BEHAVIORAL AND BRAIN SCIENCES (1989) 12, 741-794
Printed in the United States of America

Multiple Book Review of *Speech perception by ear and eye: A paradigm for psychological inquiry*

Dominic W. Massaro
Program in Experimental Psychology, University of California, Santa Cruz, CA 95064
Electronic mail: *psych36@d.ucsc.edu*

Abstract: This book is about the processing of information in face-to-face communication when a speaker makes both audible and visible information available to a perceiver. Both auditory and visual sources of information are evaluated and integrated to achieve speech perception. The evaluation of the information source provides information about the strength of alternative interpretations, rather than just all-or-none categorical information, as claimed by "categorical perception" theory. Information sources are evaluated independently; the integration process insures that the least ambiguous sources have the most influences on the judgment. Similar processes occur in a variety of other behaviors, ranging from personality judgments and categorization to sentence interpretation and decision making. The experimental results are consistent with a fuzzy logical model of perception, positing three operations in perceptual (primary) recognition: feature evaluation, feature integration, and pattern classification. Continuously valued features are first evaluated, then integrated and matched against prototype descriptions in memory; finally, an identification decision is made on the basis of the relative goodness-of-match of the stimulus information with the relevant prototype descriptions.

Keywords: attention; categorical/continuous perception; connectionism; decision theory; fuzzy logic; hearing; lipreading; modularity; psychophysics; speech perception; vision

In face-to-face communication, the perceiver both sees and hears the speaker. The goal of the research reported in *Speech perception by ear and eye: A paradigm for psychological inquiry* (hereafter *Paradigm*) is to develop a psychological description of how the perceiver recognizes what the speaker says. In addition to examining the information and processes supporting speech perception, the book confronts fundamental issues in experimental psychology and cognitive science.

Modularity of mind

The research described in *Paradigm* addresses an important issue in current psychological theory, referred to as modularity by Fodor (1983; see also multiple book review in BBS 3 (l), 1980). According to the modularity principle, there are independent psychological systems that obey different laws. For example, the system underlying the recognition of visual objects differs fundamentally from the one underlying word recognition. Modularity is assumed in a variety of domains ranging from Chomsky's language organ (1975) and Pylyshyn's (1984) impenetrable processes to Forster's (1985) modular lexicon. Liberman and Mattingly (1985) adopted Fodor's modularity principle in their revision of the "motor theory of speech perception." According to their view, speech is "special" in that a specialized module is directly responsible for the perception of phonetic information, rather than operating on preliminary input from the auditory system. Speech perception results from the speech module's specialization for phonetic gestures. Because of the putative biological link between perception and production, listeners are prevented from hearing speech as ordinary sound. Articulatory gestures are perceived because of the lawful dependencies among gestures, articulatory movements, vocal-tract shapes, and signals. Along similar lines, Eimas (1985) has proposed that infants are born with innate perceptual mechanisms to detect discrete phonemic categories. These claims about specialization for speech would offer little hope for our research framework on two counts. First, approaching the study of "bimodal" (visual and auditory) speech perception from the perspective of prototypical pattern recognition must fail. By definition, a specialized process should not follow general principles. Second, the law uncovered for bimodal speech perception should have little applicability to other domains of human performance.

The research reported in *Paradigm* goes against both expectations from the modularity principle. Modularity predicts that the processes responsible for bimodal speech perception should be unique and unlike those involved in other forms of human performance. The research reported in *Paradigm* instead reveals that the processes involved in bimodal speech perception are similar to those involved in a number of other domains of perceptual and cognitive functioning. The research has succeeded in uncovering principles of bimodal speech perception that prove to be relevant to a variety of behaviors ranging from personality judgments to the

 0140-525X/89 $5.00+.00

learning of arbitrary categories. In each domain, the perceiver evaluates multiple sources of both sensory-driven and contextually-driven information, integrates them with respect to representations in memory, and classifies the pattern on the basis of the relative goodness-of-match among the relevant categories.

The findings reported in *Paradigm* also offer a much more optimistic view of language use by the hearing-impaired. To the extent that speech perception is not specialized, we should be able to substitute other functionally valid cues for the sounds of speech. Lip-read information combined with the transformation of acoustic signals into tactile, visible, or electrical forms should be sufficient to support language perception and understanding.

Research framework

The research is carried out following the procedures of falsification and strong inference (Platt 1964; Popper 1959). Popper (1959) sees the scientist's goal as one of falsifying hypotheses and Platt advocates a strong-inference strategy of testing hypotheses. We follow these guidelines in our experimental tests of alternative hypotheses about bimodal speech perception. In addition, we exploit the idea of converging operations (Garner et al. 1956), or in the strong-inference framework, converging tests that lead to the same outcome. When very different tests lead to the same result, we are less likely to be wrong.

Psychophysical and information-processing methods

The research reported in *Paradigm* utilizes both psychophysical and information-processing methods. Most studies in experimental psychology are based on one of two metatheoretical approaches. The psychophysical approach attempts to discover laws relating the physical world to observable behavior. The value of psychophysics requires little emphasis and is illustrated by progress in the domains of color perception (Pokorny & Smith 1986), auditory perception (Scharf & Houtsma 1986), and the match with Weber's law. [See also Krueger: "Reconciling Fechner and Stevens: Toward a Unified Psychophysical Law" BBS 12(2)1989.] Psychophysical methods have also been successful in determining the influence of higher-order visual properties on visual perception and action (Hochberg 1986; Kelso & Kay 1987; Shebilske 1987). The information-processing approach attempts to discover how the stimulus world is processed to generate some observable behavior. The value of this approach is apparent in the progress that has been made in understanding pattern recognition, memory, and decision making (Massaro 1989).

Multiple sources of information

Paradigm's approach to pattern recognition is concerned with how multiple sources of information are evaluated and integrated. Consider the recognition of the word *performance* in the written sentence "The actress was praised for her outstanding performance." Its recognition is achieved by integrating a variety of bottom-up and top-down sources of information. Top-down sources include semantic and syntactic constraints; bottom-up sources include visual features and letters making up the word. Orthographic constraints have also been shown to contribute to perceptual recognition at the word level. *Paradigm* examines the evaluation and integration of audible and visible sources in bimodal speech perception. Although my focus is on bimodal speech, a central assumption is that evaluation and integration processes are basic to all forms of pattern recognition and that there should be considerable overlap in the processes across different domains.

Once it is recognized that pattern recognition involves multiple sources of information, the goal is to define the sources of information and to assess how they are integrated in perceptual recognition. To pursue this goal, I build on the developments in information-integration theory (Anderson 1981; 1982) and mathematical model testing (Townsend & Ashby 1983). I also develop expanded factorial designs that independently manipulate multiple aspects of the environment jointly and in isolation.

The availability of auditory and visual information in face-to-face speech perception illustrates the multiplicity of cues in a natural situation. One might expect the validity and reliability of the cues to be perfectly correlated. However, variation and noise in the environment and in the sensory systems involved could alter these two sources of information differentially because they are somewhat independent. For example, the perceiver might have a varying view of the speaker's face and extraneous background noise might vary randomly over time. Thus, for any particular speech event, its auditory quality is not necessarily correlated with its visual quality. In experiments, one can manipulate the auditory and visual sources independently of one another. Given that the two cues are not perfectly correlated in the natural situation, the factorial manipulation of the cues will not be completely new to the perceiver.

A theoretical framework for pattern recognition

Well-learned patterns seem to be recognized in accordance with a general algorithm, regardless of their modality or specific nature. One model that accurately describes pattern recognition is our fuzzy logical model of perception (abbreviated FLMP). Its name is derived partly from fuzzy logic – a continuously valued logic that represents the truth of propositions in terms of truth values that range between zero (false) to one (true) (Goguen 1969; Zadeh 1965). The model has received support in several domains, including letter and word recognition in reading (Massaro 1979; Massaro & Hary 1986; Oden 1979), syllable and word recognition in speech perception (Massaro & Cohen 1976; Oden & Massaro 1978), visual depth perception (Massaro 1988a), and the categorization of visual objects (Oden 1981; Thompson & Massaro 1989). The model assumes three operations in perceptual (primary) recognition: feature evaluation, feature integration, and pattern classification. Continuously valued features are first evaluated, then

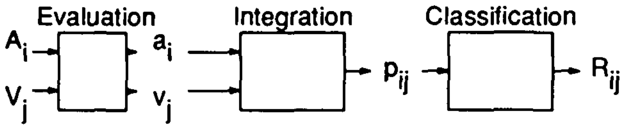

Figure 1. Schematic representation of the three operations involved in perceptual recognition. The evaluation of an auditory source of information A_i produces a value a_i, indicating the degree of support for alternative R. Evaluation of a visual source occurs in an analogous way. The integration of the truth values gives an overall goodness-of-match p_{ij}. The classification of a speech event A_iV_j is based on the relative goodness-of-match of the response alternatives. The probability of a response R_{ij} to a speech event A_iV_j is equal to the value p_{ij} relative to the goodness-of-match of all response alternatives.

integrated and matched against prototype descriptions in memory; finally, an identification decision is made on the basis of the relative goodness-of-match of the stimulus information with the relevant prototype descriptions. An attractive feature of the FLMP is that it generates many quantitative experimental predictions. It can be tested not only directly against a broad range of results, but also against other quantitative models of performance. The primary level of data analysis in *Paradigm* accordingly involves quantitative testing of models of performance.

Figure 1 illustrates three operations involved in the pattern recognition of bimodal speech. The auditory and visual sources of information are represented by uppercase letters. The evaluation process transforms each source of information into feature values (indicated by lowercase letters). The outcome of feature evaluation indicates the degree to which each source of information supports each of the relevant alternatives. For example, the input might be a visual /ba/ paired with an auditory /da/; subjects might be given /ba/ and /da/ as response alternatives. The integration process combines the feature values to give an overall goodness-of-match between all the sources of information and each alternative. The classification operation maps these values into some response, such as a discrete decision or a rating. I suggest that any model of recognition would have to assume stages that were functionally equivalent to these three. This theoretical framework sets the stage for the experimental work.

Binary oppositions

Following the logic of falsification and strong inference, I formulated a set of alternative hypotheses about speech perception by eye and ear. These alternative hypotheses – called binary oppositions – were considered to be central to a complete description of the phenomenon. The binary oppositions considered in *Paradigm* are presented in Figure 2. In some cases, the question at one level is dependent on the answers to the question at higher levels. As an example, the issue of whether or not multiple sources of information are integrated (combined) in perception requires that multiple sources rather than just a single source be available to the perceiver. Similarly, questions about the nature of integration are meaningful only if integration occurs. On the other hand, whether sources of data provide continuous or categorical information can be assessed regardless of whether or not they are processed independently. Each question in Figure 2 demands one or two chapters.

Five of the seven binary contrasts address how bimodal speech is evaluated and integrated in speech perception. The first binary contrast asks whether only auditory or both auditory and visual sources of information influence speech perception. That is, are both auditory and visual properties evaluated in perceptual recognition of speech in face-to-face communication? The second binary contrast addresses whether the perceiver integrates the auditory and visual information made available by feature evaluation. The third binary contrast tests the categorical versus the continuous nature of the information supporting speech perception. This question of categorical versus continuous perception can be asked both with respect to information made available by feature evaluation and with respect to the output of the integration process.

The fourth binary contrast concerns whether the evaluation of one source of information is dependent on another source or whether evaluation of one source can occur independently of others. Dependent evaluation indicates that the output of the evaluation of one source is contaminated by another. The fifth and most difficult binary contrast assesses the form of the integration process. The experiments test between two general classes of

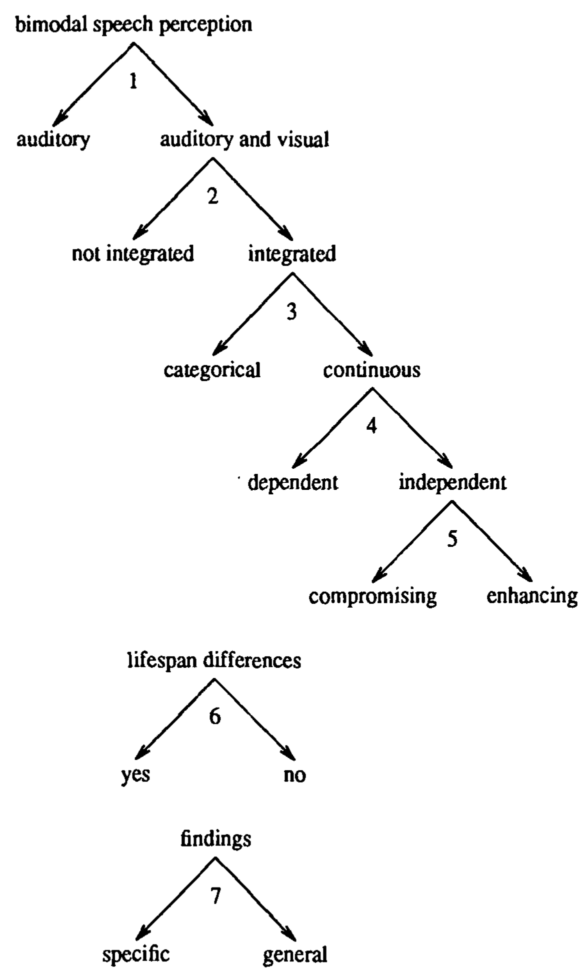

Figure 2. Seven alternative hypotheses, called "binary oppositions," about speech perception by ear and eye.

integration algorithms: "compromising" and "enhancing" algorithms. When the evaluations of two sources of information point ambiguously to the same alternative, a compromising integration would give an outcome somewhere between the separate evaluations of the two sources. An enhancing integration, on the other hand, would give an outcome more extreme than the separate evaluations of the two sources. Given evaluations of .7 and .8, for example, compromising and enhancing might give .75 and .85, respectively.

The sixth binary opposition concerns whether there are differences in bimodal speech perception across the lifespan. This opposition is actually multiple because it concerns how each of the first five oppositions interact with development and aging. In addition, the similarities and differences across the lifespan can be defined in terms of (1) the information that is available to the perceiver and (2) the processing of that information. Changes across the lifespan appear to be due to informational rather than information-processing differences. The seventh and last binary opposition has to do with the generality of the findings. Do the processes involved in speech perception by ear and eye occur in other types of behavior? The research reported in *Paradigm* seems to provide fairly unambiguous answers to each of these seven questions.

Auditory and visual sources of information

The evidence for multiple auditory sources of information in speech perception comes from the discovery of many different cues or features that contribute to the discriminable contrasts found in speech. The phonetic difference between voiced and voiceless stop consonants in medial position appears to have up to 18 acoustic characteristics that could function as acoustic features. For example, the perceived distinction between /aga/ and /aka/ can be influenced by the preceding vowel duration, the silent closure interval, the voice-onset time, and the onset frequency of the fundamental. In addition to these bottom-up sources, phonological, lexical, and sentential constraints function as top-down sources of information in recognition.

An interesting question is whether speech also has visual information when the face of the speaker is visible to the perceiver. Although visual speech does not distinguish among all speech contrasts, it is ecologically valid (provides information) to some degree. It is of psychological interest whether visible speech is functionally valid. Anecdotal evidence comes from the extra effort that seems to be involved in conversing by telephone or listening to narrative over the radio. Further support for the substantial contribution from the visible domain of speech comes from the habit of watching a person speaking and relying more on his visible articulations when the environment is noisy and distracting or if one has a hearing loss. There is evidence that the deaf gain phonological information primarily from a code based on how words look when they are said. Similar phonological information is obtained by both deaf and hearing children, suggesting that information can be derived from visual as well as auditory speech (O'Connor & Hermelin 1981).

In most of the research demonstrating a significant contribution of visible speech, the auditory signal is either absent or degraded by white noise or it is presented to hearing-impaired listeners. The ingenious manipulation of McGurk and MacDonald (1976) showed that visual articulation has an important influence even when paired with perfectly intelligible speech sounds. We have all complained abut the discrepancy between sight and sound in dubbed movies; McGurk and MacDonald modified the situation to illustrate the power of visible speech. They dubbed a visible articulation such as /pa-pa/ with the speech sounds /na-na/. Perfectly intelligible auditory speech is thereby presented with a contradictory visual articulation. Even though subjects were asked to indicate only what they heard, a strong effect of the visual source of information was observed. Faced with the visible articulation /pa-pa/, paired with the sounds /na-na/, subjects often reported hearing /ma-ma/. This surprising perceptual experience has come to be known as the McGurk effect.

Visible speech has a strong influence even when combined with unambiguous auditory speech, but because it is not capable of distinguishing all speech contrasts, it is not a sufficient source. Silent televisions displaying only lip movements are not feasible for communication in the way faceless telephones providing only speech sounds are. Visible speech supplemented with other information, such as low bandpass auditory speech, tactile speech, or electrical speech delivered to a cochlear implant, offers a potentially viable medium of communication for the hearing-impaired. Although previous research has documented a contribution of visible speech, it is important to determine whether it is integrated with audible speech.

Integration

Integration is defined as a process that combines two or more sources of information. The exact nature of the information and how the two sources are integrated are relevant questions once integration is established. We know that both auditory and visual sources of information are utilized in speech perception and it is only natural to believe that the two contributing sources must be integrated. This belief may be false, however, and demonstrating that two sources are integrated in perceptual recognition is no easy matter. Perceivers may be using just a single source of information on each trial, even though the source being used varies from trial to trial. The process of using just a single source on each trial is called nonintegration. The commonest test of integration is to vary the two sources of information independently of one another in a factorial design. In the study of speech perception by ear and eye, a set of bimodal speech events is made up by independently varying the auditory and visual sources of information. Subjects are asked to identify each of the bimodal speech events repeatedly over a series of trials. However, it is difficult to test between integration and nonintegration if subjects are given only a factorial combination of the dimensions. A more powerful method is to use an expanded factorial design in which each information source is presented in isolation, in addition to presenting the sources in factorial combination.

Consider the perception of bimodal speech events created by the factorial combination of synthetic speech sounds along an auditory /ba/ to /da/ continuum paired with /ba/ or /da/ visual articulations. By also including the auditory-alone and visual-alone conditions, it is at least logically possible to reject nonintegration. What is necessary is to find judgments of certain bimodal speech events that cannot be equivalent to some simple mixture of judgments of either the visual or the auditory dimensions presented alone. For example, if subjects usually identify a visual /ba/ paired with an auditory /da/ as /bda/, and this response is seldom given to either the auditory or visual events presented alone, then the results are incompatible with a nonintegration strategy. To make this result possible, it is necessary to allow responses other than /ba/ and /da/. Although it is not common in experiments using synthetic speech, we allow the subject a set of open-ended response alternatives. This expanded factorial design with open-ended response alternatives makes it possible to show the integration of auditory and visual information in speech perception.

The experimental evidence suggests that there is an integration of auditory and visual information in speech perception. Some of this evidence comes from the /ða/ (pronounced *the*) and /ba/ judgments. Figure 3 illustrates that an auditory syllable somewhat between the /ba/ and /da/ endpoints is sometimes perceived as /ða/. This perceptual judgment is more frequent when the auditory syllable is paired with a visual /da/ and less frequent when paired with a visual /ba/. These results make good sense in terms of the characteristics of both the auditory and visual dimensions of these syllables. Spectrograms of these sounds are similar to those of a natural /ða/ of the speaker. A /da/ articulation is also similar to a /ða/ articulation in that they both are articulated with a relatively open mouth. A /ba/ articulation is very different from that for /ða/. Relative to the no-articulation condition, a visual /da/ combined with this auditory information increases the likelihood of a /ða/ response. On the other hand, the visual information corresponding to /ba/ eliminates a /ða/ response. Visual /da/ provides some support for the alternative /ða/, and visual /ba/ provides no support or negative support for this alternative. Given that the response /ða/ seldom occurs on visual-alone trials, the increase in /ða/ responses with visual /da/ probably represents true integration of the auditory and visual information.

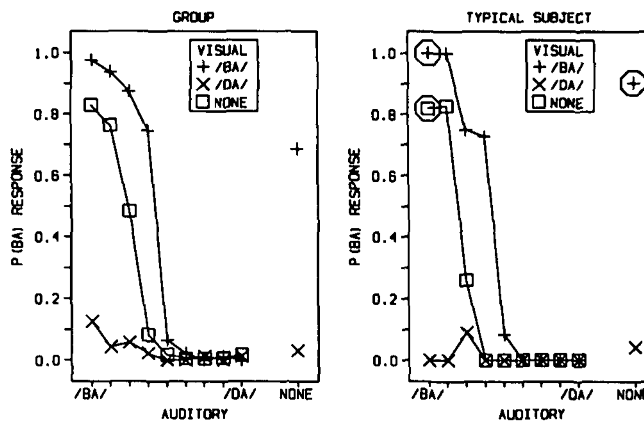

Figure 4. Probability of /ba/ judgments to auditory, visual, and bimodal speech events. The left panel gives the group results, and the right panel gives the results for a typical subject. The nine levels along the auditory continuum represent speech sounds varying in equal steps between /ba/ and /da/. The three circled points in the right panel provide evidence for "enhancing" integration.

The proportion of /ba/ judgments to auditory and visual /ba/ shown in Figure 4 cannot be accounted for by nonintegration or the identification of only the auditory or visual source on a given trial. Consider the three conditions that are circled for a typical subject plotted in the right panel of Figure 4. The speech sound at the /ba/ end of the auditory continuum is heard as /ba/ about 80% of the time when presented alone. A visual /ba/ presented alone is identified as /ba/ about 90% of the time. The same speech sound at the /ba/ end of the auditory continuum, however, paired with the same visual /ba/, is identified as /ba/ 100% of the time. This, the visual /ba/ adds to the /ba/-ness of a syllable with an auditory /ba/ more than can be predicted by nonintegration.

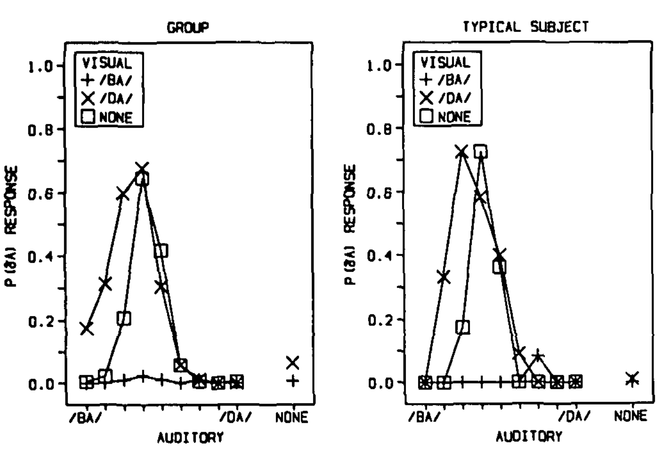

Figure 3. Probability of /ða/ judgments to auditory, visual, and bimodal speech events. The left panel gives the group results, and the right panel gives the results for a typical subject. The nine levels along the auditory continuum represents speech sounds varying in equal steps between /ba/ and /da/.

Attention, consciousness, and information integration

Other experimental results show that the dramatic influence of visible speech occurs regardless of the perceiver's attentional set. The influence of both sources of information is another case in which perceivers are not able to selectively process a single aspect of their perceptual world. This result is reminiscent of the Stroop (1935) effect in which a word influences the naming of the color that the word is printed in. With respect to perception and action in the real world, this result might first seem an inefficient way to design a perceptual process. One would expect selectivity to be more useful in the analysis of our perceptual world; if the perceptual world contains multiple sources of information, however, then selectivity can only be damaging to perceptual interpretation.

Given multiple sources of information, the better strategy would be to evaluate and integrate all sources, even though at any one time many of them may be ambiguous.

Thus, our subjects naturally integrate the auditory and visual speech and are not able to reliably report one dimension without influence from the other. The cost of a system thus designed would be the relative inability to process selectively single aspects or single dimensions of the perceptual world. In natural situations, the lack of selectivity is not usually a problem if the separate information sources are consistent with one another. Watching a speaker is not usually detrimental to hearing the message because a speaker's auditory and visual messages are necessarily consistent with one another. When separate information sources convey different messages, gross motor movements such as head and eye movements are available, allowing us to selectively expose various sources of information. Tunnel vision and eye movements in reading, for example, reduce the interference of surrounding words with the words being read. What the comparison between integration and nonintegration demonstrates is that within a given view of the perceptual world, we find it difficult to selectively process one dimension independently of the others.

On the basis of the findings that perceivers integrate auditory and visual speech, one view of attention might be the following. Multiple sources of information are available to the perceiver. An intention is generated for the task at hand, and as many sources of information are utilized as possible. In many cases, some sources of information will work against the intended action, such as watching the speaker's lips when the observer intends to report what he heard the speaker say. In this case, the lip information may bias the judgment relative to what would be perceived given only the sound. Thus, we have some leakage, in that attention or intention cannot be completely focused. Integrating and utilizing multiple sources of information may hence be the rule rather than the exception. The exceptions occur primarily in experimental tasks that require subjects to do specific things, that is, to have specific intentions. To what extent intentions are consistent with the multiple sources of information in the typical environmental situation is an empirical question.

The results in Figures 3 and 4 are also relevant to consciousness and phenomenal experience. We seem to be conscious at the highest level possible. We tend to experience a particular speech event, not separate auditory and visual dimensions of that event. As Marcel (1983a) proposes, we become conscious of the perceptual hypothesis that is consistent with as much sensory data as possible. The speech perception results also imply that there are rules of parsimony and ecological validity with respect to the perceptual hypotheses. Contradictory auditory and visual events are not a viable interpretation of contradictory auditory and visual sensory information. Rather, the perceptual hypothesis is considered to be a single speech event supported by both auditory and visual sources of information.

Although our findings provide a good example of the phenomenological unity of perception, we must be careful to distinguish between the processes that bring this experience about and the experience itself. The phenomenal experience is influenced by both the auditory and visual sources of information. However, given this experience, it is relatively difficult for the perceiver to tease out the separate auditory and visual contributions. In apparent contrast to our phenomenal experience, according to the FLMP, the auditory and visual sources of information are independent before integration. Thus, bimodal speech perception is an example in which one process leading to the phenomenal experience treats the auditory and visual sources as independent, but the outcome and phenomenal experience give the impression of dependence.

Continuous perception

The study of speech perception has been almost synonymous with the study of "categorical perception," a putative mode of perception by which changes along a stimulus continuum are not perceived as continuous, but as discrete (Studdert-Kennedy et al. 1970). According to categorical-perception theory, it is claimed that listeners are limited in their ability to discriminate between different speech sounds that belong to the same phoneme category. The sounds within a category can be identified only absolutely, and discrimination is possible for only those sounds that are identified as belonging to different categories. In much of the current literature in psychology, it is stated that speech is perceived categorically (J. R. Anderson 1985; Eimas 1985; Flavel 1985; Miller 1981). I examine the evidence for this and find it unconvincing. Previous results and more recent studies are better described in terms of continuous perception – a relatively continuous relationship between changes in a stimulus and changes in perception.

In addition to revealing the weaknesses in previous evidence, I present several approaches to testing for categorical versus continuous speech 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 have been interpreted as showing categorical perception because discrimination performance was reasonably predicted by identification performance (Studdert-Kennedy et al. 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 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 1987a; Massaro & Oden 1980). Two new tests are proposed that use formal techniques in the areas of psychophysics and mathematical psychology.

In most studies of categorical perception, only a single property of a speech syllable is varied; theorists have not addressed whether it is this property or the syllable that is perceived categorically. Categorical perception of the property and categorical perception of the syllable are two possible forms of categorical perception. When there is independent variation in two properties of the speech stimulus, one must distinguish between these two forms of categorical perception. Making this distinction, two different categorical models of bimodal speech perception are assessed. In the first model, the information obtained from each modality is assumed to be categorical. In the second model, the information from each modality need not be categorical. What is categorical is the percept resulting from the integration of the information from the separate modalities. Because previous advocates of categorical perception were not clear about which alternative they supported, we tested both.

The most convincing evidence against categorical perception comes from direct experimental comparisons 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. Subjects were also asked to make repeated ratings of how well a stimulus represents a given category. 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 degree to which the information represents the category *A*. The left panel of Figure 5 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

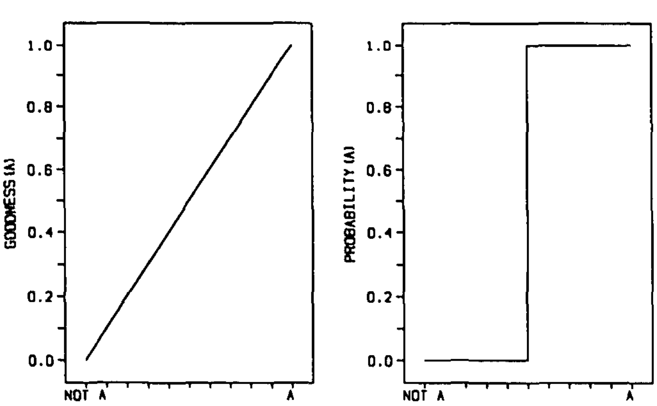

Figure 5. Left panel: The degree to which a stimulus represents the category *A*, called goodness(*A*) as a function of the level along a stimulus continuum between *not A* and *A*. Right panel: The probability of an *A* response, Probability(*A*), as a function of the stimulus continuum if the subject 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.

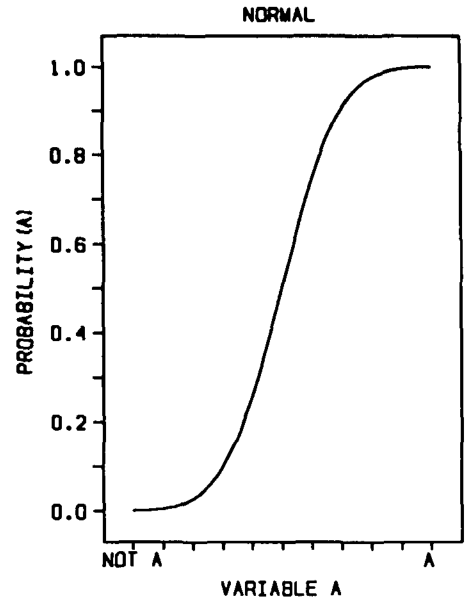

Figure 6. Probability(*A*) as a function of variable *A* given the linear relationship between goodness(*A*) and variable *A* and the decision criterion represented in Figure 5, but with normally distributed noise added to the mapping of variable *A* to goodness(*A*).

is 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 5. 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* produces 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 away from the criterion value. Figure 6 illustrates the expected outcome for identification if there is normally distributed noise with the same criterion value assumed in Figure 5.

If the noise is normal and has the same mean and variance across the continuum, a stimulus whose mean goodness 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, our 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 have been impressed with the fact that discrimination did not exceed that predicted by categorization if the discrimination task resembled something 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 1987a).

At the theoretical level, it is necessary to distinguish between sensory and decision processes in the categorization task. What is central 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 others 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 can easily discriminate among different exemplars of the same speech category. 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.

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 F2 and F3 transitions were 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 between a rising and a falling transition more easily than 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 the exact magnitude of change. This weak form of categorical perception would arise from a property of auditory processing rather than a special characteristic of speech categories. Thus similar results would be found in humans, chinchillas, and monkeys as well as for nonspeech analogs (as they are, 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. We must hence explain 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 /pa/ 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. (This author has had similar experiences, hearing certain German phonological categories in terms of similar English ones.) Our 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 have been 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.

Despite our phenomenal experience and the three decades of misinterpreting the relationship between the identification and discrimination of auditory speech, we must conclude that speech is perceived continuously, not categorically. Our work shows that visible and bimodal speech are also perceived continuously. This observation also seems to pull the carpet from under 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.

Independence

Given the evidence for the integration of continuous sources of information in bimodal speech perception, the next binary opposition concerns whether the two sources of information are dependent or independent. As with other binary oppositions, the question of independence is to some extent model dependent (Ashby & Townsend 1986).

The distinction between independence and dependence is more easily made if it is assumed (as in the FLMP) that three operations – feature evaluation, integration, and pattern classification – are involved in perceptual recognition. Assuming that the evaluation and integration operations are separate, independent sources of information imply that the information value of one source remains independent of the information value of the other. Dependent sources imply a violation of this principle in that the information value of one source is changed by another source. For example, a categorical model of perception (CMP) also has these three operations. This CMP derives discrete categorizations for the visual and auditory sources (feature evaluation), chooses one of the two categorizations with some probability (integration), and always responds with the alternative corresponding to the chosen categorization (decision). The primary difference between the CMP and the FLMP is that the CMP has binary outputs from the feature evaluation stage, whereas the FLMP has continuous outputs.

I have taken three approaches to investigating independence versus dependence. One is model-free and two are model-dependent. (Strictly speaking, no test can be completely model-independent; what is meant is less dependence on specific models of performance.) In the model-free test (Cohen 1984), RTs to the single-dimension and bimodal events are assessed to determine whether the two dimensions show some form of dependent interaction. If they do, it should not be possible to account for the RT to a bimodal /ba/ simply in terms of the RT to a visual /ba/ and the RT to an auditory /ba/. If the two dimensions are independent, we might expect RTs to the bimodal event to be somewhat faster than those to the single-dimension events, but the advantage should be completely accounted for by the statistical advantage of having two horses in the race rather than just one.

For the model-dependent tests, independence and dependence models are tested against the perceptual judgments of speech by ear and eye. The previous discovery of a particular kind of dependence between two auditory dimensions serves as the basis for the development and testing of a dependence model of bimodal speech perception. We also ask whether a model assuming independence of auditory and visual information provides an adequate account of bimodal speech perception. The hypothesis of dependence must predict a failure of any model assuming independence (unless the independence model happens to be mathematically equivalent to the dependence that exists). Independence models must assume that the information obtained from one source is independent of the information obtained from the other source. The models described in the comparison between categorical and continuous sources of information are independence ones. Both must fail if the auditory and visual sources are dependent. However, the adequate description of the FLMP provides evidence for the independence and against the dependence of auditory and visual information in speech perception.

In the third approach, the independence model is extended to account for identification judgments obtained using an expanded factorial design that includes conditions in which the sources of information are presented alone. As will be demonstrated in the derivation of the model, very strong predictions are made about the relationship between the classification of the single sources of information and the classification of speech events composed of both auditory and visual sources of information. The model predicts that the information available from just a single source of information is identical to the information from that source when both sources of information are available. This is a case of complete independence in that the feature value of a given level of a source is identical when presented alone and in the context of an additional source of information. If the model is able to describe the identification of speech events composed of single sources and composed of both sources given this constraint, it provides very strong evidence for the idea of independence. Evidence in favor of independence, as stated previously, automatically contradicts the idea of dependence.

Three different types of tests support independent feature evaluation. Of course, it is logically possible that dependence exists but that an independence model is able to capture that particular form of dependence. Until an adequate dependence model is presented, however, the most parsimonious description is that of independence. Given the evidence for the integration of independent and continuous auditory and visual sources of information, we now turn to the challenging question of how they are integrated.

Optimal integration

The binary contrasts up to this point have provided evidence for the integration of independent and continuous sources of information in bimodal speech perception. Determining the nature of the integration process is the most difficult binary opposition to resolve, primarily because of its intricate dependence on quantitative models and tests (Falmagne 1985). One can distinguish two general classes of integration algorithms: "compromising" and "enhancing" algorithms. When two sources of information point ambiguously to the same alternative, a compromising integration, such as averaging, would give an outcome somewhere between the separate evaluations of the two sources. For example, if one source gives .6 support and the other source gives .8, then averaging gives .7. An enhancing integration, such as that given by the FLMP, would give an outcome more extreme than the separate evaluations of the two sources. In our example, an enhancing integration would give a value greater than .8.

The experimental evidence supports an enhancing integration process. The same three points in Figure 4 used to support integration support enhancing over compromising integration. The judgment given two sources is more extreme than that given either one presented alone. In addition to the experimental evidence, a theoretical reason against adding or averaging the evaluations of the separate sources of information is that this combination rule is nonoptimal. If perceptual systems evolved to integrate multiple sources of information, we would expect the integration to be highly efficient and productive (i.e., optimal). Averaging an ambiguous source of information with an informative source will tend to neutralize the judgment relative to the informative dimension being presented alone. Optimal integration requires that more ambiguous sources be given less of a say in the decision. Some time after the development of the FLMP it was determined that it corresponded to an optimal integration rule. Multiplying the sources of information within the context of the FLMP (Massaro 1984; Oden & Massaro 1978) is mathematically identical to Bayes' theorem.

The mathematical equivalence of the FLMP and Bayes' theorem poses a new dilemma for the FLMP. Given that previous research has rejected Bayes' theorem in a variety of judgmental situations, why has the mathematically equivalent FLMP been so successful? Two possibilities must be considered. The first is that previous rejections of the Bayesian model have been premature and the second is that the models have been tested in very different domains. Both possibilities appear to obtain to some extent. The rejection of Bayes' theorem in many experiments has been a rejection of a normative rather than a psychological form of the model. Predictions have been derived on the basis of the objective rather than the subjective sources of information. In contrast, tests of the FLMP allow for subjective values for the various objective sources of information. Consider the test of the model in situations in which subjective base rates are assumed to be equal to objective base rates. In these cases, performance falls short of the predictions of Bayes' theorem (Leon & Anderson 1974). Central to the FLMP, however, is the evaluation stage that transforms the objective source of information into some subjective value. Thus, performance could still fall short of the optimally objective prediction but it could be described by the same algorithm if subjective values are assumed.

