COGNITION

CONCEPTUAL AND METHODOLOGICAL ISSUES

HERBERT L. PICK, JR.
PAULUS VAN DEN BROEK
AND DAVID C. KNILL
EDITORS

AMERICAN PSYCHOLOGICAL ASSOCIATION, WASHINGTON, DC

CHAPTER 3

Broadening the Domain of the Fuzzy Logical Model of Perception

Dominic W. Massaro

Introduction

This chapter describes a theoretical framework, the fuzzy logical model of perception (FLMP) and applies it to several aspects of speech perception. In contrasting this framework with other frameworks, such as the well-accepted one of categorical perception, I use a number of prescriptions that I have proposed for psychological inquiry (Massaro, 1987). I begin with a brief description of the theoretical framework, the FLMP, and its application to speech perception by ear and eye.

A Theoretical Framework for Speech Perception

Speech perception is a human skill that rivals our other impressive achievements. Even after decades of intense effort, speech recognition by machine remains far inferior to human performance. The central thesis of the present framework is that there are multiple sources of information supporting speech perception, and the perceiver evaluates and integrates all of these sources to achieve perceptual recognition. There are four central assumptions to the FLMP: (a) Each source of information is evaluated to give the degree to which that source specifies various alternatives, (b) the sources of information are evaluated independently of one another, (c) the sources are integrated to provide an overall degree of support for each alternative, and (d) perceptual identification follows the relative degree of support among the relevant alternatives.

Well-learned patterns are recognized in accordance with a general algorithm, regardless of the modality or particular nature of the patterns (Massaro, 1987). The FLMP has received support in a wide variety of domains and consists of three operations: feature evaluation, feature integration, and decision. 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 to the relevant prototype descriptions. Figure 1 illustrates the three stages involved in pattern recognition.

Central to the FLMP are summary descriptions of the perceptual units of the language. These summary descriptions are called *prototypes* and contain a conjunction of various properties called *features*. A prototype is a category, and the features of the prototype correspond to the ideal values that an exemplar should have if it is a member of that category. The exact form of the representation of these properties is not known and may never be known. However, the memory representation must be compatible with the sensory representation resulting from the transduction of the audible and visible speech. Compatibility is necessary because the two representations must be related to one another. To recognize the syllable /ba/, the perceiver must be able to relate the information provided by the syllable itself to some memory of the category /ba/.

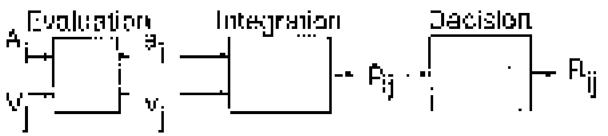

FIGURE 1 Schematic representation of the three operations involved in perceptual recognition. (The sources of information are represented by uppercase letters. Auditory information is represented by A_i and visual information by V_j. The evaluation process transforms these sources of information into psychological values (indicated by lowercase letters a_i and v_j). These sources are then integrated to give an overall degree of support for a given alternative p_{ij}. The decision operation maps this value into some response (R_{ij}), such as a discrete decision or a rating.)

Prototypes are generated for the task at hand. In speech perception, for example, we might envision activation of all prototypes corresponding to the perceptual units of the language being spoken. For ease of exposition, consider a speech signal representing a single perceptual unit, such as the syllable /ba/. The sensory systems transduce the physical event and make available various sources of information called *features*. During the first operation in the model, the features are evaluated in terms of the prototypes in memory. For each feature and for each prototype, featural evaluation provides information about the degree to which the feature in the speech signal matches the featural value of the prototype.

Given the necessarily large variety of features, it is necessary to have a common metric representing the degree of match of each feature. The syllable /ba/, for example, might have visible featural information related to the closing of the lips and audible information corresponding to the second and third formant transitions. These two features must share a common metric if they are eventually going to be related to one another. To serve this purpose, fuzzy truth values (Zadeh, 1965) are used because they provide a natural representation of the degree of match. Fuzzy truth values lie between 0 and 1, corresponding to a proposition being completely false and completely true, respectively. The value .5 corresponds to a completely ambiguous situation, whereas .7 would be more true than false and so on. Fuzzy truth values, therefore, can represent not only continuous rather than just categorical, information, but they also can represent different kinds of information. Another advantage of fuzzy truth values is that they couch information in mathematical terms (or at least in a quantitative form). This allows the natural development of a quantitative description of the phenomenon of interest.

It should be noted that fuzzy truth values are not probabilities, although both lie between 0 and 1. To say that "a penguin is a bird to degree .85" is not the same as saying that "the probability that a penguin is a bird is .85." The former represents some measure of the degree to which the concept *penguin* matches the concept *bird*, whereas the latter gives the probability that any given penguin exactly matches the concept *bird*. Thus, equivalent numerical values can correspond to different psychological representations. Although the FLMP might be mathematically equivalent to Bayes's theorem (Massaro & Friedman, 1990), the two formalizations are not equivalent psychological models.

Feature evaluation provides the degree to which each feature in the syllable matches the corresponding feature in each prototype in memory. The goal, of course, is to determine the overall goodness-of-match of each prototype to the syllable. All of the features contribute to this process, and the second operation of the model is called *feature integration*. That is, the features (actually the degrees of matches) corresponding to each prototype are combined (or conjoined in logical terms). The outcome of feature integration consists of the degree to which each prototype matches the syllable.

The third operation during recognition processing is decision. During this stage, the merit of each relevant prototype is evaluated relative to the sum of the merits of the other relevant prototypes. This decision operation is modeled after Luce's (1959) choice rule, called a *relative goodness rule* (RGR) by Massaro and Friedman (1990). In pandemonium-like terms (Selfridge, 1959), we might say that it is not how loud some demon is shouting but rather the relative loudness of that demon in the crowd of relevant demons. This relative goodness-of-match gives the proportion of times the syllable is identified as an instance of the prototype. The relative goodness of match can also be mapped into a rating judgment indicating the degree to which the syllable matches the category. An important prediction of the model is that one feature has its greatest effect when a second feature is at its most ambiguous level. Thus, the most informative feature has the greatest impact on the judgment.

The FLMP describes how multiple influences come together to influence behavior. Expanding an example from Jackendoff (1988), one can illustrate this principle in the domain of perceptual organization. It is known that similarity in shape and size and proximity in distance contribute to perceptual grouping. Each row of five objects in Figure 2 illustrates some degree of grouping that separates the two objects on the left from the three objects on the right. In the first row, the grouping is supported by size. In the second row, the grouping is supported by proximity. The grouping in the third row is stronger than that in the first two rows because it is supported by both proximity and size. The grouping is diminished in the fourth row because the influences of proximity and size are opposed to one another. These examples illustrate how multiple influences come together to influence perception.

The FLMP confronts several important issues in describing speech perception. One issue has to do with whether multiple sources of information are evaluated in speech perception. Two other issues have to do with the evaluation of the sources in that we ask whether continuous information is available from each source and whether the output of evaluation of one source is modified by the other source. The issue of categorical versus continuous perception also applies to the output of the integration process. A question about integration is whether the outcomes passed on by evaluation are integrated into at the higher order representation. If the two sources of information are integrated, it is important to determine the nature of the integration process.

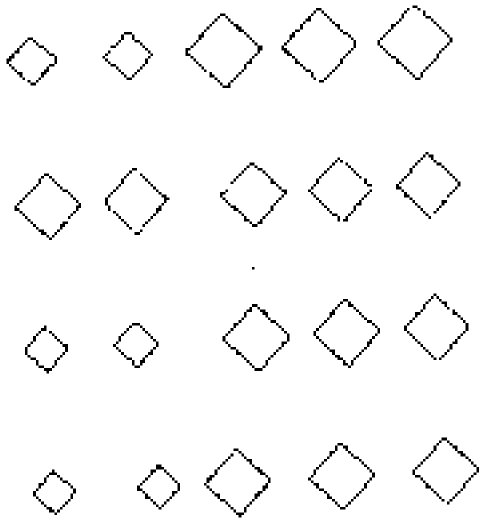

FIGURE 2 Four rows of objects varying in shape, size, and proximity. (See text for the description of the perceptual grouping within each row.)

A Prototypical Experiment

An expanded factorial design offers the potential of addressing important issues in speech perception. I describe an experiment manipulating auditory and visual information in a speech perception task. The novel design, illustrated in Figure 3, provides a unique method to address the issues of evaluation and integration of audible and visible information in speech perception. In this experiment, five levels of audible speech varying between /ba/ and /da/ are crossed with five levels of visible speech varying between the same alternatives. The audible and visible speech also are presented alone, giving a total of 35 (25 + 5 + 5) independent stimulus conditions.

