In: Michael J. Wenger, Ed; James T. Townsend, Ed. Computational, geometric, and process perspectives on facial cognition: Contexts and challenges. Lawrence Erlbaum Associates, Inc. Publishers: Mahwah, NJ, p. 285-345.

8

Face Perception: An Information Processing Framework

Christopher S. Campbell
University of California at Santa Cruz

Gudrun Schwarzer
University of Tübingen, Germany

Dominic W. Massaro
University of California at Santa Cruz

The chapters in the first half of this book have focused almost exclusively on the representations used in facial perception. This research has resulted in the creation of computational models that specify the process by which pattern spaces are constructed from sensory input. These representational models have been shown to account for a wide variety of experimental data including distinctiveness (Valentine, chap. 3, this volume), similarity, typicality (Busey, chap. 5, this volume; Steyvers & Busey, chap. 4, this volume), and generalization across viewpoint and lighting (Edelman & O'Toole, chap. 10, this volume). The purpose of this chapter, in contrast, is to discuss mathematical models for processing the psychological evidence resulting from these pattern spaces (O'Toole, Wenger, & Townsend, chap. 1, this volume; Townsend, Solomon, & Smith, chap. 2, this volume). Our approach is aimed at tackling problems associated with the hypothesized

P1: FWK Full service
CHAPT-08 LE008/Wenger July 29, 2000 17:39 Char Count= 0

rules and processes that operate on psychological evidence spaces (O'Toole et al., chap. 1, this volume). Whereas representational models are mainly concerned with information, this work focuses on information processing. Information processing models of facial perception specify the operations or procedures by which facial representations produce behavioral decisions. Our assumptions concerning the nature of information processing have been formalized in a mathematical model called the fuzzy logical model of perception (FLMP).

In this chapter, we present an information processing framework for inquiry and show how this framework can inform our understanding of facial perception. A formal modeling approach, experimental paradigm, and facial animation technology are the core components of this framework. We believe our model of inquiry provides a number of advantages for research in face processing, namely, (a) the formalization of information and information processing assumptions, (b) a common mathematical language for specifying and comparing alternative hypotheses, (c) a level of theoretical specification sufficient to falsify assumptions and hypotheses, (d) a formal distinction between representation and process or information and information processing, (e) a modeling approach flexible enough to explore a wide range of assumptions, and (f) unparalleled control and consistency of experimental stimuli through high-quality computer animation. One of the most successful models, the FLMP, is extensively reviewed because it formalizes the assumptions of this framework. The FLMP approach specifies a strong distinction between information and information processing which illustrates that this model is a powerful tool for the analysis of behavior. We show throughout that the distinction between information and information processing is central for the understanding of face perception.

The FLMP approach provides a level of specification that allows for tests of long-standing issues in the domain of facial perception. Three issues that we explore in this chapter are modularity, categorical perception, and holistic processing. Modularity is evaluated by testing the FLMP across three domains of face processing: facial affect, face identification, and facial speech. Categorical perception and holistic processing are tested through a formal modeling approach. What follows now is a brief history behind our general framework for psychological inquiry.

FUNCTIONALISM AND INFORMATION

Our approach to psychological inquiry has a long history based on the early functionalism of James (1890) and Dewey (1886) and the probabilistic functionalism of Brunswik (1956). Brunswik's lens model of perception outlines the process by which environmental stimuli are transduced by the sensory system and then evaluated and integrated. A strong distinction is made between two types of representations in this process. Ecologically valid features give reliable information about the structure of the world. Although these features are potentially useful, they may not actually be used by individuals within a given task context. Functionally valid information, on the other hand, includes only those ecologically valid features that are actually used in perceptual processing. Uncovering ecologically valid information does not inform the issue of its functional validity.

Within the present framework, we make a distinction between data and information that runs parallel to the distinction between ecological and functional validity. The sensory system transduces physical stimulation and makes available a multitude of data for further processing. Only a subset of the data in this pattern space, however, is used by the organism for a given task. These functional data are called *information*.

Brunswik proposed that functional features are only probabilistically related to perceptual categories. Thus, in any given situation there is a certain probability that a given feature will be a reliable indicator of a perceptual category. With new insights from fuzzy set theory (Zadeh, 1965) and support for continuous information in perception (Swets, 1998), the all-or-none principle of feature–category relations is no longer needed. Rather, functional features are informative to varying degrees and can therefore be represented by truth values. For example, height of an object in the vertical plane is neither a necessary nor sufficient cue for the perception of a given depth (Cutting, 1998). Height in the vertical plane only provides information about the degree to which a certain depth is present.

