE
FLMP	.0346	.0637	.0973	.0444	.0312	.0543	.0487	.0284	.0539	.0369	.0507
BPM	.0346	.0637	.0973	.0444	.0312	.0543	.0487	.0284	.0539	.0369	.0507
GMM-E	.0346	.0637	.0973	.0444	.0312	.0543	.0487	.0284	.0539	.0369	.0507
GMM-CB	.0432	.0637	.0974	.0449	.0290	.0629	.0485	.0302	.0542	.0371	.0527
EMM-E	.0346	.0636	.0973	.0444	.0312	.0543	.0487	.0284	.0539	.0369	.0507
EMM-CB	.0346	.0637	.0973	.0444	.0312	.0543	.0487	.0284	.0539	.0369	.0507
TSD-L	.0346	.0637	.0973	.0444	.0312	.0543	.0487	.0284	.0539	.0369	.0507
TSD-N	.0354	.0624	.0961	.0446	.0322	.0561	.0483	.0292	.0542	.0384	.0509
FCM	.0346	.0637	.0973	.0444	.0312	.0543	.0487	.0284	.0539	.0369	.0507
IAC-RGR5	.0518	.0719	.1014	.0610	.0482	.0569	.0645	.0487	.0611	.0445	.0628
IAC-RGRK	.0327	.0634	.0991	.0445	.0297	.0493	.0476	.0286	.0536	.0350	.0498
IAC-BOW	.0358	.0604	.0968	.0487	.0329	.0571	.0480	.0296	.0525	.0378	.0513

For the fit of the FLMP, 14 parameters were used, 7 for the *c* feature values and 7 for the *o* features. The average RMSD for the model was .0507. Figure 15.18 shows the observed and predicted data for the typical subjects with the abscissa scaled according to the obtained obliqueness parameters. Table 15.3 shows that all models of the LPM class achieved the same goodness of fit. The related GMM-CB, EMM-E, TSD-N, IAC-RGR15 did not give a significantly different fit. Of particular interest is the goodness of fit of the IAC-RGR5 — the only model different from the FLMP and related models in fitting the hypothetical data in Section 15.10. The RMSDs of the IAC-

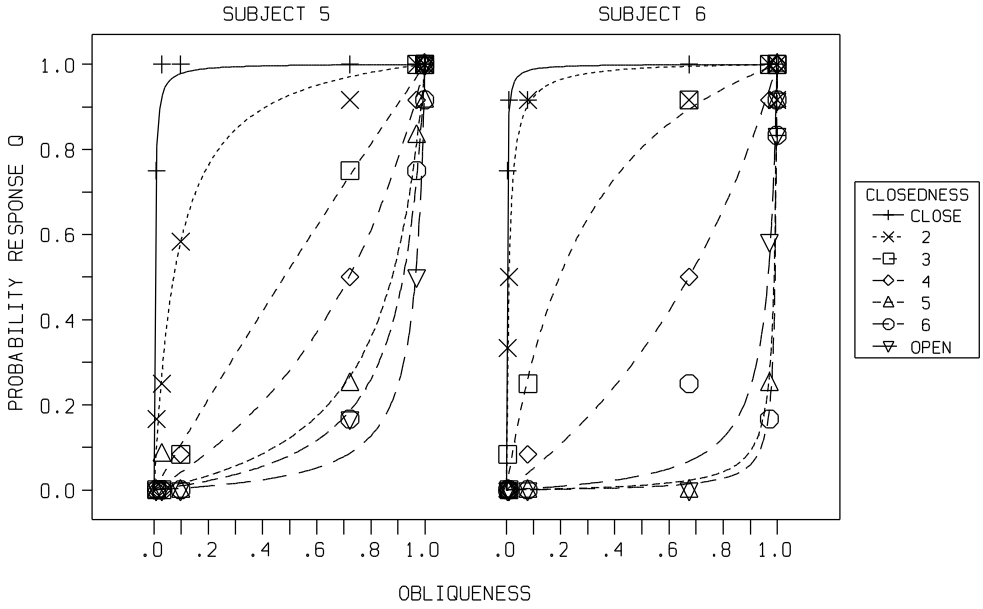

Figure 15.18. Observed (points) and predicted (lines) probability of Q responses for the forty-nine test letters presented in Figure 15.1 for two typical subjects. The predictions are for the FLMP model and the locations on the abscissa are scaled according to the estimated value of the obliqueness feature.

RGR5 were compared with those of the FLMP using an ANOVA for the 9 subject fits and found to be significantly worse for the IAC-RGR5, ($F(1,8)=30.64$, $p=.001$). Therefore, at least one of the models can be distinguished on the basis of actual results and the IAC-RGR5 is falsified by the fit of the data from the QG experiment of Massaro and Hary (1986).

15.12 Prototypical Four-Response Categorization Task

Given the successful reconciliation of a number of models for the two response task, it is worthwhile to extend our analysis to a prototypical four response alternative task, carried out by Massaro, Tseng, and Cohen (1983). The four responses in the experiment were four words in Mandarin Chinese. The experiment used a factorial design with seven levels of each of two factors. These factors were the formant structure of the vowel in the monosyllabic words and the fundamental frequency (F_0) contour (tone) during the vowel. Mandarin Chinese is a tone language and both of these sources of information are functional to distinguish different words. The formant structure was varied to make a continuum of vowel sounds between /i/ and /y/. (The vowel /y/ is articulated in the same manner as /i/, except with the lips rounded.) The F_0 contour varied between falling-rising to falling during the vowel. Six native Chinese speakers participated for four days, with each subject giving a total of 48 responses to each of the 49 test stimuli. The subjects identified each of the 49 test stimuli as one of the four words. Figure 15.19 shows the data for a single typical subject as a function of vowel and tone level for the four responses.

15.13 Four-response FLMP

In the four response version of the FLMP the prototypes are defined by:

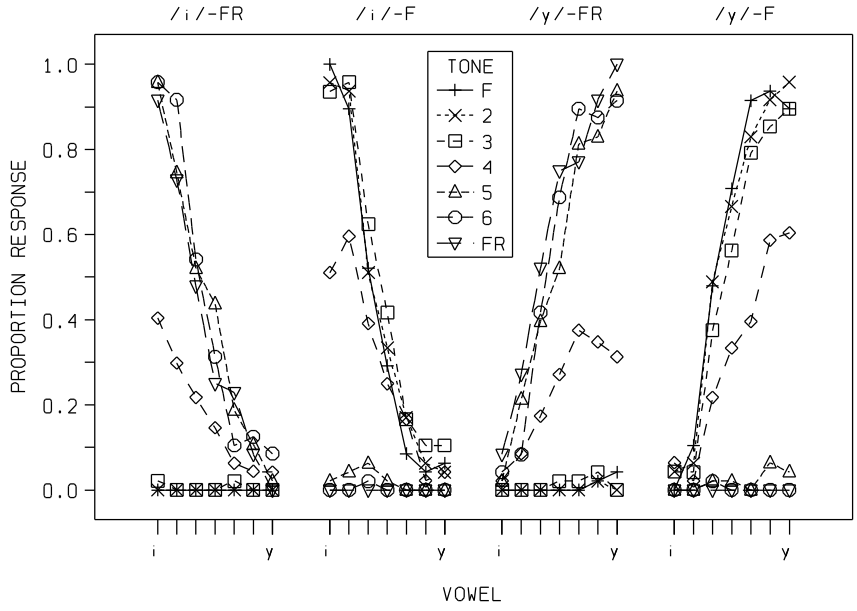

Figure 15.19. Mean observed proportion of responses for a typical subject as a function of vowel and tone level for the four responses. (After Massaro, Tseng & Cohen, 1983).

i-FR: vowel /i/ & falling-rising tone
i-F : vowel /i/ & falling tone
y-FR: vowel /y/ & falling-rising tone
y-F : vowel /y/ & falling tone

In the implementation of the model, it is assumed that the vowel /i/ and vowel /y/ features are opposites (or negations) of one another, as are falling-rising and falling tones. Thus, we can represent the prototypes' goodnesses in terms of the degree to which the vowel is /y/ (y) and the tone is falling (F), and with the multiplicative definition of conjunction:

i-FR: $(1\text{-}y) \times (1\text{-}F)$
i-F : $(1\text{-}y) \times F$
y-FR: $y \times (1\text{-}F)$
y-F : $y \times F$