A five-step /ba/ to /da/ auditory continuum was synthesized by altering the parametric information specifying the first 80 ms of the consonant–vowel syllable. Using an animated face, control parameters were changed over time to produce a realistic articulation of a consonant–vowel syllable. By modifying the parameters appropriately, an analogous

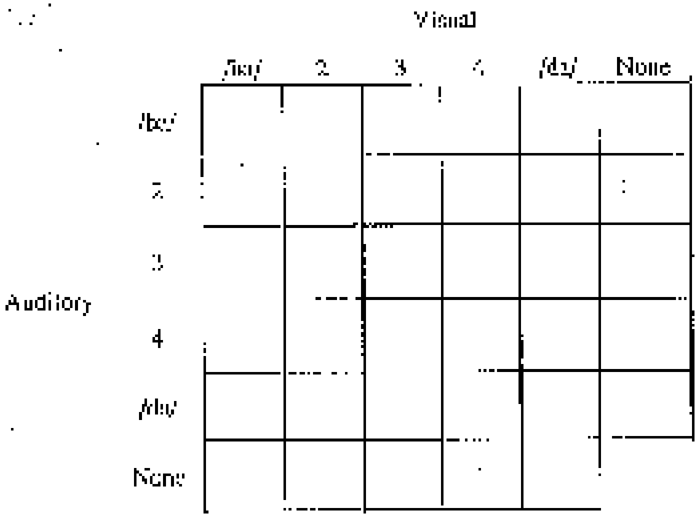

FIGURE 3 Expansion of a typical factorial design to include auditory and visual conditions presented alone. (The five levels along the auditory and visible continua represent auditory and visible speech syllables varying in equal physical steps between /ba/ and /da/.)

five-step /ba/ to /da/ visible speech continuum was synthesized. The presentation of the auditory synthetic speech was synchronized with visible speech for the bimodal stimulus presentations. All of the test stimuli were recorded on videotape for presentation during the experiment. Six unique test blocks were recorded, with the 35 test items presented in each block.

Five college students were tested in the experiment. Subjects were instructed to listen and watch the speaker and to identify the syllable as either /ba/ or /da/. Each of the 35 possible stimuli were presented a total of 12 times during two sessions, and the subjects identified each stimulus during a 4-s response interval.

The points in Figure 4 give the mean proportion of identifications across subjects. The identification judgments change systematically with changes in the audible and visible information sources. The likelihood of a /da/ identification increased as the auditory speech changed from /ba/ to /da/, and analogously for the visible speech. Each source had a similar effect in the bimodal conditions relative to the corresponding unimodal condition. In addition, the influence of one source of information was greatest when the other source was ambiguous.

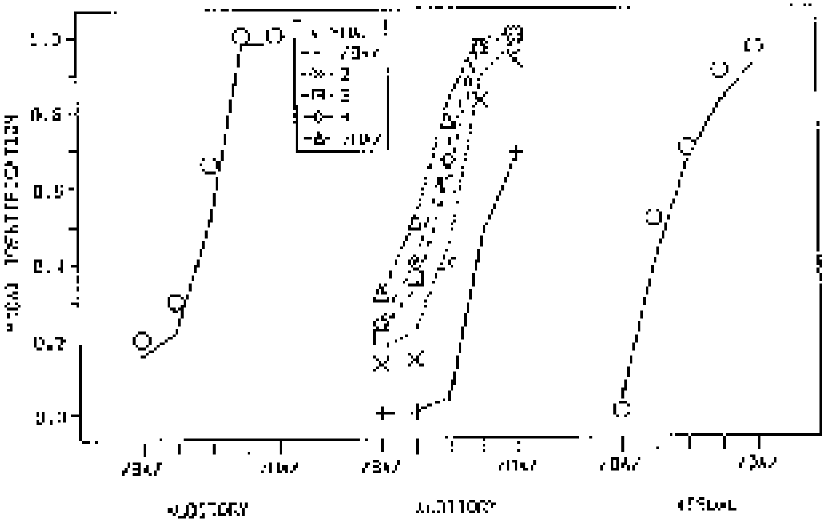

FIGURE 4 Observed (points) and predicted (lines) proportion of /da/ identifications for the auditory-alone (left panel), the factorial auditory-visual (center panel), and the visual-alone (right panel) conditions as a function of the five levels of the synthetic auditory and visual speech varying between /ba/ and /da/. (The lines give the predictions for the fuzzy logical model of perception.)

Applying the FLMP, both sources were assumed to provide continuous and independent evidence for the alternatives /ba/ and /da/. Defining the onsets of the second and third formants (F_2 and F_3, respectively) as the important auditory features and the degree of initial opening of the lips as the important visual feature, the prototype for /da/ was: slightly falling F_2–F_3 and open lips. The prototype for /ba/ was defined in an analogous fashion: rising F_2–F_3 and closed lips.

Given a prototype's independent specifications for the auditory and visual sources, the value of one source cannot change the value of the other source at the prototype matching stage. The integration of the features defining each prototype was evaluated according to the product of the feature values. If a_i represents the degree to which the auditory stimulus (A_i) supports the alternative /da/ (i.e., has slightly falling F_2–F_3) and v_j represents the degree to which the visual stimulus (V_j) supports the alternative /da/ (i.e., has open lips), then the outcome of prototype matching for /da/ would be $a_i v_j$, where the subscripts i and j index the levels of the auditory and visual modalities, respectively.

The outcome of integration would be computed in an analogous way for the other relevant alternatives.

The decision operation would determine their relative merit leading to the prediction that

$$P(/da/|A_iV_j) = \frac{a_iv_j}{\Sigma} \qquad (12)$$

where Σ is equal to the sum of the merit of all relevant response alternatives.

The important assumption of the FLMP is that the auditory source supports each alternative to some degree, as does the visual source. Each alternative is defined by ideal values of the auditory and visual information. Each level of a source supports each alternative to differing degrees represented by feature values. The feature values representing the degree of support from the auditory and visual information for a given alternative are integrated following the multiplicative rule given by the FLMP. The model requires five parameters for the visual feature values and five parameters for the auditory feature values, for each response alternative.

The FLMP was fit to the individual results of each of the 5 subjects. The quantitative predictions of the model were determined using the program STEPIT (Chandler, 1969). The model is represented to the program in terms of a set of prediction equations and a set of unknown parameters. By iteratively adjusting the parameters of the model, the program minimizes the squared deviations between the observed and predicted points. The outcome using STEPIT is a set of parameter values that, when put into the model, come closest to predicting the observed results. Thus, STEPIT maximizes the accuracy of the description of each model. The goodness-of-fit of the model is given by the root mean square deviation (*RMSD*)—the square root of the average squared deviation between the predicted and observed values.

The lines in Figure 4 give the average predictions of the FLMP. The model provides a good description of the identifications of both the unimodal and bimodal syllables (an average *RMSD* of .0791 across the individual subject fits). Figure 5 shows the best fitting parameters for each subject. As can be seen in Figure 5, the parameter values differed for the different subjects but for each subject changed in a systematic fashion across the five levels of the audible and visible synthetic speech.

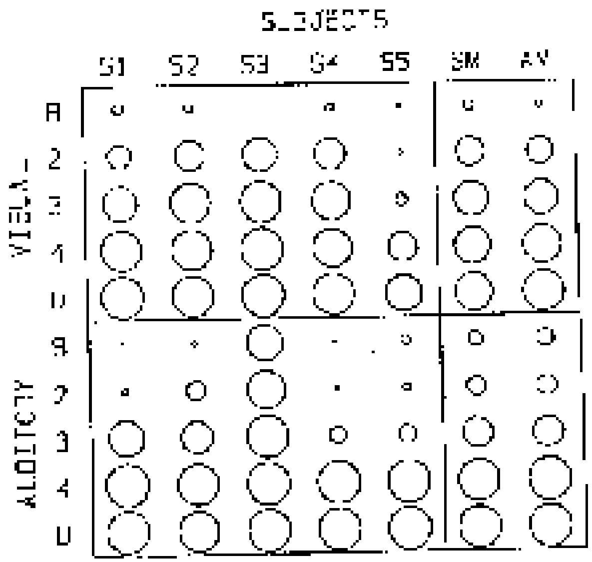

FIGURE 5 Parameters values for 5 subjects, the subject mean (SM), and the average (AV) of the 5 subjects for the support of the response alternative /da/ as a function of the five levels along the visual and auditory speech continua. (The parameter value is given by the area of each circle.)