The second possibility is also realized to some extent because most tests of Bayes' theorem have focused on estimates of probability in some variant of the two-urn task. Subjects see two urns and are told the proportion of red and blue beads in each urn. One urn in picked with some probability and a sample of beads is drawn. Given the sample, the subject estimates what urn was in fact picked. The probability of picking an urn, the relative proportion of beads in each urn, the sample size, and the sample makeup can be varied. Even in these situations, however, the normative Bayesian model has been rejected (e.g., Leon & Anderson 1974). It remains an empirical question whether the FLMP can survive in situations in which the Bayesian model has supposedly failed. Experiments and model tests will have to be analogous to those described in *Paradigm.*

Development and aging

Any theory of language processing must eventually consider the acquisition of the processes involved in this skill. It is surprising how little research on speech perception across development has been carried out relative to the large number of studies of infants and young adults. Lifespan studies are central to evaluating hypotheses about the processes responsible for observed differences and similarities in language processing with age. Experiments reported in *Paradigm* were designed to assess how evaluation and integration of audible and visible speech occur across a broad range of ages. The experiments addressed the same binary oppositions that were tested with adults except that interactions with development were of primary interest.

To illustrate the value of the FLMP framework in the study of developmental and lifespan changes, consider a subset of an experimental study reported in *Paradigm.* An expanded factorial design was used: Preschool and college subjects identified auditory, visual, and bimodal speech. The three tasks taken in combination assess the role of auditory and visual information in speech perception. One question is how bimodal speech perception results from some combination of the auditory and visual sources of information. Because the FLMP and the CMP make definite predictions about performance across the three tasks, the results were tested against these models.

Three experimental conditions were tested between blocks of trials: bimodal, visual, and auditory. During each trial of the bimodal condition, one of the five auditory stimuli on the continuum from /ba/ to /da/ was paired with one of the two visual stimuli, a /ba/ or a /da/ articulation. Trials in the visual condition used the same video tape but without the speech sounds. In the auditory condition, the TV screen was covered so that only the auditory information was presented. In the bimodal condition, subjects were instructed to watch and to listen to the /man on the TV" and to tell the experimenter whether the man said /ba/ or /da/. Before the visual condition, each child watched the experimenter's mouth as she demonstrated silent articulations of the two alternatives. In this condition, children were instructed to report whether the speaker's mouth made /ba/ or /da/. The college students were told simply to lip-read. In the auditory condition, the subjects were instructed to listen to each test sound to indicate whether the man said /ba/ or /da/. Because the task required a two-alternative forced-choice judgment, a single dependent variable, the proportion of /da/ responses, provides all the information about choice performance. For both groups, the auditory and visual variables produced significant differences as did the interaction between these two variables in the bimodal condition. Figures 7 and 8 give the results for the three conditions for two groups of subjects.

The FLMP gave a much better description of performance than did the CMP for both groups of subjects. This advantage of the FLMP held for both the results of individual subjects and the results of the average subject. The fit of the FLMP to the average subject was about three or four times better than was the description given by the CMP. The much better description given by the FLMP is especially impressive given that this model required one fewer free parameter than did the CMP. The good fit of the FLMP suggests that preschool children integrate independent and continuous sources of information in the same manner as adults. Using the FLMP, we ask whether the information values for the auditory and visual sources change with age and whether the processes involved in the perceptual recognition of speech differ with age. The better fit of the FLMP than the CMP for both age groups argues against a developmental change from one type of process to another in the perceptual recognition of speech. At every age, performance is more appropriately described as following the operations of the FLMP, which accordingly provides a framework for assessing lifespan differences in information value.

A reasonable measure of information value is the degree to which a subject discriminated the different levels of the speech dimensions. An index of discrimination can be determined by taking the difference in parameter values for two different levels of a speech dimension.

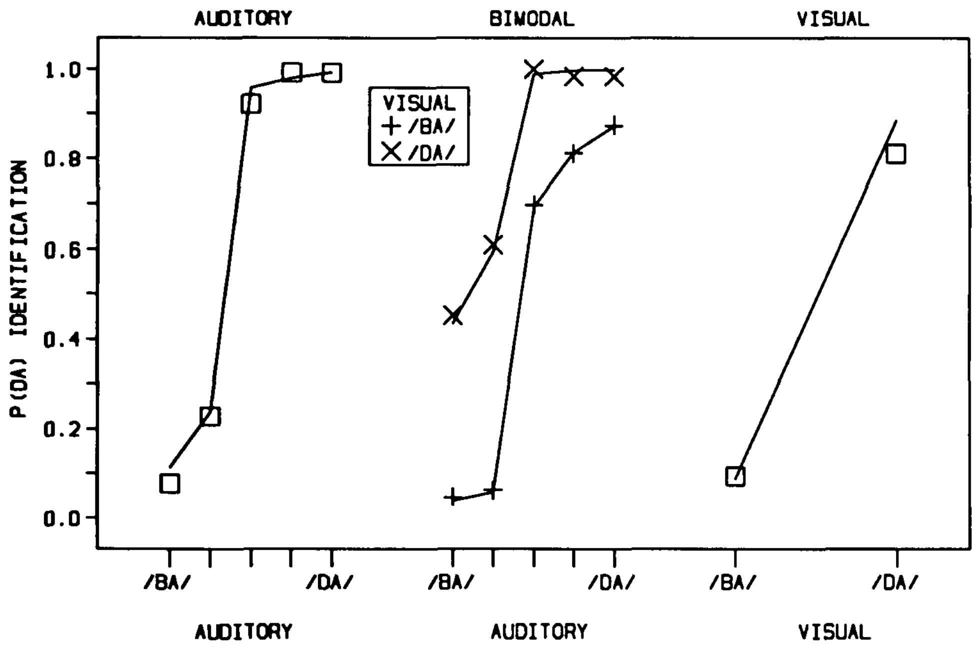

Figure 7. Observed (points) and predicted (lines) proportion of /da/ identifications for auditory (left panel), bimodal (center panel), and visual (right panel) trials as a function of the auditory and visual levels of the speech event for the experienced preschool children. The predictions are for the FLMP.

Table 1 gives the average parameter values for the two groups of subjects. In this case, visual discrimination is given by the degree of /da/-ness of a visual /da/ minus the degree of /da/-ness of a visual /ba/. As an example, visual discrimination for the preschoolers would be .885 minus .087 or .792. The auditory discrimination is given by the degree of /da/-ness of the most /da/-like auditory syllable minus the degree of /da/-ness of the most /ba/-like auditory syllable (that is, the two end points along the auditory continuum). This auditory discrimination value is .992 minus .114 or .878 for the preschoolers. Corresponding values for the college students were .962 for visual discrimination and .975 for auditory discrimination. Thus, the significantly larger discrimination values for the college students indicates that they had significantly more information about both auditory and visual speech than did the preschool children. This result and the good description by the FLMP of both groups suggest

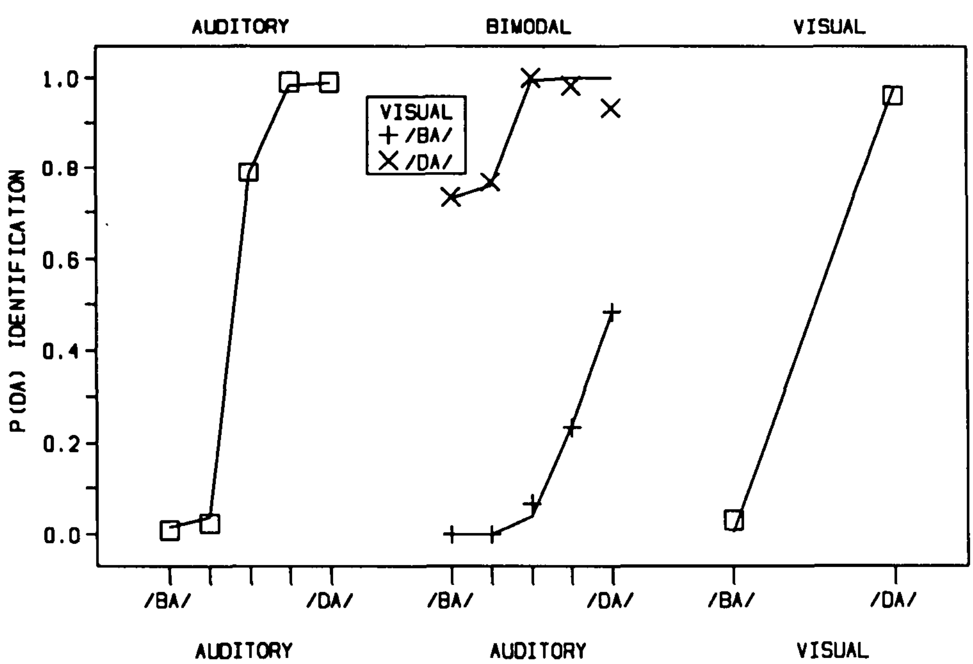

Figure 8. Observed (points) and predicted (lines) proportion of /da/ identifications for auditory (left panel), bimodal (center panel), and visual (right panel) trials as a function of the auditory and visual levels of the speech event for the college students. The predictions are for the FLMP.

that the developmental differences are due only to information differences. The processing of the information appears to follow the operations of the FLMP for both groups.

The developmental and lifespan study of speech perception by ear and eye revealed remarkable similarities and yet some reasonable age-differences. With respect to the binary oppositions described in Figure 2, the outcomes of the tests did not change across development and aging. The fundamental processes involved in pattern recognition as described by the FLMP appear to exist at age 3 and to remain constant for the next seven decades. (Other experimental paradigms than those described in *Paradigm*, such as habituation and preferential looking, will have to be used to investigate the first three years of language perception.)

Although the processes underlying speech perception did not appear to change across the lifespan, there were significant differences in performance. A distinction between information and information processing is central to understanding these lifespan changes. There were significant differences in the informational value of audible and visible speech as a function of age, but no differences in information processing. These changes are readily explained in terms of increasing experience as well as changes in the sensory systems with aging. Preschool children are still acquiring speech-perception skills. They do not lip-read as well as adults and are less able to discriminate changes along an auditory speech continuum. The differences in performance are accurately described in terms of the feature evaluation stage of the FLMP. A given source of information is less informative for preschool children than for adults. This is not surprising given that it is experience with speech that permits speech data to be treated as information. We can expect that the prototype descriptions of the distinguishing characteristics of speech will increase in resolution with experience. Aging, on the other hand, can decrease the resolution of the sensory systems, resulting in less accurate speech perception. Luckily, the availability of multiple sources of information should prevent a catastrophic breakdown even with the fairly severe loss of a given source. For example, visible speech from the speaker's lips appears to compensate for high-frequency hearing loss with age.

The value of the FLMP is that it not only describes how speech perception might be accomplished, it provides a framework for understanding how it changes with development and aging. These findings have developmental implications both within and outside the field of speech perception. Five different areas of developmental research are related to the present results. First, infants appear to perceive the joint occurrence of the auditory and visual dimensions making up bimodal events (Spelke 1979a; 1979b). Kuhl and Meltzoff (1982) provided some evidence that infants can spot the relationship between visible and audible speech complements. This finding is consistent with our finding of a substantial contribution of lipreading to children's speech perception. The integration of auditory and visual speech requires that they be evaluated and integrated with respect to the same speech event. Noticing the correspondence between these two modalities allows the infant and child to learn about both characteristics. In terms of the FLMP, the prototypes corresponding to speech segments contain visual as well as auditory information, and the resolution of these prototypes increases with experience.

Table 1. *Average parameter values indicating degree of /da/-ness given the auditory and visual sources for the FLMP for the preschool and college subjects*

	Auditory (a_i)					Visual (v_j)	
Group	/ba/	2	3	4	/da/	/ba/	/da/
Preschool	.114	.237	.959	.980	.992	.087	.885
College	.015	.037	.786	.983	.990	.014	.976

Second, our finding that young children perceive speech continuously and not categorically stands in sharp contrast to conclusions reached in previous speech research with infants. Infant speech perception has been interpreted as being categorical (Eimas et al. 1971; Gleitman & Wanner 1982). The apparent contradiction would be resolved by assuming that speech perception develops from categorical to continuous and is continuous by age three. However, some of the infant research can be interpreted in terms of possible irregularities in the discrimination of some speech dimensions rather than in terms of categorical perception (Massaro & Oden 1980). An "irregularity" occurs when the same amount of stimulus change gives rise to different amounts of perceptual change at different places along the stimulus continuum. Thus, infants might be better at discriminating some changes than others, but this does not mean that they have only categorical information available. Perception would still be continuous in that changes along the stimulus dimension produce noticeable changes in perception, even when these changes occur within a speech category. Evidence for this view comes from more recent studies that have shown within-category discriminations with infants. There is now evidence that infants can make within-category discriminations along nasal-stop (Eimas & Miller 1980), voiced-voiceless (Aslin et al. 1981) and stop-glide (Miller & Eimas 1983) continua. Continuous perception appears to describe infant, child, and adult speech perception.

Third, how does the child project the environment onto the utterance? That is, how does the child project the appropriate interpretation on the perceptual form of the utterance? As an example, when an adult points to a cat and says "The cat is on the mat," the child supposedly has some perception of an utterance and some perception of a cat on a mat. I do not plan to solve this projection problem, but simply point out that it might differ significantly from what has been stated previously. Because of the belief in categorical perception it has usually been assumed that the perception of the utterance contains ordered categories bracketed by stress into words and by intonation into phrases (Gleitman & Wanner 1982). In this case, the projection problem only involves pairing the appropriate semantic interpretation of the environmental event with the specific surface form being heard. If categorical perception does not exist, however (and there is little reason to believe that it does), then the problem is that the perceived utterance is also somewhat ambiguous. Experience with language must reduce ambiguity about both the utterance, the environment, and their pairing. This does not appear to be a formidable task. Analogous to Premack's (1986) faith in the power of behavioral methods in the study of animal communication, the child's experience will allow the code to be broken.

I view language acquisition as highly interactive in that the meaning of the situation can help disambiguate what is being heard, not simply the interpretation of the categories being heard. For example, if the child hears an utterance that narrows down the speech alternatives to *ball* and *doll,* the physical presence of one referent or the other can contribute to word recognition. In this case, the child can obtain feedback about the auditory properties of the words and can build prototype descriptions that differentiate them. Analogous to development in other domains (Kendler 1983), language acquisition at all levels (even the feature and segmental levels) is continuous and gradual. The overthrow of categorical perception offers a new outlook and a new challenge for theories of language acquisition.

Fourth, some developmental theory is centered around the idea that the child progresses from holistic to dimensional processing (Shepp 1978; Smith & Kemler 1977; Werner 1957). To accommodate the present negative results one can always claim that preschool children only perceive some stimulus events holistically; other events would be perceived in terms of independent attributes, features, or dimensions. Alternatively, one can reinterpret the previous studies used to support holistic perception. Consider the perceptual categorization of objects varying in size and brightness (Smith & Kemler 1977). Older children and adults will tend to group two objects together if they are the same size even if they differ greatly in brightness. Younger children, on the other hand, will group two objects together if they differ by relatively small amounts in both dimensions. Rather than accepting these results as implying holistic perception in young children, however, the dimensions of size and brightness might be evaluated independently at all developmental levels. The different results may simply reflect different strategies in the grouping task at the different developmental levels. Consistent with this interpretation, Kemler and Smith (1978) found that young children could treat size and brightness as independent dimensions to learn a higher-order conceptual rule. Furthermore, contrary to what might be expected from holistic perception, Ward (1980) found that 5-year-olds had no difficulty making dimensional judgments given length and density.

Fifth, popular explanation of declines in perceptual and cognitive functions with aging is based on the notion of processing capacity. The assumption is that age-related changes in performance result from reduced processing capacity or resources with aging (Salthouse 1985). In further support of the well-founded skepticism about the resources concept that already exists (Navon 1984; Neumann 1987), it appears to offer little insight into understanding performance differences across the lifespan. The results described in *Paradigm* illustrate that a process model is essential to understand aging differences in bimodal speech perception. With such a model, the fundamental differences are revealed to be a result of the information available rather than a result of the processing of the information. What is important is that information differences can lead to performance differences that could easily be mistaken for processing differences. Having less information available would only naturally lead to an increase in the time to perform various tasks. As an example, having both auditory and visual sources of information speeded up speech recognition relative to having just one of these sources. Investigators of lifespan changes will have to tease apart informational differences and information-processing differences to arrive at a better understanding of development and aging. We turn now to the generality of our findings.

Generality

The question of generality concerns the uniqueness of the processes that have been uncovered in bimodal speech perception. These processes seem to be similar to those involved in a number of other domains of perceptual and cognitive functioning, such as "person impression," the learning of arbitrary categories, sentence interpretation, probability judgments about possible events, and judgments of category membership. To the extent that we can provide a unified account for this broad range of phenomena, modularity is not a reasonable guideline for psychological inquiry. According to Fodor's (1983) classification, these five domains include both input modules and central systems. It should not be possible to provide the same process description across these domains, especially a description developed from another unique input domain of bimodal speech perception.

In person impression, subjects are given descriptions of a hypothetical person and asked how likeable (or extroverted, etc.) they judge the person to be (Anderson 1981; Asch 1946). If the bimodal speech framework can be generalized to this domain, then the judgments should reflect the integration of multiple, continuous, and independent attributes and the integration should be enhancing rather than compromising. In the learning of arbitrary categories, subjects are given examples of members of the categories along with their category status (Medin et al. 1984). The exemplars are composed of a set of features that are independently varied. After the learning phase of the experiment, subjects are asked to categorize these exemplars and other new ones. By testing the FLMP against the results of these experiments, we can determine the extent to which the categorization of arbitrary object categories resembles the categorization of speech. In sentence interpretation, subjects are asked to interpret sentences that are varied along a number of dimensions (MacWhinney et al. 1984). The experimental question is how the various syntactic and semantic properties of the sentence, such as word order and animacy, influence its interpretation. The fourth domain involves the judged probability of some event given hypothetical descriptions of the situation (Tversky & Kahneman 1983). Does this kind of decision-making fall outside the regularity we have observed in speech perception by ear and eye? Finally, subjects judge the degree to which a pictured object represents a given category (Smith & Osherson 1984). For example, to what degree does a picture of a colored object represent the categories as apple, red, and red apple?

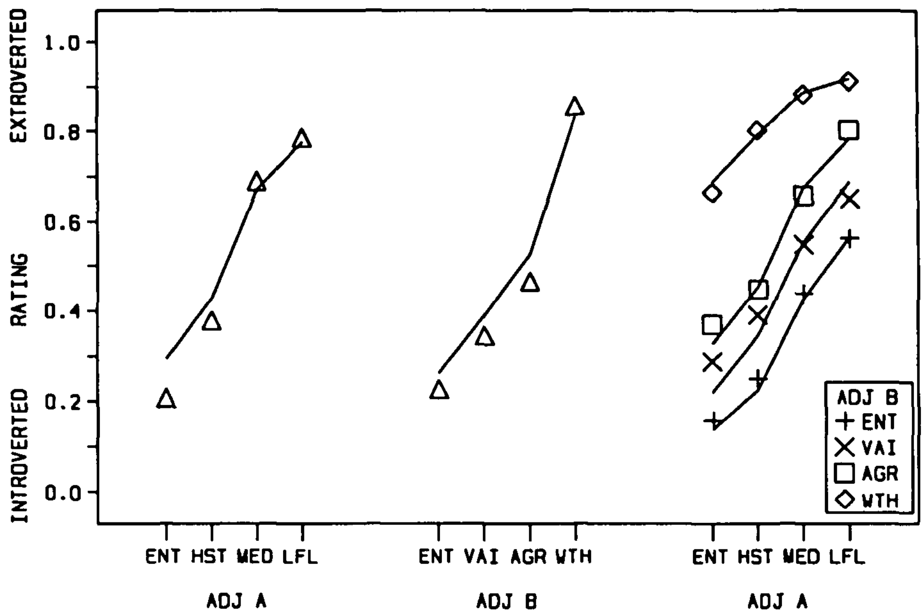

Figure 9. Observed (points) and predicted (lines) judgments for the single adjectives and pairs of adjectives. The four A adjectives were: entertaining, high-strung, meditative, and lifeless. The four B adjectives were: withdrawn, agreeable, vain, and enthusiastic. Predictions given by the FLMP.

All five of these disparate areas of inquiry have proved to be surprisingly consistent with the general framework developed in *Paradigm*. To illustrate the application of the FLMP to other areas, consider one of our experiments on person impression. Subjects judged how introverted-extroverted a hypothetical person was, given one or two traits. Four traits were factorially combined with four other traits, yielding 16 possible pairs. In addition, subjects judged a hypothetical person described by each of the eight unique adjectives. The subjects were instructed to rate the degree of introversion-extroversion of the hypothetical person on a scale between zero and one. Two sets of four adjectives were used. The four A adjectives were: *entertaining, high-strung, meditative,* and *lifeless*. The four B adjectives were: *withdrawn, agreeable, vain,* and *enthusiastic.*

Figure 9 shows the observed rating judgments. An analysis of variance indicated a highly significant interaction between the *A* and *B* variables. This interaction is consistent with the FLMP but not with a simple adding or averaging model. The FLMP and the averaging model were fitted to the average ratings of the 16 unique pairs of adjectives and the eight adjectives presented alone. The FLMP did a somewhat better job in describing the results than did the averaging model. The standard deviations were .039 for the FLMP and .035 for the averaging model. The predictions for the FLMP are shown in Figure 9. We conclude that the framework of the FLMP provides an account of the cognitive processes involved in person impression, contrary to what might be expected from modularity.

Relationship to connectionist models

Paradigm has anticipated many of the distinguishing properties of a recent but influential movement in psychological inquiry called connectionism. [See Smolensky "On the Proper Treatment of Connectionism," *BBS* 11 (1) 1988.] Connectionism shares several fundamental properties with the current theoretical framework as instantiated in the FLMP. First, both theories assume continuous rather than discrete representations; the fuzzy truth values of the FLMP are analogous to the continuous levels of activation and inhibition of connectionist models. Second, both theories acknowledge the existence of multiple constraints on human performance and provide an account of the evaluation and integration of multiple sources of information in pattern recognition. Third, there is the parallel assessment of multiple candidates or hypotheses at multiple levels in both models. Fourth, both provide a common metric for relating qualitatively different sources of information. In the FLMP, each source of information is represented by fuzzy truth values indicating the degree to which alternative hypotheses are supported. Activation levels play the analogous role in connectionist models. Fifth, the automatic categorization of a novel instance can be accomplished in both models. Finally, in both theories, pattern recognition involves finding the best fit between the relevant constraints and the pattern that is perceived. Particularly striking examples of this phenomenon were observed in bimodal speech perception, when, for example, subjects perceived /ða/ given an auditory /ba/ paired with a visual /da/ (see Figure 3).

The close fit between the FLMP and connectionism calls for an exploration of their similarities and differences. Although the two approaches appear to agree on all metatheoretical criteria, the specific models to date differ in terms of the amount of connectivity in the system. The FLMP assumes no top-down influences from a higher-level unit on the activation of a lower-level unit and no inhibition among units at a given level. Connectionist models usually assume both, and experimental tests have shown these assumptions to be wrong (Massaro 1988b). These differences are at the level of specifics and not at the level of metatheory. In fact, Massaro and Cohen (1987) proved that a connectionist model with input and output layers is mathematically equivalent to the FLMP. As presently formulated, however, many of the connectionist models with two-way connections among different levels of units and connectivity among units at a given level are too powerful (Massaro 1988b). They are capable of predicting not only observed results but also results that do not occur. That is, some connectionist models can simulate both correct and incorrect process models of performance. Although I am sympathetic to the central assumptions of the connectionist models, my worry is that the formal power of the models may distract us from the business at hand – trying to find regularities and laws and to eliminate alternative explanations of behavior (Massaro 1986a; 1986b).

Conclusion

The research described in *Paradigm* casts some light on how speech is perceived by eye and ear. In addition, the same processes that are involved in speech perception appear to occur in a wide variety of perceptual and cognitive domains. The results provide evidence against modularity because processing in horizontal (cognitive) tasks engages modules that are supposedly encapsulated and specific to a vertical process (input module) such as speech perception. Accordingly, Fodor's (1983) pessimism about understanding complex cognitive functioning seems unjustified. Given the contribution of vertical processes to cognitive skills, insights into the former clarify the latter. The alternative proposed in *Paradigm* is that the same basic processes contribute to an impressive variety of perceptual and cognitive actions. Given the value and prevalence of parsimony in scientific inquiry, we should not be surprised that laws uncovered in one behavioral domain apply to other domains.

ACKNOWLEDGMENT
The research reported in this article and the writing of the article were supported, in part, by grants from the Public Health Service (PHS R01 NS 20314), the National Science Foundation (BNS 8812728), the graduate division of the University of California, Santa Cruz, and a James McKeen Cattell Fellowship. The author would like to thank Michael Cohen for eclectic assistance.

Open Peer Commentary

Commentaries submitted by the qualified professional readership of this journal will be considered for publication in a later issue as Continuing Commentary on this article. Integrative overviews and syntheses are especially encouraged. All page references are to Massaro's Speech Perception by Ear and Eye *unless otherwise indicated.*

Speech perception as information integration

Norman H. Anderson
Department of Psychology, University of California, La Jolla, CA 92093-0109

Dominic Massaro's book is a glittering paradigm of scientific inquiry. Massaro takes a central problem in speech perception, introduces a theoretical position at odds with the dominant view, and subjects both views to careful, detailed conceptual scrutiny and indefatigable experimental analysis.

Speech perception. The central problem and basic theme of the book is that language processing involves the integration of information. Multiple sources of information are typically relevant, even for the perception of a simple phoneme, and these sources should ideally be integrated into a unitary perception. Much of Massaro's extensive experimental analysis is aimed at establishing that a certain quantitative model governs the information integration.

The integration view, however, differs from the dominant view in speech perception: "The study of speech perception has been almost synonymous with the study of categorical perception" (p. 90). The categorical view is strongly supported by phenomenal experience. If a sound is varied continuously between /ba/ and /da/, the listener does not experience a continuous change, but rather an abrupt shift from one percept to the other. The categorical quality of this phenomenal experience is compelling, and the standard interpretation has been that it mirrors a corresponding categorical quality of perception. Categorical perception would have fundamental implications for the nature of perceptual and cognitive representations.

Massaro argues that perception is continuous, that the phenomenal discreteness is imposed only at a later stage of decision. This becomes more specific in the distinction between valuation and integration: Valuation processes map the stimulus continuum into a continuum of perceptual information; perceptual information from multiple sources is integrated to produce a (sometimes) discrete percept.

This integration approach is pursued with insight and determination. Massaro conceptualizes continuous information in terms of fuzzy logic, whence the acronym FLMP (fuzzy logic model of perception) for his basic model. Choice response is assumed to obey Luce's choice model, analyzed with functional measurement (p. 51). Massaro exploits the bimodal paradigm of speech perception by ear and eye, which represent two independent sources of sensory information that need to be integrated in perceiving speech. He implements this conceptual approach with ingenious integration designs in careful, extensive experimental studies.

Massaro thus presents a unified approach to a succession of basic issues. Chapter 3 takes up integration versus nonintegration. Chapter 4 shows that previous studies of identification-discrimination do not actually support categorical perception, as had been claimed, and presents new studies to support the continuous view. Chapter 5 applies the categorical-continuous distinction to integration models. Chapter 6 discuss the basic issue of independence-nonindependence.

All these issues become complex as Massaro attempts to clarify and pin down the categorical view. At every turn, however, Massaro's FLMP seems to win the day. Categorical perception still lives (e.g., Harnad 1987), and these chapters will no doubt be scrutinized in other reviews by workers in speech perception. It seems clear, however, that Massaro has transformed the outlook in this field.

General judgment theory. The last chapter considers the question of generality: Are the ideas, methods, and results of the previous chapters specific to certain problems of speech perception? Or will they apply generally across psychology? Massaro believes the latter, and illustrates applications in a number of other areas. Among these is my own area of judgment-decision, including social judgment, and I question whether Massaro's formulation will have as much generality as he suggests.

There are, to be sure, similarities and parallels between Massaro's view and my own theory of judgment-decision. The basic issue of independence, to which Massaro gives so much consideration, was also central in social judgment, as Massaro points out (pp. 171f.). Phenomenological report was as mistaken in this area as in categorical perception. The use of variability in rating responses to refute a categorical model (pp. 131ff.) was originally developed to test the similar Piagetian hypothesis of centration (p. 56; see, e.g., Anderson 1982, p. 172). The expanded factorial design (pp. 183, 248) has been used in a number of previous judgmental studies (e.g., Anderson 1981, Figures 1.22 and 4.21); such designs provide the basis for uniqueness of weights and scale values in the averaging model. Underlying all this is the common basic concern with multiple determination and with processes of valuation and integration.

Massaro's specific concern, however, is whether his FLMP will apply to a spectrum of judgment-decision tasks that have heretofore been interpreted in terms of an averaging model. One new experiment on social judgment is presented in which FLMP is claimed to be superior to the averaging model. This test is theoretically inappropriate, however, because it did not allow differential weighting. Differential weighting for extremity effects has been an integral aspect of averaging theory almost since the beginning (see Anderson 1981, section 4.4.2), and would clearly account for Massaro's data.

More important, there is a conceptual objection to the FLMP as a general model for judgment-decision theory. The focal concern of the model is with categorical choice in uncertain, threshold-type situations with a probabilistic response. Many judgment tasks, however, ask for numerical response in suprathreshold situations that do not involve uncertain information. In serial number averaging, for example, subjects form an intuitive cumulative average of a sequence of numbers presented one at a time. Behavior in this task is well described by an analog averaging model, with serial weights to account for the serial position effects (Anderson 1981, section 2.5.4; Busemeyer, in press). But this number averaging task has no place in a categorical decision model. This task is a prototype of a large class of judgment-decision tasks.

It may be useful, therefore, to distinguish two general classes of tasks, corresponding to this distinction between continuous and choice response. Averaging models, as well as subtraction and multiplication models, may be generally useful for the former, whereas a ratio model may be appropriate for the latter.

In favor of Massaro's approach, everyday behavior may impose a pervasive go/no-go, choice framework underlying various tasks that ostensibly ask for continuous response. A pertinent example comes from a study of the prototypical Bayesian task, in which subjects rate the likelihood that a sample of red and white beads comes from one of two urns, each with specified red : white proportions. This task imposes a choice framework, corresponding to the two urns, and in fact the results were interpreted in terms of a ratio model for competing response tendencies (Leon & Anderson 1974). Similar choice constraint may be found more generally with continuous response, as Massaro suggests for his experiment on social judgment.

The ratio model used by Leon and Anderson for the Bayesian task is similar in form to Luce's choice model, which is used by Massaro. This ratio model has the advantage that it applies to numerical judgments of likelihood in particular, whereas Luce's model is derived from probability assumptions and only applies to probabilistic threshold-type decisions. Moreover, the ratio model may be derived from the general averaging model (Leon & Anderson 1974; Anderson 1981, p. 66). Thus, there may indeed be greater underlying unity.

The subtitle of Massaro's book, "A paradigm for psychological inquiry," is appropriate in two senses. Massaro presents a conceptual paradigm with concepts and methods that provide new analytical power for the study of speech perception. The demonstrated effectiveness of this approach in his extensive experimental studies shows that this new look will not fade away. At the same time, Massaro's work is a paradigm of productive experimental science, one that may stand as an ideal for other investigators.

Independent or dependent feature evaluation: A question of stimulus characteristics

Lynne E. Bernstein

Center for Auditory and Speech Sciences, Gallaudet University, Washington, DC 20002

Electronic mail: *lebernstein@gallau.bitnet*

Massaro finds little in his own work or the literature to diminish his enthusiasm for and confidence in the fuzzy logical model of perception, which *Paradigm* explicates. This reviewer is less convinced of the adequacy of the author's theory to account for certain results in the literature on bimodal speech perception. Although theory can be discussed abstractly, this commentary will make use of published data to challenge one of Massaro's several theoretical claims, namely, that "the feature value assigned to one source of information remains independent of the value assigned to another source of information" (p. 152). The contrasting dependence view, which Massaro rejects, is that "the feature assigned to one source is some function of the feature value for the level of information or the other source" (p. 152). The review draws on studies of cross-modal speech perception that examined the extent to which certain transformations of the acoustic speech signal can improve intelligibility during lipreading by profoundly deaf individuals. These studies provide evidence for dependent evaluation of speech information across sensory/perceptual channels.

The reason for focusing narrowly on only one of the major issues of the Massaro volume is to illustrate a more general point. A body of work for which the status of paradigm is claimed should be capable of accounting for important classes of results, although it may falter in specific instances. Massaro bases his conclusions on his own audiovisual studies of a small selection of speech categories ("ba," "da," "va," and "tha") and a somewhat selective reading of the speech research literature.

The audiovisual stimuli on which Massaro has focused provide place of articulation information to both sensory/perceptual channels. These stimuli are well suited to the conclusion that independent feature evaluation by each of the sensory/perceptual channels precedes integration. In contrast, if the feature is voicing, the question about independence versus dependence can be decided differently.

Research in cross-modal perception for the deaf. Several research efforts had the goal of discovering how to provide (via a cochlear implant, vibrotactile device, or digital hearing aid) a maximum amount of speech information to individuals who have little or no hearing. There is great interest in lipreading, because as yet none of the sensory aids can provide a substitute for normal hearing so that lipreading is not, in general, required for verbal communication. An explicit goal for sensory aids for deaf individuals has been to devise signals that provide information complementary to vision.

Cross-modal voicing identification. The voicing distinction between stops with equivalent place of articulation (e.g., /b/ vs. /p/) depends primarily on the temporal relationship between the onset of glottal pulsing and the release of the tongue from its position of closure in the vocal tract (Lisker & Abramson 1964). The sounds /p,t,k/ are said to be "voiceless," because articulatory release precedes voicing onset; whereas /b,d,g/ are said to be "voiced," because voicing onset closely approximates the moment of release. Chance levels of visual voicing identification are typically reported (e.g., Boothroyd 1988), as it is not possible to observe the vocal folds vibrating. Several studies have focused on providing voicing information to lipreaders (Booth royd 1988; Breeuwer & Plomp 1986; De Filippo 1984; Erber & De Filippo 1978; Grant et al. 1986; Rosen et al. 1981).

Identification data on the phonemes /b,p,m/ from Breeuwer and Plomp (1986) illustrate both the lipreader's difficulty in obtaining voicing from vision and the accuracy of judgments with audiovisual stimulation (Tables 1a–b). The frequency data in Table 1 were submitted to an analysis of transmitted information (Miller & Nicely 1955). Virtually no voicing information was obtained from the visual stimuli alone, only 3.4%. The information that was transmitted may possibly be related to perceptible cheek expansion, which may be slightly greater for voiceless versus voiced stops, since more pressure builds up behind the lips in the former case.

In order to provide the voicing information, Breeuwer and Plomp extracted the glottal pulse periods (or voice fundamental frequency, FO) of the female speaker in this experiment. The FO information was presented over headphones to normally hearing subjects as a sequence of filtered pulses, one pulse per period, in synchrony with the original visual speech. This auditory stimulus has a nonspeech, buzz-like quality. Small absolute differences in stimulus duration and perturbations in FO rate are predictable as a function of original consonant identity (House & Fairbanks 1953). For this auditory signal alone, the percentage of voicing information transmitted was reported to be 28.9% across all the voicing contrasts in the larger set of 18 consonants that were tested. (Raw data were not given for the auditory-alone stimuli.) When vision and audition were combined, the percentage of transmitted information was 84%, with virtually perfect identification of /p/ and /m/ (see Table 1b).

Does this effect result from the combination of two independent feature evaluation processes? Since voicing is virtually invisible, independence seems unlikely. What then is the explanation? The lipreader, with the help of the auditory FO signal, can obtain voice onset time information by observing the point of lip opening relative to the onset of the auditory FO signal. Feature evaluation depends on the *relative* timing of events perceived via two different channels. An evaluation that results from a relationship between two sources logically must be dependent.

A similar result supporting dependent evaluation was reported by Erber and De Filippo (1978), who synthesized stimuli that varied from /m/ to /b/ to /p/. A Lissajous figure displayed on an oscilloscope simulated a mouth opening and then closing, and a hand-held vibrator simulated voicing. In this experiment, the percentage of phoneme identification information for the separate sensory channels was necessarily zero. Subjects used "b," "p," and "m" to label stimuli, and responses were distributed in a manner predicted by the relative onset characteristics of the natural consonants upon which the synthesized stimuli were based.

Conclusions. The audiovisual or tactual-visual voicing distinction seems a strong counterexample to a strict independence model of feature evaluation, since there is no visual voicing information available in the absence of either auditory or tactual information. A consideration of the nature of the speech code makes voicing an obvious candidate factor for further study of this question.

Table 1a (Bernstein). *Confusions among /b,p,m/ when identified in a vision-alone procedure*

	Response		
Stimulus	p	b	m
p	31	10	7
b	22	20	6
m	19	16	13

Source: From Breeuwer & Plomp 1986.

Table 1b (Bernstein). *Confusions among /b,p,m/ when identified in an audiovisual procedure with auditory FO*

	Response		
Stimulus	p	b	m
p	45	1	0
b	2	39	7
m	0	0	46

Source: From Breeuwer & Plomp 1986.