These prototypes are then evaluated by a RGR. The probability of an y-F response, for example, is given by:

$$P(y\text{–}F \mid Y_j \text{ and } F_k) = \frac{(y_j)(f_k)}{(1-y_j)(1-f_k) + (1-y_j)(f_k) + (y_j)(1-f_k) + (y_j)(f_k)} \quad . \tag{15.78}$$

Noting that the denominator of Equation 15.78 is always the quantity 1, we can simplify the response predictions to:

$$P(i\text{–}FR \mid Y_j \text{ and } F_k) = (1-y_j)(1-f_k) \tag{15.79}$$

$$P(i\text{–}F \mid Y_j \text{ and } F_k) = (1-y_j)(f_k) \tag{15.80}$$

$$P(y\text{–}FR \mid Y_j \text{ and } F_k) = (y_j)(1-f_k) \tag{15.81}$$

$$P(y\text{–}F \mid Y_j \text{ and } F_k) = (y_j)(f_k) \tag{15.82}$$

To generate the FLMP predictions, the following hypothetical feature values were used for vowel /y/, going from /i/ to /y/: .01, .10, .30, .50, .70, .90, .99, and for tone falling, going from falling-rising

to falling: .03, .20, .40, .60, .80, .92, .95. Figure 15.20 shows the support for the four alternatives given these hypothetical feature values (and also the probability of responding with those alternatives since the denominator of the RGR is one).

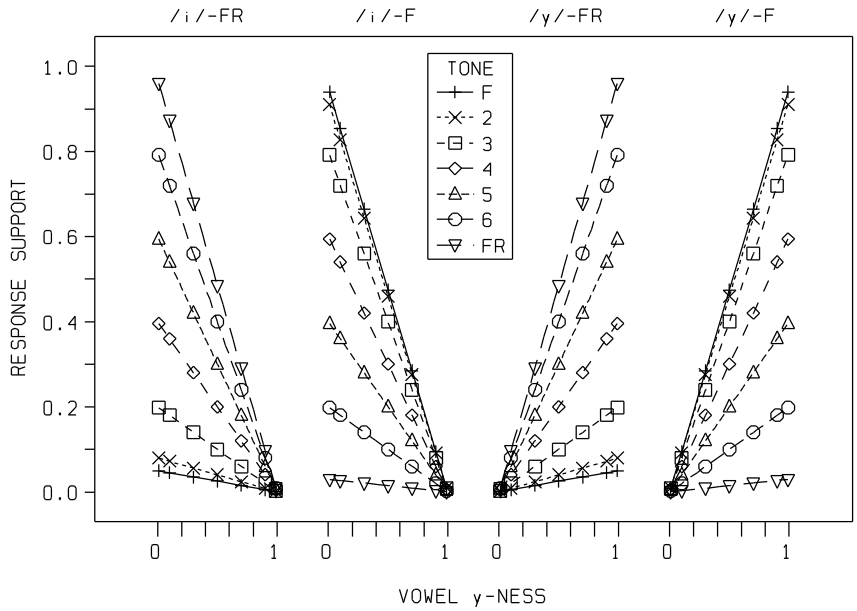

Figure 15.20. Support for alternative i-FR, i-F, y-FR, and y-F prototypes based on hypothetical vowel and tone features given in the text. The locations on the abscissa are scaled according to the y-ness of the vowel feature.

15.14 Four-response BPM

For the four response BPM, the probabilities of the four responses are computed from the probabilities of evidence $P(Y_j \mid Y)$, $P(Y_j \mid I)$, $P(F_k \mid FR)$, $P(F_k \mid F)$. The probability of a y-F response, for example would be given by:

$$P(y\text{–}F \mid Y_j \text{ and } F_k) = \frac{P(Y_j \mid Y)\,P(F_k \mid F)}{P(Y_j \mid I)\,P(F_k \mid FR) + P(Y_j \mid I)\,P(F_k \mid F) + P(Y_j \mid Y)\,P(F_k \mid FR) + P(Y_j \mid Y)\,P(F_k \mid F)} \quad . \tag{15.83}$$

As with the 2 response case, the FLMP and the BPM differ in that for the FLMP the alternatives are defined as using *complementary* features. Thus, in the FLMP, i and FR are assumed to be complements of y and F, respectively. For the BPM, however, the probabilities $P(e \mid h_1)$ and $P(e \mid h_2)$ need not sum to one. Let us consider again Equation 15.83 which, by dividing top and bottom by $P(Y_j \mid Y)\,P(F_k \mid F)$ can be rewritten as:

$$P(y\text{–}F \mid Y_j \text{ and } F_k) = \frac{1}{\left[\frac{P(Y_j \mid I)\,P(F_k \mid FR)}{P(Y_j \mid Y)\,P(F_k \mid F)}\right] + \left[\frac{P(Y_j \mid I)\,P(F_k \mid F)}{P(Y_j \mid Y)\,P(F_k \mid F)}\right] + \left[\frac{P(Y_j \mid Y)\,P(F_k \mid FR)}{P(Y_j \mid Y)\,P(F_k \mid F)}\right] + 1} \tag{15.84}$$

$$= \frac{1}{\left[\frac{P(Y_j \mid I)}{P(Y_j \mid Y)}\right]\left[\frac{P(F_k \mid FR)}{P(F_k \mid F)}\right] + \left[\frac{P(Y_j \mid I)}{P(Y_j \mid Y)}\right]\left[\frac{P(F_k \mid F)}{P(F_k \mid F)}\right] + \left[\frac{P(Y_j \mid Y)}{P(Y_j \mid Y)}\right]\left[\frac{P(F_k \mid FR)}{P(F_k \mid F)}\right] + 1}$$

$$= \frac{1}{\left[\frac{P(Y_j \mid I)}{P(Y_j \mid Y)}\right]\left[\frac{P(F_k \mid FR)}{P(F_k \mid F)}\right] + \left[\frac{P(Y_j \mid I)}{P(Y_j \mid Y)}\right] + \left[\frac{P(F_k \mid FR)}{P(F_k \mid F)}\right] + 1}$$

or the equivalent:

$$= \frac{1}{l_{IY}(Y_j)\,l_{FR\,F}(F_k) + l_{IY}(Y_j) + l_{FR\,F}(F_k) + 1} \quad . \tag{15.85}$$

where $l_{IY}(Y_j)$ is the likelihood ratio of vowel i to vowel y given evidence Y_j and $l_{FR\,F}(F_k)$ is the likelihood ratio of tone FR to tone F given evidence F_k. The *likelihood product form* (i.e. Equation 15.85) for other responses also requires $l_{YI}(Y_j)$ which is the reciprocal of $l_{IY}(Y_j)$, and $l_{F\,FR}(F_k)$ which is the reciprocal of $l_{FR\,F}(F_k)$. Thus, for each combination of closedness and obliqueness conditions, only two likelihood ratios (one for each source of evidence) are needed. For the BPM, the likelihood ratios are given by:

$$l_{IY}(Y_j) = \frac{P(Y_j \mid I)}{P(Y_j \mid Y)} \tag{15.86}$$

and

$$l_{FR\,F}(F_k) = \frac{P(F_k \mid FR)}{P(F_k \mid F)} \quad . \tag{15.87}$$

Similarly, for the FLMP, the likelihood ratios are given by:

$$l_{IY}(Y_j) = \frac{1-y_j}{y_j} \tag{15.88}$$

and

$$l_{FR\,F}(F_k) = \frac{1-f_k}{f_k} \tag{15.89}$$

As in the two-alternative case, both the FLMP and BPM are members of the LPM class.