The meaningfulness of the parameter values justify an important distinction between information and information processing. The parameter values represent how informative each source of information is. The integration and decision algorithms specify how this information is processed. This distinction plays an important role in locating several sources of variability in this inquiry. It is my contention that the variability in information is analogous to the variability in predicting the weather. There are just too many previous contributions and influences to allow an exact quantitative prediction. In addition, small early influences can lead to dramatic consequences at a later time (e.g., the butterfly effect in chaos theory). However, once this variability was accounted for (e.g., by estimating free parameters in the fit of the model), we were able to provide a convincing description of how the information was processed and mapped into a response. Although we could not predict a priori how /da/-like was a particular level of audible or visible speech for a given individual, we could predict how the two sources of information were integrated. In addition, the model does take a stand on the evaluation process in that it assumes that the sources of information are evaluated independently of one another.

Prescriptions for Psychological Inquiry

The information-processing approach is an established and accepted paradigm that has generated a variety of valuable models and experimental tasks, but current research within this domain might profit from a set of prescriptions that could easily be formulated as general guidelines, independent of theory. The first set of prescriptions concerns theoretical and experimental principles of scientific inquiry. Investigators should follow the principles of falsification and strong inference. In the falsification framework, developed by Popper (1959), experimental tests decide whether a hypothesis or theory survives. If a theory survives experimental tests, it should not be discarded. On the other hand, if the experimental tests falsify conclusions drawn from the theory, then the theory should be rejected or modified accordingly.

Expanding on Popper's (1959) approach, Platt (1964) developed the concept of *strong inference*. Rather than test a single hypothesis, the scientist should test multiple hypotheses. Each experimental test would be designed to eliminate (or in Popper's words, falsify) as many of these hypotheses as possible. The results of the experimentation would allow the generation of new hypotheses that could be subjected to further tests. Using this strategy, a scientist could not easily confirm a single pet hypothesis. In addition, at least one of the multiple hypotheses under test should fail and therefore be rejected.

Carrying out research in this manner ensures that the research is theory driven. Not only should theories guide our research, but theories should be articulated in formal or quantitative form. Verbal theories are too ambiguous and not easily distinguished from one another. Strong inference generates precise experiments and a fine-grained analysis of the results. Viable hypotheses usually make highly similar predictions, and they can be differentiated only in highly constrained and sometimes artificial (laboratory) situations. When the opposing predictions of the hypotheses differ only quantitatively and not qualitatively, only precise quantitative analyses of the results are informative.

A second set of prescriptions involves analyzing and manipulating additional well-known variables having to do with individual variability. These variables differ from the typical independent variables that are aimed at influencing psychological processes and are usually considered to be orthogonal to the questions of interest. Individual variability plays a central role in evolutionary theory and inquiry, and psychology—the study of behavior that has necessarily evolved—should be no different. A third set of prescriptions involves analyzing and manipulating additional well-known variables having to do with the tasks. Finally, it is important to study and predict the dynamics of information processing.

My observations are much more optimistic than those of Jenkins (1979), who pointed out that generalizations about memory processes will have to include interactions among four variables: subjects, materials, orienting tasks, and criterion tasks. However, interactions are usually thought of as qualifications of the main effects. For example, dissociations in memory research are usually used to qualify explanations based on a single memory mechanism, rather than to illuminate how mnemonic processing occurs. What I have found, however, is that with the appropriate process model, the interactions allow more, rather than less, generality. Like Jenkins (1980), I advocate a close interaction between experimental investigations and model development.

Falsification and Strong Inference

The principles of falsification and strong inference are central to scientific inquiry. Without this research strategy, it is impossible to reduce uncertainty about the mechanisms involved. More important, testing opposing hypotheses for the experimental investigations helps offset a confirmation bias. The failure to use falsification and strong inference is apparent in the study of categorical perception. Categorical perception occurs when changes along the stimulus continuum are not perceived continuously but in a discrete manner. Listeners are supposedly limited in their ability to discriminate among different sounds belonging to the same phonetic category. The sounds within a category are only identified absolutely, and discrimination is possible for only those sounds that can be identified as belonging to different categories.

Consider the seminal study carried out by Liberman, Harris, Hoffman, and Griffith (1957). The authors generated a series of 14 consonant–vowel syllables ranging from /be/ to /de/ to /ge/ (/e/ as in *gate*). The syllables were used in identification and discrimination tasks. The results were used to test the hypothesis that listeners can discriminate the syllables only to the extent that they recognize them as belonging to different phonemic categories. The hypothesis was quantified to predict discrimination performance from identification judgments. The authors took the correspondence between the observed and predicted results as evidence for the hypothesis, and this has been the major source of support during the past three decades. As an example, Repp (1984) concluded that "the perception of these syllable initial stops was invariably quite categorical" (p. 282).

The persistent observation that the observed discrimination is better than that predicted identification has never been seen as embarrassment for the theory nor has it ever led to consideration of an alternative view. My proposal is that Liberman et al. (1957) failed to consider alternative hypotheses, beginning an unfortunate tradition in this research. No effort has been made to evaluate the actual goodness-of-fit between the predicted and observed results or to contrast this hypothesis with alternative hypotheses. In fact, one can simply formulate the null hypothesis that predicts chance performance at each condition (Massaro, 1987). This null hypothesis gives roughly the same goodness-of-fit as does the categorical model (Massaro, 1987). Clearly, the authors' conclusion that the results were evidence for categorical perception was not justified.

With hindsight, the confirmation bias used in categorical perception research seems difficult to understand. Unfortunately, this realization has not led to an immediate reconsideration of the idea. Thus, the work of Eimas (1985) and Bornstein (1987) are recent examples of failures to incorporate falsification and strong inference concerning the issue of categorical perception. In both cases, the authors failed to consider the possibility that other descriptions provide equally good or better accounts of the phenomena of interest. The latest treatise, on the Haskins approach to speech perception, still failed to consider alternatives to categorical perception (Mattingly & Studdert-Kennedy, 1991).

The relation between discrimination and identification performance has been a central source of evidence for categorical perception. As stated previously, however, investigators did not test alternative theories. Massaro (1987) contrasted the predictions of the classical categorical model with the FLMP in the discrimination/identification paradigm. In this comparison, which confronted the categorical model on its home turf, the FLMP gave a superior prediction of the observed results.

Testing Among Alternative Models

Given the value of falsification and strong inference, it is essential to contrast one model with other models that make alternative assumptions. Given the present state-of-the-art, we must formalize our hypotheses and theories in specific models that make quantitative predictions. Hintzman (1991) also advocated the use of formal models in scientific inquiry but was content with qualitative predictions. However, formal models cannot be distinguished from one another on qualitative grounds alone. Hintzman reminds us how exemplar models can also produce outcomes that were previously thought to be necessarily limited to prototype models. This is true but still leaves the issue of discriminating between these two classes of models. Success in this enterprise requires quantitative predictions and a fine-grained analysis of the results. Hintzman demanded explanatory value from models, not just the ability to predict data. However, the history of science has seemed to show that explanatory value is more a sign of taste than substance, and what has seemed explanatory at one stage has been viewed as curve fitting at a later time when new formal models are discovered and tested.

An important alternative to the FLMP is the theory of categorical perception, which Massaro (1987) formalized as the categorical model of perception (CMP). The model assumes that only categorical information is available from the auditory and visual sources and that the identification response is based on separate decisions given to the auditory and visual sources. Given the /da/–/ba/ identification task, the visual and auditory decisions could be /ba/–/ba/, /ba/–/da/, /da/–/ba/, or /da/–/da/. If the two decisions concerning a given speech event agree, the identification response can follow either source. When the two decisions disagree, it is assumed that the subject will respond with the auditory decision on some proportion *p* of the trials and with the visual decision on the remainder (1 − *p*) of the trials. The weight *p* reflects the relative dominance of the auditory source.

The probability of a /da/ identification response, *P*(/da/), given a particular auditory/visual speech event, A_iV_j, would be

$$\begin{aligned} P(/da/|A_iV_j) = {} & (1)\,aD_i\,vD_j + (p)\,aD_i(1 - vD_j) \\ & + (1 - p)(1 - aD_i)\,vD_j + (0)(1 - aD_i)(1 - vD_j), \end{aligned} \qquad (2.1)$$

where subscripts *i* and *j* index the levels of the auditory and visual modalities, respectively. The aD_i value represents the probability of a /da/ decision given the auditory level *i*, and vD_j is the probability of a /da/ decision given the visual level *j*. The value *p* reflects the bias to follow the auditory source. Each of the four terms in the equation represents the likelihood of one of the four possible outcomes multiplied by the probability of a /da/ identification response given that outcome.