Discrimination and categorization across the life span

Marc H. Bornstein

*National Institute of Child Health and Human Development, Bethesda, MD 20892**

Electronic mail: *br2@nihcu.bitnet*

Perceiving constitutes the "first step" in experiencing and interpreting the world, and for this reason studies of perceiving have long attracted psychological researchers. Discrimination and categorization are two prominent modes of perceiving which address different levels of analysis. Discrimination concerns how many different stimuli (say) on a continuum, an observer can distinguish. Categorization concerns whether an observer perceives qualitative similarities or coordination among those stimuli; by convention, categorizing implies the treatment of a set of indiscriminable or discriminable stimuli as equivalent or coordinate in some way. Evidence attests that the reciprocal processes of discriminating and categorizing are widespread in human perceptual-cognitive, developmental, comparative, and physiological functioning.

Against normal discrimination, the fact that some dimensions or domains of experience are perceived categorically has provoked numerous interesting and paradoxical questions about existing relations among the world, the brain, and the mind. In addition, developmental studies of discrimination and categorization have provided many important kinds of information about normative processes at different points in the life span, as they have permitted special comparisons of immature with mature perceptual function.

What are the origins of perceptual discrimination and categorization abilities, and what developmental course do they follow over the life cycle? Most research in perceptual development has focused on the related goals of, first, determining whether or not single particular perceptual capacities are present and, second, tracing the emergence and the stability or change in those capacities over time. Some perceptual capacities are given congenitally – even, apparently, in the basic wiring and functioning of the sensory systems – whereas other perceptual capacities depend on experience. Ontogenetic change itself has several possible sources: Development may be genetically motivated and may transpire largely as a reflection of maturational forces; it may be experiential and reflect the influences of the environment or particular events, or both. Normal perceptual development over the life cycle also often entails a certain amount of fine-tuning: In categorization, for example, boundary regions narrow, and plateaus broaden (Bornstein 1979; 1987; Raskin et al. 1983). Whether such developmental fine-tuning reflects maturation, experience, or their interactions, is not yet well understood.

Against this backdrop of important questions, the integration of information from different physical sources, and the course of the development of integrative abilities over the life cycle, present an attractive coordinated set of problems. Seeing and hearing, for example, represent different percepts, which often work in concert and may be perceived as coordinate. There are many events about which the two modalities provide integrated and consonant information. Massaro, for example, has developed such a psychological description of how a perceiver recognizes what a speaker says. Multiple audible, visible, and contextual sources of information must be integrated to support speech perception. To perceive speech bimodally means that the auditory and visual information can be related to the same event.

A dispute that currently occupies students of perceptual development concerns the origins of such multimodal sensitivity: Are different sensations initially integrated or initially differentiated? Integrationists (e.g., Bower 1977) propose that newborns actually cannot distinguish among coordinate inputs arriving at the brain via different sensory modalities. Some research supports this integrationist view, at least insofar as newborns and very young infants respond in a similar fashion to, or treat as equivalent, select visual and auditory stimuli. Though the fact that very young babies match cross-modally supports the integrationist position, it does not supply evidence that infants do not or could not distinguish sensations – that they actually "think" that a sound and a light are the same. By contrast, differentiationists (e.g., E. J. Gibson 1982; J. J. Gibson 1979) argue that the ability to coordinate information across senses develops early in life, that perceptual growth involves increasingly efficient abstraction of invariants or constant stimulus features from an environmental array. Perceptual life begins as diffuse and through experience it differentiates, becoming more selective and acute.

Beyond questions of early ontogeny, how does information processing of this sort change in old age? Life-span studies are central to evaluating theories of the processes responsible for observed similarities and differences in aging. Although multiple sources of information are available regardless of age, the relative quality of the various sources might vary systematically throughout life. Does the informational value for auditory versus visual sources change with age, and do the processes involved in perceptual recognition of speech differ with age? Research in bimodal speech perception suggests that the acquisition of visible speech distinctions occurs gradually in development, as does the acquisition of audible distinctions. Experience with speech enhances the quality of the information but does not appear to influence how the information is evaluated, integrated, and utilized. These observations may be explained by increased experience with age, and by changes in the sensory systems with aging. In this domain, fundamental differences result from the information available rather than the processing of that information. Surely understanding the availability versus processing distinction is central to understanding the growth of perceptual capacity. But so too are the numerous questions about the ontogeny of processing that linger in the background. Coordination of visible and audible information in the speech domain exemplifies an underresearched though potentially fertile (and perhaps unique) focus of inquiry for the developmental investigation of discrimination and categorization.

NOTE

*Address correspondence to Marc H. Bornstein, Child and Family Research, National Institute of Child Health and Human Development, Building 31, Room B2-B15, 9000 Rockville Pike, Bethesda, MD 20892.

Seeing speech is special

Ruth Campbell
Department of Experimental Psychology, University of Oxford, Oxford OX1 3UD, United Kingdom
Electronic mail: *lipread@ox.vax.ac.uk*

Massaro's *Paradigm* makes many large claims: All hang on his demonstrations of the effects of adding a seen face to a continuum of heard monosyllabic speech sounds. Peoples' reports of what they hear are, under these conditions, influenced by what they see. For Massaro this is but one, very clear, example of FLMP at work – a general purpose integration rule that applies to the combination of any and all featural inputs.

Modularity: What constitutes disproof? It is unequivocally a good thing to demonstrate that lip movements and heard speech sounds combine in an effective way to produce a speech percept. Accounts of modular subsystems rarely consider how information from different modalities might be processed integratively. Because of its emphasis on the input imperative, the modularity hypothesis could be taken to imply sense-specificity. Lipreading studies alert us to the possibility that this might not be the case, and this is an issue that modularity supporters must deal with. This is not the basis for Massaro's rejection of modularity, however. Rather, he concludes that the processes that underlie audiovisual speech integration underlie the combination of any features in any domain. At this point I part company with him. Modularity needs stronger disproof. An argument against Massaro's conclusions is the one familiar (to undergraduate teachers) that what one does in the lab may not apply in the world. There is some force to it in this context, for Massaro's task forces one to make categorical decisions on stepped "unnatural" stimuli: Under these conditions I can envisage that one would use a paradigm-dependent processing mode; FLMP could then indeed be general, yet specific to such artificial tasks.

This argument may be considered jejune, however. McGurk and others (e.g. McGurk & MacDonald 1976) have shown that audiovisual speech produces percepts that can differ from those produced by sight or hearing alone. (These studies, incidentally, are well reviewed in *Paradigm*, which introduces, among other things, a range of experimental findings and approaches on the uses of lipreading to the interested reader.) These audiovisual speech illusions are compelling, instantaneous, and relatively unmodifiable by instructions to attend to one or another input modality. Their boundary conditions are still relatively unexplored. (Do we "McGurk" heard monosyllables when they are synthesized with seen speakers of the other gender; with mammals making lip-smacking movements; with birds opening and closing their beaks – even trumpets being muted – the doo-wap-ler effect, perhaps?) Yet there is little doubt that we perceive a single speech event – across modalities. Perceptually, this seems to be a quite different quality of experience compared (say) to concurrently reading and listening to a speech sound.

According to Massaro, McGurk effects arise because of the inherent ambiguity of seen speech; it is only when seen faces combine with less ambiguous heard speech that enhanced integration occurs. But if this is not a speech-special process then identical functions must be predicted when equivalently ambiguous written syllables co-occur with heard tokens. If this were the case then my faith in modularity might be shaken. But would even this really constitute disproof? A process rule like FLMP may apply to all behaviors, without compromising modularity. Unless I have missed an argument (for example, one based on resource limitation) I cannot see why different modules could not work (independently) using identical integration rules – or even why integration across defined domains could not make use of such a decision rule.

Modelling speech sound integration. Is FLMP necessarily the best model for McGurk effects? One aspect of interactive activation models, such as TRACE, is their explicit temporal characteristics. Perceptual identification occurs when the model arrives at a stable state; since activation spreads through the system in real time, one could predict how long the system would take to identify an item. TRACE, adapted to include the seen phonetic feature of mouth closing, could accommodate McGurk effects as effectively as FLMP (see Campbell, 1988, for a sketch) and moreover makes predictions about the relative speed of resolution of the perceived identification.

This may be crucial to distinguishing between different models of audiovisual speech and does not seem to be captured by FLMP. In Massaro's task auditory ba/da tokens are combined with seen ba's and da's. In the "classical" McGurk illusion a heard bilabial (e.g. ba) combined with a seen velar consonant (like ga) can give rise to a powerful illusory da. This is an immediate percept and its phenomenal quality is obscured in Massaro's careful graded studies. By contrast, seen ba combining with a heard ga is a weaker percept (less reliable, more effortful). Is this for local phonotactic reasons (there are no English words that start or end with bga), or is the underlying integration process different? According to FLMP there are no differences in the integration rule and there should be no differences in the reliability or speed of achieving the percept. According to TRACE, by contrast, it is likely that the biconsonantal blend would take longer to achieve and may be less reliable, involving the resolution of the feature array into different time-slots. This supposition is hypothetical; direct simulations are needed – of both types of integration scheme.

Fates. What happens to perceived speech events? Are the independent sources recoverable? We know that to some extent the input modality leaves a trace on later processing. Thus Roberts and Summerfield (1981) showed that repeated presentation of a fused speech sound did not shift the reported categorical boundary of an auditory continuum; only the heard speech sound did that. Yet the auditory-visual and the auditory adaptor stimulus were *perceived* to be identical. In our own studies on short-term memory for lip-read lists (Campbell 1987; Campbell et al. 1988), we find that lip-read material is remembered much as if it has been heard, but that the trace is more fragile to heard speech disruption – whether or not it is combined with sound. A theory of integrated speech perception should set us speculating about how the precursors as well as the products of integrated perception affect cognitive processing.

Massaro's *Paradigm* offers us a powerful, unified view of one aspect of perceptual processing, based on a single, fruitful paradigm. I am not convinced that it is on the right track in its conviction that FLMP can topple the idea that Speech is Special. Nor does it (yet) help me resolve the relation between perceptual judgments and postperceptual processing.

Speech perception by ear, eye, hand, and mind

Nelson Cowan

Department of Psychology, University of Missouri, Columbia, MO 65211
Electronic mail: *psycowan@umcvmb.missouri.edu*

In general, I am sympathetic to Massaro's theoretical framework. My 7-month-old son loves to put his fingers in my mouth and gaze at my mouth while I talk: He is surely gaining multimodal information on how mouth movements correspond to the speech sounds I produce, no matter whether this result is intentional or accidental.

One can view speech as an articulatory act that can be perceived through cues ranging from proximal to remote. Auditory cues typically convey the most detailed information but, paradoxically, they are the most remote from the speech act. This seems fitting, given that the main purpose of speech is to project one's ideas, sometimes over substantial distances, or concurrently to several spatial locations. Visual cues are less remote. They are available only relatively near the speaker and cannot be attributed to the wrong speaker as auditory cues can. On the other hand, as Massaro suggests, they are usually less precise than auditory cues. Finally, manual contact with the speaker's vocal apparatus (i.e., "speech perception by hand") provides cues so proximal that they could not be concealed even by a ventriloquist. However, this perceptual mode is obviously both cumbersome and imprecise.

Ultimately, speech perception appears to depend on the listener's belief about what must have been said (i.e., "speech perception by mind"). It may be for this reason that, given an audiovisual mismatch of cues, one can learn to ignore the more remote, auditory information with practice, whereas one cannot learn to ignore the visual cues even though they are less precise (e.g., see Massaro 1987, p. 77). Speech perception by mind occurs also when there are only contextual cues to guide perception, as in the phoneme restoration effect first described by Warren (1970).

The hypotheses that multiple sources of information are automatically integrated is consistent with my own research. Cowan and Barron (1987) report a cross-modal version of the Stroop effect (i.e., interference with color-naming by conflicting, spoken color words), and Cowan (1989) found that this cross-modal effect did not disappear with practice. In the model of information processing that Cowan (1988) suggested, some of the physical features of each new stimulus are automatically analyzed.

Despite this general agreement, I do have some concerns about Massaro's approach. Foremost among them is the way *quantitative modeling* was used. On p. 222 of the book, Figure 7 (left side, top function) shows that the fuzzy logical model of perception (FLMP) is capable of conforming closely to a "zig-zag" pattern that probably reflects noise rather than a replicable cubic trend. It seems important to impose additional, reasonable modeling constraints (e.g., perhaps monotonicity of the performance function across cue values, outward in either direction from an empirically established optimal cue value) and then see how FLMP fares. In the absence of additional constraints, we don't know the extent to which the advantage of FLMP is attributable to a superior ability to conform to random variations.

The manner in which the *three-stage model* was discussed and illustrated appears to imply that feature evaluation, integration, and pattern classification operate in series, but this implication was not explored. Note that Stages 2 and 3 were both incorporated in a single equation (illustrated, for example, on p. 202 of the book), which suggests that they might be thought of instead as a single stage. I expect that contextual information could also be included directly into the integration equation, in the form of a priori probabilities for each particular phoneme's occurrence.

The assumption of serial operation of the three stages should be supported with evidence or disclaimed.

Massaro's discussion of *bottom-up versus to-down processing* is oversimplified. The statement that lipreading provides bottom-up information seems fair within the task of identifying articulatory motor movements, but lipreading would seem to provide top-down or contextual information for the task of deciding what was heard. This line of reasoning suggests that a visual speech display should be adequate to take the place of lexical context within the phoneme restoration effect. That is, if a brief segment of the sound track of a videotape were replaced by noise, it should be perceptually restored by the uninterrupted visual display even if the missing segment were embedded in a nonsense string.

The debate about *categorical perception* seems correct but incomplete. It ignores the fact that there generally are peaks in the discrimination function at the category boundaries, and that the origin of these peaks is controversial. Theoretically, they could be evidence for an even weaker, but more viable, form of categorical perception. In addition, the developmental discussion should interpret the finding that discrimination of sounds within a phonetic category in the infant's native language declines markedly across the first year of life (see Werker 1989 for a review).

The debate about *modularity in speech perception* also seems incomplete. The support for modularity (e.g., Liberman & Mattingly 1989) comes not only from trading relations among cues, which Massaro discussed in detail, but also from duplex perception (speech and nonspeech percepts arising concurrently from the same stimuli). It can probably be accounted for by learned perceptual modes rather than innate modules, but it needs to be interpreted within FLMP.

Finally, although this is not truly a criticism, the locus of integration relative to the conventional, generic descriptions of information processing remains unspecified. These are interesting types of intermodal effects that should occur only if integration occurs sufficiently late in processing. For example, what if two complementary stimuli were presented within the McGurk paradigm (e.g., both visual /ga/ matched with auditory /ba/, and the converse), either by a male and a female speaking concurrently or by one person pronouncing two syllables in rapid succession? In which situations would the auditory and visual cues be recombined according to their phonetic identities? If recombinations occurred with successive presentation, would this suggest that perceptual syntheses of visual and auditory cues can be revised retroactively? Massaro's book is most useful in provoking such research questions.

Categorical perception of speech: A largely dead horse, surpassingly well kicked

Robert G. Crowder

Department of Psychology, Yale University, New Haven, CT 06520
Electronic mail: *rgcrow@yalevm.bitnet*

Twenty years ago I enjoyed startling friends and frightening undergraduates with the strong form of the claim that speech perception was categorical. Soon, studies by Barclay (1972) and Pisoni (1973) among several others, convinced me that people, after all, had access to continuous information, even for continua of stop-consonant syllables. Although quibbling has continued about what Massaro (1987, p. 118) calls the weak form of the hypothesis, I assumed that many informed bystanders, like me, really didn't *need* convincing of the general point. Thus, to find so much of the present book devoted to a relentless assault on categorical perception struck me as anticlimactic. But I leave to others the debate whether this issue was already nearly closed.

If categorical perception, in its original sense, was too solidly discredited to have been taken seriously in the late 1980s, almost the opposite can be said of the Fuzzy Logical Model of Perception (FLMP), which is just as monotonously upheld by the research reported here. Not only is it victorious in the speech data, but in such diverse applications as person impressions, category judgments, and sentence interpretation as well. The FLMP simply works too well, everywhere it has been applied. Ironically, the case for this model would actually be enhanced by the discovery of some domain in which it *failed* to beat all contenders in sight. This would be the exception that proved the rule, a concept I used to think silly, but now appreciate. The FLMP attracts too much attention in this book for my taste.

Which is a pity to the extent that the FLMP and the argument about categorical perception get in the way of other commanding virtues in Massaro's approach. I count three of these, and *they* form the paradigm promised by the title of the book:

First, the book is data-intensive, without once losing sight of important ideas. Good ideas are, after all, cheap, and what distinguishes us as experimental psychologists from others in the cognitive sciences is that we demand public, empirical tests of our ideas, whereas others use different criteria for justifying their belief. Massaro's book affirms the centrality of carefully arranged and abundant evidence to our science. In an area where some main ideas have derived from linguistics, this emphasis is all the more welcome.

Second, and more specifically about the testability of ideas, Massaro makes it clear that a hypothesis is not properly confirmed or falsified qualitatively, in a vacuum, but must rather be tested against a viable alternative hypothesis, both of them preferably worked out in enough detail to make quantitative predictions. I don't claim, by the way, that my own hands are clean by this criterion, but I wonder if some of the false starts and arguing-past-one-another that have occurred in my specialty would have been avoided if we had adopted the model-comparison principle years ago.

Finally, although Massaro's book is closely focused on audiovisual speech perception, it is also seriously attentive to a wide range of other applications in psychology, including the development of perception, decision-making in general, and social impression-formation, as well as the other applications mentioned above. Of course, this catholicism should not be interpreted as a purely transcendental virtue, since one of Massaro's strong contentions is that speech perception shares information-processing characteristics with other human activities, outside of language. Thus he is advancing this case when he reports parallel findings in other areas. More research programs, perhaps similar to this one in method, will be necessary to decide whether speech is "special" because (a) the stimuli themselves are special, because (b) the brain locus is demonstrably different from other kinds of processing, or because (c) the kind of information processing applied to speech signals is different from other modalities. Even scholars who might disagree vigorously on these three alternatives might nevertheless agree that Massaro's paradigm provides a good way of finding out.

Straw modules

Anne Cutler

MRC Applied Psychology Unit, 15 Chaucer Rd., Cambridge CB2 2EF, United Kingdom
Electronic mail: *anne@mrc-apu.cam.ac.uk*

Massaro's enterprise is impressive, and there are many ways he could have brought it to a resounding conclusion. But the final paragraph he chose begins "If the present generalizations of the current research are reasonable, they strike a blow to the concept of modularity" (p. 281); his conclusion thus returns to the theme of his introduction (pp. 2–4), in which the concept of modularity was described, and its falsification in the following chapters prophesied. Massaro apparently sees modularity as the principal objective of his theoretical attack. Note that Massaro is not loath to tilt at sacred cows. Categorical perception is another prominent target of his lance. But modularity is the target he chose to set his sights upon in opening and in closing his book.

This is all very curious, because nothing in Massaro's work is crucially damaging to modularity; rather, his results offer the concept of modularity significant support. Consider the case for modularity as put by Fodor (1983; see also multiple book review in *BBS* 8(1) 1985). Fodor proposed nine properties of input systems by which their modularity was defined (Massaro lists these properties on pp. 2–3). In fact, these nine properties have not all been debated to the same extent. Most of the action in psycholinguistic argument has centred on principle five, the informational encapsulation of modular systems; this was perhaps only to be expected from a field which has disputed autonomy versus interaction above all else for the better part of two decades. To a lesser extent the principle of mandatoriness of operation has also proved contentious.

Now how does Massaro's research address the issues with which psycholinguistics is obsessed? In one major way, at least, Massaro allies himself firmly with the autonomous camp. A crucial aspect of Massaro's Fuzzy Logical Model of Perception is the independence of separate sources of information in perception and recognition. Independence of information derived from separate sources is precisely what encapsulation requires; if the evaluation of one source of information is affected by what is coming in from another source of information, then the capsule has been penetrated and the principle is violated. Fodor argues that such dependent processing does not occur in modular input systems. Massaro is equally adamant that it does not occur in any of the domains of his research. Insofar as Massaro has made this case, he has made about the most important case for modularity.

That Massaro is perfectly aware of this is clear from his brief allusions to connectionist models. Interactive-activation models allow activation from higher levels of processing to feed back to lower levels of processing, thereby altering the operation of the lower-level processes and hence obliterating their independence. As Massaro points out, such models would make predictions about the listening-plus-lipreading situation which are not borne out by his data. For instance, it would be reasonable to expect an unambiguous lipreading cue (such as closed lips for a bilabial) to affect the relative activation of auditory features for various phonemes, rendering the features of bilabial phonemes more active and of nonbilabial phonemes less active. Neither the facilitative nor the inhibitory prediction finds support in Massaro's results; he argues that this outcome "contradicts a fundamental assumption of TRACE" (p. 171) and "provides important constraints on the assumptions that are viable for connectionist models" (p. 281). Thus he shows no reluctance to draw out the implications of his findings with respect to autonomy versus interaction of processing levels.

The autonomy-interaction debate has become far more sophisticated in recent years. It is no longer assumed that the mere presence of context effects poses problems for an autonomous view; context effects can occur when multiple outputs of lower-level processing – each even equiprobable given a degraded input – are evaluated at a higher level (see e.g. Norris 1986; Marslen-Wilson 1987). Again, this is a position of which Massaro approves: "the bottom-up information and the top-down information . . . function as independent sources . . . [which] illustrates once again that a positive effect of context does not necessarily imply that low-level processes . . . are dependent on higher-level processes" (p. 269). Clearly, he is an autonomist through and through.

Why, then, does Massaro consider himself to be arguing against modularity at all? The answer appears to be that he takes issue with another characteristic ascribed to modular systems, namely, domain-specificity. This is the principle that modular systems operate only upon their particular class of inputs, and are closely tuned to the specific properties of those inputs.

Domain-specificity is not one of the central pillars of the modularity temple. As Fodor points out, in one sense it is entirely trivial to say that the mechanisms which process language are domain-specific because language is what they process. But there are certain nontrivial aspects of domain-specificity. One is that processing by a modular system should be initiated only by input of its particular kind. Another is that the operations of the modular system should be sensitive to the particular characteristics of what is being processed.

What domain-specificity does *not* entail is that the operations of a modular system need be unlike the operations of other systems in every respect. Indeed, it would be astonishing if this were so. General constraints on processing by systems realised in neurological material will of course apply. Although a modular system will not draw upon resources shared with other systems, there is no reason why it should not carry out operations on its own type of input which are similar in kind to the operations carried out by other systems on their particular inputs. This is where Massaro appears to have misled himself: "by definition, a specialized process should not follow general principles" (p. 4). Massaro seems to have got hold of the wrong definition.

To clarify this distinction, imagine, a coin-sorting device which consigns coins of different denominations to different processing operations on the basis of weight. Since each operation is initiated by the weight characteristics specific to a particular denomination of coin, the restrictive initiation principle of domain-specificity is satisfied. We can also build in lots of denomination-specific characteristics of the processing that goes on – pennies get painted black and shot out of a barrel, dimes get smashed to smithereens and mixed with sulphuric acid, quarters get baked in a cake, and so on. And we can have all these specific processes go on in entirely separate and noncommunicating compartments. But absolutely none of this alters the fact that the physical laws which govern the assessment of each coin's weight, and which further govern the way the coins tumble into the compartments, are the same for all coins; nor does the applicability of the laws of physics compromise the domain-specificity of the coin-processing. In fact, domain-specificity would not be compromised even if there were considerable similarity between the operations – if some part of each denomination's processing included a painting operation, for example.

Just the same sort of distinction is true in perceptual processing. It is never going to be an argument against modularity to point to commonalities between different input systems either in general constraints or in operating characteristics. Yet that seems to by why Massaro thinks he is against modularity: "principles . . . of bimodal speech perception prove to be relevant to a variety of behaviors" (p. 4). But the principles in question are extremely general – feature evaluation, integration, and pattern classification; "any model of recognition would have to assume stages functionally equivalent to these" (p. 152). Quite so. Thus Massaro's chief case against modularity is null. In his final paragraph he makes, for the first time, a slightly different claim, namely, that processing of a general cognitive kind has been shown to engage supposedly encapsulated modules; but this only widens the definition of general cognitive processing, quite apart from the fact that the violation of encapsulation requires engaging general cognitive resources by an encapsulated module rather than vice versa. He follows this, however, by reiterating the principal point of his research: "the same fundamental processes contribute to an impressive variety of perceptual and cognitive actions" (p. 281). Since there is nothing in this general claim which is incompatible with modularity, it is a pleasure to welcome Massaro to the modular fold.

Models in the mind, modules on the lips

Beatrice de Gelder and Jean Vroomen
Departments of Psychology and Philosophy, Tilburg University, 5000 LE Tilburg, The Netherlands
Electronic mail: *beadegel@htikub5.bitnet*

As a consequence of studies of both sign language and lipreading it is now widely agreed that a comprehensive theory of speech perception (and production) must capture the common properties of the different modalities in which speech can be perceived. Generally speaking, evidence from sign language and from lipreading has been welcomed by modularist speech theorists. In contrast, Massaro's contribution builds on more than a decade of his research on speech perception in the auditory and visual modality and arrives at a refutation of modularism. Modularists welcome lipreading evidence because they regard it as support for the view that speech perception draws upon a representational repertoire which is both language specific and amodal or abstract. Massaro believes that the bimodality of speech perception offers a spectacular insight into the general way the organism processes information, for example, by integrating multiple sources of information. The existence of this general process is the core of his refutation of modularism.

We believe that the issue of information integration is orthogonal to the matter of modularity of speech perception. What matters is the domain specificity of the representations over which integration processes operate. Next, we consider the relevance of language-specific deficits in the visual and auditory modality observed in developmental dyslexia.

1. Modularity is about representations. Massaro claims that information integration is the basic principle of information processing in all domains of cognitive functioning. What aspects of the claims of modularism (as epitomized in Fodor 1983) does this refute? The core of modularism is the view that all cognitive processes are inferential but that some of them are subserved by a special purpose computational system having a restricted database. So both modularism and Massaro agree that information processing is hypothesis testing and decision making. Going beyond this, modularism proposes a cognitive geography of decisions. Some decision processes are local because they are based on limited knowledge (e.g., they are domain specific and encapsulated). Massaro claims that a general model of decision making is enough – that one does not need to go through the trouble of designing a (modular) model of the internal knowledge environment of the organism to understand how decisions on inputs are reached. The crux of the disagreement is categorical versus continuous perception. In Massaro's view, categorical identification results from an integrative decision about a previous continuous evaluation process taking place separately in various dimensions. Modularism claims that evaluation itself is a decision process located in the language module and resulting in phonological representations. (This view leaves room for postmodular decisions about modular phonological representations).

How does evidence for modality specific coding square with either view? The answer is complicated by the fact that the notion of dimension and that of modality are orthogonal. Massaro uses the multimodality of speech input as a privileged example to show that information processing consists of integrating the input coming from various dimensions. This way he indicates that he treats modalities as dimensions and implies that phonological category decisions require integrating bimodal information. Obviously, in normal circumstances both the visual and the auditory information are present. Still, when auditory information is not deteriorated, it is perfectly possible to identify speech in the absence of information from the visual modality. But perception in the auditory modality still results from the integration of multiple features. Perception of sign language in congenitally deaf subjects offers a similar example of multiple dimensions of a stimulus within one modality. The real issue is categorical perception within one and the same modality.

2. Developmental dyslexia. Detailed studies of acquired dyslexia support the view that we are dealing with a specific impairment in the domain of written language skills (Shallice 1988) which leaves intact decision processes in other domains of linguistic and nonlinguistic information processing. Such disorders support modularism because as noted above, modularism is a thesis about the nature of representational resources required in decision processes. There is a similar consensus about the language specificity of the deficits observed in developmental reading disorders. We have explored the idea that reading acquisition disorders might be related to spoken language deficits. By looking also at speech perception in the visual modality one might get a fuller picture of the speech processing abilities of young dyslexics. Alternatively, this approach could reveal the existence of bimodal integration difficulties.

One could use a stimulus tape made by Massaro (Massaro & Cohen 1983a) to examine this issue. Our results (de Gelder & Vroomen 1988) show that young dyslexics have less robust auditory speech categories when their performance is compared with that of control groups (both reading age and chronological age). This confirms the findings of Werker and Tees (1986). Our subjects also lag in processing speech information in the visual modality. These modality-based deficits show a significant correlation which could suggest a deficit in representational resources shared by visual and auditory information processing. Does the McGurk illusion also suggest a common resource explanation? does it show late categorization within each modality? or does it illustrate late integration of both modalities, as Massaro believes it does?

When presented with auditory /ba/ and a visual /ga/, subject's report hearing a phoneme that is a fusion of both, a /da/. On presentation of an auditory /da/ and a visual /ba/, the subject reports a blend, /bda/. Massaro's model (which does not give the full details of the representations activated at the various stages in processing) explains the former case but might have difficulty with the latter. He treats the perception of /bda/ as the identification of a single integrated speech event which is no exception to the general integration formula. The /b/ in /bda/ is indeed an integrated percept because subjects never report /mda/ or /pda/. There is, however, no visual influence in the /da/ part of /bda/. In other words, there is conflict plus integration.

The extent to which there can be conflict between modalities would appear to be closely related to the robustness of the representations in each modality. Indirect evidence about robustness of modality-specific coding has been obtained by using a serial recall paradigm (Massaro 1987, p. 50). Adapting the paradigm of the auditory recency effect in serial recall, Campbell and Dodd (1980) found that lip-read lists show recency just as auditory presented lists do. This suggests that there is a common language source for both modalities.

In a recent study using a large population of subjects we find that a visual suffix has no influence on the recency effect of stimuli presented auditorily. More surprisingly, an auditory suffix with no visual articulation does not affect recency in a visually presented list (de Gelder & Vroomen 1989). This suggests that besides shared resources for both modalities (modularism) or late integration of visual and auditory information (Massaro), there is still room for modality specific codes. One would need to know more about the influence of linguistic experience on speech perception to have strong views in these matters (Bertelson & de Gelder 1989).

ACKNOWLEDGMENTS
Research reported was supported by research grant (VF-KA) to the first author. We thank Jose Morais and Paul Bertelson for comments on the manuscript.

Categorical/continuous perception: A phenomenon pressed into different models

Günter Ehret
Abteilung Vergleichende Neurobiologie, Universität Ulm, D-7900 Ulm, Federal Republic of Germany

The continuous thread through Massaro's book, as it seems to me, is the author tilting at windmills that are permanently droning: Speech is perceived categorically and by specialized phonetic modules. Unlike Don Quixote, who trusted idealism, Massaro has compiled a wealth of facts and arguments that might eventually take the wind out of these sails.

In my commentary, I will concentrate on aspects of categorical/continuous perception, which is the topic of the central two chapters and repeatedly occurs at other places throughout Massaro's book. In the classical and, as it is obvious now, limited view (Liberman et al. 1957; Studdert-Kennedy et al. 1970), speech phonemes are perceived categorically if stimuli from an acoustic continuum (e.g. voice onset time) are labeled (or identified) as belonging to different classes and discrimination of within-class stimuli is poor, whereas between-class stimuli are well discriminated. Thus labeling functions are expected to predict the outcome of discrimination tests and vice versa, with both kinds of experiments producing boundaries between categories at the same regions of the acoustic continuum. On the basis of his own work and studies of others, Massaro replaces categorical by continuous perception followed by a categorical decision stage. Experimental evidence and arguments such as within-category discrimination, integration of multiple auditory and visual properties *before* phonemes are identified, and model-fitting calculations that favour continuous evaluation of parameters are detailed in support of a model including continuous sensory analyses of stimulus properties in appropriate channels, integration of the valued information from the channels, and a classification of the integrated information by matching it against prototypes in memory (Fig. 4, Chapter 1; Fig. 1 of the précis). The decisive difference between Massaro's model of continuous speech perception plus categorical recognition and categorical perception concerns the level in the hierarchy of proposed elements of building blocks where the categorization of the speech percept (i.e. the identification of phonemes) should take place. The term categorical "perception" implies that categorization is expected to occur within the *sensory* network of the brain, whereas Massaro places categorization on a supposedly higher level of "recognition" (whatever that means in terms of neuronal processes).

An elegant way of bringing both views in harmony with each other is to shift labels, that is, to extend the term "perception" to include Massaro's recognition or decision level (the reason for doing this should become clear below). Thus speech would be perceived categorically as long as both labeling and discrimination behaviour indicate a consistent and statistically significant but not necessarily pronounced boundary in the physical continuum of a given speech (or sound) parameter. This operational definition of categorical perception stands without any implications about levels, mechanisms, or theories of signal processing and recognition. Categorical/continuous perception understood in this way would be just a *descriptive term* about the outcome of the overall operation made by the brain which could comprise any kind of handling of information (continuous and/or categorical) in any channel at any level, including that of recognition. A good example for multi-channel continuous feature evaluation, feature integration, and pattern recognition is Ewert's (1987) model of visual pattern recognition in toads. This model has no defined decision level but nevertheless can produce categorical and continuous perception of a prey stimulus depending on the visual parameter varied in perception tests. Experiments suggest that stimulus recognition is accomplished by the coordinated activity in parallel, hierarchical, and loop-organized neuronal networks in which different classes of neurons contribute to the processing of different aspects of the stimulus, of learned associations, and of arousal. I mention Ewert's work because it demonstrates nicely that sensory perception, pattern recognition, and response generation is represented by a continuum of operations in the brain that can be divided into sections only arbitrarily. Thus, the problem of the categorical versus continuous perception of speech representing patterns may be a superficial one created by attaching different labels to different levels of the continuous process of speech evaluation which is able to produce both categorical and continuous results depending on test conditions.

Several studies have demonstrated continua between categorical and continuous speech evaluation. For example, categorical perception of voice onset time can be changed to continuous perception by specific training and stimulus presentation techniques (Carney 1977; Pisoni & Lazarus 1974; Samuel 1977). In addition, third formant transitions are categorized as /ra/ and /la/ after learning to "hear" the difference by English speakers; Japanese speakers lack this experience and thus perceive the /ra/-/la/ continuum continuously (Miyawaki et al. 1975). Again, brain mechanisms in speech perception produce a variable output in that they discriminate what is biologically (or culturally) important and dismiss potentially available distinctions if they have no semantic value. A similar phenomenon is reported in a study of tone discrimination (Spiegel & Watson 1981). Training and knowledge of an important acoustic feature of a stimulus improves the discrimination ability of that feature. In Massaro's view, these examples demonstrating a continuum between categorical and continuous stimulus evaluation are clear evidence for continuous information processing which is followed by a categorical decision process in the case of category recognition.

As I have argued before, we cannot unambiguously attribute categorization to the level of sensory processing, or to that of feature integration, or to recognition, or to processes at all levels together, until the neuronal (physiological) mechanisms underlying the categorization process are known. Thus, there seems to be little reason for changing names merely because categorical perception in the classical sense can be modeled by a theory that divides the entire process into continuous perception and categorical recognition. A theory of continuous speech perception as proposed by Massaro may easily lose sight of categorical phenomena. I am not an advocate of categorical speech or sound perception. However, lacking physiological evidence for making adequate distinctions between levels and mechanisms of continuous and categorical processing in speech, we should stay with the phenomena, which are interesting enough because they are continuous and categorical.

I find the classical concept of categorical perception a useful working hypothesis for designing experiments on stimulus perception, especially for communication sounds of animals, and a helpful framework for critically interpreting the results. There are even some cases where there exists experimental evidence for category formation within the sensory pathway leading to the categorical perception of communicative sounds, as we have demonstrated for ultrasound perception in house mice (Ehret 1987; Ehret & Haack 1981; 1982). The physiological basis for category formation in the frequency domain is the critical band phenomenon (e.g. Scharf 1970), an auditory filter mechanism established at or below the inferior colliculus of the midbrain (Ehret 1988; Ehret & Merzenich 1985; 1988). The category boundary in the temporal domain close to 25 ms may reflect absolute limits of temporal resolution in the auditory pathway because it coincides with the shortest boundary of voice onset time perception in man, monkey, and chinchilla and with thresholds of perception of temporal order in humans (compare Ehret 1987). These cases show that categorical perception in Massaro's sense is likely to exist and should not be abandoned as a possibility for human speech unless falsified by physiological evidence. Finally, one has to consider the efferent auditory system that can influence tone discrimination even at the cochlear level and may be expected to influence speech processing as well (e.g. Winslow & Sachs 1988). Thus top-down conditioning of the ascending auditory pathway is present and one might hypothesize that whenever categorical *information* in terms of speech phonemes is expected in a communicative context, peripheral processing is conditioned to accentuate these natural categories in order to preserve the semantic contents of the message optimally. If this is true, the categorization of speech information could gradually emerge in the auditory system and would hence be an inherent property of the perceptual process.

A general algorithm for pattern recognition?

Gerd Gigerenzer

Department of Psychology, Universität Konstanz, 7750 Konstanz, West Germany

Electronic mail: *sygiger@dknkurz1.bitnet*

How does the brain make sense of the world? "In the same way that scientists do, and with the same tools," answer an increasing number of cognitive psychologists. The eighteenth-century mathematicians Laplace and Condorcet used their "probability of causes" to model the way scientists reason (Daston 1988); Dominic Massaro now proposes the identical formula as an algorithm for pattern recognition in general and speech perception in particular. To show the generality of the algorithm (the "fuzzy logical model of perception [FLMP]") is the ambitious goal of Massaro's book: "well-learned patterns are recognized in accordance with a general algorithm, regardless of the modality or particular nature of the patterns" (p. 16). The project of reducing pattern recognition to any algorithm, much less a single one, may strike many as overly ambitious. For the purposes of this commentary, however, I will accept Massaro's goal as in principle attainable and will try to invite him to clarify one of his major arguments by revealing its conceptual difficulties.