15.15 Four-response GMM

The GMM for the four response task is similar to that for two responses except that there are four prototype locations. We assume that the prototypes are symmetrically centered in a multidimensional space with dimensions vowel and tone, at $[P, P]$ for y-F, at $[P, -P]$ for y-FR, at $[-P, P]$ for i-F, and at $[-P, -P]$ for i-FR. Figure 15.21 shows the arrangement of the prototype distributions in the multidimensional space. The Euclidean distances from stimulus S at $[y_j, f_k]$ to the four prototype locations are given by:

$$d(S, y\text{–}F) = \sqrt{(P-y_j)^2 + (P-f_k)^2} \tag{15.90}$$

$$d(S, y\text{–}FR) = \sqrt{(P-y_j)^2 + (-P-f_k)^2} \tag{15.91}$$

$$d(S, i\text{–}F) = \sqrt{(-P-y_j)^2 + (P-f_k)^2} \tag{15.92}$$

and

$$d(S, i\text{–}FR) = \sqrt{(-P-y_j)^2 + (-P-f_k)^2} \quad . \tag{15.93}$$

Given the Gaussian similarity function, we can compute the similarities

$$s(S, y\text{–}F) = e^{-[(P-y_j)^2 + (P-f_k)^2]} \tag{15.94}$$

$$s(S, y\text{–}FR) = e^{-[(P-y_j)^2 + (-P-f_k)^2]} \tag{15.95}$$

$$s(S, i\text{–}F) = e^{-[(-P-y_j)^2 + (P-f_k)^2]} \tag{15.96}$$

$$s(S, i\text{–}FR) = e^{-[(-P-y_j)^2 + (-P-f_k)^2]} \quad . \tag{15.97}$$

These similarities can then be factored into:

$$s(S, y\text{–}F) = e^{-(P-y_j)^2}\, e^{-(P-f_k)^2} \tag{15.98}$$

$$s(S, y\text{–}FR) = e^{-(P-y_j)^2}\, e^{-(-P-f_k)^2} \tag{15.99}$$

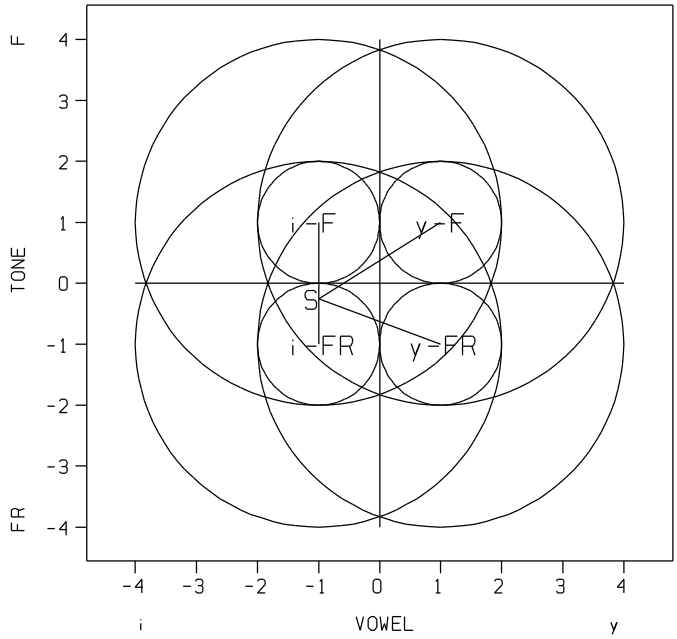

Figure 15.21. Gaussian multidimensional model of four response paradigm. The labels y-F, y-FR, i-F and i-FR represent the center of the of the probabilistic prototype distributions corresponding to the four response categories. The concentric circles indicate 1, and 3 standard deviations from the center, as described by the bivariate normal.

$$s(S, i\text{–}F) = e^{-(-P-y_j)^2}\, e^{-(P-f_k)^2} \tag{15.100}$$

$$s(S, i\text{–}FR) = e^{-(-P-y_j)^2}\, e^{-(-P-f_k)^2}. \tag{15.101}$$

from which we can draw, using the terminology of the FLMP and BPM:

$$p(Y_j \mid Y) = e^{-(P-y_j)^2} \tag{15.102}$$

$$p(Y_j \mid I) = e^{-(-P-y_j)^2} \tag{15.103}$$

$$p(F_k \mid F) = e^{-(P-f_k)^2} \tag{15.104}$$

$$p(F_k \mid FR) = e^{-(-P-f_k)^2} \tag{15.105}$$

With these probabilities obtained, we can use Equation 15.83 for the rest of the derivation. For the *likelihood product form* we derive the likelihood ratios:

$$l_{IY}(Y_j) = e^{[(P-y_j)^2 - (-P-y_j)^2]} = e^{-4\,P\,y_j} \tag{15.106}$$

$$l_{FR\,F}(F_k) = e^{[(P-f_k)^2 - (-P-f_k)^2]} = e^{-4\,P\,f_k} \tag{15.107}$$

Figure 15.22 shows how our 49 hypothetical stimuli are arranged in the multidimensional space.

For the alternate GMM-CB model with the city-block distance measure, the distance between the stimulus S and, for example, y-F would be:

$$d(S, y\text{–}F) = |P-y_j| + |P-f_k| \tag{15.108}$$

for which the similarity function would be:

$$s(S, y\text{–}F) = e^{-((P-f_k)^2 + 2|P-y_j||P-f_k| + (P-y_j)^2)} \tag{15.109}$$

As with the two-response case, these similarities are not factorable, and thus the GMM-CB is not a

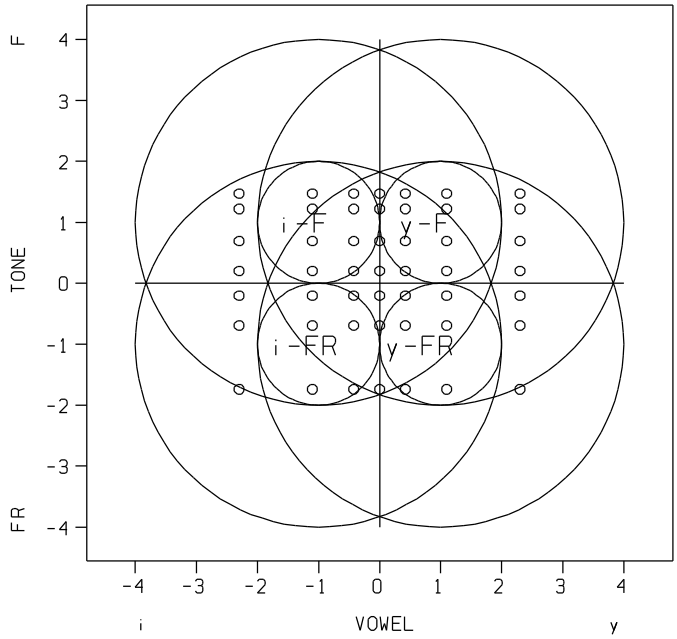

Figure 15.22. Gaussian multidimensional model of four response paradigm. The labels y-F, y-FR, i-F and i-FR represent the center of the of the probabilistic prototype distributions corresponding to the four response categories. The concentric circles indicate 1, and 3 standard deviations from the center, as described by the bivariate normal. The small circles give the locations of the hypothetical stimuli.

member of the LPM class.