Some advocates of categorical perception might respond that this model is not representative of their notion of categorical perception. This is a reasonable response, but the advocates must then propose a testable model in its place. In fact, the theory implicated by the CMP is not inconsistent with previous notions of categorical perception. For example, advocates of categorical perception have based their conclusions on auditory speech perception. That is, auditory speech is putatively perceived categorically. It does

It is not unreasonable to assume that auditory speech in bimodal speech perception is categorically perceived. In fact, McGurk and MacDonald (1976) originally interpreted their results in this way. Voicing and manner were determined by auditory speech, and place was determined by visible speech. Advocates of categorical perception must specify how the auditory and visual information is combined. Until they do, the CMP is the best known exemplar of categorical perception.

An additional inducement for testing the CMP against empirical results is that this model actually represents two other theories of speech perception. Equation 2.1 can be reduced to Equation 2.2, which shows that the model also quantifies a weighted averaging model and a single channel model.

$$P(/da/|A_i, V_j) = (p)\,ad_i + (1 - p)\,vd_j \qquad (2.2)$$

The weighted averaging model is exactly analogous to the FLMP in every respect except for the integration algorithm (Massaro, 1987, Chapter 7). The sources of information are added, which means that the contribution of a given source of information is independent of its ambiguity. The parameter p in Equation 2.2 indexes the overall weight given the auditory source; however, the integration is always additive. In the FLMP, on the other hand, the least ambiguous source has a greater impact on the final percept.

Equation 2.2 also formalizes a single-channel model, the central assumption of which is that the percept is based on a *single* source of information. The percept is determined by either the auditory or the visual source of information, but not both. The determining source can change from trial to trial. The parameter p indexes the proportion of trials determined by the auditory source.

To fit this model to the results, each unique level of the auditory stimulus requires a unique parameter ad_i, and analogously for vd_j. The modelling of /da/ responses thus requires five auditory parameters plus five visual parameters. The additional p value would be fixed across all conditions and responses. Thus, we have a fair comparison to the FLMP that requires one fewer parameter than the CMP. The CMP was fit to the individual results in the same manner as that of the FLMP. Figure 6 gives the average observed results and the average predicted results of the CMP. As can be seen in Figure 5, the CMP gave a poor description of the observed results. The *RMSD* was .114, compared with the average *RMSD* of .061 for the FLMP. Given that the CMP represents three different hypotheses, its falsification is a particularly informative outcome.

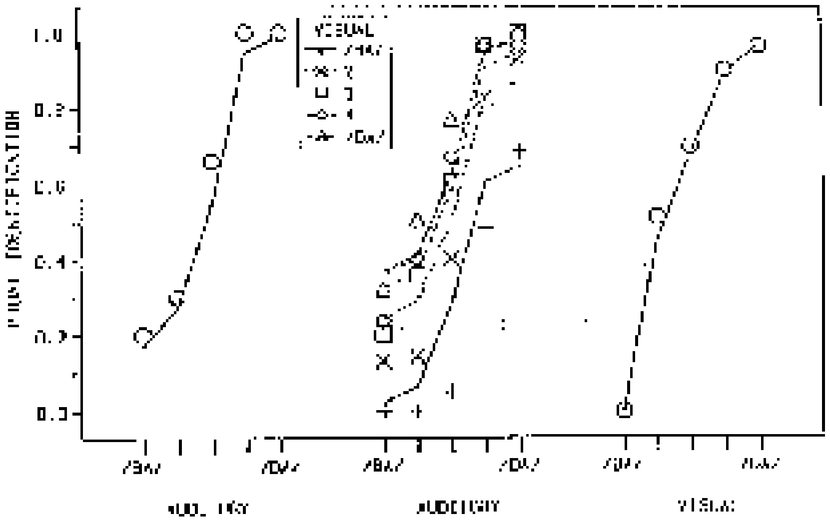

FIGURE 6 Observed (points) and predicted (lines) proportion of /da/ identifications for the auditory-alone (left panel), the factorial auditory-visual (center panel), and the visual-alone (right panel) conditions as a function of the five levels of the synthetic auditory and visual speech varying between /ba/ and /da/. (The lines give the predictions for the categorical model of perception.)

The good fit of the FLMP also provides evidence against categorical perception because the assumptions of the FLMP would contradict those of any categorical perception model. The FLMP assumes that continuous, not categorical, information is available both from the separate sources of information and from the outcome of integration of all sources. The good description given by the FLMP makes it less likely that a model with opposing assumptions would provide an equally good fit. Thus, although it is possible that another model is mathematically equivalent to the FLMP, it is less likely that the model could embody assumptions that are opposite those of the FLMP.

It is also important to determine the falsifiability of models. I am certainly sensitive to this issue, having been concerned about the falsifiability of connectionist models with hidden units (Massaro, 1988). It is necessary to show that the FLMP can be falsified. I see the contrast between the FLMP and the CMP as particularly fair because the latter model has one additional free parameter. Kanevsky (1990), on the other hand, observed that the FLMP might have an advantage because its predictions are more complex than those of the CMP. The CMP predicts additive results, whereas the FLMP predicts nonadditive results. Although this is true, I have not been able to determine why this is necessarily an a priori advantage for the FLMP. On the basis of simulated results and model fits, it is clear that the FLMP can be falsified by additive data in the same manner that the additive model can be falsified by nonadditive data.

The Demise of Categorical Perception

It has become obvious that categorical perception is a belief that will not die or fade away easily. Many textbooks and tutorial articles also state that speech is perceived categorically (J. R. Anderson, 1990; Eimas, 1985; Flavell, 1985; Miller, 1981). However, I have argued that previous results and more recent studies are better described in terms of continuous perception—a relatively continuous relation between changes in a stimulus and changes in perception (Massaro, 1987).

There are severe weaknesses in previous evidence for categorical perception. One approach—the traditional one used throughout the almost three decades of research on categorical perception—concerns the relation between identification and discrimination. In the typical experiment, a set of speech stimuli along a speech continuum between two alternatives is synthesized. Subjects identify each of the stimuli as one of the two alternatives. Subjects are also asked to discriminate among these same stimuli. The results of such experiments have been interpreted as showing categorical perception because discrimination performance is reasonably predicted by identification performance (Studdert-Kennedy, Liberman, Harris, & Cooper, 1970). It turns out that this relation between identification and discrimination provides no support for categorical perception for two reasons. First, the categorical model usually provides an inadequate description of the relation between identification and discrimination, and it has not been shown to provide a better description than continuous models. Second, even if the results provided unequivocal support for the categorical model, explanations other than categorical perception are possible (Massaro, 1987; Massaro & Oden, 1980).

As described in the previous section, evidence against categorical perception comes from a direct experimental comparison between categorical and continuous models of perception. Subjects asked to classify speech events independently varying along two dimensions produce identification results consistent with the assumption of continuous information along each of the two dimensions. A model based on categorical information along each dimension gives a very poor description of the identification judgments. In other research, we asked subjects to make repeated ratings of how well a stimulus represents a given category (Massaro & Cohen, 1983). The distribution of the rating judgments to a given stimulus is better described by a continuous model than a categorical one. The best conclusion is to reject all reference to categorical perception of speech and to concentrate instead on the structures and processes responsible for categorizing the world of speech.

Most readers will remain unconvinced as long as no satisfying explanation is given for the sharp category boundaries found in speech perception research. However, it is only natural that continuous perception should lead to sharp category boundaries along a stimulus continuum. Given a stimulus continuum from A to not A that is perceived continuously, goodness(A) is an index of the goodness of fit of Variable A to the Category A. The left panel of Figure 7 shows goodness(A) as a linear function of Variable A.