This argument is central to the issue of generality, and runs like this: (i) The FLMP is not only general but also optimal since it is mathematically equivalent to Bayes' theorem; (ii) here Bayes' theorem is implemented as it was by Laplace, that is, with the assumption of uniform prior probabilities and independence of events (features), but (iii) this equivalence poses a dilemma for the FLMP, since previous research, in particular on intuitive probabilistic reasoning, has rejected Bayesian reasoning as a general mental algorithm. In analyzing this argument, I shall proceed from the general to the specific.

1. Is the general algorithm a Bayesian one? The FLMP assumes that pattern recognition occurs in three sequential stages. I shall consider the simplest case, with only two features and two prototypes. In the feature evaluation stage, the match $t(E_1/H_1)$ between a feature E_1 and a prototype H_1 is calculated; in the feature integration stage the overall match $t(E/H_1) = t(E_1/H_1)t(E_2/H_1)$ between the two features and the prototype H_1 is calculated; and in the pattern classification stage the probability that the pattern will be identified as H_1 is given by $p(H_1/E) = t(E/H_1)/(t(E/H_1) + t(E/H_2))$. Massaro (pp. 196–98) says that this algorithm is mathematically equivalent to Bayes' theorem, assuming uniform prior probabilities and independent events, and replaces the above t (i.e. truth) values by p (i.e. probability) values. However, Massaro repeatedly (e.g. pp. 21, 166, 202) asserts that $t(E_1/H_1) + t(E_1/H_2) = 1$, which is not true for the corresponding probabilities $p(E_1/H_1)$ and $p(E_1/H_2)$ in Bayes' theorem. According to standard probability theory, which mathematically implies Bayes' theorem, the sum of these probabilities can be either less or more than 1. Thus, I doubt that the proposed general algorithm is in fact mathematically equivalent to Bayesian probabilities, and I therefore also doubt the claim that "either of these two models is adequate to account for the results" (p. 198).

2. A general pattern recognition algorithm with uniform priors? To keep the next points separate from the first, let me assume that I have overlooked something and that Massaro is right in pointing to the equivalence of the FLMP and Bayes' theorem. Laplace's urn analogy and Bayes' billiard table suggested uniform prior probabilities on the grounds that our ignorance gives us no reason to expect one urn or one area on the table to be a priori more likely than any other. But should we assume that a general pattern recognition algorithm also works on the principle of ignorance and uses uniform priors? An algorithm with uniform priors may be sufficient for the experimental designs reported in the book, in which the prototypes to be identified, such as /ba/ and /da/, are equally likely in the laboratory. But in everyday speech, just as in many other domains, different patterns have different prior probabilities depending on context. Where expectation plays a role, nonuniform priors seem to be indispensable for improving the perceptual "bet" in situations with uncertain information.[1]

3. Does intuitive probabilistic reasoning challenge the generality of the algorithm? Massaro argues that the mathematical equivalence of the FLMP and Bayes' theorem "poses a new dilemma" for the generality claim, since previous research has rejected Bayes' theorem as a model of intuitive probabilistic reasoning. His major defense is that most previous researchers used "objective" rather than "subjective" probabilities to calculate the so-called normative Bayesian outcome. Massaro's reply is correct: There are many ways to be a Bayesian, and such experiments do not rule out that intuitive reasoning is Bayesian by using subjective probabilities. But there are designs, such as in Kahneman and Tversky's (1973) Engineer-Lawyer study, which allow for subjective likelihoods and which still lead the authors to conclude that reasoning is not Bayesian, since base rates are ignored due to a representativeness heuristic. Massaro has to deal with these kinds of experiments. Moreover, even in the Engineer-Lawyer Problem, the neglect of base rates can easily be eliminated if one crucial structural assumption (random sampling of description) is made vivid to subjects, although the feature values remain constant (Gigerenzer et al. 1988). Such systematic changes in reasoning indicate that neither representativeness (i.e. uniform-prior Bayesianism, see below) nor Bayes' theorem is a general algorithm of the mind. Massaro's "dilemma," in my opinion, is not that intuitive reasoning ignores base rates (as does the FLMP in the book under review), but rather that intuition seems to have a whole toolbox of algorithms available.

Two things puzzle me concerning the relation between reasoning and the FLMP. First, as noted above, Massaro says that the pattern recognition algorithm is a Laplacean uniform-prior variant of Bayes' theorem. Why then does he believe it is a "dilemma" that intuitive reasoning seems to violate Bayesian reasoning by neglecting base rates and using uniform priors? Base rate neglect is exactly what his algorithm predicts – just as Baconian probability would (Cohen 1986). [See also Cohen: "Can Human Irrationality Be Experimentally Demonstrated?" *BBS* 4(3) 1981 and Kyburg: "Rational Belief" *BBS* 6(2) 1983.]

Second, what is the relationship between Kahneman and Tversky's representativeness heuristic and the FLMP as models of pattern recognition? Massaro says they are fundamentally different (p. 273). However, since in this context "representativeness" can be shown to mean Bayesian reasoning with uniform priors (Gigerenzer & Murray 1987, Chapter 5), I understand both the FLMP and the heuristic to refer to the same strategy – although the FLMP is a model and representativeness is just a word.

4. The "conjunction fallacy" and the FLMP. What has been called a "conjunction fallacy" is a judgment of the following kind: A fictitious person named Linda is more likely to be a feminist and a bank teller than just a bank teller. I agree with Massaro that this cannot be called a "fallacy" unless one willfully ignores the fact that the term "likely" has several meanings in everyday language. But I part ways with him when he extends the generality of the FLMP to conjunction judgments and claims that "within Bayes theorem or the FLMP, we can predict that Linda will be rated as being more likely [to be] a feminist and bank teller than just a bank teller" (p. 275). Massaro's argument is that the subjects behave as if they were carrying out pattern recognition, evaluating the features against alternative prototypes. In fact, the FLMP predicts that the probability that a pattern is recognized as prototype H_1 can be larger for two features E_1 and E_2 than for just E_1 alone. In formal terms, this means that $p(H_1/E_1\&E_2) > p(H_1/E_1)$. But this is not the "conjunction fallacy," which is to judge $p(H_1\&H_2/E_1) > p(H_1/E_1)$. Only the latter contradicts standard probability theory, and thus cannot be derived from either Bayes' theorem, which is a consequence of standard probability theory, or the FLMP, insofar as it is claimed to be mathematically identical to the former.[2]

5. Why only one algorithm? Scientific reasoning is a many-splendored thing, encompassing many and diverse forms of inference. We might want to carry the analogy between scientific reasoning and cognition far enough to extend this multiplicity to the mind. Thus, instead of one general, all-purpose algorithm, we would expect to find several or even many algorithms. There is an additional Darwinian argument against the view that evaluation has given us only a single algorithm for all cases of pattern recognition. Since each algorithm assumes a specific structure (e.g. in Massaro's model, independence of features and a mutually exclusive and exhaustive set of prototypes), it is well equipped for specific tasks that have these structures, but less so for other tasks. In order to survive in a changing environment, the mind would be better off if outfitted with a whole toolbox of algorithms, and with an evaluation program that first checks the structure of the environment before it selects a particular algorithm to apply to that environment.

I have concentrated in this commentary on conceptual issues that need further clarification. These have to do with the claims for a general algorithm for pattern recognition, but do not touch the valuable experimental work that the book presents, which I have not here addressed.

NOTES

1. However, nonuniform priors may not be required to improve the fit of the FLMP to experimental data, since it is already excellent. My point here is a conceptual one. It is in fact hard to judge to what degree the FLMP is supported by the empirical data, because of the large number of free parameters (sometimes close to 50% of the data points) that can be fitted to the data. The true degree of support could be revealed by a step-wise cross-validation procedure: Use the best-fitting feature values for given prototypes and a given subject in a new experiment with the same or an enlarged set of features. If features are evaluated independently from other features present, as the FLMP proposes, then the feature values should be stable, and the fit in the second experiment would provide a stronger test for the validity of the FLMP.

2. In his own experiments, Massaro (submitted) showed that the FLMP gives an excellent fit to judgments in Linda-type tasks. However, judgments of the type p(Linda/vocation&avocation) were compared with p(Linda/vocation) and p(Linda/avocation), which, as mentioned above, cannot violate the conjunction rule.

From speech perception to person perception? Not quite yet

Sam Glucksberg
Department of Psychology, Princeton University, Princeton, NJ 08544
Electronic mail: *samg@pucc.bitnet*

Massaro's basic argument is that people integrate information from multiple sources in order to decide among alternatives, whether those alternatives are two syllables, or two types of people. Massaro's strong claim is that the specific domain of alternatives is irrelevant: The processes that are used to perceive speech sounds are also used to perceive people, or to interpret sentences, or to learn and use natural categories. The specific processing model is not at issue here. Instead, I want to examine two claims. The first claim is that the theoretical conclusions concerning bimodal speech perception generalize to quite disparate domains, including such domains as person perception and sentence interpretation. The second claim is that the findings and the overall theoretical approach are incompatible with modularity on the one hand, but compatible with connectionism on the other hand.

With respect to the first claim, at some level it must be true. People must integrate information from various sources irrespective of stimulus domain – providing, of course, that more than one information source is available. Because such variables as the intensity and frequency of an auditory stimulus are considered to be two information sources, virtually all percepts and judgments would perforce depend upon multiple information sources. The generality claim, however, is stronger than this. Person perception and sentence interpretation, among others, "might follow the algorithm of the FLMP [fuzzy logical model of perception] that has proven appropriate for speech perception" (Massaro 1987, p. 245). This is an interesting hypothesis, but I seriously question the choice of experimental paradigms for evaluating it.

For the domain of person perception, Massaro relies on an experimental paradigm developed by Anderson (1973) and his colleagues (e.g., Birnbaum 1974). People are given two or more adjectives in every possible combination, and for each combination rate some person-characteristic. Ratings of social desirability, likableness, and introversion are examined and found to be consistent with model predictions. Massaro concludes that person perception is accomplished the same way that speech perception is: by integrating multiple, continuous, and independent stimulus attributes. The data and model fits are consistent with Massaro's claim. But do these data reflect person perception processes, or is it more likely that they reflect the demands of the experimental task?

Massaro reports a typical experiment on "person perception." College students rated how introverted or extraverted an unspecified, hypothetical person was on the basis of either a single descriptor, such as *entertaining*, or a pair of descriptors, such as *entertaining and withdrawn*. Each student rated each of eight such adjectives one at a time, and then rated all 16 pairs, one pair at a time. The results were in accord with Massaro's FLMP. Each adjective contributed independently to the judgments, leading Massaro to conclude that the contribution of each adjective was independent of the other adjective, Asch's classic demonstration (1946) of meaning interaction in person impression notwithstanding.

It may be misleading to interpret such studies as reflecting person perception processes, however. To begin with, there are no persons to be perceived, just a hypothetical someone who is, say, entertaining. Second, the repeated ratings of a pool of eight adjectives singly and in pairs could easily produce experimenter and task demands that yield simple additive or multiplicative effects. It is also absurdly easy to provide examples of adjective meaning interactions: The meaning of *wise* in the contexts of *wise man* and *wise guy* is clearly context sensitive.

Just as in the person impression analysis, where no persons were involved as objects of perception, no sentences were involved in Massaro's analysis of sentence interpretation. Data from studies reported by Bates et al. (1982) and by MacWhinney et al. (1984) were reanalyzed and shown to fit the FLMP. The original experimental task was to judge which of two nouns performed an action, given three-word phrases consisting of two nouns and a verb. Word order, animacy of nouns, prosodic stress, and number agreement between nouns and verbs each contributed to people's judgments. Each of these variables is considered to be a cue, and the way these cues were integrated and used was consistent with Massaro's general model of information processing. Massaro takes this outcome to pose "a challenge to current theories of language processing" (p. 272); "If the present generalizations of the current research are reasonable, they strike a blow to the concept of modularity" (p. 281).

In my view, Massaro misreads his own position. Contrary to his assertion, the FLMP is logically inconsistent with the basic assumptions of connectionist models, yet perfectly consistent with any reasonable notion of modularity. The FLMP posits sequential information processing, "no top-down influences of a higher-level . . . . on a lower level" (p. 281), and local symbolic rather than distributed associative representations. Each of these three model characteristics is explicitly rejected in connectionist theory. These same three model characteristics are perfectly in keeping with a modular approach, unless one restricts a modular system to single rather than multiple information sources for any given module. There is no principled reason, however, to impose such a restriction. Even a speech module, as characterized by such theorists as Liberman and Mattingly (1989), could in principle integrate information from both auditory and visual sources in order to identify articulatory gestures.

The contribution of Massaro's work then is not (yet) a general theory of perception and categorization, but rather an impressive demonstration of how visual and auditory information sources may be integrated in speech perception. The premature claims of wide generality and of the demise of modularity should not detract from this important accomplishment.

Speech perception from a Hungarian perspective

Mária Gósy

Phonetics Department, Linguistics Institute of the Hungarian Academy of Sciences, Budapest, Szentháromság u. 2. H-1014, Hungary

Massaro's excellent book is unique in this scientific field. I appreciate *BBS*'s effort to obtain comments on parts of the book from various experts, but first I would also like to make two general remarks concerning the book as a whole: (i) the author boldly overthrows and contradicts some seemingly well-founded traditions (e.g. categorical and continuous perception) and is able to support his hypothesis with his own results. (ii) On the other hand, it is always disturbing to read speech perception literature in which the authors – often using only English-speaking subjects and English material – make judgments about the speech perception process *in general*. There is no doubt that this speech perception is successful with actual language and speech peculiarities (there is a huge number of papers confirming the filtration function of the mother tongue), but no one has tried to define or distinguish what is "general" and what is "specific" in the process. This problem seems to be particularly important when two sources of information – audible and visual – are concerned. More specifically, my question concerns whether conclusions made on the basis of experimental data from one language might be relevant for the perception of other languages as well and if they are, then to what extent? (Cf. English /ba, va, ða, da/ on pages 54 and 234 and, for example, Hungarian, where the sounds /a/ and /ð/ do not occur, /b/ has a very different VOT, /v/ has a different frequency pattern, and all have a different articulation – including lip movement – compared to the appropriate English consonants.) This problem appears also in Massaro's book with the description of Mills and Thiem's study (1980) concerning German observers on pages 48–49 – but no further explanation is offered.

Chapter 2 discusses important studies and results concerning two sources of information; then the author analyzes his own experimental findings. It was surprising, however, to read that word recognition (pp. 34–36) is included in the speech "perception process." Word recognition should contain the "upper" levels of the bottom-up analysis (meaning, syntax, etc.) and all consequences of word recognition operations. This might confuse the reader, and the analysis of "context" does not seem to be acceptable evidence for its inclusion.

On the basis of my experiments on bimodal speech perception (carried out with Hungarian listeners using nonsense syllables with both conflicting and nonconflicting auditory and visual information) I obtained both similar and contradictory results. For example, my data on the proportion of correct identification for each consonant for the auditory, bimodal, and "conflicting" conditions are (in %): /p/ .37 – .88 – .20; /v/ .58 – .857 – .40; /f/ .15 – .97 – .10; or /ʒ/ .44 – .71 – .285 (cf. Table 3, p. 43, for English "equivalents"). Interestingly, our subjects only partly replicated McGurk and MacDonald's (1976) findings (pp. 46–47); when labial /ba/ was dubbed onto the articulation /ga/ they also reported hearing /da/, but this was not the case for voiceless stops! As to the identification of vowels: When labial /u/ was dubbed onto the articulation /i/, subjects reported hearing /y/, but not in the case of midvowels. These data also seem to support the claim that the visual perception of some speech sounds is integrated with audible speech process much more deeply. Further experiments should be carried out to explain the "different" behavior of similar sounds (like voiced and voiceless stops). It is likely – as suggested by Massaro as well – that the use of the two sources of information is more complex in normal understanding conditions (cf. context-dependencies, number of syllables, the paradox of "time", etc.). Although my own subjects were carefully told to identify nonsense syllables, they recognized real Hungarian words. The more difficult the identification task, the bigger the proportion of real words recognized.

It is very exciting to hear how visual perception is integrated with auditory perception during language acquisition. Massaro writes that "children might not have the same access to the sources as do adults. For example, children are short and adults are tall and this difference in height might be expected to limit visible relative to audible speech" (p. 208). On the other hand, on page 214 Massaro's conclusion – based on his experimental data – is that "children behave similarly to adults" concerning the integration of audible and visual information. I am not convinced, however, that the identification of syllables varying acoustically step by step along a continuum provides information about the actual existence of audible and visible information for children in normal communication, and if it doesn't, then how can the perceptual results be used for comprehension? I have found normal hearing 8-year-old children with excellent lipreading ability and fairly low performance in auditory speech perception, while other children showed very good auditory speech perception and poor lipreading ability. (In the experiments real Hungarian words and sentences were used.) In addition, in Massaro's book the differentiation of lip- and "speech-reading" can implicitly be found but the two expressions are generally used as synonyms. I think a sharp differentiation is necessary: "Lipreading" is the ability to integrate the visual experience of speakers' lip movements as a "simple" type of visual source of information when listening to speech, whereas "speech-reading" concerns, in addition to lip movements, several other factors (paralinguistic, extralinguistic gestures) that can be seen during speaking in the form of other, more complex types of visual information. For language-acquisition both types of visual sources should be assigned crucial importance particularly in the development of speech perception (cf. pp. 239–40).

I fully agree with Massaro's conclusion that his results offer a new challenge for theories of language acquisition (especially for the less-examined age groups from 3 to 10). His findings open new vistas for practical applications as well. For example, subtle operations of the speech perception process can be analyzed to detect dyslexic children or provide evidence for or against using a hearing aid in childhood in uncertain cases.

A multiple source, or, is a striped apple more striped than a striped orange?

Dimitri Kanevsky

Speech Recognition Group, Department of Computer Sciences, IBM Watson Research Center, Yorktown Heights, NY 10598

Electronic mail: *dimitri@ibm.com*

This comment is divided into two sections. In the first we describe some results from our experiments on the use of supplementary tactile information to enhance lipreading performance. This work offers some support for one of Massaro's recurrent themes: that the combined use of several different sensory sources can influence perception to a greater extent than one would expect on the basis of the effects of any one of them alone. This dominating influence of the multiple source over its components is a critical factor in the apparent advantage of continuous models over categorical ones in Massaro's book. In the second section we analyze a few technical problems with the mathematics used to derive Massaro's conclusions.

1. Experimental findings. In the past 15 years I have developed and studied various forms of tactile stimulating devices designed to enhance lipreading performance in hearing impaired subjects (Kanevsky et al. 1982; 1987); these stimulators, which transform acoustic information into tactile frequency codes, have been extensively tested experimentally (Cholewiak & Sherrick 1986; Donner & Luttwak 1984; Lancaster summer project 1985; Luttwak 1987; Rodriguez Santos 1987; Rose 1986; Tortosa 1987). The following summary is based on numerous experimental tests of these tactile aids as well as on informal discussions with their users. The findings can be compared with Massaro's observations on the perceptual effects of joint auditory and visual information. It must be stressed, however, that the primary objective of these experiments was to assess the effectiveness of the devices and to find ways to teach users to recognize the tactile patterns and their correlation with speech. The various phenomena that were repeatedly observed informally in the course of these experiments still need to be studied directly for formal experimental confirmation.

Supplementary tactile information was found to help hearing-impaired subjects improve their lipreading performance significantly, sometimes yielding almost 100% recognition rates. The subjects were given the tactile aids for everyday use and this marked improvement in lipreading performance was attained gradually after a few months of continuous experience. In contrast, the same subjects displayed almost no recognition of spoken words when only tactile input was presented, without accompanying visual (or auditory) information, even after long use of the tactile devices as supplements to lipreading or after feedback training on using the devices alone.

In experiments on the perception of phoneme combinations it was also found that supplementary tactile information produced significant improvement in lipreading performance compared to very poor performance with tactile or visual input alone. The typical task in these experiments was to indicate which of several phonemes (say, /ba/, /da/, /ga/, etc.) had been uttered; a few training sessions, usually lasting 10–20 minutes, preceded testing. Formal studies along these lines are reported by Cholewiak and Sherrick (1986), Sherrick (1984), Rose and Haymond (1986), and Potts and Weisenberger (1987). These findings are evidence of a complex interaction between the two information sources, in general agreement with Massaro's conclusions, but some further findings of ours may help to clarify the nature of this complex bimodal interaction.

First, after sustained experience with supplementary tactile input over an extended period of time (a half year or more), the performance of hearing-impaired subjects on lipreading alone improved significantly compared to their prior levels of lipreading performance. This was not a formal experimental finding but an overall impression derived from informal reports volunteered by some of the users. They indicated that the circle of individuals with whom they could interact had greatly increased after prolonged use of the stimulator and that with some additional effort they could now even communicate better without tactile aids. They reported that this level of communication had been impossible for them before they had started to use the supplementary tactile aids. An objective indirect indication of this improvement in lipreading comes from the observation that recognition performance levels with the joint use of tactile aids compared to lipreading alone remained relatively constant in time (about 20–30%). Thus one can infer that the improvement in lipreading alone was the same as the improvement in lipreading with supplementary tactile aids (for details, see Cholewiak & Sherrick 1986).

This long-term outcome seems to conflict with Massaro's general finding that performance with each input modality alone is left unchanged by bimodal training. This discrepancy can probably be explained by the fact that our tactile experiments lasted much longer (on the order of a few months) than the experiments described in Massaro's book. Our finding also suggests that some of the hidden factors in multimodal interactions might only be revealed by a sufficiently long experiment. If the kinds of long-term changes we have observed also occur with the sense modalities studied by Massaro, then some revision of his conclusions (e.g., concerning the independence of different sources) may be necessary.

Another finding in our experiments with tactile stimuli was that subjects perceived much more information than they needed in order to identify and respond to the inputs they received: Hearing-impaired subjects using supplementary tactile devices were instructed to lip-read as illumination was slowly reduced. Their performance level remained almost unchanged, and much better than that of subjects placed in a comparably dark room at the very beginning of the experiment (after sufficient time for dark adaptation). (The reason we performed these experiments was that new users of the tactile aids tend to focus on the visual rather than the tactile information source. The gradual reduction in visual information was intended to divert more attention to the tactile information.)

This adaptation effect suggests that with multimodal presentation the characteristics of each of the input modalities may change across time, even within a comparatively short period (30–45 minutes). It would be interesting to determine whether there are analogues of short-term adaptation effects under the conditions studied in Massaro's book and what modifications they would dictate in the various proposed models, in which all parameters were assumed to be constant throughout the duration of the experiments. For example, one can jointly present to a subject a long string of identical synthesized visual and auditory stimuli – say, an unambiguously pronounced /da/ is seen while something lying between /da/ and /ba/ on the formant-transition continuum is heard. This procedure may well prove to prime or potentiate the /da/ response to the auditory stimulus subsequently presented in isolation; at any rate, short-term adaptation effects as well as long-term learning effects should be investigated further.

Further observations of ours suggest that different tactile codes can give rise to the same performance levels in the recognition of various types of nonspeech sounds (say, pure tones generated by frequency generators, or environmental information) as well as phoneme combinations, while giving rise to different performance levels in enhancing the recognition of words through lipreading (see also Potts & Weisenberger 1984). This suggests that studying human perception at a very elementary level (including syllable recognition) may not capture certain features that appear at higher levels such as word and sentence recognition. One can accordingly expect multimodal interactions to become even more complex for higher level cognitive tasks.

Finally, we note that the special discrimination task used in our experiments was much easier than the identification task: Subjects (both hearing and hearing-impaired) viewed the lips of a speaker soundlessly pronouncing a word and were then either asked to identify the pronounced word or were shown a printed word and asked to say whether it was the same or different. Discrimination performance was much better than identification performance. This capability is important for some experiments we are performing in which the real-time printed output of an IBM automatic speech recognizer is being investigated as a visual aid for the hearing impaired. The output of the speech recognizer is imperfect, but we have evidence that it improves the level of lipreading, probably by supplementing discrimination performance, thereby disambiguating and improving identification performance. This is another example of a potential application of multimodal findings, as recommended by Massaro. Our experiments also suggest that discrimination ability is not directly related to identification performance, a conclusion consistent with Massaro's findings.

2. The formal model. In this section we discuss how Massaro applies his findings on multimodal perception to perception in general (with respect to factors such as categorical vs. continuous information and independent vs. dependent evaluation of sources).

Given that it was Massaro who set the parameters of the formulas describing the models under comparison, as well as the criteria for comparing the predicted and experimental outcomes, it is perhaps less surprising that he obtained such overwhelming evidence for the superiority of continuous models over categorical ones, for the relative independence of components of multiple sources, and so forth. Hence the only way we can do a critical analysis of these models is to study their relation to the mathematical formulas chosen for their description.

Our first example concerns formula (1) on page 122 and formula (2) on page 123, which represent categorical and continuous models, respectively. The first formula is linear in its arguments, which are values representing the probability of /da/ decisions as a function of auditory and visual levels. The second formula is the ratio of linear functions to a quadratic function. Hence we are in fact comparing the performance of linear and nonlinear functions. Even without doing experiments one can predict that a linear function will yield a worse fit to data than a nonlinear one (with the same number of parameters). So the question is whether it is valid to compare a categorical model with a continuous one by attaching a linear equation to the first model and a nonlinear equation to the second.

Moreover, the fit with the linear model is actually not so bad; and if we modified the equation for the categorical model to make it nonlinear we would get a still better fit, perhaps no worse than the continuous model. The linear formula in (1) was derived from a quadratic function by choosing coefficients to eliminate the squared terms. It would not be difficult to modify the quadratic slightly so as to avoid reducing it to such a simple linear function.

The same observation about comparing formulas of essentially different algebraic complexity applies elsewhere in the book. Massaro's continuous models are represented mostly by rational functions, the categorical ones by polynomials. Rational functions can be represented in an appropriate neighborhood as an infinite series; so perhaps the true underlying function is more complex than a polynomial. Throughout the volume an attempt is made to ensure that categorical and continuous models have a comparable number of parameters. Perhaps the formula should also be equated for algebraic complexity.

Considering formula (1) not in terms of its algebraic formalism but how plausibly it represents a (categorical) model, one wonders why the proportion p with which the identification mechanism chooses /da/ or /ba/ remains constant, independent of auditory and visual stimulus variations in a confusable source. There also seems to be a problem with the fact that the formulas in question represent the likelihood of a /da/ response given the source conditions. Even with categorical assumptions one could suggest more plausible formulas.

It may very well be that given a fixed continuous model one can always refine a categorical model to obtain a better fit to the experimental data. And vice versa, given a fixed categorical model perhaps one can always improve a continuous model over it. Massaro argues against this kind of logic in considering the possibility (Chapter 4, Section 6, p. 110) that a categorical model could be improved if we postulated additional phonetic categories. His argument is that "following this logic, it would no longer be possible to disprove the CMP, and thus the relationship between identification and discrimination cannot be used as evidence for or against categorical perception."

This logic seems to imply that we should stop trying to refine our models (e.g., the categorical model), for we'll never be able to decide when to halt this procedure and simply compare them! I think that a situation in which we have no more solid basis for halting our efforts to refine our models than the foregoing logical argument may be one in which we have an ill-posed mathematical problem. The variety of mathematical formulas that could be associated with informal psychological considerations is quite rich. One can always pick a formula to extrapolate a given model (continuous or categorical) to any desired degree of precision. So without more deeply motivated constraints a problem can easily become indeterminate.

My next example pertains to the discussion of fuzzy logic on pages 202–3. The object under analysis is a striped apple. Assuming that the object is .8 striped and .6 apple, Massaro obtains the following estimated true values:

$$J(\text{striped}) = 0.8,\ J(\text{apple}) = 0.6 \text{ and } J(\text{striped apple}) = 0.8 \times 0.6/(0.8\times0.6+(1-0.8)(1-0.6)) = 0.857$$

Massaro considers the fact that the latter number is larger than the former one consistent with psychological considerations. Let us look at these formulas more closely, however. They can be represented by the following formal algebraic expression:

$$XY/(XY + (1-X)(1-Y))$$

where X and Y are variables representing the property "striped" and "apple," respectively. The condition

$$XY/(XY + (1-x)(1-Y)) > X$$

is equivalent to the condition

$$Y > 1 - Y \text{ if } 0 < X,\ Y < 1$$

Thus a "striped apple" will be judged "more striped" only if the truth value for "apple" (Y) is judged to be more than ½. So if we observe an object (say, an orange) for which the truth value of "apple" is less than ½, does this mean that "striped orange" becomes less "striped" than a "striped" object simpliciter? I am not sure that this "number magic" provides a valid empirical description of the psychological facts.

One last observation is that it is not clear whether the sample sizes in many of Massaro's experiments are large enough to be representative of the underlying population distributions. To capture the unimodal or bimodal character of distributions in Chapter 5 one should perhaps consider samples larger by one or two orders of magnitude.

Cognitive impenetrability of perception

Lester E. Krueger

Department of Psychology, Ohio State University, Columbus, OH 43210
Electronic mail: *ts0340@ohstmvsa.ircc.ohio-state.edu*

Massaro provides some very interesting and useful data. Section 1 of this commentary discusses additional data or analyses that buttress his "fuzzy logical model of perception" (FLMP), which holds that continuous visual and auditory sources of information are independently evaluated in speech perception. Section 2 discusses the need to sharpen or replace some key terms or concepts in the model, namely "fuzzy logical" and "perception." Section 3 discusses the main implication of his model – the cognitive impenetrability of perception. His data, along with those of others, clearly point to the demise of top-down models of feature extraction.

1. Additional data and analyses. Massaro ought to have taken some sources of visual information in his experiments more seriously. He presumed that visual information was lacking if the mouth of the (videotaped) speaker remained closed and motionless when a syllable was heard. However, his data indicate that this so-called no-visual-information condition was not the neutral baseline he supposed it was. It biased subjects toward identifying the auditory /ba/ as /va/ and "the." That bias is consistent with a closed-mouth visual cue; the same bias was found with a visual /da/ (see *Paradigm,* Figure 4, Chapter 3). Thus, the greater robustness of the auditory /da/ (vs. the auditory /ba/) in the no-visual-information condition may reflect the aid /da/ received from the closed-mouth visual cue.

Visual information ought to be especially important for the hearing impaired. The approach known as "total communication" (TC), in which the teacher augments speech with a form of sign language and with fingerspelling, is used in about 65% of classes for the deaf in North America (Reich 1986, Chapter 6). "Teachers using total communication are supposed to sign every word or morpheme they say" (Reich, p. 238), and thus "the sign language used most often in total-communication classrooms is not the sign language used by most deaf people when they talk to one another" (p. 238). Various manual methods are used in TC and related approaches, and it is not yet clear which method best augments the information received by ear (Reich 1986). The fact that these methods work provides support for Massaro's point that perceptually independent information may be received by ear and by eye, and integrated and classified at a higher, cognitive level.

2. Replace "fuzzy logical" with "continuous"; define "perception" more fully. Massaro states (Chapter 7) that the "fuzzy logical model of perception" really means "continuous model of perception." For him, "an equivocal cue might be better thought of as providing information to some degree rather than providing probabilistic information" (*Paradigm,* p. 13). Thus, in his model, the representation "is not uncertain, nor is it intrinsically indeterminate, nor is it undecided. It is simply the representation of continuous grades of knowledge about set membership" (*Paradigm,* p. 204). If so, then to prevent confusion, he ought to have replaced "fuzzy logical" with "continuous," and have merely made a cross-reference to "fuzzy logical." What he really shows is that perceptual and cognitive dimensions are continuous, and that the output becomes discrete (categorical) only when certain (all-or-none) decisions are required.

An important unanswered question is whether the continuous dimensions are represented internally in an analog form, which seems more natural for continuous dimensions, or in a propositional form, which is also possible (Anderson 1978; Pylyshyn 1980; 1984). If both continuous and categorical attributes are represented propositionally, then the distinction between them becomes quantitative rather than qualitative, and thus somewhat less significant.

Massaro uses the term "perception" (e.g., in FLMP), but he does not define it fully enough. He distinguishes the "information" of Fodorian input systems from central-level "information processing," but does not further distinguish perception (hearing) from cognition (hearing as). Would Gibson's (1979) distinction between "projection" and "coding" be appropriate? Fodor's (1983) modularity principle applies only to the input systems, and not to central systems. In order to comment on this principle, as he does, Massaro ought to have stayed within the confines of early (perceptual) processing. However, the examples he gives in Chapter 9 to show the generality of the FLMP heavily involve central (cognitive) systems. The tasks he describes (sentence interpretation, probability judgments of possible events, personality judgments, the learning of arbitrary categories, and judgments of category membership) may well be similar at the central level, even if they are very different at the input level. Thus, the good fit of the FLMP over these diverse domains does not provide evidence against Fodor's view that each input system is specialized for its own particular domain.

The specialization of the input systems depends on their information encapsulation, or cognitive impenetrability (Pylyshyn 1980; 1984), and on this point, Massaro's findings actually provide a way to modify and strength Fodor's modularity principle. Rather than overturn Fodor's entire modularity principle, as Massaro proposes, a simpler and more parsimonious solution is simply to discard speech/language as a module, which is what Massaro's findings really indicate ought to be done. Thus, speech/language processing is not precognitive, as Liberman and Mattingly (1989) claim. There is already an acoustic module, so the addition of a speech/language module is not necessary. Perhaps the defining characteristic of an input system (module) ought to be the presence of sensory transducers. That would eliminate speech/language as well as the sense of time, which may be a "sense" only metaphorically, like the "sense of decorum." Time, like speech/language, may be "perceived" at a cognitive level rather than a perceptual one (Ornstein 1969).

3. Demise of top-down models of feature extraction. According to Fodor, each input system (module) can be conceptualized as a special-purpose computer with a proprietary database, which operates within a specific domain in a fast, mandatory, innately specified, and computationally autonomous manner. The input systems are informationally encapsulated, that is, "they go off largely without regard to the beliefs and utilities of the behaving organism" (1985, p. 2). However, the percept is typically richer than the proximal stimulus, and thus, "a lot of inference typically intervenes between a proximal stimulus and a perceptual identification" (1985, p. 2). To accomplish the latter, Fodor allowed for a back-door transfer of cognitive ability to the modules. Across time, particular associations built up at the central level are transformed and stored in the module itself, thus enabling the module to perform an abbreviated top-down analysis of its own when it encounters an ambiguous or distorted stimulus, as in the case of the phonemic restoration effect (Warren 1970).

The problem with allowing abbreviated top-down analysis within a module is that it is difficult to distinguish this from a full-blown top-down effect from central systems. A cleaner solution would be to show that there is no need for top-down analysis of either type, and this is what Massaro's evidence for the FLMP appears to do in the case of bimodal speech perception. Thus, Massaro has made speech perception cognitively impermeable as well as impenetrable. By stripping them of the cognitive (judgmental) component, the acoustic and visual modules underlying speech perception become truly informationally encapsulated in Massaro's model.

Fodor was not fully committed to the cognitive impenetrability of perception. He stated that "the empirical evidence for the continuity of perception with cognition is not overwhelming when contemplated with a jaundiced eye. There is, in any event, something for laboratory psychology to do for the next twenty years or so, namely, to try to develop some designs subtle enough to determine who's right about all this" (Fodor, 1985, p. 5). Actually, considerable work has already been done on this question, and Massaro's book confirms the general finding of cognitive impenetrability of perception.

In visual processing, letters are detected faster and more accurately in words than nonwords (Krueger 1975). The locus of this word-superiority effect may be perceptual (feature extraction) or cognitive (interpretation) (see Table 1, Krueger 1975). If feature extraction were the locus, that would be evidence for cognitive penetrability of perception. The evidence now clearly rules that out. Broadbent and Broadbent (1975), Estes (1975a; 1975b), Krueger and Shapiro (1979), and Massaro (1979a) each concluded that word context or bias and feature information are independent factors. Massaro, for example, deleted part of the crossbar on the lowercase *e*, so as to present a nonletter that was intermediate between *e* and *c*; he found that the amount of crossbar deleted and the word-versus-nonword context made independent contributions toward evoking the report of an *e* or *c* as present. Wandmacher, Shapiro, and Mohr (1981) likewise found that letter familiarity did not aid feature extraction. Feature deletions were no more readily detected when the mutilation of an *E* formed an intact letter (*F*, *L*) than when it did not (inverted *F*, inverted *L*).

Unfortunately, the finding of no top-down influence usually depends on the acceptance of the null hypothesis. That leaves open the possibility that a more sensitive procedure might be devised that would detect such an effect. Massaro found no interaction between visual and auditory feature extraction in his tasks and situations, but that does not mean that such effects will not be found in other situations. Massaro concedes, too, that part of the problem in obtaining evidence for dependence may be the lack of an adequate model of dependence (*Paradigm*, pp. 154 & 165). Given the variety of techniques used by Massaro and others, however, it is likely that a top-down effect on feature extraction would have been found if it existed. The fact that so many investigators have tried (and failed) to find such effects suggests that they do not exist and that perception is, to all intents and purposes, cognitively impenetrable.

Is *Paradigm* a new and general paradigm for psychological inquiry? Read my lips

Donald G. MacKay
Psychology Department, University of California, Los Angeles, CA 90024

Massaro (1987) presents some interesting empirical findings on the integration of visual and auditory information in speech perception, but wishes to make a more general claim: that this work constitutes a new and general paradigm for psychological inquiry, providing insights into "complex cognitive functioning." What is Massaro's *Paradigm*? Is it really new? And what are its limitations?