15.16 Four-response EMM

The derivation of the EMM for four responses is very similar to that for the four response GMM, except that the exponential similarity function is used. For the EMM-CB, the city-block distances from stimulus S at $[y_j, f_k]$ to the four prototype locations are given by:

$$d(S, y\text{–}F) = |P - y_j| + |P - f_k| \tag{15.110}$$

$$d(S, y\text{–}FR) = |P - y_j| + |-P - f_k| \tag{15.111}$$

$$d(S, i\text{–}F) = |-P - y_j| + |P - f_k| \tag{15.112}$$

$$d(S, i\text{–}FR) = |-P - y_j| + |-P - f_k| \tag{15.113}$$

Given the exponential similarity function, we can compute the similarities

$$s(S, y\text{–}F) = e^{-(|P - y_j| + |P - f_k|)} \tag{15.114}$$

$$s(S, y\text{–}FR) = e^{-(|P - y_j| + |-P - f_k|)} \tag{15.115}$$

$$s(S, i\text{–}F) = e^{-(|-P - y_j| + |P - f_k|)} \tag{15.116}$$

$$s(S, i\text{–}FR) = e^{-(|-P - y_j| + |-P - f_k|)} \; . \tag{15.117}$$

These similarities can then be factored into:

$$s(S, y\text{–}F) = e^{-|P - y_j|} \; e^{-|P - f_k|} \tag{15.118}$$

$$s(S, y\text{–}FR) = e^{-|P - y_j|} \; e^{-|-P - f_k|} \tag{15.119}$$

$$s(S, i\text{–}F) = e^{-|-P - y_j|} \; e^{-|P - f_k|} \tag{15.120}$$

$$s(S, i\text{–}FR) = e^{-|-P-y_j|}\, e^{-|-P-f_k|}. \tag{15.121}$$

from which we can draw:

$$p(Y_j \mid Y) = e^{-|P-y_j|} \tag{15.122}$$

$$p(Y_j \mid I) = e^{-|-P-y_j|} \tag{15.123}$$

$$p(F_k \mid F) = e^{-|P-f_k|} \tag{15.124}$$

$$p(F_k \mid FR) = e^{-|-P-f_k|} \tag{15.125}$$

Given these probabilities, we can return to Equation 15.83 in the BPM section for the rest of the derivation. For the *likelihood product form* we can derive the likelihood ratios:

$$l_{IY}(Y_j) = e^{(|P-y_j| - |-P-y_j|)} \tag{15.126}$$

$$l_{FR\,F}(F_k) = e^{(|P-f_k| - |-P-f_k|)} \tag{15.127}$$

For the alternate EMM-E model with the Euclidean distance measure, the distance between the stimulus S and, for example, y-F is:

$$d(S, y\text{–}F) = \sqrt{(P-y_j)^2 + (P-f_k)^2} \tag{15.128}$$

for which the similarity function is:

$$s(S, y\text{–}F) = e^{-\sqrt{(P-y_j)^2 + (P-f_k)^2}} \tag{15.129}$$

As with the two-response case, these similarities are not factorable, and thus the EMM-E is not a member of the LPM class.

15.17 Four-response TSD

The TSD model is illustrated for the four response task in Figure 15.23. As in the four response GMM model the prototypes are defined as distributions located symmetrically at $[P, P]$, $[P, -P]$, $[-P, P]$, and $[-P, -P]$ (with P=1) in a multidimensional feature space. We assume the covariance matrix associated with each prototype (and stimulus) distribution is the same scaler multiple of the identity matrix. We also assume that the perceiver establishes a decision rule based on the use of criterion lines of equal likelihood which separate the space into four regions, one for each response. Given the symmetric locations of the alternative distributions, the criterion lines lie on the main axes of the space, as shown in Figure 15.23.

On each trial, the subject simply determines what region the stimulus has occurred in and responds accordingly. As in the previous TSD model, we assume that stimuli are noisy. Figure 15.24 shows a typical bivariate normal stimulus distribution centered at [.7, .7]. Given a stimulus distribution S centered at $[y_j, f_k]$, we can simply determine what volume (proportion) falls in each response region. Looking at the figure, we can see that the area to the right of the vertical axis is $\Phi(y_j)$. Similarly, the area above the horizontal axis is $\Phi(f_k)$. Since the area in the upper right quadrant is that fraction of the distribution to the right, times the fraction above (and similarly for the other quadrants), we arrive at the following formulae:

$$P(i\text{–}FR \mid S) = [1-\Phi(y_j)]\,[1-\Phi(f_k)] \tag{15.130}$$

$$P(i\text{–}F \mid S) = [1-\Phi(y_j)]\,\Phi(f_k) \tag{15.131}$$

$$P(y\text{–}FR \mid S) = \Phi(y_j)\,[1-\Phi(f_k)] \tag{15.132}$$

$$P(y\text{–}F \mid S) = \Phi(y_j)\,\Phi(f_k) \tag{15.133}$$

Comparing Equations 15.130-15.133 with Equations 15.79-15.81 of the FLMP, we can draw the

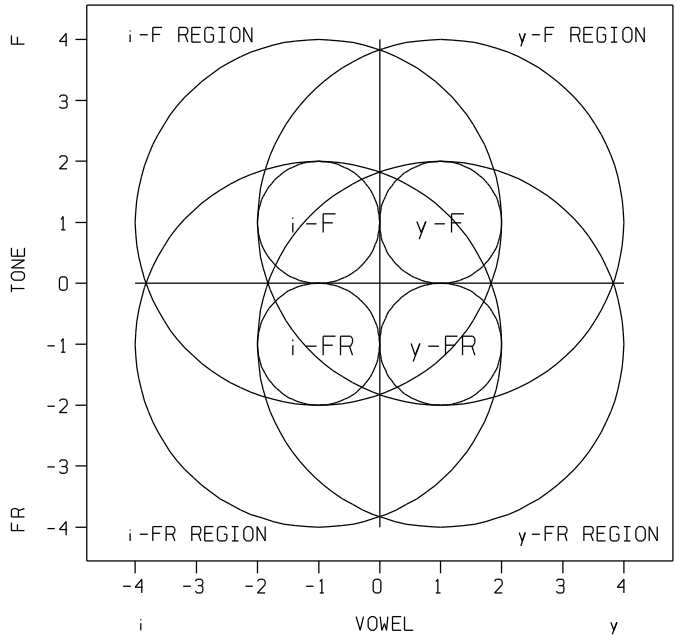

Figure 15.23. Multidimensional TSD model of four response task. The four prototypes are defined as distributions located symmetrically in a multidimensional feature space, indicated by y-F, y-FR, i-F, and i-FR corresponding to the centers. The main axes divide the space into four response regions.

simple equivalences:

$$y_{FLMP} = \Phi(y_{TSD}) \tag{15.134}$$

and

$$f_{FLMP} = \Phi(f_{TSD}) \tag{15.135}$$

Similarly, for the *likelihood product form*, we have the likelihood ratios:

$$l_{IY}(Y_j) = \frac{1-\Phi(y_j)}{\Phi(y_j)} \tag{15.136}$$

and

$$l_{FR\,F}(F_k) = \frac{1-\Phi(f_k)}{\Phi(f_k)} \tag{15.137}$$

The hypothetical four response stimuli are arranged in the multidimensional space in Figure 15.25. To summarize, we find an exact mathematical equivalence between the TSD and the FLMP when using an equal variance, uncorrelated bivariate normal stimulus distribution. To put things another way, we can interpret the fuzzy feature values to be the cumulative normal areas of the multidimensional scale values.

15.18 Four-response FCM

At first glance, the FCM for four responses appears quite similar to that for two. In a recent report, however, Massaro and Friedman (1990) demonstrated that an FCM becomes nonequivalent to the FLMP when 3 or more response alternatives exist. This FCM has a strong constraint on its predicted response probabilities when more than 2 response alternatives exist. In the 4 response task, for example, the maximum predicted probability for any response is limited to .5.

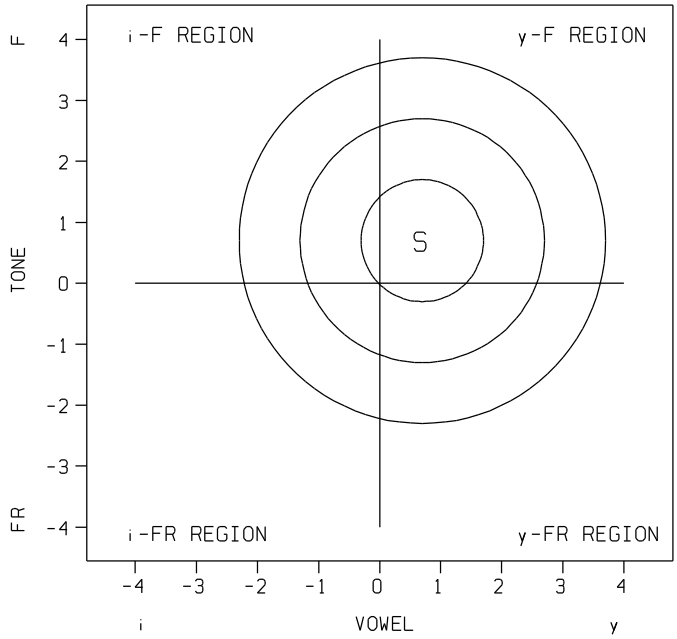

Figure 15.24. Multidimensional TSD model of four response task with space divided into four response regions. A stimulus distribution indicated by S is centered at [.7, .7].