An optimal decision rule in a discrete judgment task would set the criterion value at .5 and classify the pattern as A for any value greater than .5. Otherwise, the pattern is

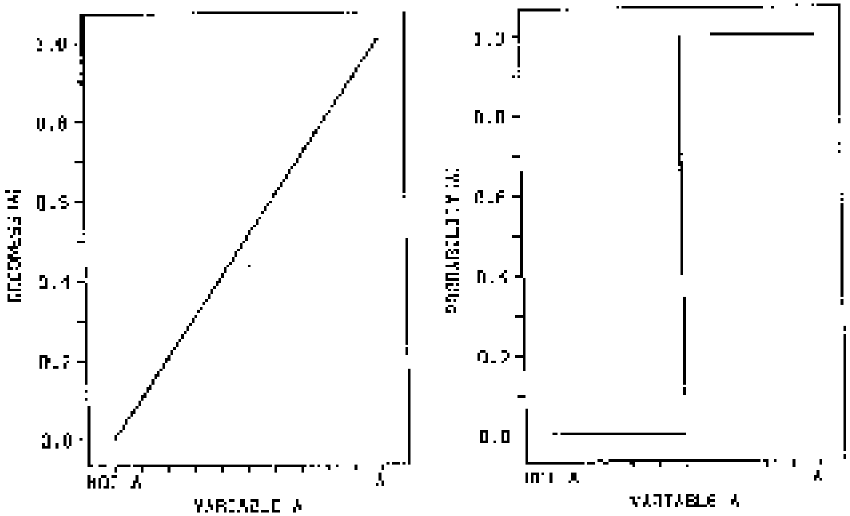

FIGURE 7 Left panel: The degree to which a stimulus (Variable A) represents the category A, called GOODNESS(A), as a function of the level along a stimulus continuum between not A and A. Right panel: The probability of an A response, Probability(A), as a function of the stimulus Variable A if the subject maintains a decision criterion at a particular value of GOODNESS(A) and responds A if and only if the GOODNESS(A) exceeds the decision criterion.

classified as not A. Given this decision rule, the probability of an A response would take the form of the step function shown in the right panel of Figure 7. That is, with a fixed criterion value and no variability, the decision operation changes the continuous linear function given by the perceptual operation into a step function. Although based on continuous perception, this function is identical to the idealized form of categorical perception in a speech identification task. It follows that a step function for identification is not evidence for categorical perception because it can also occur with continuous information.

If there is noise in the mapping from stimulus to identification, a given level of Variable A cannot be expected to produce the same identification judgment on each presentation. It is reasonable to assume that a given level of Variable A will produce a normally distributed range of goodness(A) values, with a mean directly related to the level of Variable A and a variance equal across all levels of Variable A. If this is the case, noise will influence the identification judgment for the levels of Variable A near the criterion value more than it will influence the levels further away from the criterion value. Figure 8 illustrates the expected outcome for identification if there is normally distributed noise with the same criterion value assumed in Figure 7.

If the noise is normal and has the same mean and variance across the continuum, a stimulus, the mean goodness of which is at the criterion value, will produce random classifications. The goodness value will be above the criterion on half of the trials and below the criterion on the other half. As the goodness value moves away from the criterion value, the noise will have a diminishing effect on the identification judgments. Noise has a larger influence on identification in the middle of the range of goodness values than at the extremes because variability goes in both directions in the middle and only inward at the extremes. This differential effect of noise across the continuum will produce an identification function that has a sharp boundary. Thus, the hypothetical subject giving this result appears to show enhanced discrimination across the category boundary when, in fact, discrimination was constant across the continuum. The shape of the function resulted from noise at the decision stage.

This example shows that categorical decisions made on the basis of continuous information produce identification functions with sharp boundaries, previously taken to indicate categorical perception. Strictly speaking, of course, categorical perception was considered present only if discrimination behavior did not exceed that predicted from categorization. However, one should not be impressed with the failure of discrimination to exceed that predicted by categorization if the discrimination task resembles something

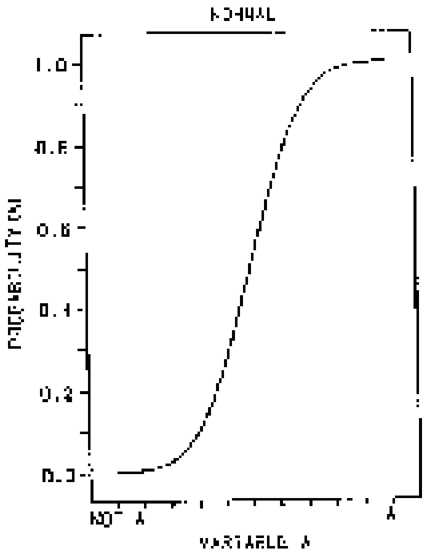

FIGURE 8 Probability(A) as a function of Variable A given the linear relation between GOODNESS(A) and Variable A and the decision criterion represented in Figure 7 but with normally distributed noise added to the mapping of Variable A into GOODNESS(A).

more akin to categorization than discrimination. That is, subjects will tend to rely on identification labels in discrimination tasks if the perceptual memory is poor (Massaro, 1987).

At the theoretical level, it is necessary to distinguish between sensory and decision processes in the categorization task. What is critical for our purposes is that decision processes can transform continuous sensory information into results usually taken to reflect categorical perception. A finding of relatively categorical partitioning of a set of stimuli in no way implies that these stimuli were perceived categorically. Tapping into the process in ways other than simply measuring the identification response reveals the continuous nature of speech perception. Perceivers can rate the degree to which a speech event represents a category, and they can discriminate among different exemplars of the same speech category. Werker (1991) has demonstrated remarkable changes in speech categorization as a function of development and native language. She and others (Kuhl, 1993) have found age-related changes in the sensitivity to nonnative contrasts. These changes are not necessarily evidence for categorical speech perception, however. As Werker (1991) stated, "The fact that adults can still discriminate the nonnative contrasts under certain testing conditions indicates that maintenance is operating at the level of linguistic categories rather than auditory abilities" (p. 101). In addition, reaction times (RT) of identification judgments illustrate that members within a speech category vary in ambiguity or the degree to which they represent the category (Massaro, 1987).

Although speech perception is continuous, there may be a few speech contrasts that qualify for a weak form of categorical perception. This weak form of categorical perception would be reflected in somewhat better discrimination between instances from different categories than between instances within the same category. As an example, consider an auditory /ba/ to /da/ continuum, similar to one used in the current experiments. The F_2 and F_3 transitions are varied in linear steps between the two endpoints of the continuum. The syllable /ba/ is characterized by rising transitions and /da/ by falling transitions. Subjects might discriminate a rising transition from a falling transition more easily than they might discriminate between two rising or two falling transitions even though the frequency difference is identical in the two cases. Direction of pitch change is more discriminable than exact magnitude of change. This weak form of categorical perception would arise from a property of auditory processing rather than from a special characteristic of speech categories. Thus, similar results are found in humans, chinchillas, and monkeys as well as in nonspeech analogs (e.g., Kuhl, 1987; Pastore, 1987). However, it is important to note that discrimination between instances within a category is still possible, and although a weak form of categorical perception might exist for a few categories, most do not appear to have this property. Hence, what is required is an explanation of continuous rather than categorical speech perception.

Psychology and the speech sciences seem reluctant to give up the notion of categorical perception, perhaps, in part, because of phenomenal experience. Our phenomenal experience in speech perception is that of categorical perception. Listening to a synthetic speech continuum between /ba/ and /da/ provides an impressive demonstration of this. Students and colleagues usually agree that their percept changes qualitatively from one category to the other in a single step or two with very little fuzziness in between. (I have had similar experiences, hearing certain German phonological categories in terms of similar English ones.) Phenomenal experience, however, is not enough to confirm the existence of categorical perception. As noted by Marcel (1983), phenomenal experience might be dependent on linking current hypotheses with sensory information. If the sensory information is lost very quickly, continuous information could participate in the perceptual

process but might not be readily accessible to introspection. Reading a brief visual display of a word might lead to recognition even though the reader is unable to report certain properties of the type font or even a misspelling of the word. Yet the visual characteristics that subjects cannot report could have contributed to word recognition. Analogously, continuous information could be functional in speech perception even if retrospective inquiry suggests otherwise. As in most matters of psychological inquiry, we must find methods to tap the processes involved in cognition without depending only on introspective reports.

Dennett (chapter 3) clarifies an important distinction between "filling in" and finding out as descriptions of a variety of experiences such as the apparent motion in the phi phenomenon. The issue for Dennett is whether it is correct to say that the sensory system accomplishes these outcomes by filling in. That is, the sensory system accomplishes an outcome identical in the phi phenomenon to that of continuous motion. It has been reported that the color of the moving object changes in midstream when a red dot at one location is alternated with a green dot at another location. Does the visual system fill in to give us the impression of a continuously moving dot that changes color? Dennett argues very forcefully that our impressions go beyond the information given but that our sensory systems do not—that is, they do not fill in.

Dennett's philosophical argument is highly relevant to categorical perception. In the categorical perception viewpoint, there seems to be significant filling in. Categorical perception accomplishes at the sensory/brain level a direct correspondence between some representation and our impression. Categorical perception supposedly occurs because the sensory/perceptual system blurs any stimulus differences within a category and perhaps sharpens stimulus differences between categories. Categorical perception in the context of filling in occurs when one perceives two different speech events as belonging to the same category because the speech-is-special module makes them equivalent at the sensory/perceptual level. Categorical perception also seems to predict filling in because sensory processing supposedly occurs in a manner that renders the stimuli within a category indiscriminable. This process would be analogous to filling in. On the other hand, as I argue, it is possible that categorization is simply finding out. That is, the goal of speech perception is categorization, and it is possible to find out which category best represents the speech event without necessarily modifying the sensory/perceptual representation of it. In terms of the FLMP, one evaluates, integrates, and makes a categorical decision if need be without necessarily modifying the sensory/perceptual representations of the speech event.