Central to *Paradigm* is information processing (IP),[1] something Massaro calls variously a theory (pp. 8–10), an approach (p. 8), a metatheory (p. 8), a paradigm (p. 9), a framework that encompasses "informative research areas" such as "the additive-factor method and backward recognition masking" (p. 9), and a methodology that "allows the experimenter to see what is not directly observable." How well can something function as a theory, as a metatheory, as an area of research, *and* as a methodology? The present comments examine one aspect of this question, with implications for all the others: How well does *Paradigm* function as metatheory?

The first point to clarify is that IP *is* metatheory (a characterization of the goals for a field) rather than a theory per se, and can actually retard the development of theory. As Newell (1973) noted, IP flowcharts can conceal our ignorance rather than promote insight into underlying mechanisms if IP stages are only described via labels such as "evaluation," "integration," and "categorization," as in *Paradigm*. Massaro's claim (p. 25) that he is clarifying the processes underlying these stages by applying additional labels such as "specific" versus "general," or "compromising" versus "enhancing," and so forth, does not help in this regard: These new labels are likewise descriptive rather than explanatory or theoretical in nature. This descriptive aspect of *Paradigm* runs throughout the book. For example, Why does vision often dominate in conflicts with audition (visual capture)? Massaro's answer (p. 84) is a vague description or empirical correlation rather than a theoretical explanation: Because "resolution of visual space is better than resolution of auditory space."

Others (including Massaro, pp. 9–10) have noted many additional limitations of the IP metatheory but have generally been ignored because of IP's assumed pragmatic value: "Despite its many faults, look how many experiments IP has inspired." However, the IP metatheory also suffers from pragmatic limitations of a very general nature that can be illustrated by considering the hypothetical possibility that physics had taken an IP approach to ballistics in the seventeenth century. Analogous to IP psychology, the goal of seventeenth-century IP physics would be the analysis of ballistics, from the construction of a projectile and launching device (Input), through a sequence of ordered stages tracing the projectile's path, until it makes impact. A typical flow chart might resemble Figure 1: Stage 1 (Orientation Stage): Position the gun relative to the earth, determining its vertical angle relative to the ground, and its horizontal angle relative to some fixed criterion such as due North. Stage 2 (Ignition Stage): Insert the projectile into the gun and fire it. Stage 3 (Ascent Stage): The projectile climbs to some determinable height (Let's provide seventeenth-century physics with twentieth-century technology) before starting to fall back to earth. Stage 4 (Descent Stage): The projectile falls back to some determinable point on the earth (Impact).

What makes such an end product so inadequate for seventeenth-century physics (and twentieth-century psychology)? It is not that the IP metatheory is *totally* useless: After all, with progressive refinement of the above stages, IP physicists might eventually be able to describe something like the actual path of a projectile. Moreover, by following exactly the same steps under exactly the same conditions, IP physicists could accurately predict the projectile's path.

The problem is that IP physics is descriptive rather than explanatory or theoretical, and has severely limited pragmatic value. At best, IP physics could describe only particular trajectories for particular projectiles fired under particular conditions, a major limitation because one isn't interested in launching a series of identical projectiles under identical conditions. Moreover, no matter how long it is followed, the IP approach provides no way of arriving at explanatory or theoretical concepts such as mass, gravitational force, inertia, kinetic energy,

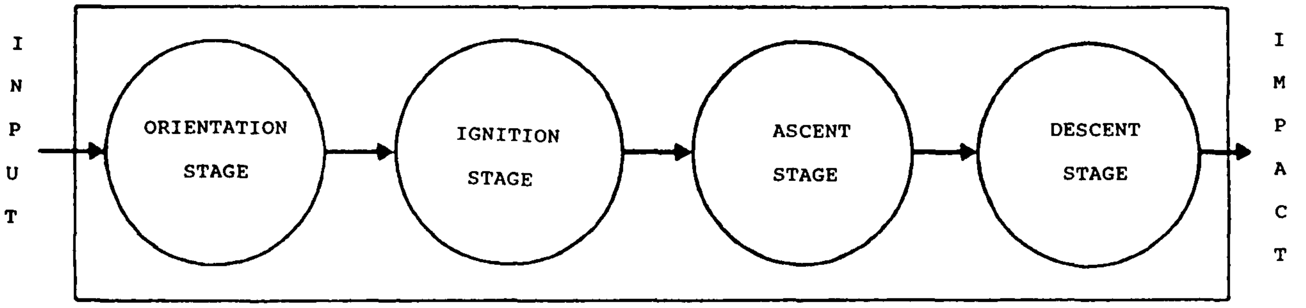

Figure 1 (MacKay). A hypothetical flow chart from seventeenth-century information processing physics tracing a projectile from construction (Input), through firing, until impact.

centrifugal force, and so on, which are the sort of concepts needed for achieving accurate predictions when firing different projectiles under variable (real world) conditions (see MacKay 1988a). In short, the IP metatheory would have retarded practical and theoretical developments in physics just as it is said to have done in psychology (see MacKay 1988a; 1988b).

A second component in Massaro's *paradigm* is Popper's (1959) falsificationism. Falsificationism is a methodological approach with well-known limitations that have been widely discussed and agreed upon in philosophy of science and elsewhere (see e.g., Kuhn 1977); and many have pointed to the dangers of overgeneralizing or overapplying falsificationism in the straightforward way that Massaro advocates. Moreover, Massaro proceeds in a manner that seems flagrantly inconsistent with falsificationism at many points throughout his book. For example, it has long been known that some aspects of language learning are gradual whereas others are not, especially those involving the acquisition of rules such as regular English plural formation (see e.g., Pinker 1984). Massaro's (p. 269) support for the already falsified claim that language learning is universally slow and gradual therefore seems to violate falsificationism, especially when Massaro fails to address the issue of rule learning.

It is sometimes difficult to imagine what evidence could possibly falsify Massaro's fuzzy logical model of perception, and his search for similar findings in various areas seems to violate falsificationism directly. For example, Massaro describes his analysis of the sentence integration results of MacWhinney et al. (1984) as direct and illuminating (p. 272), an illustration of the success of *Paradigm* in confronting underlying psychological processes, unlike most of modern "psycholinguistics [which] has been satisfied with demonstrating the psychological reality of some linguistic concept" (p. 272, brackets mine).

MacWhinney et al. (1984) showed that word order, noun-verb agreement, sentential stress, and animacy have radically different effects on guesses as to which of two nouns acts as "sentential subject" for German- versus Italian- versus English-speaking subjects. For example, English speakers were more influenced by word order than Italian speakers (who were more influenced by animacy) and German speakers (who were more influenced by noun-verb agreement and sentential stress). When Massaro "explains" these opposing differences in terms of post-hoc or assumed variations in "information values" for word order, noun-verb agreement, sentential stress, and animacy, one wonders whether his theory is too powerful, a criticism that Massaro levels against other theories (see p. 281).

What Massaro counts as evidence for "sentence integration processes" is as troublesome as what he counts as explanation. For example, MacWhinney et al. (1984, p. 139) defined "sentence subject" for their subjects as "the one who (sic) does the action": Isn't "who" an animate (human) pronoun? Couldn't this animate pronoun have influenced the "information value" for animacy? MacWhinney et al. also impoverished the normally available cues to sentence subject, for example, by deliberately excluding German words with morphological markers that distinguish nominative versus accusative case. Surely this procedure cannot qualify as ecologically valid, another basic tenent of *Paradigm* repeated throughout Massaro's book. Finally, MacWhinney et al. warned their subjects that many of their stimuli "would seem odd" and consisted of nonsentences such as "The eraser the pig chases" and "Licks the cow the goat," a methodological detail that can be discovered only by examining the original paper. Surely nonsentences cannot in principle directly illuminate sentence processing as Massaro suggests. *Paradigm* seems to tell us very little about either "sentence integration" or "complex cognitive functioning" as we know it in cognitive psychology.

Another component of *Paradigm* is Platt's strong inference (p. 6), the idea that investigators should test an array of contrasting hypotheses rather than just one; Massaro criticizes the motor theory of speech perception for failing to demonstrate that it is better than alternative models. However, contrary to the spirit of strong inference, he also ignores or downplays other theories that bear on the same phenomena as his own. For example, he downplays McClelland and Elman's (1986) theory because "the evidence points to continuous rather than categorical perception" (p. 149). However, the evidence only points that way because of an arbitrary and highly controversial definition: Massaro defines "perception" to include "sensory operations" (exhibiting continuous characteristics), but not "decision operations" (exhibiting categorical characteristics). If distinguishable sensory and decision operations exist, our metatheoretical goal should be to explain *why* sensory operations exhibit continuous characteristics and *why* decision operations exhibit categorical characteristics. The "business at hand" is *not* to "eliminate alternate explanations of behavior" (p. 281), especially not via definitional hat tricks, and not in favor of alternatives such as *Paradigm* that provide *less* detailed specifications of how information is represented and processed.

The final, implicit rather than explicit, component of Massaro's metatheory is the empirical epistemology that has dominated psychology during the last 75 years (see MacKay 1988a): *Paradigm* carries all of the earmarks of the empirical epistemology, from a concept of theory that is broad, imprecise, and descriptive in nature, to a plea for empirical paradigms. Although the promise of replicability represents its basic appeal in contemporary psychology (see MacKay 1988a), the paradigm-specific fact gathering that Massaro calls for doesn't really solve the problem of replicability (see MacKay 1988a) and carries limitations that should be stressed: It has focused the field on narrow interests that are easily ignored or forgotten; it has resulted in reduplicative rather than progressive or incremental theoretical and empirical efforts (see Cole & Rudnicky 1983); and it has interfered with rather than promoted the development of viable theory (see MacKay 1988a). A strong case can be made that the field needs a new and different sort of epistemology to supplement the empirical one (see MacKay 1988a; 1988b), along with a different sort of theory, one that is explanatory rather than descriptive, and does not originate from observational "20 questions" in the way Massaro (p. 24) suggests (see MacKay 1988a; 1988b).

In conclusion, Massaro's *Paradigm* is not new, has basic limitations, and is basically empirical and descriptive: The fact that fuzzy set mathematics can describe aspects of Massaro's results or results in other areas such as person perception is unenlightening about underlying mechanisms. Moreover, a strong case can be made that selective searches for surface similarities in findings from a variety of areas *should not* constitute a preferred paradigm for psychological inquiry.

ACKNOWLEDGMENT
The author thanks Asa Kasir for helpful comments on this commentary.

NOTE

1. In view of his advocacy of IP boxes, Massaro's very general attacks on J. A. Fodor's (1983; see *BBS* multiple book review, BBS 8(1) 1985) modularity concept are curious. Fodor was after all proposing *boxes* of a certain sort, and many investigators have seen modularity as a welcome addition, a source of new life for the old and problem-ridden "boxes and arrows" idea that Massaro advocates. Moreover, Massaro's Feature Evaluation box (Fig. 4, p. 25) seems to share many of the "vertical" properties that Fodor proposed, and his Integration and Classification boxes share many of Fodor's "horizontal" properties. In short, Massaro could have chosen to interpret his work as a valuable extension of the modularity approach, rather than "completely opposed," "the antithesis" of the modularity approach, with his results striking "a blow to the concept of modularity" (p. 281).

Continuous and discrete models and measures of speech events

Richard E. Pastore, Robert J. Logan, and Jody Kaplan Layer
Cognitive and Psycholinguistic Sciences (CaPS) Center, State University of New York at Binghamton, Binghamton, NY 13901
Electronic mail: *faco69@bingvma.bitnet*

Massaro (1987) presents an interesting and important alternative interpretation of the nature of information representation during the processing of speech and other cognitive events. In support of his conjectures, he reports some new data as well as reanalyses of extant literature. In this critical review we will evaluate the relative merits of Massaro's book and his conjectures by addressing the following questions: (1) What are the explicit and implicit goals of the book? (2) Are these goals met? (3) What are the contributions of the book apart from the goals? (4) Are the proposed methods, theory, and "paradigm" useful?

Explicit and implicit goals of the book. Its title, *Speech perception by ear and eye: A paradigm for psychological inquiry* defines the two explicit goals of Massaro's book. The first goal, which is narrow in scope, is to provide a psychological description of how a perceiver recognizes what a speaker says based on available auditory and visual sources of information. Knowledge gained through an understanding of this problem is applied to meet the second, more ambitious, goal, which is to confront fundamental issues in experimental psychology and cognitive science by providing a "paradigm" for further inquiry.

In addition to these explicit goals, there are several *related* subgoals which are essential for the advancement of Massaro's theoretical position and for achieving the general goals. One subgoal is to use evidence collected from bimodal speech perception to refute the notion of discrete perceptual processing in favor of continuous processing models. From these specific empirical demonstrations, Massaro hopes to generalize the notion of continuous processing to other domains. By establishing the generality of continuous processing, he attempts to refute the "modularity of the mind" (e.g., Fodor 1983; 1985; Liberman & Mattingly 1985) as a reasonable mental architecture. Finally, Massaro attempts to establish the Fuzzy Logical Model of Perception (FLMP) as a viable generalized continuous process model. Although Massaro treats the attainment of this subgoal as synonymous with the attainment of a new paradigm, the two goals should be kept separate. A theory or model of information processing can be related to a paradigm, but it is not necessary for the existence of the paradigm.

Discrete versus continuous. In the 1960s and 1970s the major working assumption in the speech perception literature was that consonants are perceived in terms of discrete categories, with the normally continuous variability of acoustic stimuli not perceptually available to the subjects. Although categorical perception research was a major theme in the speech perception literature of that period, many researchers rejected the notion of absolute recoding of consonants long before the appearance of Massaro's book. However, implicit assumptions about discrete perception of phonemes can still be found in the modern speech literature, especially in the areas of phonetic trading and duplex perception. This book may help convince readers that these implicit assumptions are false.

The experimental methods used to study speech perception were, and still are, typically based upon ABX discrimination tasks and labeling tasks with limited response sets. These tasks are biased toward finding discrete categories (as we have often observed in our own laboratory even with very simple stimuli) and thus tend to impede the empirical evaluation of continuous components or "cues" usually conjectured to be the basis of speech perception. With this implicit bias in the experimental paradigm, only the most continuous of perceptual dimensions can be demonstrated to be other than relatively discrete. Pastore and Layer (submitted) provide a summary of these biases in a critical evaluation of phonetic trading research.

A number of precategorical (continuous) models of auditory and visual speech integration are summarized by Summerfield (1986), although none of these models provide a means for directly evaluating continuous "cues." Massaro argues that these "cues" are best evaluated using continuous *measures*. Whereas continuous measures are most appropriate for the evaluation of continuous processes, such measures also have an implicit bias in favor of finding the appearance of continuous rather than discrete processes.

In our opinion, Massaro's book and recent work by others (e.g., Pastore & Layer, submitted) suggest that future research use a multimeasure methodology when studying the nature of information representation. By using both continuous and discrete measures on the same stimuli and subjects, one will develop a more appropriate model of information processing. Specifically, any biases introduced by categorical or continuous measures will become apparent when comparing the results.

Comparison of FLMP and CMP. There are major discrepancies between the theoretical and operational implementation of the FLMP and a categorical model of perception (CMP). In the descriptions of both models, auditory and visual speech provide the subject with independent and continuous information about the degree to which features are present. In both models, a_i and v_j theoretically represent the degree to which these features are perceived in the given auditory and visual stimuli (A_i and V_j). Operationally, Massaro estimates these parameters for both models from average response probabilities. These parameters reflect the average pooled probability that the subject (1) attended to, (2) fully evaluated, and (3) utilized the available types of auditory and visual feature information. Both the CMP and FLMP models thus compute equivalent continuous statistics. Ignoring the theoretical distinctions, the major operational difference between the models is the application in CMP of a multiplicative constant, p, which determines the relative contribution of the auditory and visual parameters. Although the purpose of p is to make CMP discrete, the addition of this fixed parameter can be expected to restrict the predictability of the model, resulting in poorer goodness-of-fit.

Contributions of the book. *Goodness of fit.* In the speech perception literature, few recent studies provide any measure of goodness-of-fit between obtained and predicted results; the unusual modern study will compute a correlation. Massaro cites problems with this use of correlation and proposes the use of root mean square deviation (RMSD) as a more desirable measure. He also warns, "There is no robust statistic that can give acceptable values of goodness-of-fit of a model relative to the number of independent observations and free parameters" (p. 20). We fully agree with this analysis of goodness-of-fit measures and add our support to the intelligent use of measures of goodness-of-fit.

Identification of features. The average reader is led to believe that FLMP (in contrast to CMP) is evaluating the contribution to perception of information about specific features. For instance, the reader is told that in one set of studies the single auditory and visual features for /ba/ are "Slightly falling F2-F3" and "Open lips" (p. 122). However, (as stated above) a_i and v_j are estimated solely from overall response probabilities and not in terms of whether the specific features (or any other features) are ever evaluated by the subjects.

Lack of specificity. In the first chapter Massaro introduces the FLMP and associated concepts. The reader is given a very fuzzy description of the terms and how the information associated with those terms is to be implemented in the prediction equation; the reader is referred to the program (STEPIT) that does the calculations. Following this broad description, Massaro provides a simple application of the FLMP (p. 20). From his description, even a dedicated researcher would find it difficult to replicate the simplest of these tests. This lack of specificity for implementing the methods of testing one model against another makes it difficult to produce consistency and replicability across laboratories. A more useful presentation would include, at a minimum, an appendix which provides an adequate description of methods for use by the more ambitious reader.

Validity of empirical support. Throughout the book, Massaro repeatedly cites his own research and that of his colleagues. A portion of this research (e.g., Cohen 1979; 1984) is unpublished, and thus not subjected to peer review and not readily available in the public domain. Moreover, some of the major conclusions in the book are based upon evidence from these manuscripts, forcing the reader to accept the results and conclusions at face value. In addition, when there were published criticisms of cited research, Massaro failed to acknowledge them. For example, Pastore et al. (1984) and MacMillian (1983) provide specific criticisms of the methods and conclusions found in Hary and Massaro (1982).

Concluding comments. Although there are some problems with the book, Massaro is reasonably successful in achieving his goals. In addition, the book makes an important contribution to the psychological literature by reinforcing the need to consider the notion of continuous processing (rather than tacitly accepting the assumption of discrete representations). It remains to be seen, however, whether Massaro's approach to human information processing research will lead to the establishment of a "paradigm" for psychological inquiry.

The fuzzy logical model of perception: A teaspoon for a pyramid

Robert F. Port

Departments of Linguistics and Computer Science, Indiana University, Bloomington, IN 47405

Electronic mail: *port@iuvax.cs.indiana.edu*

Massaro's book presents a "process model of speech perception." Although many interesting experiments are presented therein, it greatly overreaches itself in claiming to present a model of the *process* of human speech perception. Massaro poses that his Fuzzy Logical Model of Perception (FLMP) is the answer to all our prayers – the model that illuminates how perception occurs. In fact, however, it primarily reveals the limitations of research on perception that focuses on standard psychological paradigms and illuminates how impoverished our models of speech perception are. Concentration on identification experiments that use briefly presented doses of stimulus information and highly restricted response sets encourages the development of seriously inadequate models like the FLMP (among others). The FLMP model should be thought of as a convenient framework for summarizing the results of a wide variety of experiments, but it does not help us very much in developing a model of the process of speech perception. It offers us a teaspoon as a tool for building an Egyptian pyramid. Intuitively, it seems as if it should be useful, but practically it does not help much.

Certainly the FLMP has many attractive features. First, by using fuzzy truth, the model provides a metric for combining information from many sources – and one that allows the most distinctive cues to dominate recognition in a given situation. Since fuzzy truth values, like probabilities, always lie in the range [0,1], they are easily combined by multiplication. A decision is based on the relative "truth" value of the set of alternative categorical labels. That is, the product of truth values for various alternatives is divided by the sum of the truth values for the other alternatives. It is assumed that the greater relative value is the category choice.

Second, the framework is flexible enough to incorporate any form of input data that can be shown to have an effect on performance. Of course, this very flexibility means that FLMP itself imposes almost no substantive constraints on specific models to be entertained. Thus there is no constraint to discourage us from testing the appropriate values of parameters for, say, the color of the speaker's hair on the perception of speech. Perceptual systems invariably have some kinds of constraints built in – some hypotheses that are more "natural" and others that are "bizzare."

Scale-up difficulties. Since this model is applied by Massaro to the results of simple forced-choice perception experiments, it seems reasonable to use parameters for each of the response alternatives. When the number of alternative responses is fixed and small (e.g., /ba/ vs. /da/, and the number of stimulus levels is small (e.g. 9 values along a /ba-da/ continuum plus the presence vs. absence of a video-taped image of a speaker), then the model can be applied with ease. A simple learning program, like STEPIT (Chandler 1969), can find the optimum values for the 11 parameters for each of the two response categories.

But real human perception does not resemble this case in the least. When a speaker opens his mouth, we do not know which of hundreds of possible syllables will be produced, nor what aspects of the audio and video stimulation will be available before presentation is completed. For example, Massaro uses a 9-level continuum from /ba/ to /da/, but why not 18? Or why not vary many other acoustic parameters (e.g. other formants, formant amplitudes, FO, formant intensities, etc.)? The number of psychophysically distinct stimuli near /ba/ and /da/ is vast – surely in the thousands. (Massaro correctly rejects the idea that a preprocessor provides a small number of categories for us a priori.) And each of these stimuli needs a fuzzy truth value for both /ba/ and /da/ in Massaro's 2-alternative paradigm (as in the table on p. 130). But the number of alternatives in real speech perception is in the hundreds or thousands. Massaro tests the effectiveness of visual information representing /ba/ or /da/ or no visual information at all while subjects listen to members of his /ba-da/ continuum. But how many discrete "visual variables" were there in his video tape? If, as Massaro proposes, speech perception does not involve a separate module from other domains of auditory perception, then the need to compare each truth value with *every* alternative appears to require us to ask about the relative /ba/-ness of the sound of a fire engine as well as the relative woodpecker-ness of any spoken syllable. Even leaving aside what evidence these syllables might provide for approaching trucks, barking dogs, Beethoven's Fifth, squeaking chairs, my wife's footsteps, etc., it is clear that a model along these lines will be hopelessly bogged down just selecting among speech possibilities – too many parameters to estimate during perceptual learning. I seriously doubt that FLMP could be scaled up to any real recognition task. The number of parameters that would have to be estimated, equal to the number of possible stimulus levels (that is, relevant stimulus differences in all modalities, visual, auditory) times the number of alternative response categories, is staggeringly large. How could such a catalogue of parameters be estimated without exhaustive mapping of them? Surely, a simple, linear, gradient-descent algorithm like STEPIT is not up to the task. Nor is any other known learning technique.

In fact, however, the task is probably much worse. The current FLMP, like many pattern recognition models, assumes that there is a fixed set of discrete parameters, each independent of all the others. That is, the fuzzy truth value of each cue is assumed to be independent of the truth value of all the other cues. Although many synthetic speech cues may be linearly independent in this way, many other cues clearly are not (as noted by Massaro, pp. 162–63). This will greatly complicate attempts to estimate the parameters and employ these cues.

Time and parallelism. There is a second major difficulty, one that has received little attention in recent years, yet which prevents Massaro's from being a "process model." Massaro believes that it is reasonable to assume that all the information required to recognize a word (for example) is presented at once, and that it is all processed in parallel. Thus, to recognize /da/ the speaker (or at least FLMP) can examine the acoustic information from the burst, the formant transitions, and the steady state vowel *all at once*. The perceptual system can combine the fuzzy-truth values of all these discrete independent cues along with other dynamic information about lip and tongue tip motions visible in the videotape, and do a unitary operation conjoining these sources. Then a perceptual decision is made on this basis. At this point, the system is ready for another trial. It is assumed that a model developed like this can lead to a theory that can be extended to general real-time speech perception.

But real speech arrives continuously; it is not divided into trials, syllables, cues, words, or "items." The cognitively relevant sample rate is not as fast as that of the A/D converter, certainly, but something like that of the 8th nerve (that is, it is analogous to the fast Fourier transform frame at rates between 10 and 50 samples/sec). But no time window longer than the single frame of spectral information is apparently provided to the central auditory system – and single spectral frames are not long enough to permit syllable recognition. That means that listeners do not have the possibility of collecting sequentially presented information, storing it up until some prespecified point in time (e.g., the end of a trial) and then processing it all at once. This kind of "parallelism" claimed by Massaro for his model (on p. 280), like the parallelism of most models of auditory perception (connectionist or otherwise), is merely a trick – an artifact supported by looking only at data from isolated-word perception experiments. It has nothing whatever to do with the parallelism of nervous systems. Although the mathematical operations of FLMP are computed sequentially when the equations are computed, the multiplication operations could, of course, be implemented in parallel. But the only reason he can exploit this parallelism is that he has artificially forced simultaneity on sequentially presented acoustic information. He imposes long time windows on the acoustic signal and imposes word-sized experimental trials on the subject. Before we can develop a true process model of speech perception, we must address the question of how to collect information over time in a perceptual system *without* imposing windows of simultaneous information (see recent work from our laboratory, Port & Anderson, in press; Port et al. 1988; Anderson et al. 1988 for attempts to achieve this).

In general, this work points up why it is important for cognitive psychology to be married to computational modelling of perceptual processes. By actually attempting to implement recognition models, the model builder is forced to face scale-up difficulties and logical problems such as externally specified time windows.

In summary, then, Massaro's model may work fine on the results of small experiments with a small number of experimentally varied parameters and forced-choice responses. It is useful for summarizing the results of such experiments. The experiments of Massaro and his colleagues demonstrate many interesting properties about the interplay of different kinds of perceptual information. The book provides a clear and thoughtful summary of these experiments and demonstrates, among other things, the importance of multimodal information in speech perception. My objection is that FLMP itself is not an interesting model for human perception. The model will not scale up to any real perceptual tasks – or, at least, the author has given us no reason to suspect that it might. By testing subjects with brief trials of stimulus information, the experimenter can pretend that all the information is available at once, thus posing the completely artificial task of integrating simultaneously presented information. To be sure, these criticisms apply also to most other models of perception as well. Adequate models do not yet exist, of course. But it is important to appreciate the limitations of the Fuzzy Logical Model of Perception.

Paradigm lost

Michael Studdert-Kennedy
*Haskins Laboratories, New Haven, CT 06510**
Electronic mail: *haskins@yalevm.bitnet*

The theme of *Paradigm* is that perception, in general, and speech perception, in particular, entails the integration of information from many sources. The theme is scarcely new, having been a commonplace of speech research almost since its inception (e.g. Jones 1948; Miller et al. 1951; Pollack & Pickett 1963; Scholes 1968). In fact, very little in *Paradigm* is new, beyond the impressive mathematical apparatus marshalled to support the argument. Yet it is precisely the mathematics that robs the discussion of its interest – at least for a reader concerned with the biological foundations of language, for it leads Massaro to treat all sources of information as equivalent, and so to miss the true import of his own and others' work on audiovisual speech perception.

Paradigm is concerned solely with process, with *how* we perceive, and is indifferent to *what* we perceive (cf. Repp 1988, p. 253). Accordingly, the author grants exactly the same status to information stemming from the perceiver's knowledge of phonology, syntax, semantics, and pragmatics as to information carried by the physical properties of the perceived event itself. There can be no question that all these forms of knowledge contribute to speech perception, and may even be essential to later stages in the "interactive" process of language acquisition that the author favors (*Paradigm*, p. 240). Yet there can equally be no question that for the infant learning to speak, it is the physical (acoustic, optic) properties of an utterance that must first be apprehended, for these are properties that specify the articulatory organization of one language rather than another.

The central weakness of *Paradigm's* approach to speech perception, then, is its failure to acknowledge that the process of integrating auditory and visual information is qualitatively different from the process of integrating "top-down" linguistic and extralinguistic knowledge. While the latter process may be as general as *Paradigm* proposes, the former evidently rests on the specific fact that acoustic and optic information are structurally equivalent because they have a common origin in the speaker's articulation. Curiously, *Paradigm* cites (p. 51) the important experiment of Summerfield (1979) supporting the hypothesis that the integration of acoustic and optic properties in speech engages a "common metric" of articulatory dynamics, but elects not to address the issue. *Paradigm* also conspicuously fails to acknowledge that it was Summerfield (1979; see also Summerfield 1983; 1987), not Massaro, who first demonstrated that, contrary to the original hypothesis of McGurk and MacDonald (1976), acoustic and optic information are analog rather than categorically discrete at their conflux.

Unhappily, *Paradigm's* selective and self-serving treatment of Summerfield's work is typical of its treatment of other work. For example, roughly one quarter of the text is devoted to demonstrating that categorical phonetic percepts derive from a decision process based on continuous auditory information. No reader who was unfamiliar with the literature on speech perception would suspect that this conclusion has been a commonplace of that literature for nearly 20 years, and was indeed well established before Massaro even entered the field.

In fact, the hypothesis of Studdert-Kennedy et al. (1970) that stop consonants (though not vowels) could be discriminated only if they were differently identified had already been undermined before it was published. Stevens (1968) and Sachs (1969) both attributed the poor within-category discrimination of stop consonants to the transient nature of their acoustic cues. Fujisaki and Kawashima (1970) proposed a model explaining the consonant-vowel difference as due to differential decay of auditory memory during a discrimination trial. Pisoni (1973) and his colleagues (Pisoni & Lazarus 1974; Pisoni & Tash 1974) supported Fujisaki's "dual process" model in a series of elegant and decisive experiments to which *Paradigm* gives no credit. Studdert-Kennedy and Shankweiler (1970) also attributed the consonant-vowel differences in dichotic ear advantages to differential loss of auditory information during transcallosal transfer.

Moreover, all these studies and many others have been repeatedly acknowledged in review articles (e.g., Repp 1984; Studdert-Kennedy 1975; 1976; 1980; 1981). For example: "In short, consonants and vowels are distinguished in the experiments we have been considering, not by their phonetic class or the processes of assignment to that class, but by their acoustic characteristics and by the duration of their auditory stores" (Studdert-Kennedy 1976, p. 263). Compare this statement with the following from *Paradigm:* "If the sensory information is lost very quickly, continuous information could participate in the perceptual process but might not be readily accessible for introspective reports" (p. 120; p. 14 of target article). Massaro apparently does not recognize the correspondence between these two statements. But then he does not acknowledge many of the papers cited in this and the preceding paragraph, nor the many other relevant papers cited in those papers. Perhaps he has not read them. If so, one can only observe that those who do not study the literature are condemned to repeat it.

Finally, let me note that *Paradigm* grossly overrates the importance of categorical perception as support for the hypothesis that speech perception engages specialized cerebral processes. The principal supports for this hypothesis are: (1) the failure, despite attempts over many years, to devise an arbitrary acoustic substitute for speech in reading machines for the blind that would be any more efficient than Morse code, for which perceptual rates are scarcely one tenth those of speech (Liberman et al. 1967); (2) the repeated demonstration, in both brain-lesioned and normal subjects, of a double dissociation between left and right cerebral hemispheres in the perception of speech and nonspeech sounds (e.g., Blumstein 1981; Molfese & Betz 1988; Tartter 1988). That such findings have no place in the grand design of a *Paradigm for Psychological Inquiry* invites the conclusion that this is a paradigm we would do well to lose.

ACKNOWLEDGMENT

Preparation of this comment was supported in part by NIHCD Grant 01994 to Haskins Laboratories.

NOTE

*Author's additional affiliations are University of Connecticut and Yale University.

Winning "20 Questions" with mathematical models

James T. Townsend

Department of Psychological Sciences, Purdue University, West Lafayette, IN 47907

Electronic mail: *townsend@brazil.psych.purdue.edu*

Massaro's book is a milestone in perceptual/cognitive psychology. First and foremost, it shows that: (1) The information processing approach is not only not dead, but still has much to offer behavioral science. (2) It is possible to test psychologically interesting opposed or dichotomous alternatives (i.e., it is possible to successfully play the game of "twenty questions"; see e.g., Newell 1973) using creative, theoretically motivated experimental designs. (3) Mathematical modeling offers great power in this quest. (4) The discussion of such issues as modularity, though enriched by philosophical as well as physiological disputation (e.g., Fodor 1983), requires behavioral intervention for successful resolution.

All this having been said, and accepting that Massaro has amassed impressive evidence in favor of his claims, a number of the very important issues broached by Massaro are far from settled. For one thing, psychology's experience is inevitably that as investigators dig deeper into phenomena in a search for lawfulness "things are always more complex than they seemed." This has led again and again to the wholesale abandonment of topics (not just dichotomous issues either!) and research areas and has contributed to psychology's reputation of flightiness and trendiness. This problem is not one we are likely to remedy in the near future, however.

Other problems are more immediate. One is that it is extraordinarily difficult to obtain answers from nature in a model-free way (as Massaro notes on p. 152). That is, for many research questions, it would be optimal if we were able to test all models which capture the critical psychological notions in which we are interested. This is rarely feasible, for many reasons. Even the testing of two different models of the same phenomena can be pragmatically impossible within the standard paradigms. For example, Luce's choice theory (1959) predicts receiver operating curves that are, for all practical purposes, indistinguishable from the traditional theory of signal detectability (Swets et al. 1961). Parallel-serial identifiability difficulties are another example (Townsend 1971; 1972), although with effort, solutions can be discovered that permit experimental testability (e.g., Townsend 1976a; 1976b; Ross & Anderson 1981; Thomas 1969; Townsend & Ashby 1983). The main point here is that we are in most cases forced to start with a specific model or theory (that is, with assumptions and resulting predictions embedded in specific mathematical formulae) to test our hypotheses. It often takes years to work up the fully model-free tests and these may differ, depending on the empirical facts learned in the meantime. (Note that putting models in verbal form does not solve this difficulty; it merely hides it.)

This is perhaps the weakest link in Massaro's armada of evidence. He is presumably interested in testing opposing concepts such as "categorical versus continuous" and "independent versus dependent" in some supramodel sense; and this, of course, has not been done. Yet to the extent that his model is supported by many sets of data, both that model and its formalization of the various concepts (e.g., continuity) receive support. (Aside: I can't help wondering why Massaro did not use a statistical technique such as the chi-square procedure which allows statistical tests to be performed along with the fits – as opposed to least squares, which don't permit this in the present circumstances.)

Nevertheless, certain issues such as independence are particularly vulnerable to this criticism. A recent general theory of recognition (Ashby & Townsend 1986), which can be viewed as a generalization of signal detectability theory (Green & Swets 1966), portrays recognition quite differently from the way Massaro's does. It also contains apparatus for response bias, an omission, though a remediable one, in Massaro's theory. More generally, the way independence manifests itself in data could take many forms, depending on the particular model.

A final comment concerning the necessity of distinguishing the various types of independence (Ashby & Townsend 1986): In Chapter 6, for example, Massaro uses the notion of perceptual independence largely as Ashby and Townsend use it (e.g., in Massaro's horse race model of reaction time), whereas he later invokes a concept that we would refer to as "perceptual separability" (e.g., pp. 166–69). The various types of independence may be related in a particular model of perception, but they are logically distinct concepts.

The use of mathematical models in perceptual theory

Richard M. Warren

Department of Psychology, University of Wisconsin-Milwaukee, Milwaukee, WI 53201

Massaro states that "My goal in this essay is to integrate the information processing approach with the study of the world of information" (*Paradigm*, p. 10). In implementing this goal, he uses bimodal speech perception to develop a general model, or paradigm, for research in psychology. Hence, if this model is successful in dealing with the perception of speech, it would have a significance extending well beyond this particular topic. However, for the reasons given below, I believe that the ambitious goal of developing a generally useful paradigm has not been achieved.

Central to Massaro's treatment of the role of audition and vision in speech perception is a series of seven alternative hypotheses which he calls "binary oppositions" (précis, Fig. 2; *Paradigm*, p. 24). These two-valued alternatives are used to construct "trees of wisdom" (*Paradigm*, Fig. 3). Choosing the appropriate alternative leads to another binary choice along the tree trunk (the inappropriate alternative is a dead branch, leading nowhere). Examination of the tree shows that the auditory and visual information concerning the speaker's utterance can be either integrated or not integrated – then, if integrated, it can be categorical or continuous – then, if continuous, . . . and so on. The use of a succession of binary oppositions produces a fragile model, since each stage assumes the validity of all prior stages. Also, Massaro's choice of the correct alternative is, in some cases, controversial. For example, his conclusion that speech perception (whether auditory, visual, or bimodal) is continuous rather than categorical (p. 150) conflicts with the conclusions of many, if not most, of the people who have done research in this area (see Harnad 1987).

While Massaro uses binary oppositions for his "trees of wisdom," he changes to truth values which can vary continuously from zero to one when constructing his Fuzzy Logical Model of Perception (FLMP). Massaro emphasizes that "fuzzy truth" is different from probability. He gives as an example: "If we say that a whale is a fish to degree .2, that does not mean that there is a .2 probability that a particular whale is a fish. Rather it is true that a whale is a fish to a degree .2" (*Paradigm*, p. 186). The FLMP uses quantified descriptors which function as empirically determined constants. The number of such constants is quite large, varying from 11 (*Paradigm*, p. 168) to 48 (pp. 128–29).