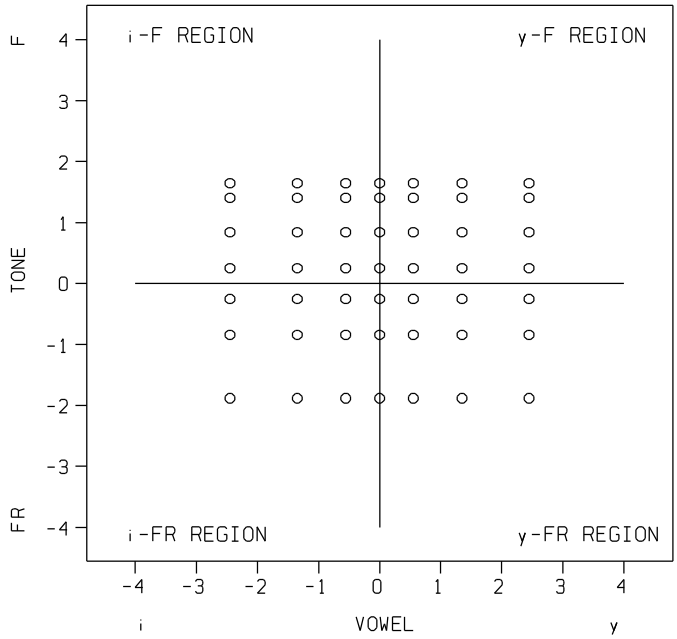

Figure 15.25. Multidimensional TSD model of four response task with space divided into four response regions. Small circles show locations of hypothetical stimuli.

To overcome this constraint, a threshold unit can be added to the network, as illustrated in Figure 15.26. This unit is connected to each of the output units with a positive weight of one—as with other excitatory connections. The revised FCM with a threshold is called the FCM-T. In this model, the additional free threshold parameter is sufficient to bring the model into close correspondence with the LPM class of models. The parameter functions in somewhat the same way as the k value in the IAC model in that it alters the effect of the function producing the response strengths in the RGR. Figure 15.27 provides a demonstration of how the threshold works for the FCM. To see

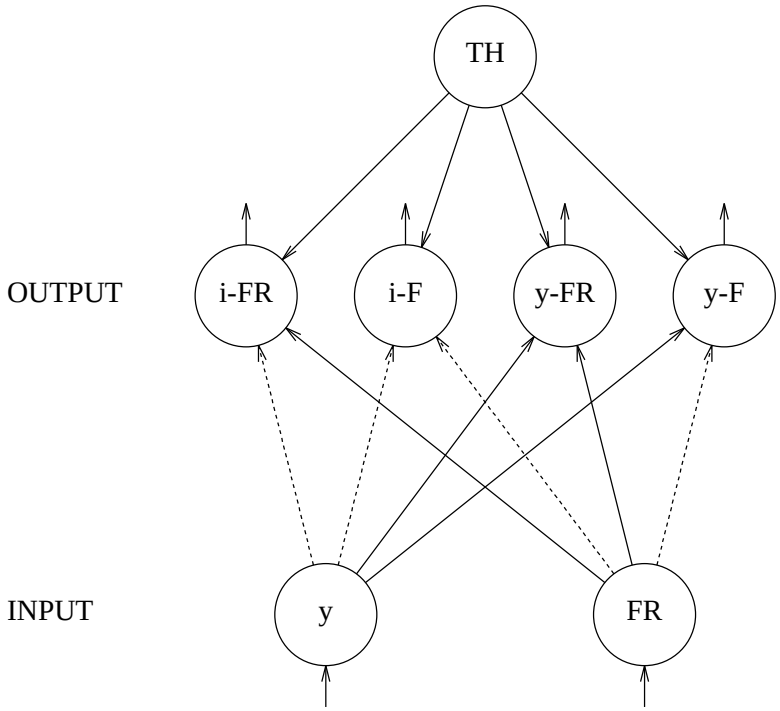

Figure 15.26. FCM-T for the four response task. Input and output layers are shown and also an additional threshold unit (TH). Solid arrows indicate connections with weight 1, and dashed arrows indicate connections with weight -1.

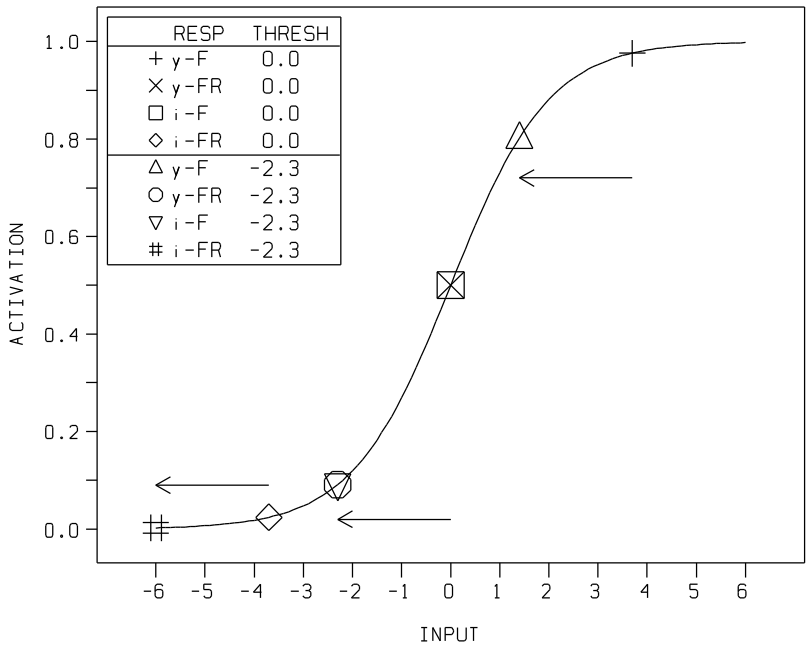

Figure 15.27. Sigmoid activation curve showing unbiased (threshold = 0.0) and negatively biased (threshold = -2.3) activations. See text for details.

more clearly the problem that the threshold corrects, let us first consider the overall pattern of predicted responses for a hypothetical pair of input features. Suppose the *y* and *f* features for the FLMP are both .9. The resulting goodnesses for y-F, y-FR, i-F, and i-FR prototypes in the FLMP would then be .810, .090, .090, and .010, respectively. For the FCM with no threshold and with the

y and f input features both at 3.7, the resulting activations for y-F, y-FR, i-F, and i-FR would then be .976, .500, .500, and .024, respectively as illustrated by the first four symbols in Figure 15.27. Two differences between the FLMP and FCM are apparent. First, while the FLMP prototype goodnesses sum to 1, the activations of the FCM sum to 2. This follows from the additive combination of the separate activations for the FCM. Since these sums are in the denominator of the RGR, and the numerator for the FCM can be at most 1, the highest proportion of response that the FCM can predict is .5 which is clearly incorrect. The second difference between the models concerns the goodness of the two responses with conflicting information. For the FCM, the goodness of the two responses are centered on the scale at .5 (because the opposing features sum to 0). For the FLMP, however, the goodness of the responses are both at .09.

Now consider what happens when we add a threshold value of -2.3 to each response unit in the FCM. The arrows in the figure indicate this leftward shift of 2.3 which results in the second set of four points in Figure 15.27. In this case, the resulting activations for y-F, y-FR, i-F, and i-FR would then be .802, .091, .091, and .002, respectively. These sum to .986 and have the center activation asymmetry (not at .5 activation for conflicting input values) of the FLMP. This biasing effect of the threshold thus allows the FCM to mimic the overall characteristics of the FLMP.