Filling in might also appear to be an attractive explanation of phenomenal experience of contradictory auditory and visual speech. When told to report what one hears, the visible speech biases one's experience relative to the unimodal case. Because it is auditory experience that one reports, it seems only natural to believe that the representation of the auditory speech has been changed—filled in—by the visual. Another interpretation, however, is that we do not have veridical access to the auditory representation. As Marcel (1983) has pointed out, we report interpretations—finding out—and not representations. Thus, one must be careful about equating phenomenal reports with representations.

As pointed out by N. McCarrell (personal communication, April, 1991), Dennett's argument is reminiscent of Gibson's (1969) argument for differentiation as opposed to enrichment. Gibson was primarily concerned with perceptual learning and argued that experience allows better differentiation among sensory/perceptual representations rather than modification of these representations. The issue of filling in versus finding out can be addressed independent of learning, however. At face value, finding out is much more compatible with the Gibsonian framework because it reduces the amount of processing that is assumed. However, a difficult point for the neo-Gibsonians is the apparently necessary admission that perceivers must at times go beyond the information given when they find out. For neo-Gibsonians, perhaps, finding out is too like inferential perception.

Categorical perception has been a popular assumption because it appears to place certain constraints on the speech perception process—constraints that make speech perception possible or easier. If the infant were limited to perceiving only the discrete categories of his or her language, then acquisition of that language would be easier. However, an inability to discriminate within-category differences can only hurt speech perception. We know that higher order sentential and lexical information contribute to speech perception. If categorical perception were the case, errors would be catastrophic because perceivers would access incorrect categories. Categorical perception would also make it difficult to integrate sentential and lexical information with the phonetic information. Continuous information is more naturally integrated with higher order sources of information (Massaro, 1987).

One of the impediments to resolving the controversy is the term *perception*. If perception simply refers to our phenomenal experience, then we cannot deny categorical perception because we naturally attend to the different categories of language. If perception refers to the psychological processing, however, then it is clear that the processing system is not limited to categorical information. One possible reason why categorical perception has been viewed so positively is that scientists have mistaken the outcome for the processes leading up to the outcome.

Despite phenomenal experience and the three decades of misinterpreting the relation between the identification and discrimination of auditory speech, we must conclude that speech is perceived continuously, not categorically. Various work shows that visible and bimodal speech are also perceived continuously. This observation also seems to undermine current views of language acquisition that attribute discrete speech categories to the infant and child (Eimas, 1985; Gleitman & Wanner, 1982). Most important, the case for the modularity or specialization of speech is weakened considerably because of its reliance on the assumption of categorical perception. We are now faced with the bigger challenge of explaining how multiple continuous sources of information are evaluated and integrated to achieve a percept with continuous information.

Individual Variability

This assumption involves looking at individual differences and similarities. We might expect large individual differences in the influence of visible speech on bimodal speech perception. On the other hand, we might expect very little variability across individuals. In either case, the question of individual variability is of interest and should be addressed by models of the phenomenon. As students, we all learned the sins of averaging results across subjects. Group results can camouflage the nature of the individual results. However, it is fair to say that our science remains mostly one of averages rather than of individuals.

Individuals

It is well-known that individual differences exist, and our experimental investigations are aimed usually at reducing them as much as possible. In accuracy studies, for example, we often adjust the stimulus conditions to give relatively equal overall performance across individuals. There is nothing wrong with this procedure, but it may preclude discovery of important properties of the processes of interest. Individual differences can be meaningless or misleading, however, unless the investigator has available a good process model of the task. We all know that individuals differ but need to know how they differ; that is, individuals might simply differ with respect to the information they use, or they might differ in how they process the information. We must verify a process model of the task as much as possible to understand individual differences.

My theoretical and experimental strategy has focused on the analysis and prediction of individual subjects' results. In contrast to most research, this requires a large number of observations from each subject to test the quantitative models against the individuals' data.

The FLMP is ideal for locating the effects of individual variability. As described by Massaro (1987, 1998a), the model allows one to make an important distinction between information and information processing. Information corresponds to the outcome of evaluation: how much a given stimulus characteristic by itself supports the various alternatives. Information processing corresponds to the process of integration: how the various sources of information are combined. Individual perceivers might differ regarding either or both of these characteristics. Consider the second level along a synthetic speech continuum between /ba/ and /da/. It is not possible to predict how much this stimulus will support the alternative /ba/ for a given subject. Subjects have unique representations of speech categories given their unique speech histories. We can guess that the stimulus will be perceived as more /ba/-like than /da/-like, but we cannot quantify how much—even given the results of hundreds of other observers.

The FLMP makes a very strong prediction, however. Regardless of the amount of /da/-ness from a given source of information, it will be combined with other sources of information as prescribed by the integration and decision operations. Thus, the model allows for individual differences at the level of evaluation but not in the processes of integration and decision. Thus, testing the FLMP against the results also tests whether individual differences can be located entirely at the evaluation stage of processing.

To test this strong prediction of the FLMP, we reanalyzed subjects from several of our early experiments in bimodal speech perception. A 5-step /ba/–/da/ auditory continuum was factorially combined with a visual /ba/, a visual /da/, or no lip movements coming from the face of the speaker. The speech events were identified as /ba/ or /da/. Thirty-nine subjects were tested across three experiments. First, we asked whether the individual differences can be accounted for simply in terms of information. This means that the model should give a good fit to the results of all subjects, not just to a subset of the subjects. Second, to locate the individual differences in terms of information, we looked at the magnitude of the effects of the auditory and visual sources of information.

Figure 9 plots the distribution of *RMSD* values from the fit of the FLMP across the 39 subjects. As can be seen in Figure 9, the values appear to make up a fairly homogeneous distribution. There is no evidence that the model accounts for the results of only a subset of the population. To provide a better test, we contrasted the predictions of the FLMP with those of the CMP. Figure 9 also plots the distribution of *RMSD* values from the fit of the CMP across the 39 subjects. The two distributions have little overlap, even though the figure is somewhat illusory in giving the appearance of more overlap than there actually is. Even the worst-fitting FLMP subject was a better fit than 35 of the 39 CMP subjects. Also shown in Figure 9, the FLMP provides a better description than does the CMP for each of the 39 subjects.

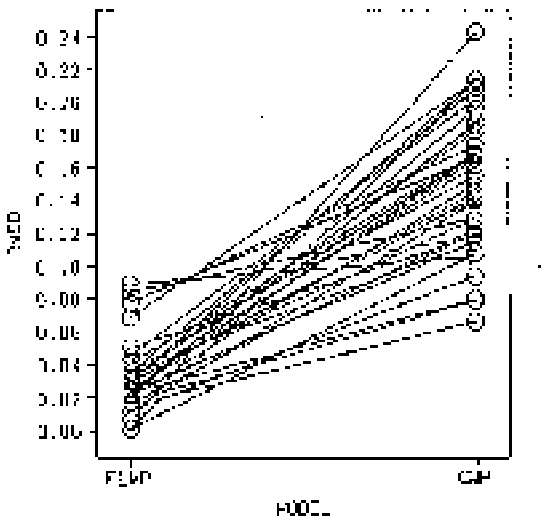

FIGURE 9 The root mean square deviation (*RMSD*) values from the fit of the fuzzy logical model of perception (FLMP; left distribution) and from the fit of the categorical model of perception (CMP; right distribution) for the 39 subjects. (The lines connect the fits of the two models for each subject.)

To assess the individual differences at the evaluation stage, the sizes of the visual and auditory effects were computed for each subject. A visual effect size (*V*) for each subject was defined by the marginal difference between the probability of a /da/ or /ba/ response to a visual /da/ or a visual /ba/, respectively:

$$V = P(/da/|\text{visual}/da/) - P(/da/|\text{visual}/ba/). \quad (3.1)$$

This value varies between 0 and 1. An analogous measure was computed for the auditory effect *A* using the probability of a /da/ response to the first and fifth levels along the auditory continuum. Figure 10 plots the size of the predicted visual effect as a function of the predicted auditory effect. As can be seen in Figure 10, there was impressive negative correlation ($r = -.925$) between these two effects. This result is expected from the FLMP integration algorithm in which the influence of one source of information is negatively correlated with the influence of another source of information. Figure 10 also illustrates the tremendous individual variability in the dominant source of information and the relative dominance of the dominant source. The goodness-of-fit of the FLMP for all subjects is all the more impressive given the large variability across subjects.