Both our experimental and modeling approaches are deeply rooted in Anderson's (1973, 1981, 1982, 1996) functional measurement theory of cognition. This theory proposes that the integration of different information sources (informs) can be understood through common algebraic operations such as multiplication, addition, and averaging. According to Anderson, the goal-directed nature of information processing results in valuations of environmental cues. These valuations combined with valid measurements of implicit responses provide a quantitative basis for the calculation of cognitive algebra. However, support for the validity of cognitive algebra can only come from evidence that valuations or subjective meanings of environmental cues are invariant across contexts and situations. Such evidence was provided by an array of experiments on person impression formation (Anderson, 1962, 1965, 1974).

Functional measurement provides a framework for measuring the valuations of each of the information sources, adjectives in this case, and simultaneously a test of how the valuations are combined. The most powerful aspect of this framework is that it allows specific tests, and therefore potential falsification, of various theories. According to cognitive consistency, for example, participants in the person impression experiments should attempt to reconcile two trait adjectives so that they are in concordance with each other. This involves changing the meaning of each adjective so that they provide a consistent overall impression of the evaluated person. Cognitive consistency, therefore, asserts that the subjective meanings of the adjectives are contextually dependent. Contrary to this idea, however, empirical results showed that the adjectives had an additive effect on likableness and therefore, no interaction. According to Anderson, this result supports contextual independence and thus, meaning invariance.

The validity of functional measurement supports the use of cognitive algebra in psychological inquiry. As it appears that we are justified in quantifying mental processes, the use of mathematical modeling as a tool for studying the mind becomes a reality. Undoubtedly, the results of any investigation are informative only to the degree that they distinguish among alternative theories. Our approach to experimental design, data analysis, and model testing has been devised specifically to attempt to reject some theoretical alternatives. Thus, we use a falsification and strong inference strategy of inquiry (Massaro, 1987b, 1989; Platt, 1964; Popper, 1959). Mathematical modeling is a powerful tool for exploring mental processes because it allows for specific predictions and therefore, decisive rejections of competing models. Likewise, the factorial and expanded factorial designs, manipulation of two or more variables independently, provides a rich and fine-grained data set to challenge and discriminate among theories.

INFORMATION PROCESSING ANALYSIS OF FACE PERCEPTION

Stages of information processing and hierarchical processing have been central to our approach to the study of psychological phenomena (Massaro, 1975a, 1975b, 1987b). In face processing, for example, there are at least three stages of processing: retinal transduction, sensory cues, and perceived attributes (DeYoe & Van Essen, 1988). Visual input is transduced by the visual system, a conglomeration of sensory cues is made available, and attributes of the visual world are experienced by the perceiver. There is no reason to assume that sensory cues directly map to perceived attributes in a one-to-one relation. Both one-to-many and a many-to-one relations are possible. As an example of the former, motion provides information about both perceived shape of an object and its perceived movement. In the case of a many-to-one relation, information about the shape of an object is enriched not only by motion, but also by perspective cues, picture cues, binocular disparity, and shading (e.g., chiaroscuro).

In face processing, sensory cues can imply more than one perceived attribute. For example, lip rounding can indicate the open mouth of surprise and also the rounding articulation of the consonant /r/ or the vowel /u/. Straight downward sloping eyebrows may give the impression of anger as well as providing a cue to person identity.

When perceiving emotion in the face, surprise is indicated by the perceived attributes of raised eyebrows, wide open eyes, and an open rounded mouth. These attributes, in turn, are made up of many sensory cues. For example, raised eyebrows include wrinkled skin on the forehead, rounded eyebrow shape, and wide eyebrow spacing. Sometimes when an expression lacks one or more of these cues or has contradictory cues we feel uneasy about the legitimacy of the underlying emotion. In this case, multiple cues may indicate an attempt to deceive (Ekman, 1992).

Similar to the mapping between sensory cues and perceived attributes, the mapping from perceived attributes to categories may be one-to-many and many-to-one. It is probably most common to think of several perceived attributes providing evidence for a single perceptual category. In the preceding example, surprise is given by raised eyebrows, open eyes, and open mouth. It is also possible, however, that a single attribute is evidence for many categories. For example, raised rounded eyebrows indicate surprise as well as identity and gender. Additionally, the openness of the mouth signals an emotional category (surprise), a speech category (vowel), and an identity category (my brother).