It appears to me that the FLMP violates Occam's razor by multiplying entities unnecessarily. It also seems to violate an even older stricture which might be called "Aristotle's constraint" (expressed in his Nicomachean Ethics, book 1, chapter 3): *Do not quantify inappropriately*. By using many empirically determined constants, each of which corresponds to a "parameter" in his FLMP, Massaro found it possible to obtain good fits to empirical data. With up to 48 constants to adjust, Massaro turned to the computer program STEPIT, which iteratively adjusted the value of each parameter so as to minimize the squared deviation of the observed and "predicted" points (I'll return to this use of "predicted" shortly). Massaro stated that, "The outcome of the program STEPIT is a set of parameter values which, when put into the model, come closest to predicting the observed results. Thus, STEPIT maximizes the accuracy of the description of each model" (*Paradigm*, pp. 19–20).

It has been stated, possibly with some exaggeration, that four constants suffice for plotting the figure of an elephant, and the introduction of a fifth can lift the tail. Consider what might be described with the numbers of constants used in the FLMP. Obtaining a fit between the output of the STEPIT program and any experimentally obtained function would not seem surprising. Since this agreement is obtained *after* fixing the constants to match the empirical data, the term "predicted" does not seem appropriate. Yet the curve fitting results of the STEPIT program are considered as a validation of predictions of the FLMP (see précis, Figs. 7 & 8).

Mathematical models have a seductive appeal. They permit experimental data to be described, not as a collection of individual values, but as measures that are linked by derivation from the same mathematical expression. However, if the agreement of the mathematical model with experiments is achieved only after the data are gathered and several constants are adjusted to fit, then this agreement would have little more significance for theory than the shaping of a flexible strip of rubber to fit an array of points on a graph.

A comparison of speech perception and spatial localization

Robert B. Welch

NASA-Ames Research Center, Mail stop 239-7, Moffett Field, CA 94035

With respect to both methodology and outcome, Massaro's impressive research program on speech perception is highly reminiscent of studies of intersensory interaction in the perception of spatial location (Welch & Warren 1980). Both areas share the assumption that perception is based on the *integration* of multiple sources of sensory information whose relative weighting is most accurately assessed by placing them in *conflict* with one another. For example, momentarily viewing the hand through a prism causes it to feel as if it were located near (although not coincident with) its seen position (e.g., Hay et al. 1965), a phenomenon referred to, in general, as "intersensory bias." From this observation it has been concluded that vision is more heavily weighted than proprioception in the everyday (i.e., sensorily congruent) perception of limb position. The present commentary is devoted to a discussion of the striking similarities between speech and spatial perception. This comparison is useful, first, because it strengthens Massaro's contention that the perception of speech is not unique and, second, for its further illumination of this perceptual capacity and suggestions for future research.

Massaro argues that the integration of sensory cues is demonstrated only if responses made by observers in the presence of the multimodal event occur (rarely, if ever) when either of the sensory modalities is presented alone. And he reports just this outcome when, for example, the simultaneous presence of a visual /ba/ and an auditory /da/ leads to the report of /bda/. Thus, one of Massaro's most theoretically significant contributions is that neither vision nor audition ever totally dominates speech perception, exactly the conclusion that emerges from the intersensory bias research with respect to spatial localization. Pick et al. (1969), for example, report not only a very strong visual bias of proprioception and audition, but a small, though reliable, effect of the "lesser" modalities on vision. The fact that "genuine" sensory integration is demonstrable in such disparate perceptual domains suggests that it is a fundamental aspect of intersensory interaction. Perhaps if such integration did not occur, and therefore the "lesser" modality were completely ignored or suppressed, the observer would be unable to respond accurately or quickly enough in those situations in which this modality was the only one available.

Massaro concludes that in visual-auditory speech perception it is the more informative of the two modalities that carries the major weight (e.g., p. 18). This is congruent with a strongly held belief in the area of intersensory bias that the weighting of a sensory cue is positively correlated with its *precision* (i.e., trial-to-trial variability). Thus, for the spatial modalities, the generally accepted descending order of precision – vision, proprioception, audition – conforms to the order of weighting that emerges from studies of intersensory bias (e.g., Pick et al. 1969).

Another well-accepted notion in the area of intersensory bias is that the relative weighting of the two sensory modalities in bimodal perception is determined by the distribution of the observer's *attention* (e.g., Canon 1970; 1971; Warren & Schmitt 1978). Thus, according to the so-called directed attention hypothesis, vision outweighs audition in spatial localization because it is more heavily attended. Massaro and Warner (1977), pp. 66–72, found apparent support for this hypothesis in a study in which subjects were instructed to attend to either one or the other (or both) of the modalities. However, the effects of this attentional manipulation were not nearly as dramatic as the *reversal* of modality weighting frequently found for intersensory bias (e.g., Warren & Schmitt 1978). One likely reason for the relatively weak effects in the Massaro (p. 66) experiment was its failure to use a task for the manipulation of attention that could only be accomplished if subjects did, in fact, direct their attention in the desired manner. Mere instructions to attend have frequently proved ineffectual (Welch & Warren 1980). A task-induced manipulation was in fact used by Massaro (1984a), pp. 216–17, to direct the subjects' attention to the visual stimulus, but the results were negative. This finding does not necessarily reveal a qualitative difference between speech and space perception, however, since previous research in intersensory bias has generally failed to obtain an effect of directing the observers' attention to the already-dominant modality (e.g., Warren & Schmitt 1978), indicative perhaps of a "ceiling effect."

One important concept that Massaro merely alludes to is that for bimodal perception to occur it is necessary for the observer to assign the separate sensory inputs to a *single* object or event (p. 236). Welch and Warren (1980) have argued that the subject's "assumption of unity" is based on how "compelling" the bisensory stimulus is. "Compelling" means that the two sensory inputs have in common a number of characteristics (e.g., motion, spatial location, temporal rate/duration). Massaro's condition of listening to and watching the face of a human speaker is thus very compelling, presumably inducing a strong assumption of unity. However, it should be possible to *weaken* this assumption by various means such as using a very incompatible relationship between lip movements and vocal stimuli, imposing a spatial or temporal conflict between the visual and auditory stimuli, or pairing a man's face with a woman's voice. Congruent with the findings in intersensory bias (e.g., Jack & Thurlow 1973; Jackson 1953; Warren et al. 1981), such reductions in compellingness should lessen the influence of *both* sensory modalities, but especially the normally dominant one. Finally, weakening the unity assumption is likely to make the manipulation of attention more effective (Welch & Warren 1980).

It would be quite incorrect to conclude that speech perception is *identical* to other forms of perception. For example, natural speech is clearly a much more complicated stimulus than is spatial location, as seen in the fact that speech has both spatial and temporal aspects and that some phonemes (e.g., /ba/) are more visually salient than others (e.g., /da/). As Massaro notes, "The degree of influence of a given modality depends on what property subjects report" (p. 84). Whereas for spatial perception vision normally dominates audition, exactly the reverse is true for the perception of temporal rate (e.g., Welch et al. 1986). This difference is most probably attributable to the fact that vision is better suited for spatial perception, whereas audition is more appropriate for temporal perception (Freides 1974; O'Connor & Hermelin 1972). It may be proposed, then, that for those aspects of speech perception that are subserved primarily by temporal perception, audition will prove the dominating modality.

Massaro has done the psychological community a great service by bringing speech perception into the mainstream of perception. How welcome it is when the conceptualization of a perceptual phenomenon is made simpler rather than more complicated!

Strong inferences about development

Friedrich Wilkening

Department of Psychology, University of Frankfurt, D-6000 Frankfurt am Main, Federal Republic of Germany

Electronic mail: *wilkening@vax1.rz.uni.frankfurt.dbp.de*

When experimental psychologists extend their paradigms to study how the behavior in question changes across development, they often find things that are at variance with mainstream beliefs held in developmental psychology. Massaro's book is very illuminating as to why this may be so. One of its many merits is that the author, who started out from experimental psychology, did not stick to adult perception but has made an attempt to apply his framework to phenomena of development and aging. I would like to direct my comments to these aspects.

Two controversial points are of particular interest. First, Massaro reports that children as young as three years old evaluate different sources of information, such as auditory and visual cues in speech perception, independently from each other. Second, he concludes that the stimulus variables are perceived continuously, not categorically, from that early age on.

These claims about nondevelopment contrast sharply with views that prevail in the field of developmental psychology. As to Massaro's independence assumption, there is the long-standing idea that perceptual development proceeds from holistic to analytic processing: Children before school age normally perceive the stimulus event as a whole and have no access to its dimensional sources (e.g., Smith & Kemler 1977). Massaro's claim about continuous perception contrasts with the popular developmental idea that young children partition stimulus dimensions into a few categories or even dichotomize them, a view that has its origin in Piagetian theory and is reflected in current research methodologies on cognitive development (e.g., Siegler 1986).

I believe that Massaro has strong arguments for his position, and I agree with his major conclusions. The power of his arguments rests on his strong-inference strategy, testing not only *his* favorite hypotheses but also alternatives that may explain the same data. This has rarely been done by those who favor the holistic-to-analytic or the categorical-to-continuous view of perceptual and cognitive development.

However, the evidence for Massaro's inferences about development comes from a relatively narrow domain, that of speech perception. There is now more general evidence for his position perhaps the converging evidence he has called for. To supplement Massaro's own account, I would like to point to some new material, most of which has became available after the book was written.

Concerning the traditional idea that young children are holistic perceivers, it is now becoming apparent that this claim was based on weak inference. Even the strongest proponents of this view have conceded that young children can, at least under certain task conditions, perceive stimulus events in terms of their independent attributes (e.g., Kemler 1983). Most recently, Smith (1989) seems to have given up the whole idea. In her new model of perceptual classification, she starts with the assumption "that objects are represented in terms of constituent dimensions and that the representation of objects changes little with development" (p. 125). This now sounds very similar to what Massaro says.

The question of interest here is why, for over a decade of extensive research, holistic processing was inferred from one particular behavior, namely, a child deciding to group together two multidimensional objects that differed slightly on all dimensions, even when there were two objects in the test set that were *identical on one dimension but differed greatly on another one.* This preference in young children's classification behavior was seen as a reflection of their inability (or disinclination) to perform a dimensional analysis. The alternative possibility that the same classification can result from an analytic perception has not been considered in a large series of studies. Only recently, Smith (1989) has recognized this in her new, immensely modified model of perceptual classification.

But even if a child attends selectively to only one dimension in the classification paradigm, the aggregated data of that child can meet the same criteria that have been used for inferring "holistic" processing. This has been shown both theoretically and empirically by Wilkening and Lange (1989). They used a task that followed a logic similar to Massaro's paradigm and found ample evidence of young children's analytic processing of size and brightness. It also deserves mention that, in a recent experiment, Thompson and Massaro (1989) arrived at basically the same conclusion: that size and brightness are initially registered separately, and then integrated.

Analytic processing in young children does not seem to be restricted to "separable" dimensions such as size and brightness or the auditory and visual cues involved in speech perception. Ward and Vela (1986) found that young children made more analytic responses than adults in classifying colors that varied in hue and brightness – dimensions that have been said to be "integral," that is, normally nonanalyzable for adults. And in an investigation of learning family-resemblance concepts, Ward and Scott (1987) showed that holistic modes were not used more often by young children than by adults. All these data converge on the same conclusion and thus support Massaro's assumption of process similarity across the life span, as far as perceptual independence is concerned.

Massaro's other assumption, that of continuous perception and its stability over age, seems to be equally well supported, especially by recent studies on children's intuitive physics. For example, Wilkening et al. (1987) have shown that 5-year-old children not only perceive but also can conceptualize time as a continuous dimension. Further examples are given by Anderson and Wilkening (in press). All these studies provide ample evidence of young children's function knowledge, in which one variable depends on another in a typically continuous way. The traditional approaches to studying cognitive development, particularly those that rely on choice tasks, lack a natural representation for function knowledge and metric concepts. Accordingly, those approaches have not found that children in the so-called preoperational years can grasp physical dimensions as continua (Wilkening 1988).

The research framework underlying the recent studies on children's intuitive physics is similar to Massaro's approach in that it is based on ideas of information integration theory and attempts to use the strategy of strong inference. Massaro has shown convincingly that this paradigm is promising for inquiry in developmental psychology. Nevertheless, I agree with his conclusion that his developmental work "has only scratched the surface of a rich mine of interesting research" (p. 235) in that field.

Author's Response

The logic of the fuzzy logical model of perception

Dominic W. Massaro

Program in Experimental Psychology, University of California, Santa Cruz, CA 95064

Electronic mail: *psych36@ucscd.ucsc.edu*

The commentators have made a variety of important points, ranging from issues central to the foundations of experimental psychology to descriptions of how speech is perceived by eye and ear. My Response to these points follows the same organization as the Précis. The bulk of the reply will necessarily address points of disagreement rather than points of agreement, even though the latter appear to outweigh the former.

Modularity of mind. Modularity is only one of several important issues addressed in *Paradigm* and, although the book began and ended with this issue, modularity was *not* meant to be the principal theory under test. There is agreement among several commentators who dispute my claims against modularity. **Campbell, Cutler, Glucksberg, de Gelder & Vroomen, Krueger,** and **MacKay** view the empirical research and the theoretical framework as generally consistent with the notion of modularity. As I have already noted elsewhere (Massaro 1986c), Fodor's modularity encapsulated much of what has already been accepted in psychological research over the last century. I wrote that Fodor's thesis "represents a cumulative statement of the state of the art integrated into a parsimonious and coherent framework" (Massaro 1986c, p. 435). Donders, Sternberg, and the theory of signal detectability are intellectual ancestors of the modularity principle. Behavior reflects the contribution of multiple systems (e.g., modules or stages), and psychological science advances when these components are isolated and assessed independently of one another. Modularity represents a specific thesis within the information-processing framework, one that can be empirically tested in the same way as other theses, such as connectionism (Massaro 1986a; 1986b; 1988b) can. Even if specific assumptions of modularity and connectionism are falsified, however, we should not be surprised if some

Table 1. *Section headings in précis and author's response indicating the questions and commentators specifically addressed*

Modularity of mind
- Modularity of FLMP? (**Campbell, Cutler, Glucksberg, de Gelder & Vroomen, Krueger, MacKay, Studdert-Kennedy, Welch**)
- Consistency of results with others? (**Anderson, Cowan, Crowder, Kanevsky, Krueger, Pastore et al., Welch, Wilkening**)

Research Framework
- Value of falsification? (**Crowder, MacKay**)
- Falsifiability of FLMP? (**Crowder, Kanevsky, MacKay, Warren**)
- Sufficient specification of FLMP? (**Pastore et al.**)
- External validity of findings? (**Campbell, Glucksberg, MacKay**)

Psychophysical and information-processing methods
- Value of information processing? (**MacKay**)
- Meaningfulness of stage analysis? (**Ehret, Townsend**)
- Information used in speech perception? (**Pastore et al., Studdert-Kennedy**)

Multiple sources of information
- Word level in speech perception? (**Gósy**)

Continuous perception
- Categorical perception as simply categorization? (**Bornstein, Ehret**)
- Existence of categorical perception? (**Anderson, Bornstein, Cowan, Cutler, de Gelder & Vroomen, Ehret, Glucksberg, Krueger, MacKay, Pastore et al., Studdert-Kennedy**)
- Validity of FLMP and CMP tests? (**Kanevsky, Pastore et al.**)

Independence
- As evidence for modularity? (**Krueger**)
- Of information about voicing? (**Bernstein**)
- Of adjectives in person impression? (**Glucksberg**)

Optimal integration
- Role of prior probabilities in FLMP? (**Gigerenzer**)
- Mathematical equivalence to Bayes theorem? (**Gigerenzer**)
- Explanation of conjunction fallacy? (**Gigerenzer**)
- Predictions for conjunction? (**Kanevsky**)
- Plethora of response alternatives? (**Port**)
- One integration algorithm or many? (**Gigerenzer**)

Development and aging
- Supporting evidence from other domains? (**Bornstein, de Gelder & Vroomen, Wilkening**)
- Originality of this theme? (**Studdert-Kennedy**)

A theoretical framework for pattern recognition
- Explanatory power? (**MacKay, Welch**)
- Quantitative modeling concerns? (**Cowan, Kanevsky, Townsend, Warren**)

Binary oppositions
- Necessarily hierarchical? (**Warren**)

Auditory and visual sources of information
- Integration of tactile and visible speech? (**Kanevsky**)

Integration
- Averaging with differential weighting? (**Anderson**)
- Task specificity of integration algorithm? (**Anderson**)
- Relation among choice rules? (**Anderson**)
- Outcome given bimodal syllables with conflicing information? (**Campbell, de Gelder & Vroomen**)
- Nature of visible information? (**Cowan, Studdert-Kennedy**)
- Explanation of integration? (**Cowan, Studdert-Kennedy**)
- Cross-linguistic differences? (**Gósy**)

Attention, consciousness, and information integration
- Cross-modal Stroop effects? (**Cowan**)
- Gradual learning of language? (**MacKay**)

Generality
- Supporting evidence from other domains? (**Wilkening, Welch**)
- Extension to continuous speech? (**Port**)

Relationship to connectionist models
- Adequacy of TRACE model? (**Campbell, MacKay**)

Conclusion

findings in *Paradigm* remain consistent with other previous findings and theory. The agreement of my findings with those of others is noted by **Anderson, Cowan, Crowder, Kanevsky, Krueger, Pastore et al., Welch,** and **Wilkening.**

My test of modularity depends on a distinction between information and information processing. As I never doubted that the information supporting different input functions would necessarily vary, I assumed that Fodor was talking about information processing. That is, the information processing (and not just the information) involved in speech perception would differ from that involved in object recognition. Fodor (1983) states that "there is a sense in which input systems are ipso facto domain specific." He describes situations in which the information in visual perception must differ from the information in language processing, concluding that "nothing useful follows" (1983, p. 48). Seeing the visual differences between a hamster and a mouse certainly does not involve the same input information as hearing the auditory differences between the syllables /ba/ from /da/. According to Fodor, it is the "idea that there are distinct psychological mechanisms – vertical facilities – corresponding to distinct stimulus domains that's now at issue" (p. 48). Following Fodor's logic, my finding that similar evaluation, integration, and decision operations in two different input domains seems to constitute a disproof of modularity. Furthermore, finding central systems operating in the same way as the input systems appears to undermine Fodor's version of modularity even more (**Krueger**). However, **Cutler** and **Campbell** point out that it is logically possible that two different modules work in identical says. One cannot deny this possibility, although it violates Occam's razor because multiple systems are assumed when one is sufficient.

Cutler appears to have an extremely broad notion of modularity, stating that independence of auditory and visual speech is exactly what is predicted by modularity. Similarly, **de Gelder & Vroomen** take sign language as support for modularity in that the speech/language module cannot be reduced to a sensory module. Fodor (1983, p. 132) viewed the influence of visible speech on speech perception as evidence for cross-modal linkages in phonetic analysis, but not as evidence *for* input systems. Similarly, I do not understand how sign language can constitute support for modularity. It is not even clear whether sign language and speech perception engage the same modules or different ones. As pointed out by **Krueger**, the evidence for independent evaluation of auditory and visible speech seems to go against the existence of a speech/language module, but it leaves intact the notion of sensory/transducer modules. On the other hand, modularity might be viewed as consistent with the FLMP's description of context effects in which top-down and bottom-up sources of information are integrated in a way that leaves the lower-level bottom-up representation intact.

In terms of specific assumptions, the FLMP might appear to have more in common with modularity than with connectionism. **Glucksberg** views the FLMP as positing three properties of modularity: (1) sequential processing, (2) no top-down influences on lower levels, and (3) local symbolic as opposed to distributed associative representations. However, each sequential stage in the FLMP may entail parallel processing as in, for example, the simultaneous feature evaluation of multiple sources of information and in the determination of goodness-of-match with all of the response alternatives at the same time. In addition, the model is clearly neither symbolic nor local because it quantifies continuous information and described units at one level in terms of smaller units at a lower level. Sequential versus parallel processing and local versus distributed implementation are, for now, orthogonal to the empirical issues being addressed by the model.

The FLMP shares assumptions with both modularity and connectionism. The autonomy of the information-processing stages in the FLMP is a formal instantiation of modularity, but there is also a close mathematical relationship between the FLMP and a specific implementation of a connectionist model (Massaro & Cohen 1987; Massaro & Friedman 1989). As clearly stated by Rueckl and Oden (1986), the FLMP can be "taken to represent a synthesis of models from both poles of the interactive/autonomous continuum. It is interactive in proposing that contextual information influences the outcome of the . . . identification process but autonomous in contending that this comes about without any dependence of featural analysis on the results of semantic analysis" (p. 459). The research reported in *Paradigm*, however, challenges several assumptions of specific models in both of these theories. As noted previously, I interpret the discovery of the same information-processing algorithm both in different input systems and in different central systems as evidence against modularity. With respect to connectionism, an important assumption falsified by the research is that of interactive activation.

De Gelder & Vroomen interpret modularity and the FLMP as consistent with one another in viewing speech perception as hypothesis testing and decision making. They go beyond noting this agreement and state that my view implies "that a general model of decision making is enough – that one does not need to go through the trouble of designing a (modular) model of the internal knowledge environment of the organism to understand how decisions on inputs are reached." As also observed by other commentators (e.g., **Welch**), however, my view does not ignore the domain specificity of the information supporting speech perception. Uncovering these sources of information requires an exploration of the terrain of speech perception; studies of person perception, or some other domain will not suffice. In the research reported in *Paradigm,* however, similar information processing is found to occur in disparate domains. Thus, the information for recognizing a cow does not overlap at all with the information for recognizing the spoken word /kaU/, but the information processing (evaluation, integration, and decision) appears to be highly similar. Information may be mostly modular, but information processing is not. Modularity as a research paradigm (see Gardner 1985) surely fails in the sense that psychologists studying widely different domains ask similar questions, carry out analogous experiments, and arrive at related conclusions. The commonality of their paradigm will relate the different domains and should prove to be more productive than the combined efforts of individuals from different disciplines and paradigms working in a single domain.

Krueger believes that we should discard the idea of a speech/language module, but not the idea of sensory transducer modules. This proposal is not unreasonable; it captures the finding of independence of auditory and visual speech at the evaluation stage of processing and it bears some similarity to the minimodularity conclusion reached by Bruno and Cutting (1988). Sensory-transducer modules would be consistent with the domain specificity of inputs; however, input information specifying the relationship between modalities is also available. One example is voicing information provided by the temporal correspondence between the audible consequences of glottal pulsing and the visible consequences of the release of the tongue from its position of closure in the vocal tract (**Bernstein**; see also my discussion of this result in the section *Independence* of this reply).

Studdert-Kennedy holds that speech perception engages specialized cerebral processes. For support, he cites the failure to find an arbitrary acoustic substitute for speech in reading machines for the blind and the cerebral specialization of the left and right hemispheres. However, I do not see the relevance of the first piece of evidence. A simple model based on localizing language processing in the left hemisphere cannot be correct because the right hemisphere appears capable of accomplishing much of language processing (Chiarello 1988). Furthermore, the specific linguistic function traditionally assigned to Broca's area may be a much more general cognitive capacity (Grodzinsky et al. 1985). **Crowder** offers several more testable possibilities: Speech stimuli are special, the locus of brain processing is unique, and/or the information processing is unique. Most important, Crowder sees "speech-is-special" as an empirical issue that can be attacked by the research strategy in *Paradigm.* Having just read a biography of Helen Keller to my children, I wonder if her achievements would have been possible given modularity's narrow view of input versus central systems, or even given speech as special.

Krueger nicely summarizes results of letter and word recognition in reading in order to provide converging evidence for the processes I have uncovered in speech perception. **Welch** gives a concise and well-documented summary of the similarities between the perception of spatial location and speech perception. Both these domains offer a valuable challenge for those interested in maintaining belief in a special modular system for speech perception.

In summary, Fodor's thesis goes beyond Donders and Sternberg in its claim that the stages of processing associated with each input module are unique. In information-processing theory, detection, recognition, and response selection processes might function similarly across different input domains. In modularity, detection, recognition, and response selection processes would differ for different modules. My discovery that the same information-processing algorithm occurs across different input systems is evidence against modularity. In contrast to modularity, the FLMP is constrained to predict similar processes across different input domains. However, the nature of the information will necessarily differ for different domains. Information is modular, but information processing is not.

Research framework. The value of falsification and strong inference is acknowledged by **Crowder,** who regrets that it was not used more often in his own area of inquiry. **MacKay,** on the other hand, believes that falsification has well-known limitations – although he is silent on the strong-inference extension of Popper's research strategy. I have noted elsewhere that the research strategy I advocate is a bit like Winston Churchill's view of democracy: You can always shoot holes in it but it's still the best game in town. I take heart from the apparent strong advocacy of falsification in well-established disciplines such as theoretical physics (Hawking 1988) and behavioral biology (Alcock 1989). Hawking values falsification and views creationism as a competing theory to be tested and falsified in the same manner that other more scientifically acceptable theories are. Alcock (1989, p. 229) stresses that most biologists "especially admire research that tests *multiple* working hypotheses in ways that permit clean discrimination between the competing alternatives.

Crowder, Kanevsky, MacKay, and **Warren** are worried that the FLMP might not be falsifiable. The model has failed empirical tests, however, when either the assumed sources of information are dependently evaluated or when the wrong sources of information are assumed. Massaro and Cohen (1977) found that the duration and intensity of the fundamental frequency were dependent cues to the voicing of fricatives in initial position and that the FLMP gave a poor description of these results. The FLMP also fails empirical test if the consonant/vowel ratio is assumed to be the functional cue to the voicing of medial stops (Derr & Massaro 1980). Thus we do have boundary conditions specifying when the model does and does not hold (Greenwald et al. 1986).

Pastore et al. believe that I have not sufficiently specified how the FLMP and other models are implemented and tested. I sometimes have the same initial reaction when I read about computer simulations. However, there is enough information given in *Paradigm* to allow others to test the models. A more comprehensive tutorial on the testing of mathematical models is given in Massaro (1989, pp. 209, 244–49). To make the model more accessible to the scientific community, we have also collaborated with other scientists to test the models against their results (*Paradigm,* pp. 34–36, 52–54; Massaro & Cohen 1983).

My impression is that remarks about external validity grow proportionately with the strength of someone's experimental findings. (Note that I am using the term *external* rather than *ecological* to describe how well an experimental situation conforms to a comparable real-world situation. I reserve the latter term for Brunswik's (1955) use, to describe how well some environmental variable predicts some particular state of affairs, e.g., how well height in the visual field predicts distance.)

External validity is an issue for **Campbell, Glucksberg,** and **MacKay** because they believe that certain of the characteristics of the experiments preclude much generalization of the results.

According to **Campbell,** my speech perception task "forces one to make categorical decisions on stepped 'unnatural' stimuli." Are my stimuli unnatural because they are synthesized or because many of them are constructed to fall between ideal tokens of the categories being tested? The stimuli are not unnatural simply because they are synthetic, rather than natural speech. We have shown that the synthetic endpoint stimuli can be identified as reliably as (and even more reliably than) natural tokens (*Paradigm,* p. 222, Figure 7). The stimuli might also be considered unnatural because the intermediate test syllables supposedly do not occur in natural speech. Given the large influences of coarticulation, rate of speaking, sex, age, and affect of the speaker on the speech signal, however, natural syllables could easily have the variation that we have in our studies (Perkell & Klatt 1986). Perhaps the most convincing answer to **Campbell**'s concerns is simply that the FLMP describes experiments using natural speech as well as those using synthetic speech (*Paradigm,* Chapter 2, section 4.2).

Campbell believes that visible speech enhances the percept "only when seen faces combine with less ambiguous heard speech." She must mean "more" rather than "less" because it is the least ambiguous source that has the biggest effect on the judgment.

For **Glucksberg,** there was no real person being perceived in the person impression task and, thus, the results might simply reflect task demands (whatever those are). Once again, external validity is an empirical question. Some effort has been made to establish the external validity of the person impression task. For example, Anderson found that the results generalized to the judgments of United States presidents about statesmanship, given paragraphs of biographical information (N. H. Anderson 1981).

For **MacKay,** the instructions in the sentence processing task could have influenced the information value for animacy, and the German sentences did not have morphological markers that distinguish nominative and accusative case (MacWhinney et al. 1984). The influence of both instructions and morphological markers in German is an empirical matter, however, and it is difficult to imagine that the theoretical framework will crumble when such questions are answered. MacKay is also bothered by the fact that many of the test items contained "nonsentences" such as "The fence the cow kicks," and concludes that "nonsentences cannot in principle directly illuminate sentence processing." MacKay seems to have forgotten his physics analogy on this one because, for example, an object falling in a vacuum (an unnatural event on earth) does illuminate objects falling in the real world. Concerns about external validity are not common in physics. Could you imagine the plight of physicists today if their requests for cyclotrons were rejected because a cyclotron had no external validity (Giere 1988). For physicists, the control of the observational situation and the extrapolation of the results are more central than external validity.

In some respects, worries about external validity overestimate the predictive power of theory – even reasonably correct theory. We can expect predictability in the appropriate laboratory situations, but not always for naturally occurring events. Theories can best be tested when the complexity of everyday life is simplified, controlled, and measured (Manicas & Secord 1983). If our experiment is simply equivalent to real life, we may gain little. Experimentation permits us to perturb real life, to observe the consequences, and to test theories. More generally, however, I would hope that the decade since Neisser's (1976) call for external validity has brought the field back to some reasonable stance on this issue. We want theories to inform us about both our experiments and the real world: to test theories, experiments must necessarily differ from the natural world.

Two issues appear to be central to the research framework: falsifiability and external validity. Falsifiability of FLMP is possible. For example, finding nonindependence between top-down and bottom-up sources would be critical evidence against it. Rather than asking how the FLMP can be falsified, however, a more appropriate research strategy is to establish what the boundary conditions really are. Given correct definition of the sources of information, the FLMP must describe how they are processed. Incorrect specification of the sources should lead to a poor description the model. Falsification would occur when the FLMP predicts poorly with appropriate specification of the sources of information. The discovery of the sources of information, on the other hand, is an empirical matter. The FLMP is neutral with respect to the issue of which ecological characteristics are functionally valid.

An important issue is how FLMP deals with more naturalistic, real-life phenomena. Methodologically, laboratory experiments must address the critical characteristics of the phenomenon of interest. An experimental task involving multiple sources of information reflects what people have available in the real world more accurately. Thus, the theory developed from this laboratory situation is more easily extended to natural situations. Earlier speech-perception experiments fell short of this goal because manipulating only one cue in an experiment could not address the important issue of how perceivers use multiple cues in everyday speech perception. Our view of everyday speech perception becomes more realistic with the finding that perceivers actually use multiple cues when they are available in the laboratory.

Can we determine whether or not the FLMP applies to how people perceive speech in the real world? I believe we can, although the critical tests of the model must be in the laboratory. The description of how the FLMP can be extended to deal with continuously arriving speech illustrates how the model can be scaled up and applied to more complex situations (see section titled "Generality").

Psychophysical and information-processing methods. Contrary to what might be implied by **MacKay**'s critique of information processing, I would never justify this approach on the basis of the number of experiments it has generated. MacKay's stage analysis of ballistics is puzzling because his stages already appear to be grounded in physical laws. In some respects, a stage analysis must have been used to arrive at the physical laws relevant to ballistics. Given that a good stage analysis requires equations for each stage, the physicist would soon discover that the descent stage was the mirror image of the ascent stage – an important constraint that would hasten the formulation of explanatory theoretical concepts.

Contrary to the position of **Ehret**, who believes that a stage analysis is not meaningful until we know the underlying neuronal mechanisms, a stage analysis is essential to behavioral research. As observed by **Townsend**, the information-processing framework has much to offer. It illuminates and organizes a broad range of apparently different findings within a single framework (Massaro 1989; in press). Ehret correctly points out that there are continuous and categorical phenomena to account for; the problem that has not been solved by the neurophysiologist is how continuous activity in the nervous system gets mapped into categorical behavior. Recurrent networks, efferent feedback, and physiological thresholds remain untested possibilities.

In our task, subjects identify speech stimuli modified in systematic ways; their responses are used to test quantitative models of the identification process. **Pastore et al.** are concerned with what features are actually being used (the *what* of **Studdert-Kennedy**). This question can be answered by using our method, however. An important distinction must be made between the stimulus characteristics that are manipulated in the experiment and the features that the perceivers actually use in their identification. Patterns can be described by an almost unlimited number of characteristics or properties (Palmer 1978), only a small set of which will be used. Thus, manipulating a particular characteristic does not insure that it is a feature actually used in pattern recognition (Cheng & Pachella 1984; Sattath & Tversky 1987). Which characteristics are features remains a psychological question to be answered.

The methods in *Paradigm* allow comparative testing among models that assume different sources of information are being used. An experiment by Derr and Massaro (1980) and the analyses by Massaro and Cohen (1983) provide an informative example of how the functional utility of different properties of speech can be assessed. In speech perception, a voicing distinction allows us to perceive a difference between the verb in the phrase *to use* and the noun in the phrase *the use*. Speech scientists traditionally believed that consonant duration relative to vowel duration was the critical cue to voicing judgments (Denes 1955; Port & Dalby 1982). However, Massaro and Cohen (1977; 1983) showed that this cue (called C/V ratio) is invalid when the results are tested against mathematical models. A model based on C/V ratio gives a much poorer description of the results than one based on independent consonant and vowel duration cues (Derr & Massaro 1980). Our research paradigm was also successful in determining whether F_0 onset frequency, F_0 contour, or both, influence the voicing of consonants in initial position in words (Massaro & Cohen 1976; 1977). These studies should be reassuring to **Crowder, Kanevsky, MacKay,** and **Warren,** who are concerned that the FLMP may not be falsifiable. Derr and Massaro (1980) showed that if the wrong sources of information are assumed, the model fails empirically. Thus, the research strategy developed in *Paradigm* not only addresses how different sources of information are evaluated and integrated, it can uncover what sources of information are actually used. The research paradigm confronts both the important psychophysical question of the nature of information and the process question of how the information is transformed and mapped into behavior.

In reply to **Studdert-Kennedy:** I treat all sources of information as equivalent because the research enterprise is indifferent to *what* we perceive. As I have just discussed, uncovering the information sources supporting our perception of speech is an empirical question. Experimental research and mathematical models are better tools for this purpose than is biology. One gets the impression from Studdert-Kennedy that Darwin (and the many evolutionists before and since) have solved the problem of speech perception – our charge is simply to reiterate their doctrine. I can only repeat the distinction that I have made between information and information processing – both ingredients are essential for a full account of speech perception. It is only natural that a research paradigm might give more emphasis to one than the other.

Multiple sources of information. The multiple-source theme in *Paradigm* has been commonplace in speech research almost since its inception. **Studdert-Kennedy** says. His impression must certainly be based on a generous interpretation of early thinking in language research rather than on the literature itself because most studies of speech have concentrated on manipulating a single source of information (Massaro 1987a; Massaro & Cohen 1976). It is true that no one would argue against the idea of multiple cues in speech perception. However, as has already been argued by Massaro and Cohen (1976), appropriate methods were not used by speech scientists to determine how multiple sources of information were processed. The few studies that manipulated more than a single cue could not address any questions about evaluation, integration, or decision. The completely erroneous conclusion reached by Denes (1955) is an example of how little the actual results can constrain their interpretation if the investigator does not use the appropriate quantitative techniques. At the theoretical level, there was no real concern with how multiple sources of information were evaluated and integrated. Miller (1962), for example, saw the contribution of grammatical context to speech intelligibility as evidence for larger decision units (e.g., phrases) than normally assumed; and not in terms of multiple sources of information simultaneously influencing decisions about smaller units (e.g., phonemes). Contrary to the position of Studdert-Kennedy, then, previous research has *not* anticipated *Paradigm* with respect to the issue of *how* multiple sources of information support speech perception (Hirst 1988). More important, no one seems to have shown how a speech module would handle multiple sources of information bombarding it more adequately than the general perceptual process described by the FLMP.

A theoretical framework for pattern recognition. Although **MacKay** is dissatisfied with how *Paradigm* asks and answers questions, he does not offer any alternative. He suggests that my interpretation of the finding that vision often appears to dominate audition when the respective inputs are made to conflict is just a vague description or empirical correlation rather than a theoretical explanation. My explanation is based on the psychophysical fact that the resolution of visual space is better than the resolution of auditory space, which entails different degrees of ambiguity at the evaluation stage of processing. **Welch** offers this same explanation of the apparent dominance of vision over audition in the perception of spatial localization. Thus, there is no need to assume that there is an attentional bias toward vision relative to audition, as does Posner (1978). These two alternative explanations could be tested, and one or both might be falsified. For example, creating a situation in which auditory resolution is better than visual should lead to auditory dominance if the resolution explanation is correct but not if the attentional bias is correct (see Welch's commentary). What more can we ask for? MacKay's (1987) own theorizing, on the other hand, appears to fall short of these goals. It does not address information either; MacKay simply takes linguistic constructs and puts them in the performer's head.

MacKay's dissatisfaction with stage models because they conceal our ignorance misses one important point. When stages are articulated in quantitative models, each is specified by an autonomous function or equation. A stage analysis helps determine what is known and unknown. On the one hand, the good fit of the FLMP supports the assumption of a multiplicative integration process. On the other hand, requiring a free parameter in a model to represent the /ba/-ness of a particular auditory syllable admits that the psychophysics and learning history of this stimulus are not completely known. MacKay's stage model of a ballistics task would be fruitful and could very well lead to the discovery of physical concepts such as mass and inertia.