It is of interest whether the fit of the FLMP varies with the influence of the two sources of information. To assess this question, a correlation was computed between the *RMSD* and the visual range and between the *RMSD* and the auditory range. The correlations were relatively small ($r = .166$ and $.224$ for the visual and auditory ranges, respectively). Thus, the goodness-of-fit of the FLMP does not vary much with the magnitude of the influence of the audible and visible speech. This generalization of the model across individual variability enhances our faith in the model.

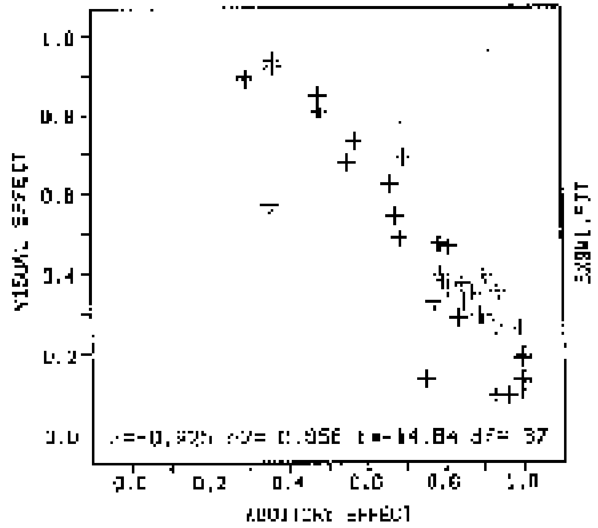

FIGURE 10 Scatterplot of the predicted visual effects as a function of the predicted auditory effect for the 39 subjects. (The ranges are computed from the parameter values of the fit of the fuzzy logical model of perception.)

Task Variability

This prescription addresses the extent to which the phenomenon of interest generalizes across task variability. Investigators should be concerned with whether their findings generalize across different variations of the experimental task. Evaluating changes in behavior as a function of variations within and between different tasks provides valuable information on several fronts. The study of performance in various tasks improves the chances of gaining insights into underlying mechanisms. We do not expect superficial results, such as whether an independent variable had a significant effect, to generalize across tasks. However, we do expect our theories to account for performance across variability in the task domain. Our understanding of psychological mechanisms is good to the extent that we can predict across different tasks.

Generalizing Across Domains of Inquiry

One of the most engaging issues of this decade has been modularity of mind (Fodor, 1983). This thesis makes the very strong prediction that mechanisms uncovered in one domain will not be adequate to describe performance in a different domain. Of course, this thesis is most directly tested by studying behavior as a function of domain variability.

Regarding speech perception by eye and ear, the question is to what extent similar processes occur in speech perception by means of other modalities. There is substantial evidence that the processes found in speech perception by ear and eye generalize to electrical stimulation of cochlear implants and tactile stimulation on the skin. Modularity of mind has been the center of much controversy, and the issue of the modularity of speech perception is equally relevant (Mattingly & Studdert-Kennedy, 1991). Do the processes uncovered in speech perception occur in other domains? Table 1 lists the different domains that have supported the processes assumed by the FLMP. As can be seen in the table, the FLMP has given a good description of various aspects of speech perception (Massaro, 1988b; Massaro & Cohen, 1990; Massaro, Cohen, & Thompson, 1988), letter and word recognition in reading (Massaro & Cohen, 1991), language processing (Massaro, 1987, chapter 8), object categorization (Oden, 1981), the visual perception of depth (Massaro, 1988), the localization of sound in space (Fisher, 1991), memory retrieval (Massaro, Weldon, & Kitzis, 1991), person impression (Massaro, 1987, chapter 9), and decision making (Massaro, 1990).

TABLE 1
Domains of Evidence for the Fuzzy Logical Model of Perception

Domain	Test situation
Speech perception	Acoustic features Phonological constraints Lexical constraints Syntactic constraints Semantic constraints Audible and visual speech Audible speech and gesture Audible speech and written letters
Reading	Letter features Orthographic constraints Lexical constraints
Language	Semantic and syntactic information
Categorization	Object classification
Visual perception	Cues to egocentric distance
Localization	Audible and visible stimuli
Memory retrieval	Letters and semantic cues
Social events	Person impression
Reasoning	Conjunction fallacy

Dynamics of Information Processing

The final prescription is to study the dynamics of information processing. This advice might seem superfluous at this time of heady dynamic modeling within the connectionist framework. It is fair to say, however, that most of the work involves learning input–output correspondences without consideration of the time course of processing a given event. In some recent work, we have contrasted the interactive activation model of word recognition with the FLMP. Although a revised form of the interactive activation model (McClelland, 1991) and the FLMP make similar asymptotic predictions, it is possible to distinguish between the models in terms of the dynamics of information processing.

Visual Information and Orthographic Context in Reading

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

The experiment factorially combined six levels of visual information with these four levels of orthographic context, giving a total of 24 experimental conditions. The bar length of the letter took on six values from a prototypical c to a prototypical e. The test letter was presented at each of the four letter positions in each of the four context types. The test string was presented for a short duration, followed after a short interval by a masking stimulus composed of random letter features. Subjects were instructed to identify the test letter on the basis of what they saw. The results of the experimental test are shown in Figure 11. As can be seen in Figure 11, both the test letter and the context influenced performance in the expected direction. Furthermore, the effect of context was larger for the more ambiguous test letters along the stimulus continuum.

This study also evaluated context effects as a function of processing time controlled by backward masking. The test stimulus was presented for 50 ms. Four masking interstimulus intervals (25, 40, 95, or 210 ms) were tested at each of the other experimental conditions. The points in Figure 11 show the probability of an e response as a function of the bar length of the test letter and the four contexts at each of the four masking intervals. Each point represents data from 176 trials (16 observations from each of 11 subjects). As can be seen in the figure, performance was more chaotic at the short masking intervals. That is, less processing time led to less orderly behavior—as expected from research on the time course of perceptual processing. Even for

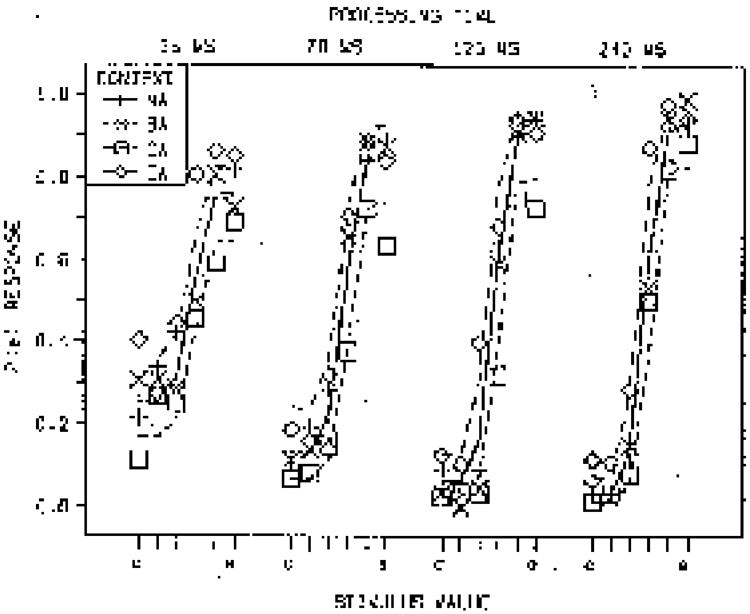

FIGURE 11 Observed (points) and predicted (lines) probability of e identification as a function of the bar length of the test letter, the orthographic context, and the processing interval between the onset of the test stimulus and the onset of the masking stimulus. (Adapted from Massaro, 1979. Copyright 1979 by the American Psychological Association.)

unambiguous test letters, subjects did not make consistent identification judgments at short masking intervals. According to perceptual processing theory, there was not sufficient time for feature evaluation and integration to take place before the onset of the masking stimulus.

Both the test letter and the context influenced performance at all masking intervals. The effect of the test letter was attenuated at the short, relative to the long, processing time. That is, the identification functions covered a larger range across the e–c continuum, with increases in processing time. Context had a significant effect at all masking intervals. In fact, the context effect was larger for the unambiguous test letters at the short than at the longer masking intervals. This result followed naturally from the trade-off between stimulus information and context in the FLMP. Context had a smaller influence to the extent that the stimulus information was unambiguous. Figure 11 also gives the predictions of the FLMP (lines) along with the observed (points) data. The RMSD between the observed and predicted points was .0501.