FLMP

The results from a wide variety of experiments have been described within the framework of the FLMP. Within this framework, facial processing is robust because there are usually multiple sources of information that the perceiver evaluates and integrates to achieve identification. When encountering a well-known person, for example, we not only use cues from the face to identify that person, but also how they walk, their vocal characteristics, and even distinctive clothing or jewelry. According to the FLMP, patterns are recognized in accordance with a general algorithm, regardless of the modality or particular nature of the patterns. The information processing

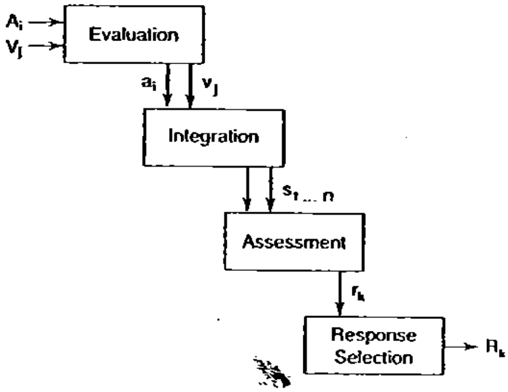

FIG. 8.1. A schematic representation of the four processes involved in perceptual recognition. The four processes are shown in sequence, left to right, to illustrate their necessarily successive, but overlapping operations. These processes make use of prototypes stored in long-term memory. Uppercase letters represent the sources of information. Auditory information is represented by A_i and visual information by V_j. The evaluation process transforms these sources of information into psychological values (indicated by lowercase letters a_i and v_j). These sources are then integrated to give an overall degree of support s_k, for each speech alternative k. The assessment operation uses these to arrive at some overall measure of support for a given alternative. The response selection process maps that value into some response alternative, R_k. The response can take the form of a discrete decision or a rating of the degree to which the alternative is likely.

assumptions central to the model are (a) each source of information is evaluated to give the degree to which that source supports the relevant alternatives, (b) the sources of information are evaluated independently of one another, (c) the sources are integrated to provide an overall degree of support for each alternative, and (d) perceptual identification follows the relative degree of support given the alternatives (Massaro & Friedman, 1990).

The FLMP assumes four stages of processing in perceptual identification: evaluation, integration, assessment, and response selection (Fig. 8.1). At the evaluation stage, input information is compared to a prototype description in memory and the degree of match is output as a fuzzy truth value between 0 and 1.[1] This input may be in the form of specific

[1] We define the output of the integration stage, s_k, as a support function, $S(k)$. The result of this support function is a real number (fuzzy truth value between 0 and 1). Alternatively, we define the output of decision, R_k, as a probability function, $P(k)$. The result of this function is a probability between 0 and 1.

features or more general dimensions of information but the model assumes all sources are independent.[2] For example, if the second (F2) and third (F3) formants are the functional auditory sources of information, and the lips are visual sources then the prototype for syllables /da/ and /ba/ are given by the following propositions:

/da/ : Slightly falling F2-F3 and open lips
/ba/ : Rising F2-F3 and closed lips

In the implementation of the model, the features for /da/ and /ba/ are mutually exclusive such that the support for one alternative can be defined as the negation of the support the other alternative:

$/da/ : a \cap v$
$/ba/ : \neg a \cap \neg v$

where a is the auditory information and v is the visual. With fuzzy truth values between 0 and 1, negation is generally defined as: $\neg x = 1 - x$ (Zadeh, 1965).[3] At the evaluation stage, support for any alternative is proportional to the degree that the stimulus matches the prototypes in memory. According to the FLMP, it is assumed that every source in the stimulus is evaluated against its prototype independently of all other sources. If the visual stimuli are denoted by V_j and the auditory stimuli are given by A_i, then

$$v_j = g(V_j)$$
$$a_i = g(A_i)$$

where i and j are levels of the auditory and visual sources of information defined in our experiment. The degree of match is given by a_i and v_j, which are continuous fuzzy truth values (real numbers) between 0 and 1. The values that a_i and v_j can obtain is some function, $g(x)$, of the stimulus

---

[2]Note that we use the terms features, dimensions, information, and sources more or less interchangeably to refer to the input to the FLMP. This information is assumed to be at the level of psychological evidence (see O'Toole et al., chap. 1, this volume). Given that the focus of this modeling approach is processing, we do not constrain the definition of a feature in the usual sense to mean a local, continuous, and holistic unit of information. Rather, each feature may be composed of many independent sources of information, include relational or configural properties, or be built from, for example, pixels distributed spatially across the stimuli.

[3]While we refer to Zadeh's work here, it should be noted that several other classes of functions have been developed for performing operations on fuzzy sets (see Dombi, 1982; Yager, 1980).

values. Because a_i and v_j are free parameters, this function is not explicitly specified in the model but, rather, is determined by the fitting process. As just seen, both sources of information are independent because their values do not depend on one another. Fuzzy truth values represent the subjective merit of each source at each level in the experiment. A value close to 0 means a good match to the /ba/ prototype, and a value close to 1 means a good match to /da/. A value around 0.5 indicates that the stimulus is ambiguous and does not support either alternative.