The mathematics of the four response FCM-T are not as tractable as the two response case. For the network shown in Figure 15.26 we have the activations:

$$a_{y-F} = \frac{1}{1+e^{-(y_j + f_k + th)}} \tag{15.138}$$

$$a_{y-FR} = \frac{1}{1+e^{-(y_j - f_k + th)}} \tag{15.139}$$

$$a_{i-F} = \frac{1}{1+e^{-(-yj + f_k + th)}} \tag{15.140}$$

$$a_{i-FR} = \frac{1}{1+e^{-(-yj - f_k + th)}} \tag{15.141}$$

Given an RGR decision rule we have (e.g. for y-F):

$$P(y-F \mid Y_j \text{ and } F_k) = \frac{\frac{1}{1+e^{-(y_j + f_k + th)}}}{\frac{1}{1+e^{-(y_j + f_k + th)}} + \frac{1}{1+e^{-(y_j - f_k + th)}} + \frac{1}{1+e^{-(-yj + f_k + th)}} + \frac{1}{1+e^{-(-yj - f_k + th)}}} \quad . \tag{15.142}$$

If $th = 0$, we get an ordinary FCM, in which the denominator always sums to 2, producing predictions no greater than .5 — clearly an incorrect result. If $th \neq 0$, we get the FCM-T, which cannot be factored and is thus mathematically different from the LPM.

15.19 Four-response IAC models

The application of the IAC to the four-response task is straightforward. Figure 15.28 shows the network we used which contains layers for vowel and tone feature inputs and a layer of memory units for the four response alternatives. Within each layer, although not shown in the figure, there is mutual inhibition. Each of the memory units receives support from (and sends support back to) one unit in each input layer. The y-F memory unit, for example, is connected to the y vowel unit and the F tone unit.

For the IAC-RGR, each activation after 60 cycles was transformed by Equation 15.76 to a strength value. For the IAC-RGR5, k was fixed at 5. For the IAC-RGRK, k was an additional free parameter allowed to vary. For both of these models, the resulting strengths were evaluated for a decision using a RGR. For a y-F response, for example, the probability of a response was:

$$P(y-F) = \frac{S_{y-F}}{S_{y-F} + S_{y-FR} + S_{i-F} + S_{y-FR}} \quad . \tag{15.143}$$

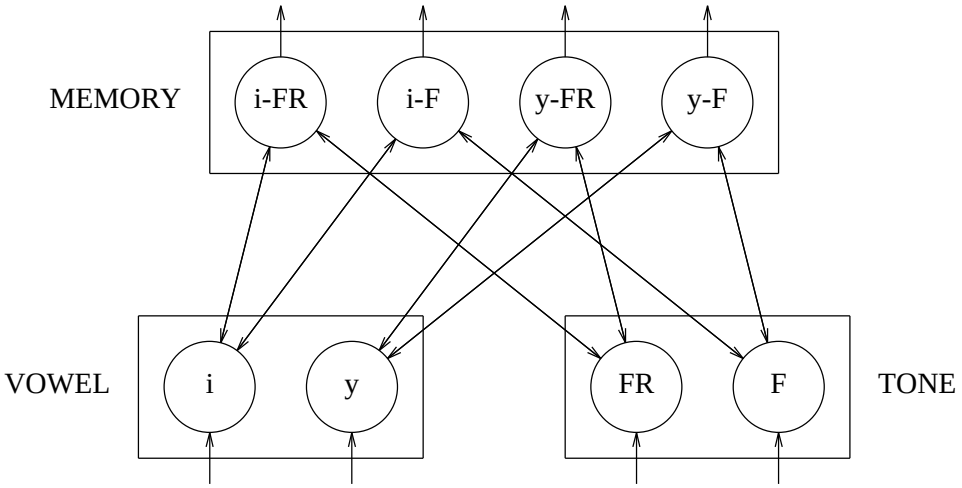

Figure 15.28. An IAC model for the four response task. The three layers of units contain input units corresponding to the vowel and tone features, respectively, and to the "prototypes" or memory of i-FR, i-F, y-FR, and y-F. Units between layers have bidirectional excitatory connections indicated by arrows. Bidirectional inhibitory connections exist between all unit pairs within each layer but have been omitted from the figure for clarity.

For the IAC-BOW model, 1000 simulated trials with Gaussian noise added to the inputs were run with 60 cycles and the response on each trial was made on the basis of which memory unit had the highest activation on each trial.

15.20 Model tests with hypothetical four-response data

The predictions of each model were fit to the predictions of the FLMP (the hypothetical 196 data points shown in Figure 15.20) using STEPIT. For the fit of the FLMP, 14 parameters were used, 7 for the y feature values and 7 for the f features. The predictions of this model are shown in Figure 15.20. Table 15.4 summarizes the fits of the various models, giving the RMSDs between the observed and predicted data for each model. As can be seen in the table, the FLMP, BPM, GMM-E, EMM-CB, and TSD-N models all had perfect fits, as predicted by the theoretical analysis. Relative to the two-response task, the alternative metric MDS model versions GMM-CB (RMSD=.0205 vs .0060 for two-response) and EMM-E (RMSD=.0093 vs .0002 for two-response) models provided slightly different predictions. As expected, the non-threshold FCM did very poorly (RMSD=.2721) while the FCM-T gave a good fit (RMSD=.0002) mimicking the FLMP with th=–8. For the IAC models, the IAC-RGR5 once again gave different predictions (RMSD=.0457) while the IAC-RGRK had a good fit to the FLMP predictions (RMSD=.0030) with k=80 and an extremely small range for the feature values. The IAC-BOW provided a good fit (RMSD=.0091) with the noise standard deviation at .1466.

15.21 Model tests with experimental four-response data

As with the two-response experimental data, it is of interest to see whether we can discriminate between these models with real experimental data. Table 15.5 gives the RMSD for each model for each of the six subjects in the QG task, a fit of a mean subject, and the mean of the subject fits. The FLMP and theoretically equivalent (BPM, GMM-E, EMM-CB, TSD) models all had fits of about RMSD=.0426. Figure 15.29 shows the fit of the FLMP for the typical subject earlier shown in Figure 15.19. Of the remaining models, none differed significantly from the FLMP in goodness of fit, except the non-threshold FCM (RMSD=.2364), $F(1,5)=371.1, p<.001$ and the IAC-RGR5 (RMSD=.0804), $F(1,5)=111.5, p<.001$. For the IAC-RGRK model, the additional k parameter

Table 15.4. Model fits of hypothetical FLMP data.

MODEL	RMSD
FLMP	.0000
BPM	.0000
GMM-E	.0000
GMM-CB	.0205
EMM-E	.0093
EMM-CB	.0000
TSD-N	.0000
FCM	.2712
FCM-T	.0002
IAC-RGR5	.0521
IAC-RGRK	.0030
IAC-BOW	.0091

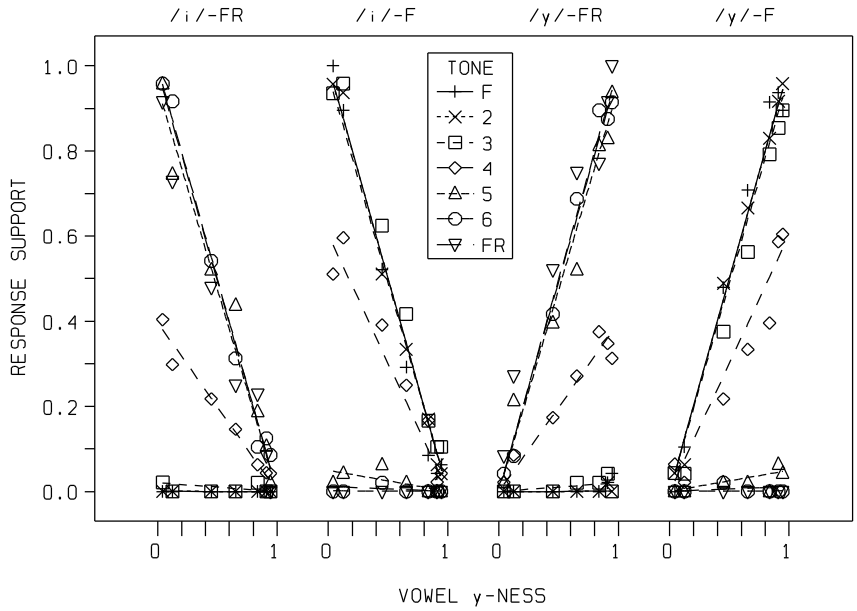

Figure 15.29. Observed (points) and predicted (lines) proportion of i-FR, i-F, y-FR, and y-F responses for a typical subject. Predictions are for the FLMP and locations on the abscissa are scaled according to the y-ness of the vowel feature parameters.

averaged 9.2 and for the IAC-BOW model the standard deviation of the input noise averaged .133.