Quantitative modeling is central to *Paradigm,* and several commentators raised important questions. **Cowan** is concerned that the FLMP (and the other models under consideration for that matter) can predict nonmonotonic functions across a speech continuum. The reason for the possibility of nonmonotonic functions is that the model does not constrain the continuum to be monotonic. Recall that the FLMP and the alternative models are neutral with respect to how /ba/-like a given level of the stimulus continuum will be for a given subject. Usually our speech continua are monotonic but it should not be surprising if we overshoot the prototypical values for an endpoint stimulus – leading to a nonmonotonic function. In most cases, the theoretician can save free parameters by assuming a specific monotonic function across a given continuum (see Oden & Massaro 1978 for an example of this analysis).

However, the point here is to test the nature of the integration rather than to determine the psychophysical relationship exactly. We would have to do this for all models and little would be gained with respect to distinguishing among different models.

Cowan also interpreted an equation of the FLMP (*Paradigm*, p. 202, Equation 16) as the blurring of two stages: integration and decision. Although expressed in a single equation, the separate terms have to be computed first before the equation is solved. That is, it is first necessary to integrate the sources of information with respect to each prototype, and only then can a decision based on the relative goodness-of-match be made. Computing the goodness-of-match for each prototype is necessarily different from and logically prior to computing the relative goodness-of-match.

Kanevsky points out that the FLMP and the categorical model of perception (CMP) equations differ in complexity. The former is nonlinear whereas the latter is linear. Sometimes linearity is a desirable property and sometimes it is not. There is no a priori reason to expect that one is more intuitively plausible than the other to describe performance in our tasks. A case might be made for a nonlinear model in discrete identification tasks, whereas a linear model might be more reasonable in rating tasks (**Anderson**). The fact that the FLMP is superior to the CMP in both types of speech perception tasks appears to argue against a complexity explanation of the differences in goodness of fit. Taking Kanevsky's advice, it might be worthwhile to increase the complexity of the models under study to determine whether their descriptions of the results can be improved. In all cases, however, the proposed mathematics should be guided by well-motivated processes (Massaro & Friedman 1989). Contrary to Kanevsky's suggestion, increasing the complexity of the models should not necessarily lead to a situation in which the most complex and refined model always wins. We have specific examples in which increasing the complexity of the models does not make the most complex model win. As an example, the TRACE model is more complex than the FLMP but gives a poorer description of the integration of top-down and bottom-up information in speech perception. **Townsend**, however, correctly describes the difficulty of testing models and the years of difficult work involved in eliminating alternatives and converging on a robust solution.

Warren cautions us not to quantify unnecessarily, but it is a fact of life that quantification is necessary, even to distinguish between qualitative hypotheses such as categorical versus continuous perception (see **Townsend**). Investigators are usually inclined to interpret observed differences as qualitative when in fact quantification reveals that they are best interpreted as simply quantitative. A good example comes from theories of perceptual development (see **Wilkening**). Warren is also worried about the number of free parameters used in the application of the models to the results and seems to imply that their number is arbitrary. However, the number of necessary parameters usually grows with the number of independent points being predicted. In model testing, it is the relative number of parameters that is important, not the absolute number. What is crucial, however, is that the models should be falsifiable and should make different predictions. One way to show this is to generate predictions from one model and fit the predictions with those from another model. Massaro and Friedman (1989) have shown that the FLMP and the CMP make different predictions from one another and that the predictions of one model cannot be described by the other even though the standard number of free parameters is used to optimize the description.

Townsend wonders why I did not use statistical tests such as chi-square – as opposed to least squares. The latter were used because these specify directly the correspondence between a model and data or the correspondence between the predictions of two models. That is, a least-squares value of .05 means that the observations and predictions are within roughly .05 of one another on the average. Other measures of goodness-of-fit, including chi-square, require knowledge about the actual frequencies in each cell. With large enough frequency, we know that any model – no matter how good the fit to data – can be rejected. In fact, having a least-squares measure of the fit of each model to each subject's results permits statistical tests to be carried out (*Paradigm*, pp. 123–24).

Binary oppositions. Although **Warren** is concerned about being led down the garden path because of the apparent hierarchical structure of the binary oppositions, I already warned that they were not strictly hierarchical. For example, in *Paradigm* (p. 23), we ask whether sources of information are continuous or categorical whether or not they are evaluated independently or dependently.

Auditory and visual sources of information. Visible information is viewed by **Cowan** as top-down in deciding what was heard. Even if visible speech has the same effect as lexical context in phonemic restoration, visible speech might be more reasonably viewed as bottom-up. For **Studdert-Kennedy**, on the other hand, the process of integrating auditory and visual information is qualitatively different from integrating top-down sources, since "acoustic and optic information are structurally equivalent because they have a common origin in the speaker's articulation." This statement would have been more convincing if it included some empirical evidence. I do not understand Studdert-Kennedy's claim that Summerfield's (1979) study supports this thesis; it can be equally well used to support many other opposing theses. Even Summerfield (1987) does not claim that his results support one thesis over others. We find similar processing of bottom-up and top-down domains – making the thesis of qualitative differences between these domains difficult to defend. Perhaps the issue can be attacked empirically, but what would it take to convince Studdert-Kennedy? The contrast between Cowan's and Studdert-Kennedy's views on this issue points to a fuzzy boundary between bottom-up and top-down information.

On the other hand, I have no argument with **Studdert-Kennedy**'s claim for the primacy of the physical sources of information for the infant learning to speak. Analogously, my research has documented the primacy of bottom-up relative to top-down sources of information in experienced language perceivers. Top-down effects (and to some extent the effects of visible speech) are large only to the extent that the bottom-up sources are ambiguous (*Paradigm*, Chapter 8, section 4.4). Related to this issue, **Cowan** incorrectly describes the phoneme restoration effect as occurring when there are only contextual cues to guide perception. Bottom-up evidence is in fact also necessary because phoneme restoration does not occur when the missing phoneme is replaced by silence. Replacing the phoneme /s/ in *legislature* with noise (Warren 1970) provides some bottom-up support for this phoneme, in addition to the top-down support of the word context.

Krueger seems to have missed my conclusion (*Paradigm*, p. 61) that the neutral articulation condition was actually more /da/-like than /ba/-like, due in part to the fact that the speaker posed with his mouth slightly open during the auditory-alone condition.

To handle **Port**'s concern that every potentially relevant response alternative has to be considered, some assumption has to be made about alternatives being activated in bottom-up fashion. That is, feature evaluation would activate only a limited number of prototypes and only those prototypes would enter into the feature integration process. Hence, fire-engines and woodpeckers would not enter into classifying a bimodal /ba/ because they would not be activated.

Gosy is concerned about some apparent differences between English and Hungarian bimodal speech perception. These differences should not be unexpected, and can be explained along lines similar to my account of cross-linguistic differences in sentence interpretation (*Paradigm*, Chapter 9, section 3). I also noted that the observed differences in bimodal speech perception in German and English speakers are to be expected, given the speech categories in those languages (*Paradigm*, p. 48). I trust that a similar explanation can account for the differences in Hungarian and English.

Kanevsky discusses evidence for a similar integration of tactile and visible speech. This result seems problematic for the thesis of specialized processes in speech perception because, for example, it conflicts with **Studdert-Kennedy**'s notion that the sounds and sights of speech are special. The learning effects that Kanevsky observed in the tactile domain are to be expected from perceptual learning theories, such as Brunswik's or J. J. Gibson's (1979). According to both theorists, ecologically valid "cues" predictive of environmentally important events can be learned and used. We found that performance on each single input modality together could account for performance on bimodal speech. Although training should improve performance, the FLMP predicts that the performance relationship in the three testing conditions should remain the same across different levels of expertise.

Integration. The extent to which the FLMP can be claimed to be superior to averaging with differential weighting is questioned by **Anderson.** There are two types of differential weighting. If one source of information (such as auditory speech) receives a different weight from another (such as visual speech), then this type of weighted averaging can be rejected (*Paradigm*, Chapter 7). If, on the other hand, differential weighting is based on the unique level of a source of information (such as the second level of a visual /ba/, then weighted averaging will give a good fit. This type of weighted averaging, however, takes twice the number of free parameters required by the FLMP (which should not be welcomed by **Pastore et al., Kanevsky,** or **Warren**). A weight and scale value must be estimated for each level of each source of information in this type of weighted averaging model whereas only a scale value need be estimated in the FLMP. More generally, the FLMP can be seen as providing optimal weights integral to the scale values. It makes the strong assumption (perhaps falsifiable in some contexts) that a distinction between scale value and weight is not necessary. Similarly, the FLMP has not needed any additional process to handle response bias (**Townsend**).

Anderson is right in observing that number averaging is outside the domain of the FLMP; subjects asked to average numbers should do so, and not integrate them as multiple sources of information specifying a category. However, I am not willing to accept Anderson's proposed taxonomy, in which he distinguishes between two general classes of tasks defined in terms of continuous-rating versus discrete-choice responses. He seems to imply that these tasks should engage different integration processes simply because of the response requirements (however, see N. H. Anderson 1974). The two classes of tasks do differ with respect to what is required from the decision stage, but this does not mean that the earlier evaluation and integration stages should differ. The beauty of the stage analysis is that changing one stage does not necessitate changes in other stages. Although we currently disagree on a few specific issues, it is important for me to acknowledge that *Paradigm* owes much of its success to the research strategy and techniques developed by Anderson and his students.

Anderson correctly points out that his ratio model (Leon & Anderson 1974) of decision is both similar in form to and more general than Luce's choice rule. A relative goodness rule of this type has remarkable robustness as well as a close mathematical relationship to a criterion rule of signal detection theory (Massaro & Friedman 1989; **Townsend**).

Campbell describes differences in the perception of two "mirror-image" bimodal syllables: auditory /ba/-visual /ga/ and auditory /ga/-visual /ba/. She asks how the underlying FLMP integration process is different (**de Gelder & Vroomen** pose a similar question). Campbell speculates that the FLMP predicts no difference in the time course of achieving the percept for these two bimodal syllables. I do not understand how she could arrive at this speculation unless she implicitly assumed that the auditory /ba/-ness would be equivalent to the visual /ba/-ness (and similarly for /ga/). In fact, I had already given a detailed description of the perceived differences in two bimodal syllables made up in this fashion (*Paradigm*, pp. 60–62). The first bimodal syllable is often perceived as /da/ because auditory /ba/ is similar to auditory /da/ and visual /ga/ is similar, if not identical, to visual /da/. The second bimodal syllable is often perceived as /bga/ because both auditory /ga/ and visual /ba/ are somewhat similar to /bga/. (Thus, de Gelder & Vroomen are mistaken when they say that "there is no visual influence in the /da/ part of /bda/.") Visual /ba/ is consistent with the alternative /bda/. That is, neither source gives a perfect match but the two mediocre matches combine to give a good fit. Other alternatives, such as /ba/ or /ga/, give a good match on one dimension and a poor match on the other – giving an overall poor fit. In addition, visual /ba/ appears to be evaluated sooner than auditory /ga/, which also contributes support for the percept /bga/ (see *Paradigm*, p. 62). Contrary to Campbell's speculation, the FLMP can predict differences in reliability and speed of identifying these two syllables.

Attention, consciousness, and information integration. His cross-modal Strop results – a nonconcordant spoken color word interfering with color naming – are seen by **Cowan** as analogous to the integration of auditory and visual information in speech perception. There is not sufficient evidence, however, that this interference reflects integration. Subjects given conflicting auditory and visual speech, for example, often report perceiving a syllable different from what they report for the corresponding unimodal syllables. Subjects never make the analogous "mistake" in the Stroop task by, for example, naming a red color purple when they hear the word blue. Bimodal speech perception appears to reflect perceptual integration whereas Stroop interference appears to reflect response competition (Massaro, in press).

On one hand, **Gosy** appears to criticize the idea that speech perception subsumes word recognition and, on the other, she describes results of her own that support it. When instructed to identify auditory and visual speech as nonsense words, subjects tended to give word responses. This result agrees with my hypothesis that perceivers naturally impose a meaningful interpretation at the highest level possible.

Continuous perception. A serious theory should define perception – which is not easy to do. However, the next best alternative might be to describe the processes subsumed by exposure to a speech stimulus and some behavior. We have learned that phenomenal experience can be misleading (for a summary, see Massaro 1989, pp. 29–31), so we should focus on behavior. As I noted in *Paradigm,* the task of the language perceiver is to categorize the input but this action cannot be considered to be identical to categorical perception. If some scientists, such as **Ehret** and **Bornstein,** want to define categorization of the input as categorical perception, then they have bypassed several important phenomena. An analogy might come from a somewhat Gibsonian treatment of visual perception in which the perceptual scientist begins with the perceiver's visual world rather than the processes leading up to it. When I say that perception is continuous (and **Krueger** may be right in his opinion that "fuzzy" in the FLMP should be replaced with "continuous"), I mean that the system doing the perceiving has access to continuous information. When **Anderson** writes that perceptual information from multiple continuous sources "is integrated to produce a (sometimes) discrete percept," it can only mean, in my view, that integration usually further disambiguates the input because it makes the continuous information more informative with respect to the category intended by the speaker. The integration of several consistent, but relatively ambiguous, sources of information presents fairly conclusive evidence.

The decision stage maps this continuous information into a categorical bet made observable by an identification decision or some other phenomenal report on the part of the perceiver. It is possible, however, that we experience and observe discrete behavioral events when, in fact, the underlying neuronal representation remains continuous. Each stage or operation need not erase or modify the representation made available by the previous stage; the later stages only generate additional representations (Young 1987). It is this representation of information in my theory that is clearly modular, as noted by **Cutler** and **Glucksberg.** Thus, perceivers can reliably rate the degree to which the speech input supports a speech category; they can also identify it categorically, or even use the separate evaluations of auditory and visual speech to discriminate two different speech inputs that give the same support for a speech category (*Paradigm,* Chapter 5).

Categorical perception might be a dead horse, as claimed by several of the commentators, but it is a horse that remains alive in the race of recent research and theory. **Mackay, Warren, Cowan,** and **Ehret** view the phenomenon as one to be accounted for. **Bornstein** writes of "the *fact* that some dimensions or domains of experience are perceived categorically" (my emphasis). **Pastore et al.** observe that categorical perception was rejected by many before my book, but they also note that categorical perception is usually assumed in contemporary studies of trading relations and duplex perception. Advocates of TRACE found it necessary to illustrate how the model produces categorical perception.

Studdert-Kennedy accuses *Paradigm* of a selective and self-serving treatment of previous research. In his view, Summerfield (1979) first demonstrated the continuous nature of auditory and visible speech in bimodal speech perception. However, Summerfield's experiment and data are indeterminate; they are equally consistent with the categorical models described in Chapters 4 and 5 of *Paradigm.* Finer-grained experiments and more refined data analyses than those given by Summerfield (1979) are needed to reject categorical models.

With respect to the general issue of categorical perception, **Studdert-Kennedy** claims that it had already been well-established and a commonplace of the literature for nearly 20 years that "categorical phonetic percepts derive from a decision process based on continuous auditory information." Furthermore, he states that *Paradigm* gives no credit to this previous work. I can easily answer both of his claims. First, there was no hint in Studdert-Kennedy et al.' (1970) seminal paper that categorical perception is wrong. Although Studdert-Kennedy may have disassociated himself from his coworkers at Haskins Laboratories, there has been a steady stream of "categorical perception" findings from them over the last two decades – even though it is sometimes found in different guises, such as "perceptual equivalence" (*Paradigm,* Chapter 4, sections 3–6).

With respect to the issue of giving credit where credit is due, my students and I had already favorably reviewed the research that discredited categorical perception soon after this work was completed (Massaro 1975a; 1975b). All the work I supposedly neglected is exhaustively covered in these textbooks. The coverage in these books and other publications (Massaro 1976) can be read to check whether I have studied the relevant literature. I have studied it, do cite it extensively, and certainly have *not* repeated it. I am sure that **Crowder** at least was relieved not to see this coverage repeated once again in *Paradigm.* On the one hand, **Studdert-Kennedy** chides me for not reviewing the early work, and on the other, he complains that too much of *Paradigm* is directed at the issue of categorical perception. What seems to have been missed by Studdert-Kennedy and others, however, is that continuous perception of information sources makes speech perception a more difficult phenomenon to describe adequately – or at least one that is more involved than would be expected from the easy solution given by categorical perception.

Kanevsky observes that our tests of categorical and continuous models of rating judgments could have benefited from larger sample sizes. We succeeded in refuting one model relative to the other, but future studies might want to use a larger number of observations. Accepting the finding that the integration of multiple information sources occurs within a modality as well as between modalities, **de Gelder & Vroomen** view the question of categorical perception within a modality as the real issue. But the evidence also supports continuous perception within a modality (*Paradigm*, Chapter 4). I agree with their observation, however, that there are probably modality-specific codes supporting memory in the suffix paradigm. This result is analogous to our finding that subjects can discriminate two different bimodal syllables even though the syllables are identified equivalently (*Paradigm*, Chapter 4, sections 4–6).

Pastore et al. are mistaken in their judgment that the implementation of the FLMP and CMP is inconsistent with the theoretical assumptions of these models. What they miss is that a given mathematical entity, such as .7 or a feature value a_i, can be given different meanings. In the CMP, it refers to the probability of some event, such as perceiving the category /ba/. In the FLMP it means the degree to which a given category is supported. The possibility of two referents for the same quantitative value is one reason why it is not so easy to test between continuous and categorical models such as the FLMP and the CMP. Thus, for example, a subject's mean rating judgments of speech stimuli are not sufficient to reject one model over another, but the distribution of these judgments might be (*Paradigm*, Chapter 5).

It is also not correct that the parameter p makes the CMP discrete and restricts the predictability of the model, as **Pastore et al.** suggest. This parameter is an edge that the CMP has over the FLMP because it is an extra free parameter. In a related point, **Kanevsky** questions why the parameter p in the CMP must remain constant across the different conditions in an identification task. Its constancy is necessary to implement the assumptions of categorical perception. The idea is that each modality is perceived categorically, the decision system therefore has no other information that could be used to modify the value of p on a given trial.

Pastore et al. also worry that some of the research in *Paradigm* is unpublished – but it is published there in enough detail to allow it to be scrutinized by peer review. I am also puzzled by Pastore et al.'s concern that I did not acknowledge the so-called criticisms of Hary and Massaro's (1982) conclusion that the identification/discrimination task is not diagnostic of categorical perception of plucks and bows. I also did not cite my replies to these criticisms (e.g., Massaro & Hary 1986). However, the only important issue is whether anyone would still argue that plucks and bows are categorically perceived (Rosen & Howell 1981; 1983).

Independence. As interpreted by **Krueger,** independence of auditory and visual sources of information and the independence of bottom-up and top-down sources are evidence for the modularity thesis that perception is cognitively impenetrable. We need to be more explicit about what we mean by perception, however. In my view, the evaluation of bottom-up information is impenetrable by top-down information whereas integration is not. Because I want to include both of these processes in perception, some aspects of perception are penetrable and some are not.

Bernstein's description of the dependence of auditory and visual information in the specification of the voicing of stop consonants is useful in several respects. First, it brings us back to the empirical question of delineating the stimulus characteristics supporting bimodal speech perception. Second, it provides an opportunity to illustrate that it is possible to falsify the FLMP prediction when the features are not, in fact, independent of one another (**Crowder, Kanevsky, MacKay,** and **Warren**). Third, it helps to impress upon the reader how little we can expect to say theoretically without the corresponding empirical evidence. Bernstein is right in pointing out the incompleteness of our empirical work and our analyses of existing literature. Our analyses (as well as most relevant empirical work) have been directed to the bimodal perception of place and manner (but not voicing) information. (Let me caution the reader now that these articulatory descriptors do not necessarily have psychological reality; they are convenient categories for the present discussion, however.) It appears that auditory and visual modalities do provide independent sources of information relevant to distinctions of place and manner. Also, the model works well in situations with 12 alternatives (see discussion in *Paradigm,* pp. 279–80).

Information about voicing can be heard but is not easily seen. Breeuwer and Plomp's (1986) results can be used to provide a quantitative test of whether the auditory and visual sources are treated independently with respect to the voicing distinction. The observed likelihood of a correct voicing judgment to a bimodal stimulus was much larger than the FLMP prediction based on independent auditory and visual sources. Thus, we are able to reject the hypothesis that auditory and visual sources are evaluated independently with respect to information about voicing. The functional value of the temporal occurrence of auditory and visible speech reveals that a source of information, even a bottom-up source, cannot be linked to a single sensory system.

Glucksberg is troubled by the independence of adjective meanings. The meaning of one should surely influence the meaning of the other. He gives the example of *wise man* and *wise guy*. Independence does not require that these two descriptions have the same meaning, for, after all, they are different (Oden & Anderson 1974). Given one description of John as *wise and serious* and another as *wise and a joker,* then *wise* is claimed to be evaluated equivalently in the two scenarios. *Wise and sage* versus *wise and crafty* should certainly be nonindependent because the second adjective should activate different meaning senses of the first. Independence certainly might fail, but it is an empirical question. If independence does fail, it is not the end of the world – nor of *Paradigm*. It only duplicates in the person impression domain what **Bernstein** has uncovered in the speech domain – certain cases in which there is a necessary dependence between properties of the input in their evaluations. What is sobering for me, and probably more palatable for most readers, is how many of the issues are empirical.

Finally, with respect to independence and separability as described by **Townsend** and Ashby and Townsend (1986), auditory and visual speech are evaluated independently but are not separable in the output of integration.

Optimal integration. Contrary to **Gigerenzer**'s suggestion, the FLMP is not a Laplacean variant of Bayes' theorem that assumes uniform equal prior probabilities. The model is neutral with respect to where the information comes from, and the prior probabilities of the alternatives appear to be reasonable sources of information. Although it is not stressed in *Paradigm*, the FLMP has been applied to a variety of tasks that vary both stimulus information and context. For example, an ambiguous letter between *e* and *c* was placed in several different letter-string contexts (Massaro 1979). In some cases, the context made a word if the letter was an *e* but not a *c*, and vice versa. The context in this case can be thought of as prior probability in the sense that subjects had experienced the test words more often than the nonwords. Consistent with the FLMP (and Bayes' theorem), word context had a positive effect on letter identification and this effect was well-described in terms of two independent sources of information.

Although the FLMP accounts for the influence of prior probabilities (base rates) in terms of an independent source of information, people's failure to use base rates in decision making does not invalidate the model. A failure to use base rates only reveals that the information value attached to base rates at the evaluation stage in the model was negligible. If subjective feature values are granted for base rates, then the magnitude of their effect cannot discredit either the FLMP or Bayes' theorem (even if base rates have no effect whatsoever). It is simply an empirical fact that base rates carry little information unless they are highlighted in the task (Gigerenzer et al. 1988). In Brunswikian terms, ecologically valid information is not necessarily functionally valid. Although context in language processing can be viewed as modulating the prior probability, as also suggested by **Cowan**, the role of prior probabilities in language processing remains uncertain. On the one hand, word frequency (actually, familiarity) is a powerful influence in several language processing tasks (Kreuz 1987). Overall letter frequency in written language, on the other hand, has no influence on performance in single-letter recognition tasks (Appelman & Mayzner 1981; Mason 1982).

Gigerenzer claims that our use of the FLMP is not mathematically equivalent to Bayes' theorem. In cases in which a feature is evaluated against two mutually exclusive alternatives, H_1 and H_2), we make the simplifying assumption that $t(H_2)$ is equal to $1 - t(H_1)$. In Bayesian terms, when one is considering only a single source of information, this equivalence means that $P(E/H_2$ is equal to $1 - P(E/H_1)$. Obviously, this assumption is an additional constraint imposed on Bayes' theorem; hence this form of the FLMP is no longer equivalent to Bayes' theorem. It turns out, however, that this constrained version of the FLMP and the general version of Bayes' theorem make the same predictions in a task with two mutually exclusive alternatives. The constrained version of the FLMP can be written as

$$\frac{x_i y_j}{x_i y_j + (1 - x_i)(1 - y_j)} \tag{1}$$

whereas Bayes' theorem can be expressed as

$$\frac{u_i v_j}{u_i v_j + (a_i)(b_j)} \tag{2}$$

A demonstration of their mathematical equivalence reduces to verifying that

$$\frac{x_i y_j}{x_i y_j + (1 - x_i)(1 - y_j)} = \frac{u_i v_j}{u_i v_j + (a_i)(b_j)} \tag{3}$$

for all values of x_i, y_j, a_i, b_j, u_i, and v_j in [0,1]. Multiplying both the numerator and denominator of the left and right terms of Equation 3 by $1/x_i y_j$ and $1/u_i v_j$, respectively, gives

$$\frac{1}{1 + \frac{(1 - x_i)(1 - y_j)}{(x_i)(y_j)}} = \frac{1}{1 + \frac{(a_i)(b_j)}{(u_i)(v_j)}} \tag{4}$$

The equivalence of the right and left sides of Equation 2 can be seen by noting the equivalence of the right hand term in the denominator of both sides of the equation. That is, it is sufficient to show that

$$\frac{(1 - x_i)(1 - y_j)}{(x_i)(y_j)} = \frac{(a_i)(b_j)}{(u_i)(v_j)} \tag{5}$$

Given that each ratio is indexed by a single subscript, a single parameter is sufficient to specify each of their values. For example, there is a value of x_i in [0,1] so that $(1 - x_i) \div x_i$ is equal to $a_i \div u_i$ for all values of a_i and u_i in [0,1]. Thus this implementation of the FLMP does not decorrelate its predictions from those given by Bayes' theorem.

Gigerenzer helps illuminate the conjunction fallacy and its explanation by the FLMP. Given a description of Linda, subjects rate it more likely that Linda is a feminist and a bank teller than that Linda is just a bank teller. In formal terms, this means that $P(H_1 \& H_2|E_1)$ is greater than $P(H_1|E_1)$, which cannot be predicted by Bayes' theorem or the FLMP. On the other hand, these same models predict that $P(H_1|E_1 \& E_2)$ can be greater than $P(H_1|E_1)$. My argument, in Gigerenzer's terms, is simply that subjects interchange the hypotheses and the evidence in their judgment algorithm. Thus, they view bank teller and feminist as two pieces of evidence for determining the goodness-of-match with the description of Linda. The integration process, along with the relative-goodness decision rule of the FLMP, can lead to a judgment that it is more likely that Linda is both a feminist and a bank teller than that Linda is just a bank teller – a conjunction fallacy (see, however, Tversky and Kahneman's (1983) putative refutation of this possibility in another test scenario).

Kanevsky is concerned with the number magic required to produce the conjunction of two properties making a better instance of a category than that given either property alone. Kanevsky's concern about fuzzy logic can be addressed by stressing that the prediction equations on pp. 202–3 give the probability that – or the degree to which – an object is judged as an instance of a specific category. Thus a striped apple will be judged as a better instance of the category "striped apple" than the category "apple" only if the truth value for "striped" is greater than .5. Given a striped orange, it will *not* be a poorer instance of the category "striped apple" than the category "apple." A striped orange makes a better instance of the category "striped apple" than the category "apple." However, a striped orange will be a poorer instance of the category "striped apple" than the category "striped."

Finally, with respect to **Gigerenzer's** argument for many rather than one computational algorithm, Darwin can be used for either side of the issue – an important limitation in evolutionary accounts.

Development and aging. It was gratifying to have **Wilkening's** summary of supporting evidence to bolster my conclusions about the consistency of information processing across development and aging. The similar results across speech perception, object classification, and intuitive physics once again argue for general processes rather than speech as a special module.

Bornstein seems to agree about the value of the distinction between information and information processing in the description of changes of perception across the lifespan. His concern for the ontogeny of processing must be addressed empirically.

de Gelder & Vroomen appear to view written language processing as a separate module. What they find empirically, however, is that dyslexics have deficits in the processing of spoken language in both its heard and lipread forms.

MacKay argues that some aspects of language learning are not gradual; he cites the learning of regular English plural formation. Given only diaries of children's utterances, it is not clear how one would test whether this is actually the case; however, the fact that performance is extremely variable would seem to argue for a gradual acquisition process (Kuczaj 1977).

Generality. Provided by **Welch** and **Wilkening** are informative summaries of two respective research domains that support *Paradigm*'s claim for generality of the findings (see discussions of these in the sections *Modularity of Mind* and *Development and Aging,* respectively).

With respect to **Port's** comment about perceiving speech across time, no one knows better than I that perceptual processing of speech occurs over time (Massaro 1975b). We cannot be expected to solve the more difficult problem of describing continuous speech perception without first solving the problem of the perception of segments. Port believes that the FLMP and models like it will not scale up to the real world of speech research very easily. He observes that the FLMP has not been articulated to do its thing across continuously arriving speech, as opposed to describing a single pass through the model. Although it has not been done, the model could be extended to describe the processing of continuous speech as it is spoken. Feature evaluation would occur during "windows" of the continuously arriving speech and its output would be continuously fed forward to the integration process. Integration would update its outcomes continuously with the onslaught of the updating of feature values, making its integration outcome continuously available for decisions. The result is, of course, that decisions are continuously being made online with the arrival of the speech stream. Although the processing of continuous speech is much more complex than a single pass through the model, the computational assumptions should not change. Increasing the complexity of the experimental situation to test the model would be a welcome extrapolation of the current research enterprise.

Port's goal seems to be a working system to recognize speech whereas mine is to understand how people do it. One goal might be achieved without the other. Physical laws fall short of predicting the weather, and cement has been made for centuries without a full understanding of how it works.

Relationship to connectionist models. The TRACE model appears to be wrong because of its assumption of interactive activation. Top-down activation from the phoneme level to the feature level enhances the discriminability between two syllables across a category boundary (McClelland & Elman 1986). Empirical evidence that discriminability does not change at the category boundary constitutes a disproof of TRACE without the necessity of magic (**MacKay**). If differences in discriminability are found, on the other hand, the reason could be nonlinearities across the speech continuum rather than phoneme categories. For example, **Ehret** believes that a category boundary exists for temporal resolution at around 25 msec (however, see Rosen & Howell 1987). This is the type of evidence that **Cowan** would like to use to support a weaker form of categorical perception. However, this weaker form of categorical perception would not be consistent with TRACE because it occurs without top-down activation from the phoneme level to the feature level. Thus, similar enhancement would be found in humans and nonhumans without the phoneme categories of interest.

Because interactive activation is central to the TRACE model, I do not hold much hope for **Campbell's** promise that TRACE can account for speech perception by ear and eye. It is also important to clarify a distinction that has been made between prediction and description (**MacKay, Pastore et al.,** and **Warren**) and to determine the extent to which models such as TRACE predict a broader range of behavior than does the FLMP. Campbell states that TRACE could predict how long the system would take to identify an item. However, TRACE operates in "perceptual cycles," and not time; any prediction would require some assumption (albeit perhaps straightforward) about how the former maps onto the latter. Furthermore, a decision rule must be assumed to map activations into responses and this decision rule falls outside the "neural processing" of the model. Contrary to the view of Campbell, the FLMP also makes predictions about the time course of the processing involved in identification, and tests of these predictions are presented in *Paradigm* (Chapter 4, section 7; Chapter 6, section 1).

Conclusion. As I have noted elsewhere, there are no easy answers to questions about how we perceive speech and other patterns. Some scientists seem to overestimate what a successful theory can accomplish, as well as the reality of that theory. Given the complexity of even the simplest behavior, a good psychological theory will be limited in its predictions. The uncertainty principle in physics dictates that one cannot predict future events exactly; the best we can do is predict a number of different possible outcomes (quantum mechanics). With respect to the issue of whether there is some independent reality of a theory, we should remember that our theories are invented rather than discovered. Hawking (1988), for example, views a theory or model of the universe as nothing more than our representation of it in the mind or on paper. A successful psychology requires us to evaluate and integrate the empirical facts in parsimonious and coherent theories. Scientific decisions will be based on the relative goodness-of-match of these facts with these theories. I look forward to new facts and theories to challenge the old.

ACKNOWLEDGMENT
The author wishes to thank Ray Gibbs for his comments on this Response.

References

Letters "a" and "r" appearing before authors' initials refer to target article and response respectively.

Alcock, J. (1989) *Animal behavior: An evolutionary approach.* Sinauer. [rDWM]
Anderson, J. R. (1978) Argument concerning representations for mental imagery. *Psychological Review* 85:249–77. [LEK]
(1985) *Cognitive psychology and its implications.* W. H. Freeman. [aDWM]
Anderson, N. H. (1973) Functional measurement of social desirability. *Sociometry* 36:89–98. [rDWM, SG]
(1981) *Foundations of information integration theory.* Academic Press. [arDWM, NHA]
(1982) *Methods of information integration theory.* Academic Press. [aDWM, NHA]
Anderson, N. H. & Wilkening, F. (in press) Adaptive thinking in intuitive physics. In: *Contributions to information integration theory,* ed. N. H. Anderson. Academic Press. [FW]
Anderson, S., Merrill, J. W. L & Port, R F. (1988) Dynamic speech categorization with recurrent networks. In: *Proceedings of the 1988 Connectionist Models Summer School,* ed. D. Touretzky, G. Hinton & T. Sejnowski. Morgan-Kaufmann. [RFP]
Appelman, I. B. & Mayzner, M. S. (1981) The letter-frequency effect and the generality of familiarity effects on perception. *Perception & Psychophysics* 30:436–46. [rDWM]
Asch, S. E. (1946) Forming impressions of personality. *Journal of Abnormal and Social Psychology* 41:258–90. [aDWM,SG]
Ashby, F. G. & Townsend, J. T. (1986) Varieties of perceptual independence. *Psychological Review* 93:154–79. [arDWM, JTT]
Aslin, R. N., Pisoni, D. B., Hennessy, B. L. & Perey, A. J. (1981) Discrimination of voice-onset-time by human infants: New findings concerning phonetic development *Child Development* 52.1135–45. [aDWM]
Barclay, J. R. (1972) Noncategorical perception of a voiced stop: A replication. *Perception & Psychophysics* 11:269–73. [RGC]
Bates, E., McNew, S., MacWhinney, B., Devescovi, A. & Smith, S. (1982) Functional constraints on sentence processing: A cross-linguistic study. *Cognition* 11:245–99. [SG]
Bertelson, P. & de Gelder, B. (1989) Learning about reading from illiterates. In: *From neurons to reading,* ed. A. Galaburda. MIT Press/Bradford Books. [BDG]
Birnbaum, M. H. (1974) The nonadditivity of personality impressions. *Journal of Experimental Psychology* 102:543–61. [SG]
Blumstein, S. E. (1981) Phonological aspects of aphasia. In: *Acquired aphasia,* ed. M. T. Sarno. Academic Press. [MS-K]
Boothroyd, A. (1988) Perception of speech pattern contrasts from auditory presentation of voice fundamental frequency. *Ear and Hearing* 9:313–21. [LEB]
Bornstein, M. H. (1979) Perceptual development: Stability and change in feature perception. In: *Psychological development from infancy· Image to intention,* ed. M. H. Bornstein & W. Kessen. Lawrence Erlbaum. [MHB]
(1987) Perceptual categories in vision and audition. In: *Categorical perception: The groundwork of cognition,* ed. S. Harnad. Cambridge University Press. [MHB]
Bower, T. G. R. (1977) *A primer of infant development.* Freeman. [MHB]
Breeuwer, M. & Plomp, R. (1986) Speechreading supplemented with auditorily presented speech parameters. *Journal of the Acoustical Society of America* 79:481–99. [rDWM, LEB]
Broadbent, D. E. & Broadbent, M. H. P. (1975) Some further data concerning the word frequency effect. *Journal of Experimental Psychology: General* 104:297–308. [LEK]
Bruno, N. & Cutting, J. E. (1988) Minimodularity and the perception of layout. *Journal of Experimental Psychology: General* 117:161–70. [rDWM]
Brunswik, E. (1955) Representative design and probabilistic theory in a functional psychology. *Psychological Review* 62:193–217. [rDWM]
Busemeyer, J. R. (in press) Intuitive statistical estimation. In: *Contributions to information integration theory,* ed N.
Campbell, R. (1987) Lipreading and immediate memory processes. In: *Hearing by eye· The psychology of lipreading,* ed. B. Dodd & R. Campbell. Lawrence Erlbaum. [RC]
(1988) Tracing lip movements: Making speech visible. In: *Visible language* [a special issue on lipreading], ed. R. Campbell. [RC]
Campbell, R. & Dodd, B. (1980) Hearing by eye. *Quarterly Journal of Experimental Psychology* 32:509–15. [BDG]
Campbell, R., Garwood, J. & Rosen, S. (1988) Adding sound to lipread lists: The effects on serial recall of adding an auditory pulse train and a pure tone to silently lipread lists. *Memory and Cognition* 16:210–19. [RC]
Canon, L. K. (1970) Intermodality inconsistency of input and directed attention as determinants of the nature of adaptation. *Journal of Experimental Psychology* 84.141–47. [RBW]
(1971) Directed attention and maladaptive "adapatation" to displacement of the visual field. *Journal of Experimental Psychology* 88:403–8. [RBW]
Carney, A. E., Widin, G. P. & Viemeister, N. F. (1977) Noncategorical perception of stop consonants differing in VOT. *Journal of the Acoustical Society of America* 62.961–70. [GE]
Chandler, J. P. (1969) Subroutine STEPIT – Finds local minima of a smooth function of several parameters. *Behavioral Science* 14:81–82. [RFP]
Cheng, P. W. & Pachella, R. G. (1984) A psychophysical approach to dimensional separability. *Cognitive Psychology* 16:279–304. [rDWM]
Chiarello, C., ed. (1988) Right hemisphere contributions to lexical semantics. Springer-Verlag. [rDWM]
Cholewiak, R. W. & Sherrick, C. E. (1986) Tracking skill of a deaf person with long-term tactile aid experience: A case study. *Journal of Rehabilitation Research and Development* 23(2)BPR 10–44:20–26. [DK]
Chomsky, N. (1975) *Reflections on language.* Pantheon. [aDWM]
Cohen, L. J. (1986) *The dialogue of reason.* Clarendon Press. [GC]
Cohen, M. M. (1979) Three cues to voicing of intervocalic velar stops. Unpublished master's thesis, University of Wisconsin, Madison. [REP]
(1984) *Processing of visual and auditory information in speech perception.* Unpublished doctoral dissertation, University of California, Santa Cruz. [aDWM, REP]
Cole, R. A. & Rudnicky, A. I. (1983) What's new in speech perception? The research and ideas of William Chandler Bagley, 1874–1946. *Psychological Review* 90.94–101. [DGM]
Cowan, N. (1988) Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information processing system. *Psychological Bulletin* 104:163–91. [NC]
(1989) The reality of cross-modal Stroop effects. *Percetion & Psychophysics* 45:87–88. [NC]
Cowan, N. & Barron, A. (1987) Cross-modal, auditory-visual Stroop interference, and possible implications for speech memory. *Perception & Psychophysics* 41:393–401. [NC]
Daston, L. (1988) *Classical probability in the enlightenment.* Princeton University Press. [GC]
De Filippo, C. L. (1984) Laboratory projects in tactile aids to lipreading. *Ear and Hearing* 5:211–27. [LEB]
de Gelder, B. & Vroomen (1988) Cross-modal speech perception in young dyslexics. Paper presented at the Sixth Language and Speech Conference, Sydney. [BDG]
(1989) Seeing lips and hearing voices separately. Paper presented at the Workshop on Lipreading, Oxford. [BDG]