At the integration stage, total support, s_k, for each alternative, k, is defined as:

$$S(/da/) = a_i \cap v_j$$
$$S(/ba/) = (1 - a_i) \cap (1 - v_j)$$

According to this equation, the manner in which auditory and visual sources of information are integrated is defined as the conjoint of two fuzzy truth values (Massaro & Friedman, 1990; Massaro & Oden, 1980). One of the chief assumptions of the FLMP is that information sources are combined according to a multiplicative rule. Generally, the conjunction of fuzzy truth values is defined as: $x \cap y = x \times y$. Given this, the combined support for /da/ and /ba/ for each of the $i \times j$ conditions is:

$$S(/da/) = a_i v_j \tag{1}$$
$$S(/ba/) = (1 - a_i)(1 - v_j) \tag{2}$$

More generally, the multiplicative support for any alternative, k, given n sources of information is:

$$S(k) = \prod_{x=1}^{n} f_x \tag{3}$$

where f_x is the evaluated sources of information, x indexes over all sources, and n is the total number of sources. This represents the general form of the equation for combining independent sources of information.

After the support for each alternative is found, a final decision is performed with two operations: assessment and response selection. The assessment operation finds the total support for some alternative relative to the support for all relevant alternatives. Response selection follows a probability matching rule in which the likelihood of a given response is equal to its relative goodness of match to the input. These two operations are summarized in the relative goodness rule (RGR), which is closely related to Luce's (1959, 1977) choice axiom. The RGR gives the probability of responding for each alternative in each condition of the experiment. The general form of the RGR is the probability:

$$P(k) = \frac{S(k)}{\sum_{k=1}^{m} S(k)} \tag{4}$$

where k is the alternative and m is the number of relevant alternatives. Applying this equation to our example with just two alternatives we have:

$$P(/da/|A_i V_j) = \frac{S(/da/)}{S(/da/) + S(/ba/)}$$
$$= \frac{a_i v_j}{a_i v_j + (1 - a_i)(1 - v_j)}$$

This equation states the probability of /da/ given the auditory and visual stimuli is equal to the support for /da/ divided by the support for /da/ plus the support for /ba/. Using this equation we can predict the probability of responding in each of the conditions of the experiment. Each level of the a and v stimuli is treated as a free parameter. Thus, for a factorial experiment we have $i + j$ free parameters to predict $i \times j$ conditions. For an expanded factorial experiment we have $i + j$ unimodal levels to predict $i \times j$ bimodal conditions plus $i + j$ unimodal conditions.[4]

Although the FLMP represents the assumptions of our framework, it is relatively straightforward to formulate alternative hypotheses. As stated earlier, Anderson (1962) first proposed that information integration for person impression formation is additive. Changing multiplicative to additive integration involves merely adding the support from each source instead of multiplying. Thus, Equation 3 now becomes:

$$S(k) = \sum_{x=1}^{n} f_x \tag{5}$$

For the preceding example, the combined support for auditory and visual sources of information becomes:

$$S(/da/) = a_i + v_j$$
$$S(/ba/) = (1 - a_i) + (1 - v_j)$$

[4]For both the factorial and expanded factorial experimental designs, every level of one stimulus is presented with every level of another in all possible combinations. For the expanded factorial design, however, every level of both stimuli is also presented alone.

The probability of responding for /da/ and /ba/ is now:

$$P(/da/|A_i V_j) = \frac{a_i + v_j}{a_i + v_j + [(1 - a_i) + (1 - v_j)]} = \frac{a_i + v_j}{2}$$

$$P(/ba/|A_i V_j) = \frac{(1 - a_i) + (1 - v_j)}{a_i + v_j + [(1 - a_i) + (1 - v_j)]} = \frac{(1 - a_i) + (1 - v_j)}{2}$$

Other assumptions about perceptual identification can also be represented. For example, some theories assume that perceptual information is not integrated at all. Rather, responses are based only on one source of information at a time. This single channel model (SCM) is constructed by assuming that the individual has some bias β for using one source or the other. To compute the influence from one source, simply multiply it by the bias parameter. The general form of the SCM equation for n sources of information is:

$$S(k) = \sum_{x=1}^{n} \beta_x f_x \tag{6}$$

Using p as the bias parameter for auditory information and $1 - p$ as the bias for visual, Equations 1 and 2 become:

$$S(/da/) = pa_i + (1 - p)v_j$$
$$S(/ba/) = p(1 - a_i) + (1 - p)(1 - v_j)$$

The probability of responding for /da/ and /ba/ is now:

$$P(/da/|A_i V_j) = \frac{pa_i + (1 - p)v_j}{pa_i + (1 - p)v_j + [p(1 - a_i) + (1 - p)(1 - v_j)]} = pa_i + (1 - p)v_j$$

$$P(/ba/|A_i V_j) = \frac{p(1 - a_i) + (1 - p)(1 - v_j)}{pa_i + (1 - p)v_j + [p(1 - a_i) + (1 - p)(1 - v_j)]} = p(1 - a_i) + (1 - p)(1 - v_j)$$