15.22 Summary

Several models of categorization were developed and analyzed within the context of prototypical pattern recognition tasks. These tasks involve the independent manipulation of two sources of information. The subject categorizes the stimulus event by choosing among two or four response alternatives. This task has been used in a variety of empirical settings, and a set of prototypical results has been observed. Given that a fuzzy logical model of perception (FLMP) has consistently provided a good description of results in this paradigm, the model's predictions were used as the target predictions for the other models. Evaluation, integration, and decision processes are considered for each model. Important features are whether evaluation is noisy, whether integration follows Bayes's theorem, and whether decision consists of a criterion rule or a relative goodness rule.

Table 15.5. RMSDs for model fits of four response experimental data for 6 individual subjects, the fit of the mean subject, and the average of the 6 subject fits.

MODEL	S1	S2	S3	S4	S5	S6	SM	AVE
FLMP	.0331	.0397	.0300	.0450	.0360	.0715	.0265	.0426
BPM	.0326	.0393	.0295	.0449	.0359	.0714	.0262	.0423
GMM-E	.0331	.0397	.0300	.0450	.0360	.0715	.0265	.0426
GMM-CB	.0322	.0387	.0319	.0433	.0348	.0693	.0273	.0417
EMM-E	.0333	.0373	.0293	.0435	.0353	.0705	.0254	.0415
EMM-CB	.0331	.0397	.0300	.0450	.0360	.0715	.0265	.0426
TSD	.0331	.0397	.0300	.0450	.0360	.0715	.0265	.0426
FCM	.2392	.1994	.2203	.2609	.2561	.2423	.2210	.2364
FCM-T	.0309	.0388	.0300	.0425	.0362	.0664	.0264	.0408
IAC-RGR5	.0770	.0686	.0605	.0927	.0816	.1018	.0714	.0804
IAC-RGRK	.0394	.0379	.0300	.0521	.0375	.0776	.0455	.0458
IAC-BOW	.0329	.0395	.0283	.0455	.0358	.0723	.0291	.0424

The models developed and compared to the FLMP include a model based on Bayes's theorem (BPM), models based on multidimensional scaling (MDS), the Theory of Signal Detection (TSD), a feed-forward two-layer Connectionist Model (FCM), and an interactive activation connectionist model (IAC). Theoretical analysis of the models reveals that most can be reduced to a canonical *likelihood product form*. We call this class of models Likelihood Product Models (LPM). The required likelihood ratios for this form are given in Tables 15.6 and 15.7, for the two- and four-response tasks, respectively. Additionally, model fits were carried out to determine to what extent the non-equivalent models can make similar predictions.

With two response alternatives, the results show that all of the models, except the IAC-RGR model, can be brought into line with the predictions of the FLMP and the observed results. Even the IAC model can be saved if activations are multiplied by a very large constant which has the effect of making activations additive when transformed to logistic values or if a BOW decision rule is used instead of a RGR. With four response alternatives, similar outcomes are observed except for the FCM. The FCM cannot match the target results unless an additional threshold unit is assumed. The additional free parameter given by the threshold unit makes it possible for a FCM-T to predict the target results.

Given the equivalences found among the models, simple predictions of response probabilities in these categorization tasks are not sufficient to distinguish among the models. Some of these models (e.g. the EMM-CB and GMM-E) can be discriminated when more complex prototype structures are considered (Nosofsky, 1985, 1987). Other dependent measures, such as ratings, similarity judgments and reaction times, might also permit tests among the models. For example, it has been shown that the FLMP can account for the dynamics or time-course of processing when a categorization task is embedded in a backward masking task (Massaro & Cohen, submitted). The IAC model had difficulty predicting these same results. To date, the other models have not been systematically developed to make predictions about the time course of processing. Future work should address this issue because psychologists should be concerned with differentiating among models of human performance (Townsend, 1990). Many of the models presented here assumed equal, uncorrelated distributions. We acknowledge that this is not always the case (e.g. with the correlated stimulus dimensions used by Ashby and Gott, 1988), and the theoretical comparison of models should be extended to more general distributional models (see Chapter 16).

Table 15.6. Table of likelihood ratios for *likelihood product form* of two-response models

MODEL	$l_{GQ}(C_j)$	$l_{GQ}(O_k)$
FLMP	$\frac{1-c_j}{c_j}$	$\frac{1-o_k}{o_k}$
BPM	$\frac{P(C_j \mid G)}{P(C_j \mid Q)}$	$\frac{P(O_k \mid G)}{P(O_k \mid Q)}$
GMM-E	$e^{-4P\ c_j}$	$e^{-4P\ o_k}$
GMM-CB	—	—
EMM-E	—	—
EMM-CB	$e^{(\lvert P-c_j \rvert - \lvert -P-c_j \rvert)}$	$e^{(\lvert P-o_k \rvert - \lvert -P-o_k \rvert)}$
TSD-N	—	—
TSD-L	$e^{-\frac{\pi}{\sqrt{6}}c_j}$	$e^{-\frac{\pi}{\sqrt{6}}o_k}$
FCM	e^{-c_j}	e^{-o_k}
IAC-RGR	—	—
IAC-BOW	—	—

Table 15.7. Table of likelihood ratios for *likelihood product form* of four-response models

MODEL	$l_{IY}(Y_j)$	$l_{FR\,F}(F_k)$
FLMP	$\frac{1-y_j}{y_j}$	$\frac{1-f_k}{f_k}$
BPM	$\frac{P(Y_j \mid Y)}{P(Y_j \mid I)}$	$\frac{P(F_k \mid FR)}{P(F_k \mid F)}$
GMM-E	$e^{-4P\ y_j}$	$e^{-4P\ f_k}$
GMM-CB	—	—
EMM-E	—	—
EMM-CB	$e^{(\lvert P-y_j \rvert - \lvert -P-y_j \rvert)}$	$e^{(\lvert P-f_k \rvert - \lvert -P-f_k \rvert)}$
TSD-N	$\frac{1-\Phi(y_j)}{\Phi(y_j)}$	$\frac{1-\Phi(f_k)}{\Phi(f_k)}$
FCM	—	—
FCM-T	—	—
IAC-RGR	—	—
IAC-BOW	—	—

15.23 References

Ashby, F. G., & Gott, R. E. (1988). Decision rules in the perception and categorization of multidimensional stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*, 33-53.

Ashby, F. G., & Perrin, N. A. (1988). Toward a unified theory of similarity and recognition. *Psychological Review, 95*, 124-150.

Ashby, F. G., & Townsend, J. T. (1986). Varieties of perceptual independence. *Psychological Review, 93*, 154-179.

Chandler, J. P. (1969). Subroutine STEPIT - Finds local minima of a smooth function of several parameters. *Behavioral Science, 14*, 81-82.

Ennis, D. M. (1988). Confusable and discriminable stimuli: Comment on Nosofsky (1986) and Shepard 1986). *Journal of Experimental Psychology: General, 117*, 408-411.

Goguen, J. A. (1969). The logic of inexact concepts. *Synthese, 19,* 325-373.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics.* New York: Wiley.

Hinton, G. E. & Sejnowski, T. J. (1983) Analyzing cooperative computation. *Proceedings of the Fifth Annual Conference of the Cognitive Science Society*

Hinton, G. E. & Sejnowski, T. J. (1986) Learning and relearning in boltzmann machines. In D. Rumelhart, J. L. McClelland & the PDP research group (Eds.), *Parallel Distributed Processing: Explorations in the microstructure of cognition. Volume I.* Cambridge, MA: Bradford Books.