Denes, P. (1955) Effects of duration on the perception of voicing. *Journal of the Acoustical Society of America* 27:761–64. [rDWM]

Derr, M. A., & Massaro, D. M. (1980) The contribution of vowel duration, Fo contour, and frication duration as cues to the /juz/-jus/ distinction. *Perception & Psychophysics* 27:51–59. [rDWM]

Donner, K. & Luttwak, J. (1984) The uses of the Kanievski vibro-tactile articulator (preprint). Descriptions of tests of KS devices performed in Israel in 1984. [DK]

Ehret, G. (1987) Categorical perception of sound signals: Facts and hypotheses from animal studies. In: *Categorical perception. The groundwork of cognition*, ed. S. Harnad. Cambridge University Press. [GE]

(1988) Frequency resolution, spectral filtering and integration on the neuronal level. In: *Auditory function*, ed. G. M. Edelmann, W. E. Gall & W. M. Cowan. Wiley. [GE]

Ehret, G. & Haack, B. (1981) Categorical perception of mouse pup ultrasounds by lactating females. *Naturwissenschaften* 68:208. [GE]

(1982) Ultrasound recognition in house mice: Key-stimulus configuration and recognition mechanism. *Journal of Comparative Physiology* 148:2245–51. [GE]

Ehret, G. & Merzenich, M. M. (1985) Auditory midbrain responses parallel spectral integration phenomena. *Science* 227.1245–47. [GE]

(1988) Complex sound analysis (frequency resolution, filtering, and spectral integration) by single units of the inferior colliculus of the cat. *Brain Research Reviews* 13:139–63. [GE]

Eimas, P. D. (1985) The perception of speech in early infancy. *Scientific American* 252(1):46–52. [aDWM]

Eimas, P. D. & Miller, J. L. (1980) Discrimination of information for manner of articulation. *Infant Behavior and Development* 3:367–75. [aDWM]

Eimas, P. D., Siqueland, E. R., Jusczyk, P. & Vigorito, J. (1971) Speech perception in infants. *Science* 171:303–6. [aDWM]

Erber, N. P. & De Filippo, C. L. (1978) Voice/mouth synthesis and tatual/visual percception of /pa,ba,ma/. *Journal of the Acoustical Society of America* 64:1015–19. [LEB]

Estes, W. K. (1975a) The locus of inferential and perceptual processes in letter identification. *Journal of Experimental Psychology: General* 1104:122–45. [LEK]

(1975b) Memory, perception, and decision in letter identification. In: *Information processing and cognition: The Loyola Symposium*, ed. R. L. Solso. Erlbaum. [LEK]

Ewert, J. P. (1987) Neuroethology of releasing mechanisms. Prey-catching in toads. *Behavioral and Brain Sciences* 10:337–405. [GE]

Falmagne, J. (1985) *Elements of psychophysical theory*. Oxford University Press. [aDWM]

Flavell, J. H. (1985) *Cognitive development*. Prentice-Hall. [aDWM]

Fodor, J. A. (1983) *Modularity of mind*. Bradford Books/MIT Press. [arDWM, AC, BDG, LEK, DGM, REP, JTT]

(1985) Précis of *The modularity of mind*. *Behavioral and Brain Sciences* 8:1–42. [LEK, DGM, REP]

Forster, K. I. (1985) Lexical acquisition and the modular lexicon. *Language and Cognitive Processes* 1:87–108. [aDWM]

Freides, D. (1974) Human information processing and sensory modality: Cross-modal functions, information complexity, memory, and deficit. *Psychological Bulletin* 81:284–310. [RBW]

Fujisaki, H. & Kawashima, T. (1970) Some experiments on speech perception and a model for the perceptual mechanism. *Annual Report of the Engineering Institute* (University of Tokyo) 29:207–14. [MS-K]

Gardner, H. (1985) *The mind's new science*. Basic Books. [rDWM]

Garner, W. R., Hake, H. W. & Eriksen, C. W. (1956) Operationism and the concept of perception. *Psychological Review* 63:149–59. [aDWM]

Gibson, E. J. (1982) The concept of affordances in development: The renascence of functionalism. In: *The Minnesota symposia on child psychology*, ed. W. A. Collins. Lawrence Erlbaum. [MHB]

Gibson, J. J. (1979) *The ecological approach to visual perception*. Houghton Mifflin. [rDWM, MHB, LEK]

Giere, R. N. (1988) *Explaining science: A cognitive approach*. University of Chicago Press. [rDWM]

Gigerenzer, G., Hell, W. & Blank, H. (1988) Presentation and content: The use of base rates as a continuous variable. *Journal of Experimental Psychology: Human Perception and Performance* 14:513–25. [rDWM, GG]

Gigerenzer, G. & Murray, D. J. (1987) *Cognition as intuitive statistics*. Lawrence Erlbaum. [GG]

Gleitman, L. R. & Wanner, E. (1982) Language acquisition: The state of the state of the art. In: *Language acquisition: The state of the art*, ed. E. Wanner & L. R. Gleitman. Cambridge University Press. [aDWM]

Goguen, J. A. (1969) The logic of inexact concepts. *Synthese* 19:325–73. [aDWM]

Grant, K. W., Ardell, L. A. H., Kuhl, P. K. & Sparks, D. W. (1986) The transmission of prosodic information via an electrotactile spreechreading aid. *Ear and Hearing* 7:328–35. [LEB]

Green, D M. & Swets, J. A. (1966) *Signal detection theory and psychophysics*. Wiley. [JTT]

Greenwald, A. G., Pratkanis, A. R., Leippe, M. R. & Baumgardner, M. H. (1986) Under what conditions does theory obstruct research progress? *Psychological Review* 93:216–29. [rDWM]

Grodzinsky, Y., Swinney, D & Zurif, E. (1985) Agrammatism: Structural deficits and antecedent processing disruptions. In: *Agrammatism*, ed. M. L. Kean Academic Press. [rDWM]

Harnad, S., ed. (1987) *Categorical perception: The groundwork of cognition*. Cambridge University Press. [NHA, RMW]

Hary, J. M. & Massaro, D. W. (1982) Categorical results do not imply categorical perception. *Perception & Psychophysics* 32:409–18. [rDWM, REP]

Hawking, S. W. (1988) *A brief history of time: From the big bang to black holes*. Bantam Books. [rDWM]

Hay, J. C., Pick, H. L., Jr., & Ikeda, K. (1965) Visual capture produced by prism spectacles. *Psychonomic Science* 2:215–16. [RBW]

Hirst, W. (1988). *The making of cognitive science: Essays in honor of George A. Miller*. Cambridge University Press. [rDWM]

Hochberg, J. (1986) Representation of motion and space in video and cinematic displays. In: *Handbook of perception and human performance*, vol. 2, *Sensory processes and perceptions*, ed. K. R. Boff, L. Kaufman & J. P. Thomas. Wiley. [aDWM]

House, A. S. & Fairbanks, G. (1953) The influence of consonant environment upon the secondary acoustical characteristics of vowels. *Journal of the Acoustical Society of America* 25:105–13. [LEB]

Jack, C. E. & Thurlow, W. R. (1973) Effects of degree of visual association and angle of displacement on the "ventriloquism" effect. *Perceptual and Motor Skills* 37:967–79. [RBW]

Jackson, C. V. (1953) Visual factors in auditory localization. *Quarterly Journal of Experimental Psychology* 5:52–65. [RBW]

Jones, D. (1948) *Differences between spoken and written language*. Association Phonetique Internationale. [MS-K]

Kahneman, D. & Tversky, A. (1973) On the psychology of prediction. *Psychological Review* 80:237–51. [GG]

Kanevsky, D. & Skurkovich, G. (1982) Vibrotactile articulator. In: *Uses of computers in aiding disabled*, ed. J. Raviv. North-Holland. [DK]

Kanevsky, D., Skurkovich, G. & Yabrov, A. (1987) Consideraciones sobre el desarrollo perceptivo en rehabilitacion logopedica. In: *Logopedia y Nuevas Tecnologias, Ponencias y Communicaciones del Congreso Hispanoamercano de Asociaciones de Padres de Ninos Deficientes Auditivos y Symposium Internacional sobre Nuevas Tecnologias y Lenguaoes Alterntivos*. [DK]

Kelso, J. A. & Kay, B. (1987) Information and control: A macroscopic basis for perception-action coupling. In: *Perspectives on perception and action*, ed. H. Heuer & A. F. Sanders. Lawrence Erlbaum. [aDWM]

Kemler, D. G. (1983) Holistic and analytic modes in perceptual and cognitive development. In: *Perception, cognition, and development: Interactional analyses*, ed. T. J Tighe & B. E. Shepp. Erlbaum. [FW]

Kemler, D. G. & Smith, L. B. (1978) Assessing similarity and dimensional relations: Effects of integrality and separability on the discovery of complex concepts. *Journal of Experimental Psychology: General* 108:133–50. [aDWM]

Kendler, T. S. (1983) Labeling, overtraining, and levels of function. In: *Perception, cognition, and development: Interactional analyses*, ed. T. J. Tighe & B. E. Shepp. Lawrence Erlbaum. [aDWM]

Kreuz, R. J. (1987) The subjective familiarity of English homophones. *Memory & Cognition* 15:154–68. [rDWM]

Krueger, L. E. (1975) Familiarity effects in visual information processing. *Psychological Bulletin* 82:949–74. [LEK]

Krueger, L. E. & Shapiro, R. G. (1979) Letter detection with rapid serial visual presentation: Evidence against word superiority at feature extraction. *Journal of Experimental Psychology: Human Perception and Performance* 5.657–73. [LEK]

Kuczaj, S. A. (1977) The acquisition of regular and irregular past-tense forms. *Journal of Verbal Learning and Verbal Behavior* 16:589–600. [rDWM]

Kuhl, P. K. (1987) The special mechanisms debate in speech research: Categorization tests on animals and infants. In: *Categorical perception: The groundwork of cognition*, ed. S. Harnad. Cambridge University Press. [aDWM]

Kuhl, P. K. & Meltzoff, A. N. (1982) The bimodal perception of speech in infancy. *Science* 218:1138–41. [aDWM]

Kuhn, T. S. (1977) *The essential tension: Tradition and innovation in scientific research*. University of Chicago Press. [DGM]

Lancaster Summer Project (1985) Preprint. (Summary report on testing KS device on a group of children in the USA). [DK]
Leon, M. & Anderson, N. H. (1974) A ratio rule from integration theory applied to inference judgments. *Journal of Experimental Psychology* 102:27–36. [arDWM, NHA]
Liberman, A. M., Cooper, F. S., Shankweiler, D. & Studdert-Kennedy, M. (1967) Perception of the speech code. *Psychological Review* 74:4311–61. [MS-K]
Liberman, A. M., Harris, K. S., Hoffman, H. & Griffith, B. C. (1957) The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology* 54.358–68. [GE]
Liberman, A. M. & Mattingly, I. G. (1985) The motor theory of speech perception revised. *Cognition* 21:1–36. [aDWM, REP]
(1989) A specialization for speech perception. *Science* 243:489–94. [NC, SG, LEK]
Lisker, L. & Abramson, A. (1964) A cross-language study of voicing in initial stops: Acoustical measurements. *Word* 20:384–422. [LEB]
Luce, R. D. (1959) *Individual choice behavior: A theoretical analysis*. Wiley. [rDWM, JTT]
Luttwak, J. (1987) El sistema de entrenamiento para el uso del A.V.K. In: *Logopeida y Nuevas Tecnologias, Ponencias y Communicaciones del Congreso Hispanoamercano de Asociaciones de Padres de Ninos Deficientes Auditivos y Symposium Internacional sobre Nuevas Technologias y Lenguaoes Alterntivos* [DK]
MacKay, D. G. (1987) The organization of perception and action: A theory for language and other cognitive skills. Springer-Verlag. [rDWM]
(1988a) Under what conditions can theoretical psychology survive and prosper?: Integrating the rational and emprical epistemologies. *Psychological Review* 95(4):559–65. [DGM]
(1988b) Practical applications and theories of memory: A new epistemology to supplement the old. In: *Practical aspects of memory*, vol. 2, ed. M. M. Gruneberg, P. E. Morris & R. N. Sykes. Wiley. [DGM]
MacMillian, N. A. (1983) A psychophysical interpretation of a "categorical perception" experiment by Hary and Massaro. *Perception & Psychophysics* 34:494–98. [REP]
MacWhinney, B., Bates, E. & Kliegl, R. (1984) Cue validity and sentence interpretation in English, German, and Italian. *Journal of Verbal Learning and Verbal Behavior* 23:127–50. [arDWM, SG, DGM]
Manicas, P. T. & Secord, P. F. (1983) Implications for psychology of the new philosophy of science. *American Psychologist* 38:399–413. [rDWM]
Marcel, A. J. (1983) Conscious and unconscious perception: An approach to the relations between phenomenal experience and perceptual processes. *Cognitive Psychology* 15:238–300. [aDWM]
Marslen-Wilson, W. D. (1987) Functional parallelism in spoken work recognition. *Cognition* 25:71–102. [AC]
Mason, M. (1982) More about the letter-frequency effect. *Perception & Psychophysics* 31:589–90. [rDWM]
Massaro, D. W. (1975a) Experimental psychology and information processing. Rand-McNally. [rDWM]
(1975b) Understanding language: An information-processing analysis of speech perception, reading, and psycholinguistics. Academic Press. [rDWM]
(1976) Auditory information processing. In: *Handbook of learning and cognitive processes*. Vol. 4: *Attention and memory*, ed. W. K. Estes. Lawrence Erlbaum. [rDWM]
(1979) Reading and listening (tutorial paper). In: *Processing of visible language*, vol. 1, ed. P. A. Kolers, M. Wrolstad & H. Bouma. Plenum. [arDWM]
(1979a) Letter information and orthographic context in word perception. *Journal of Experimental Psychology: Human Perception and Performance* 5:595–609. [LEK]
(1984) Building and testing models of reading processes. In: *Handbook of reading research*, ed. P. D. Pearson. Longman. [aDWM]
(1984a) Children's perception of auditory and visual speech. *Child Development* 55:1777–88. [RBW]
(1986a) *Connectionist models of mind*. Paper given at the 27th annual meeting of the Psychonomic Society, New Orleans, November. [aDWM]
(1986b) The computer as a metaphor for psychological inquiry: Considerations and recommendations. *Behavior Research Methods, Instruments, & Computers* 18.73–92. [aDWM]
(1986c) Book review of J. A. Fodor *The modularity of mind: An essay on faculty psychology*. *American Journal of Psychology* 99.435–42. [rDWM]
(1987) *Speech perception by ear and eye: A paradigm for psychological inquiry*. Lawrence Erlbaum.
(1987a) Categorical partition: A fuzzy logical model of categorization behavior. In: *Categorical perception: The groundwork of cognition*, ed. S. Harnad. Cambridge University Press. [arDWM]
(1988a) Ambiguity in perception and experimentation. *Journal of Experimental Psychology: General* 117:417–21. [aDWM]
(1988b) Some criticisms of connectionist models of human performance. *Journal of Memory and Language* 27:213–324. [arDWM]
(1989) *Experimental psychology: An information processing approach*. Harcourt Brace Jovanovich. [arDWM]
(in press) An information-processing analysis of perception and action. In: *Relationships between perception and action. Current approaches*, ed. O. Neumann & W. Prinz. Springer. [rDWM]
(in press) A pattern recognition account of decision making. In: *Human information processing*, ed. D. Vickers & P. L. Smith. [GG]
Massaro, D. W. & Cohen, M. M. (1976) The contribution of fundamental frequency and voice onset time to the /zi/-/si/ distinction. *Journal of the Acoustical Society of America* 60:704–17. [arDWM]
(1977) Voice onset time and fundamental frequency as cues to the /zi/-/si/ distinction. *Perception & Psychophysics* 22:373–82. [rDWM]
(1983) Consonant/vowel ratio: An improbable cue in speech. *Perception & Psychophysics* 33:501–5. [rDWM]
(1983a) Evaluation and integration of visual and auditory information in speech perception. *Journal of Experimental Psychology: Human Perception and Performance* 9:753–71. [BDG]
(1987)*Process and connectionist models of pattern recognition*. Proceedings of the Ninth Annual Conference of the Cognitive Science Society. Lawrence Erlbaum. [aDWM]
Massaro, D. W. & Friedman, D. (1989) Models of integration given multiple sources of information. Unpublished manuscript. [rDWM]
Massaro, D. W. & Hary, J. M. (1986) Categorical results, categorical perception, and hindsight. *Perception & Psychophysics* 35:586–88. [rDWM]
Massaro, D. W. & Oden, G. C. (1980) Speech perception: A framework for research and theory. In: *Speech and language: Advances in basic research and practice*, vol. 3, ed. N. J. Lass. Academic Press. [aDWM]
Massaro, D. W. & Warner, D. S. (1977) Dividing attention between auditory and visual perception. *Perception & Psychophysics* 21:569–74. [RBW]
McClelland, J. L. & Elman, J. L. (1986) The TRACE model of speech perception. *Cognitive Psychology* 18:1–86. [rDWM, DGM]
McGurk, H. & MacDonald, J. (1976) Hearing lips and seeing voices. *Nature* 264:746–48. [aDWM, RC, MG, MS-K]
Medin, D. L., Altom, M. W. & Murphy, T. D. (1984) Given versus induced category representations: Use of prototype and exemplar information in classification. *Journal of Experimental Psychology: Learning, Memory and Cognition* 10:333–52. [aDWM]
Miller, G. A. (1962) Decision units in the perception of speech. *IRE Transactions in Information Theory*, IT-8:81–83. [rDWM]
(1981) *Language and speech*. W. H. Freeman. [aDWM]
Miller, G. A., Heise, G. A. & Lichten, W. (1951) The intelligibility of speech as a function of the context of the test materials. *Journal of the Acoustical Society of America* 41:329–35. [MS-K]
Miller, G. A. & Nicely, P. E. (1955) An analysis of perceptual confusions among some English consonants. *Journal of the Acoustical Society of America* 27:338–52. [LEB]
Miller, J. L. & Eimas, P. D. (1983) Studies on the categorization of speech by infants. *Cognition* 13:135–65. [aDWM]
Mills, A. E. & Thiem, R. (1980) Auditory-visual fusions and illusions in speech perception. *Linguistische Berichte* 68/80:85–108. [MG]
Miyawaki, K., Strange, W., Verbrugge, R., Liberman, A. M., Jenkins, J. J. & Fujimura, O. (1975) An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English. *Perception & Psychophysics* 18:331–40. [GE]
Molfese, D. H. & Betz, J. C. (1988) Electrophysiological indices of the early development of lateralization for language and cognition, and their implications for predicting later development. In: *Brain lateralization in children*, ed. D. H. Molfese & S. J. Segalowitz. Guilford. [MS-K]
Navon, D. (1984) Resources: A theoretical soup stone? *Psychological Review* 91:216–34. [aDWM]
Neisser, U. (1976) *Cognition and reality*. W. H. Freeman. [rDWM]
Neumann, O. (1987) Beyond capacity: A functional view of attention. In: *Perspectives on perception and action*, ed. H. Heuer & A. F. Sanders. Lawrence Erlbaum. [aDWM]
Newell, A. (1973) You can't play 20 questions with nature and win: Projective comments on the papers of this symposium. In: *Visual information processing*, ed. W. E. Chase. Academic Press. [DGM, JTT]
Norris, D. G. (1986) Word recognition: Context effects without priming. *Cognition* 22:93–136. [AC]
O'Connor, N. & Hermelin, B. (1972) Seeing and hearing in space and time. *Perception & Psychophysics* 11:46–48. [RBW]

(1981) Coding strategies of normal and handicapped children. In: *Intersensory perception and sensory integration*, ed. R. D. Walk & H. L. Pick, Jr. Plenum. [aDWM]

Oden, G. C. (1979) A fuzzy logical model of letter identification. *Journal of Experimental Psychology: Human Perception and Performance* 5:336–52. [aDWM]

(1981) A fuzzy propositional model of concept structure and use: A case study in object identification. In: *Applied systems and cybernetics*, vol. 6, ed. G. W. Lasker. Pergamon Press. [aDWM]

Oden, G. C. & Anderson, N. H. (1974) Integration of semantic constraints. *Journal of Verbal Learning and Verbal Behavior* 13:138–48. [rDWM]

Oden, G. C. & Massaro, D. W. (1978) Integration of featural information in speech perception. *Psychological Review* 85:172–91. [arDWM]

Ornstein, R. E. (1969) *On the experience of time*. Penguin. [LEK]

Owens, E. & Blazek, B. (1985) Visemes observed by hearing-impaired and normal-hearing adult viewers. *Journal of Speech and Hearing Research* 28:381–93. [LEB]

Palmer, S. E. (1978) Fundamental aspects of cognitive representation. In: *Cognition and categorization*, ed. E. Rosch & B. B. Lloyd. Erlbaum. [rDWM]

Pastore, R. E. (1987) Categorical perception: Some psychophysical models. In: *Categorical perception: The groundwork of cognition*, ed. S. Harnad Cambridge University Press. [aDWM]

Pastore, R. E. & Layer, J. K. (submitted) Phonetic trading relations: Theoretically important but empirically deficient concepts. [REP]

Pastore, R. E., Szczesiul, R., Wielgus, V., Nowikas, K. & Logan, R. (1984) Categorical perception, category boundary effects, and continuous perception: A reply to Hary and Massaro. *Perception & Psychophysics* 35:583–85. [REP]

Perkell, J. S. & Klatt, D. H., eds. (1986) *Invariance and variability in speech processes*. Erlbaum.

Port, R. F. & Dalby, J. (1982) Consonant/vowel ratio as a cue for voicing in English. *Perception & Psychophysics* 32:141–52. [rDWM]

Pick, H. L., Jr., Warren, D. H. & Hay, J. C., Jr. (1969) Sensory conflict in judgments of spatial direction. *Perception & Psychophysics* 6:203–5. [RBW]

Pinker, S. (1984) *Language learnability and language development*. Harvard University Press. [DGM]

Pisoni, D. B. (1973) Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception & Psychophysics* 13:253–60. [RGC, MS-K]

Pisoni, D. B. & Lazarus, J. H. (1974) Categorical and noncategorical modes of speech perception along the voicing continuum. *Journal of the Acoustical Society of America* 55:328–33. [GE, MS-K]

Pisoni, D. B. & Tash, J. (1974) Reaction time to comparisons within and across phonetic categories. *Perception and Psychophysics* 15:285–90. [MS-K]

Platt, J. R. (1964) Strong inference. *Science* 146:347–53. [aDWM]

Pollack, I. & Pickett, J. M. (1963) The intelligibility of excerpts from conversation. *Language and Speech* 6:165–72. [MS-K]

Popper, K. R. (1959) *The logic of scientific discovery*. Basic Books. [aDWM, DGM]

Porkorny, J. & Smith, V. C. (1986) Colorimetry and color discrimination. In: *Handbook of perception and human performance*, vol. 2, *Sensory processes and perception*, ed. K. R. Boff, L. Kaufman & J. P. Thomas. Wiley. [aDWM]

Port, R. F. & Anderson, S. (in press) Recognition of continuously performed melodies. *Proceedings of the Cognitive Science Society*, vol. 2. Lawrence Erlbaum. [RFP]

Port, R. F., Anderson, S. & Merrill, J. (1988) *Temporal information and memory in connectionist networks*. Technical Report No. 265, Department of Computer Science, Indiana University. [RFP]

Posner, M. I. (1978) *Chronometric explorations of mind*. Erlbaum. [rDWM]

Potts, S. M. & Weisenberger, J. M. (1987) Evaluation of two multichannel tactile aids for the hearing-impaired. Paper presented at a meeting of the Acoustical Society of America, Indianapolis. [DK]

Premack, D. (1986) *Gavagai! Or the future history of the animal language controversy*. MIT Press. [aDWM]

Pylyshyn, Z. N. (1980) Computation and cognition: Issues in the foundations of cognitive science. *Behavioral and Brain Sciences* 3:111–69. [LEK]

(1984) *Computation and cognition. Toward a foundation for cognitive science*. MIT Press. [aDWM, LEK]

Raskin, L. A., Maital, S. & Bornstein, M. H. (1983) Perceptual categorization of color: A life-span study. *Psychological Research* 45:135–45. [MHB]

Reich, P. A. (1986) *Language development*. Prentice-Hall. [LEK]

Repp, B. H. (1984) Categorical perception: Issues, methods, findings. In: *Speech and language: Advances in basic research and practice*, vol. 10, ed. N. J. Lass. Academic Press. [MS-K]

(1988) Integration and segregation in speech perception. *Language and Speech* 31:239–71. [MS-K]

Roberts, M. & Summerfield, Q. (1981) Audiovisual presentation demonstrates that selective attention to speech is purely auditory. *Perception & Psychophysics* 30:309–14. [RC]

Rodriguez, Santos, J. M. (1987) Evolucion y desarrollo del lenguaje mediante el A.V.K. In. *Logopedia y Nuevas Tecnologias, Ponencias y Comunicaciones del Congreso Hispanoamercano de Asociaciones de Padres de Ninos Deficientes Auditivos y Symposium Internacional sobre Nuevas Tecnologias y Lenguaoes Alterntivos*. [DK]

Rose, D. E., Haymond, J. & Facer (1987) Vibro-tactile results using the SENSORAID with successful single channel cochlear implant patients. *Mayo Clinic Proceedings*, ed. Robert G. Siekert. [DK]

Rosen, S. M., Fourcin, A. J. & Moore, B. C. J. (1981) Voice pitch as an aid to lipreading. *Nature* 291:150–52. [LEB]

Rosen, S. & Howell, P. (1981) Plucks and bows are not categorically perceived. *Perception & Psychophysics* 30:156–68. [rDWM]

(1983) Sinusoidal plucks and bows are not categorically perceived, either. *Perception & Psychophysics* 34:233–36. [rDWM]

(1987) Auditory, articulatory, and learning explanations of categorical perception in speech. In: *Categorical perception. The groundwork of cognition*, ed. S. Harnad. Cambridge University Press. [rDWM]

Ross, B. H. & Anderson, J. R. (1981) A test of parallel versus serial processing applied to memory retrieval. *Journal of Mathematical Psychology* 24(3):183–223. [JTT]

Rueckl, J. G. & Oden, G. C. (1986) The integration of contextual and featural information during word identification. *Journal of Memory & Language* 25:445–60. [rDWM]

Sachs, R. M. (1969) Vowel identification and discrimination in isolation vs. word context. *Quarterly Progress Report* (Research Laboratory of Electronics, I.T.) 93:220–29. [MS-K]

Samuel, A. G. (1977) The effect of discrimination training on speech perception: Noncategorical perception. *Perception & Psychophysics* 22:321–30. [GE]

Salthouse, T. A. (1985) *A theory of cognitive aging*. North-Holland. [aDWM]

Sattath, S. & Tversky, A. (1987) On the relation between common and distinctive feature models. *Psychological Review* 94:16–22. [rDWM]

Scharf, B. (1970) Critical bands In. *Foundations of modern auditory theory*, vol. 1, ed. J. V. Tobias. Academic Press. [GE]

Scharf, B. & Houtsma, A. J. M. (1986) Audition II: Loudness, putch, localization, aural distortion, pathology. In. *Handbook of perception and human performance*, vol. 2, *Sensory processes and perception*, ed. K. R. Boff, L. Kaufman & J. P. Thomas. Wiley. [aDWM]

Scholes, R. J. (1968) Phonemic interference as a perceptual phenomenon. *Language and Speech* 11.86–103. [MS-K]

Shallice, T. (1988) *From cognitive neuropsychology to mental structure*. Cambridge University Press. [BDG]

Shebilske, W. L. (1987) An ecological efference theory of natural event perception. In: *Perspectives on perception and action*, ed. H. Heuer & A. F. Sanders. Lawrence Erlbaum. [aDWM]

Shepp, B. E. (1978) From perceived similarity to dimensional structure. In: *Cognition and categorization*, ed. E. Rosch & B. Lloyd. Lawrence Erlbaum. [aDWM]

Sherrick, C. E. (1984) Basic and applied research on tactile aids for deaf people: Progress and prospects. *Journal of the Society of America* 75:1325–42. [DK]

Siegler, R. S. (1986) *Children's thinking*. Prentice-Hall. [FW]

Smith, E. E. & Osherson, D. N. (1984) Conceptual combination with prototype concepts. *Cognitive Science* 7:337–61. [aDWM]

Smith, L. B. (1989) A model of perceptual classification in children and adults. *Psychological Review* 96:125–44. [FW]

Smith, L. B. & Kemler, D. G. (1977) Developmental trends in free classification: Evidence for a new conceptualization of perceptual development. *Journal of Experimental Child Psychology* 24:279–98. [aDWM, FW]

Smolensky, P. (1988) On the proper treatment of connectionism. *Behavioral and Brain Sciences* 11:1–74. [aDWM]

Spelke, E. S. (1979a) Exploring audible and visible events in infancy. In: *Perception and its development: A tribute to Eleanor J. Gibson*, ed. A. D. Pick. Lawrence Erlbaum. [aDWM]

(1979b) Perceiving bimodally specified events in infancy. *Developmental Psychology* 15:626–36. [aDWM]

Spiegel, M. F. & Watson, C. S. (1981) Factors in the discrimination of tonal patterns. III. Frequency discrimination with components of well-learned patterns. *Journal of the Acoustical Society of America* 69:223–30. [GE]

Stevens, K. N. (1968) Acoustic correlates of place of articulation for stop and fricative consonants. *Quarterly Progress Report* (Research Laboratory of Electronics, M.I.T.) 92:199–205. [MS-K]

Stroop, J. R. (1935) Studies of interference in serial verbal reactions. *Journal of Experimental Psychology* 18:643–62. [aDWM]

Studdert-Kennedy, M. (1975) From continuous signal to discrete message: Syllable to phoneme. In: *The role of speech in language*, ed. J. F. Kavanagh & J. E. Cutting. MIT Press. [MS-K]

(1976) Speech perception. In: *Contemporary issues in experimental phonetics*, ed. N. J. Lass. Academic Press. [MS-K]

(1980) Speech perception. *Language and Speech* 23:45–65. [MS-K]

(1981) The beginnings of speech. In: *Behavioral development*, ed. K. Immelmann, G. Barlow, L. Petrinovich & M. Main. Cambridge University Press. [MS-K]

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* 77:234–49. [arDWM, GE, MS-K]

Studdert-Kennedy, M., & Shankweiler, D. (1970) Hemispheric specialization for speech perception. *Journal of the Acoustical Society of America* 48:579–94. [MS-K]

Summerfield, A. G. (1979) Use of visual information in phonetic perception. *Phonetica* 36:314–31. [rDWM, MS-K]

(1983) Audio-visual speech perception, lipreading, and artificial stimulation. In: *Hearing science and hearing disorders*, ed. M. E. Lutman & M. P. Haggard. Academic Press. [MS-K]

(1987) Preliminaries to a comprehensive account of audio-visual speech perception. In: *Hearing by eye: Experimental studies in the psychology of lipreading*, ed. B. Dodd & R. Campbell. Lawrence Erlbaum. [rDWM, MS-K, REP]

Swets, J. A., Tanner, W. P., Jr., & Birdsall, T. G. (1961) Decision processes in perception. *Psychological Review* 68:301–40. [JTT]

Tartter, V. (1988) Acoustic and phonetic feature effects in dichotic listening. In: *Handbook of dichotic listening: Theory, methods, and research*, ed. K. Hugdahl. John Wiley. [MS-K]

Thomas, E. A. C. (1969) Alternative models for information processing: Constructing non-parametric tests. *The British Journal of Mathematical and Statistical Psychology* 22(2):105. [JTT]

Thompson, L. A. & Massaro, D. W. (1989) Before you see it, you see its parts: Evidence for feature encoding and integration in preschool children and adults. *Cognitive Psychology* 21:334–62. [aDWM, FW]

Tortosa Peydro, F. (1987) El articulador vibrotactil Kanievski, tras un ano de experiecia en espana. In: *Logopedia y Nuevas Tecnologias, Ponencias y Comunicaciones del Congreso Hispanoamercano de Asociaciones de Padres de Ninos Deficientes Auditivos y Symposium Internacional sobre Nuevas Tecnologias y Lenguaoes Alterntivos*. [DK]

Townsend, J. T. (1971) A note on the identifiability of parallel and serial processes. *Perception & Psychophysics* 10(3):161–63. [JTT]

(1972) Some results concerning the identifiability of parallel and serial processes. *British Journal of Mathematical and Statistical Psychology* 25:168–99. [JTT]

(1976a) A stochastic theory of matching processes. *Journal of Mathematical Psychology* 114:1–52. [JTT]

(1976b) Serial and within-stage independent parallel model equivalence on the minimum completion time. *Journal of Mathematical Psychology* 14:219–38. [JTT]

Townsend, J. T. & Ashby, F. C. (1983) *Stochastic modeling of elementary psychological processes*. Cambridge University Press. [aDWM, JTT]

Tversky, A. & Kahneman, D. (1983) Extension versus intuitive reasoning: The conjunction fallacy in probability judgement. *Psychological Review* 90:293–315. [arDWM]

Wandmacher, J., Shapiro, R. G. & Mohr, W. (1981) Letter familiarity does not aid feature extraction. *Psychological Research* 43:33–43. [LEK]

Ward, T. B. (1980) Separable and integral responding by children and adults to the dimensions of length and density. *Child Development* 51:676–84. [aDWM]

Ward, T. B. & Scott, J. (1987) Analytic and holistic modes of learning family-resemblance concepts. *Memory & Cognition* 15:42–54. [FW]

Ward, T. B. & Vela, E. (1986) Classifying color materials: Children are less holistic than adults. *Journal of Experimental Child Psychology* 42:273–302. [FW]

Warren, D. H. & Schmitt, T. L. (1978) On the plasticity of visual-proprioceptive bias effects. *Journal of Experimental Psychology: Human Perception and Performance* 4:302–10. [RBW]

Warren, D. H., Welch, R. B. & McCarthy, T. J. (1981) The role of visual-auditory "compellingness" in the ventriloquism effect: Implications for the transitivity among the spatial senses. *Perception & Psychophysics* 30:557–64. [RBW]

Warren, R. M. (1970) Perceptual restoration of missing speech sounds. *Science* 167:392–93. [NC, LEK]

Welch, R. B., DuttonHurt, L. D. & Warren, D. H. (1986) Contributions of audition and vision to temporal rate perceptions. *Perception & Psychophysics* 39:294–300. [RBW]

Welch, R. B. & Warren, D. H. (1980) Immediate perceptual response to intersensory discrepancy. *Psychological Bulletin* 88:638–67. [RBW]

Werker, J. F. (1989) Becoming a native listener. *American Scientist* 77:54–59. [NC]

Werker, J. F. & Tees, R. C. (1986) Speech perception in severely disabled and average reading children. *Canadian Journal of Psychology* 41:48–61. [BDG]

Werner, H. (1957) *Comparative psychology of mental development*. International Universities Press. [aDWM]

Wilkening, F. (1988) A misrepresentation of knowledge representation. *Developmental Review* 8:361–67. [FW]

Wilkening, F. & Lange, K. (1989) When is children's perception holistic? Goals and styles in processing multidimensional stimuli. In: *Cognitive style and cognitive development*, ed. T. Globerson & T. Zelniker. Ablex. [FW]

Wilkening, F., Levin, I. & Druyan, S. (1987) Children's counting strategies for time quantification and integration. *Developmental Psychology* 23:823–31. [FW]

Winslow, R. L. & Sachs, M. B. (1988) Single-tone discrimination based on auditory-nerve rate responses in backgrounds of quiet, noise, and with stimulation of the crossed olivocochlear bundle. *Hearing Research* 35:165–90. [GE]

Young, P. (1987) *The nature of information*. Praeger. [rDWM]

Zadeh, L. A. (1965) Fuzzy sets. *Information and Control* 8:338–53. [aDWM]