These same results entailed the falsification of stochastic interactive activation (SIAC) models. A strong context in the FLMP will not override a relatively weak stimulus as it can in the SIAC models. Given the assumption of interactive activation, context can sometimes overwhelm stimulus information about the target as additional processing occurs. This prediction is contradicted by both phenomenological experience and experimental results. The more carefully one reads a word, the more likely one is to notice a misspelling. In experiments varying target information, context, and processing time, stimulus effects were larger to the extent that the processing time was substantial (Massaro & Cohen, 1991). These SIAC models could not simultaneously predict the observed stimulus and context effects with increasing processing time. Although the FLMP and SIAC models made similar predictions for asymptotic performance, we were able to discriminate between the models by attending to the dynamics of information processing.

Summary

This chapter has described the FLMP account of speech perception by eye and ear and has illustrated several prescriptions for psychological inquiry. The idea is that the information-processing framework can be broadened to make it much more valuable and general. The first prescription is that investigators should follow the principles of falsification and strong inference. Strong inference should be used in conjunction with specific models, precise experiments, and a fine-grained analysis of the results. A second prescription involves analyzing individual variability; a good theory accounts for performance across a wide range of individual differences. A third prescription consists of determining to what extent the processes uncovered in the task of interest generalize across different domains. Finally, the fourth prescription centers around the importance of accounting for the dynamics of information processing. These prescriptions were discussed within the context of the theoretical framework of the FLMP and its application to speech perception by ear and eye. I look forward to the next 25 years of our inquiry.

References

Anderson, J. R. (1990). *Cognitive psychology and its implications.* San Francisco: Freeman.

Bornstein, M. H. (1987). Perceptual categories in vision and audition. In S. Harnad (Ed.), *Categorical perception: The groundwork of cognition* (pp. 287–300). Cambridge, England: Cambridge University Press.

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

Eimas, P. D. (1985, January). The perception of speech in early infancy. *Scientific American, 252,* 46–52.

Fisher, B. (1991). *Integration of visual and auditory information in perception of speech events.* Unpublished doctoral dissertation, University of California, Santa Cruz.

Flavell, J. H. (1985). *Cognitive development.* Englewood Cliffs, NJ: Prentice-Hall.

Fodor, J. A. (1983). *Modularity of mind.* Cambridge, MA: Bradford Books.

Gibson, E. J. (1969). *Principles of perceptual learning and development.* New York: Appleton-Century-Crofts.

Gleitman, L. R., & Wanner, E. (1982). Language acquisition: The state of the state of the art. In E. Wanner & L. R. Gleitman (Eds.), *Language acquisition: The state of the art* (pp. 3–48). Cambridge, England: Cambridge University Press.

Hintzman, D. L. (1991). Why are formal models useful in psychology? In W. E. Hockley & S. Lewandowsky (Eds.), *Relating theory and data: Essays on human memory in honor of Bennet B. Murdock* (pp. 39–56). Hillsdale, NJ: Erlbaum.

Jackendoff, R. (1988). Conceptual semantics. In M. Santambrogio & P. Violi (Eds.), *Meaning and mental representations* (pp. 81–97). Bloomington: Indiana University Press.

Jenkins, J. J. (1979). Four points to remember: A tetrahedral model of memory experiments. In L. S. Cermak & F. I. M. Craik (Eds.), *Levels of processing in human memory* (pp. 429–446). Hillsdale, NJ: Erlbaum.

Jenkins, J. J. (1980). Can we have a fruitful cognitive psychology? In H. E. Howe, Jr., & J. H. Flowers (Eds.), *Nebraska symposium on motivation* (pp. 211–238). Lincoln: University of Nebraska Press.

[illegible], D. (1989). A multiple source, or, is a striped apple more striped than a striped orange? *Behavioral and Brain Sciences, 12,* 757–758.

Kuhl, P. K. (1987). The special-mechanisms debate in speech research: Categorization tests on animals and infants. In S. Harnad (Ed.), *Categorical perception: The groundwork of cognition* (pp. 355–386). Cambridge, England: Cambridge University Press.

Kuhl, P. K. (1990). Towards a new theory of the development of speech perception. *Proceedings of 1990 International Conference on Spoken Language Processing, 2,* 745–748.

Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology, 54,* 358–368.

Luce, R. D. (1959). *Individual choice behavior.* New York: Wiley.

Marcel, A. J. (1983). Conscious and unconscious perception: An approach to the relations between phenomenal experience and perceptual processes. *Cognitive Psychology, 15,* 238–300.

Massaro, D. W. (1979). Letter information and orthographic context in word perception. *Journal of Experimental Psychology: Human Perception and Performance, 5,* 595–609.

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

Massaro, D. W. (1988). Ambiguity in perception and experimentation. *Journal of Experimental Psychology: General, 117,* 417–421.

Massaro, D. W. (1989a). Multiple book review of *Speech perception by ear and eye: A paradigm for psychological inquiry. Behavioral and Brain Sciences, 12,* 741–794.

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

Massaro, D. W. (1990). *A pattern recognition account of decision making.* Paper presented at the meeting of the XXIV International Congress of Psychology, Sydney, Australia.

Massaro, D. W., & Cohen, M. M. (1983). Evaluation and integration of visual and auditory information in speech perception. *Journal of Experimental Psychology: Human Perception and Performance, 9,* 753–771.

Massaro, D. W., & Cohen, M. M. (1990). Perception of synthesized audible and visible speech. *Psychological Science, 1,* 55–63.

Massaro, D. W., & Cohen, M. M. (1991). Integration versus interactive activation: The joint influence of stimulus and context in perception. *Cognitive Psychology, 23,* 558–614.

Massaro, D. W., Cohen, M. M., & Thompson, L. A. (1988). Visible language in speech perception: Lipreading and reading. *Visible Language, 22,* 8–31.

Massaro, D. W., & Friedman, D. (1990). Models of integration given multiple sources of information. *Psychological Review, 97,* 225–252.

Massaro, D. W., & Oden, G. C. (1980). Speech perception: A framework for research and theory. In N. J. Lass (Ed.), *Speech and language: Advances in basic research and practice* (Vol. 3, pp. 129–165). San Diego, CA: Academic Press.

Massaro, D. W., Weldon, M. S., & Kitzis, S. N. (1991). The integration of orthographic and semantic information in memory retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17,* 277–287.

Mattingly, I. G., & Studdert-Kennedy, M. (1991). *Modularity and the motor theory of speech perception.* Hillsdale, NJ: Erlbaum.

McClelland, J. L. (1991). Stochastic interactive processes and the effect of context on perception. *Cognitive Psychology, 23,* 1–44.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature, 264,* 746–748.

Miller, G. A. (1981). *Language and speech.* San Francisco: Freeman.

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. 6, pp. 2890–2897). Elmsford, NY: Pergamon Press.

Pastore, R. E. (1987). Categorical perception: Some psychophysical models. In S. Harnad (Ed.), *Categorical perception: The groundwork of cognition* (pp. 29–52). Cambridge, England: Cambridge University Press.

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

Popper, K. (1959). *The logic of scientific discovery.* New York: Basic Books.

Repp, B. H. (1984). Categorical perception: Issues, methods, findings. In N. J. Lass (Ed.), *Speech and language: Advances in basic research and practice* (Vol. 10, pp. 243–335). San Diego, CA: Academic Press.

Selfridge, O. G. (1959). Pandemonium: A paradigm for learning. In *Mechanisation of thought processes.* London: Her Majesty's Stationery Office.

Studdert-Kennedy, M., Liberman, A. M., Harris, K. S., & Cooper, F. S. (1970). Motor theory of speech perception: A reply to Lane's critical review. *Psychological Review, 77,* 234–249.

Werker, J. (1991). The ontogeny of speech perception. In I. G. Mattingly & M. Studdert-Kennedy (Eds.), *Modularity and the motor theory of speech perception* (pp. 91–109). Hillsdale, NJ: Erlbaum.

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

CHAPTER 4

Ecological Foundations of Cognition: Invariants of Perception and Action

M. T. Turvey

What is required to build a theory of cognition? What kinds of concepts are needed? What classes of empirical phenomena should lead the way? There is overwhelming consensus in contemporary science that theorizing on cognition means giving an account of mental processes and mental models. In my opinion, the search for mental mechanisms (of either the symbolic or subsymbolic kind) is overvalued. The challenges facing cognitive theory are considerably more profound, having to do with laws and principles formative at the functional order characterizing nature's ecological scale—the scale at which animals and their environments are defined. I believe that the major concepts needed to address cognition will not be found in the concepts provided by formal logics, computational languages, or network architectures. Rather, the kinds of concepts needed will be developed in the context of an emerging ecological physics and must include a physical notion of information that

Preparation of the manuscript and much of the research reported was supported in part by the National Science Foundation Grants BNS [illegible], DIR [illegible], and BNS [illegible].