Indow, T. (1974). Applications of multidimensional scaling in perception. In E. C. Carterette, & M. P. Friedman (Eds.), *Handbook of perception. Vol. II. Psychophysical judgement and measurement.* New York: Academic Press.

Massaro, D. W. (1984). Building and testing models of reading processes. In P. D. Pearson (Ed.), *Handbook of reading research* (pp. 111-146). New York: Longman.

Massaro, D. W. (1987a). *Speech perception by ear and eye: A paradigm for psychological inquiry*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Massaro, D. W. (1987b). Integrating multiple sources of information in listening and reading. In D. A. Allport, D. G. MacKay, W. Prinz, & E. Scheerer (Eds.), *Language perception and production: Shared mechanisms in listening, speaking, reading and writing* (pp. 111-129). London: Academic Press.

Massaro, D. W. (1989). Testing between the TRACE model and the fuzzy logical model of speech perception. *Cognitive Psychology, 21,* 398-421.

Massaro, D. W., & Cohen, M. M. (1987). Process and connectionist models of pattern recognition. *Proceedings of the Ninth Annual Conference of the Cognitive Science Society* (pp. 258-264). Hillsdale, NJ: Lawrence Erlbaum Associates.

Massaro, D. W., & Cohen, M. M. (submitted). Interactive Activation and the Interaction of Sensory Information and Context in Perception. *Cognitive Psychology,* submitted.

Massaro, D. W. & Friedman, D. (1990). Models of integration given multiple sources of Information. *Psychological Review*, in press.

Massaro, D. W., & Hary, J. M. (1986). Addressing issues in letter recognition. *Psychological Research, 48,* 123-132.

Massaro, D.W., Tseng, C. & Cohen, M.M. (1983). Vowel and lexical tone perception in Mandarin Chinese: Psycholinguistic and psychoacoustic contributions. *Quantitative Linguistics, 19,* 76-102.

McClelland, J. L. (1990). Stochastic interactive processes and the effect of context on perception. *Cognitive Psychology*.

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

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

McClelland, J. L. & Rumelhart, D. E. (1988). *Explorations in Parallel Distributed Processing: A Handbook of Models, Programs, and Exercises*, Cambridge, MA: MIT Press.

Minsky, M., & Papert, S. (1968, 1988). *Perceptrons.* Cambridge, MA: MIT Press.

Nosofsky, R. M. (1985). Overall similarity and the identification of separable-dimension stimuli: A choice model analysis. *Perception & Psychophysics, 38,* 415-432.

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General, 115,* 39-57.

Nosofsky, R. M. (1985). Further tests of an exemplar-similarity approach to relating identification and categorization. *Perception & Psychophysics, 45,* 279-290.

Oden, G. C. (1981). A fuzzy propositional model of concept structure and use: A case study in object identification. In G. W. Lasker (Ed.), *Applied systems and cybernetics: Vol. VI* (pp. 2890-2897). Elmsford, NY: Pergamon Press.

Oden, G. C. (1984). Dependence, independence, and emergence of word features. *Journal of Experimental Psychology: Human Perception and Performance, 10,* 394-405.

Peterson, W. W., Birdsall, T. G., & Fox, W. C. (1954). The theory of signal detectability. *Transactions of the IRE Professional Group on Information Theory, PGIT-4.* 171-212.

Platt, J. R. (1964). Strong inference.
Science, 146, 347-353.

Press, W. H., Flannery, B. P., Teukolsky, S. A., & Vetterling, W. T. (1988) *Numerical Recipes: The Art of Scientific Computing*. Cambridge, UK: Cambridge University Press.

Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review, 65,* 386-407.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning internal representations by error propagation. In D. E. Rumelhart & J. L. McClelland (Eds.) *Parallel distributed processing, Vol. 1: Foundations.* Cambridge: MIT press.

Rumelhart, D. E., & McClelland, J. L. (Eds.) (1986). *Parallel distributed processing, Vol. 1: Foundations.* Cambridge: MIT press.

Sejnowski, T. J., & Rosenberg, C. R. (1987). Parallel networks that learn to pronounce English text. *Complex Systems,* 1, 145-168.

Shepard, R. N. (1957). Stimulus and response generalization: A stochastic model relating generalization to distance in psychological space. *Psychometrika, 22,* 325-345.

Shepard, R. N. (1984). *Similarity and a law of universal generality*. Paper presented at the annual meeting of the Psychonomic Society, San Antonio, TX.

Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science, 237,* 1317-1323.

Stigler, S. M. (1986). *The history of statistics: The measurement of uncertainty before 1900.* Cambridge, MA: Belknap Press of Harvard University Press.

Swets, J. A. (1961). Is there a sensory threshold? *Science, 134,* 168-177.

Swets, J. A., Tanner, W. P., & Birdsall, T. G. (1961). Decision processes in perception. *Psychological Review, 68,* 301-340.

Townsend, J. T. (1990). Serial vs. parallel processing: Sometimes they look like Tweedledum and Tweedledee but they can (and should) be distinguished. *Psychological Science,* 1, 46-54.

Zadeh, L. A. (1965). Fuzzy sets. *Information and Control, 8,* 338-353.

Appendix of Symbols

act_i	activation of unit i
a_n	support for prototype n in FLMP
α	excitatory strength parameter in IAC
A_n	degree of support for response n in FLMP
BPM	Bayesean Probability Model
c_{EMM}	closedness coordinate in EMM
c_{FCM}	closedness activation in FCM
c_{FLMP}	closedness feature value in FLMP
c_{GMM}	closedness coordinate in GMM
c_{TSD}	closedness coordinate in TSD
C_j	closedness stimulus level j
c_j	closedness feature value of C_j
$d(a,b)$	distance from a to b in space
$decay$	rate of activation decay parameter in IAC
d'	d prime metric
e	2.71828
e	event or source of evidence
$estr$	external strength parameter in IAC
EMM-CB	Exponential Multidimensional Model - City Block Metric
EMM-E	Exponential Multidimensional Model - Euclidean Metric
exc_i	excitatory inputs to unit i in IAC
ext_i	external inputs to unit i in IAC
F_0	fundamental frequency or pitch
FCM	Feed-forward Connectionist Model
FCM-T	Feed-forward Connectionist Model with Threshold
F	chinese falling tone
FR	chinese falling-rising tone
F_k	falling-tone stimulus level k
f_k	falling-tone feature value from F_k
FLMP	Fuzzy Logical Model of Perception
f_{FLMP}	falling tone feature value of FLMP
f_{TSD}	falling tone coordinate of TSD
γ	inhibitory strength parameter in IAC
GMM-CB	Gaussian Multidimensional Model - City Block Metric
GMM-E	Gaussian Multidimensional Model - Euclidean Metric
h_i	hypothesis i
i	vowel /i/ as in "she"
IAC-BOW	Interactive Activation and Competition Model - Best One Wins Rule
IAC-RGR	Interactive Activation and Competition Model - Relative Goodness Rule

inh_i	inhibitory inputs to unit i in IAC
$L(x)$	cumulative logistic function of x
l_{kj}	likelihood ratio of hypothesis k to hypothesis j
LPM	Likelihood Product Model
M	maximum activation level parameter in IAC
m	minimum activation level parameter in IAC
net_i	summed input activation to unit i in IAC
o_{EMM}	obliqueness coordinate of EMM
o_{FCM}	obliqueness activation of FCM
o_{FLMP}	obliqueness feature value of FLMP
o_{GMM}	obliqueness coordinate of GMM
o_{TSD}	obliqueness coordinate of TSD
O_k	obliqueness stimulus level k
o_k	obliqueness feature value from O_k
$\Phi(x)$	cumulative normal function of x
rest	resting activation level parameter in IAC
$s(a,b)$	similarity of a and b
S_i	strength of response i in IAC
TSD-L	Theory of Signal Detection Model - Cumulative Logistic Distribution
TSD-N	Theory of Signal Detection Model - Cumulative Normal Distribution
w_{ij}	weight to unit i from unit j in IAC
y	chinese vowel /y/, rounded version of /i/
y_{FLMP}	y-ness feature value of FLMP
Y_j	y-ness stimulus level j
y_j	y-ness feature value from Y_j
y_{TSD}	y-ness coordinate of TSD