Once formulated, models are tested by fitting their predictions to observed data. Fitting is typically performed by iteratively adjusting the parameter values until the difference between the observed and predicted data is minimized. To measure this difference we use the root mean squared deviation (RMSD). The model with the lowest RMSD is assumed to fit or describe the data better than all competing models. Thus, our criteria for falsification is a quantitative measure as given by goodness of fit. Although we cannot prove conclusively that the best fitting model is true, we can rule out poorer fitting models since they give a less accurate description of the data.

Whenever possible we also use a qualitative falsification strategy like that outlined by Wenger and Townsend (chap. 7, this volume). According to this strategy models are falsified when their predictions are inconsistent with the observed data. Here there is no need for goodness-of-fit measures. Either the pattern of results predicted by the model is seen in the data or it is not. The FLMP makes a qualitative prediction that the combined support from multiple sources will be superadditive or some value greater than the sum of support. In contrast, the weighted averaging model (WTAV) predicts that the combined support cannot be greater than the sum of support from all sources. Because these predictions are mutually exclusive, the data must falsify one model. As demonstrated later, the observed data clearly show superadditivity, thereby falsifying the WTAV.

Relative RMSD values are one way to compare models, but to obtain a more absolute measure of performance, one must calculate how well the model would fit under ideal conditions. This type of fit is called a *benchmark* and indicates the best possible accuracy of the model given a certain number of observations or samples per condition. As the number of observations increases, the benchmark RMSD, RMSD(b), approaches 0.0. For example, an RMSD(b) of .0238 for model A and .0492 for model B indicates that we would expect more accuracy (less sampling variability) for model A and possibly better RMSDs.

To compute the benchmark we first generate an ideal set of data using the model under consideration. The model is fit to the observed data, yielding a set of predicted data points. Because the model generated the predicted data, refitting the model to these data would result in a perfect fit (RMSD = 0.0). Using Monte Carlo simulation we resample each predicted data point as follows. A random number between 0.0 and 1.0 is selected from a uniform distribution. If we have two alternatives, A and B for example, a random number below the predicted proportion is recorded as an A response. Otherwise, a B response is recorded. This Monte Carlo resampling is performed *N* times for each predicted data point where *N* is the number of observations in the experiment. Given some number of observations less than infinity, sampling variance will be introduced into the data by the simulation. The RMSD(b) is then calculated by fitting the model to this simulated data set. For a more detailed explanation of benchmarking methods see Massaro (1998).

Our view of facial perception is specified within the formulation of the FLMP, but the modeling approach presented here is a valuable tool for inquiry more generally. Modeling provides the opportunity to formalize our theories using a common mathematical language. This in turn allows for the testing and even falsification of competing hypotheses. Formal modeling also forces researchers to produce a fairly detailed, well-developed account of their theoretical positions. Verbal theories that are vague or incomplete cannot be easily formalized. The exercise of formalizing theories may highlight these difficulties and in itself prove to be useful for theory development.

SYNTHETIC STIMULI: BALDI

For much of our research we use a computer animated talking head called Baldi (see Fig. 8.2) instead of a natural person. Synthetic stimuli provide the precise control and standardization needed in psychophysical experimentation. Many times, using synthetic stimuli is the easiest if not the only way to manipulate the variable of interest. For example, if we wish to test speechreading accuracy without jaw rotation we can either wire a natural speaker's mouth shut or simply disable the jaw rotation parameter in Baldi. Additionally, if the rate of speaking needs to be increased or decreased only Baldi can change his speaking rate consistently. Finally, Baldi can be used to create conflicting features or ambiguous stimuli much more easily than with humans.

Baldi is constructed from about 900 triangular polygons joined at the edges to form the three-dimensional head with eyes, pupil, eyebrows, nose, skin, lips, tongue, and teeth. Baldi's name stems from the fact that Baldi has no hair. Generating hair would require additional polygons or some type of texture mapping process and would significantly slow down the facial animation. To give Baldi a more natural appearance, the surface of his skin is smooth shaded using the Gouraud method. The head shape and

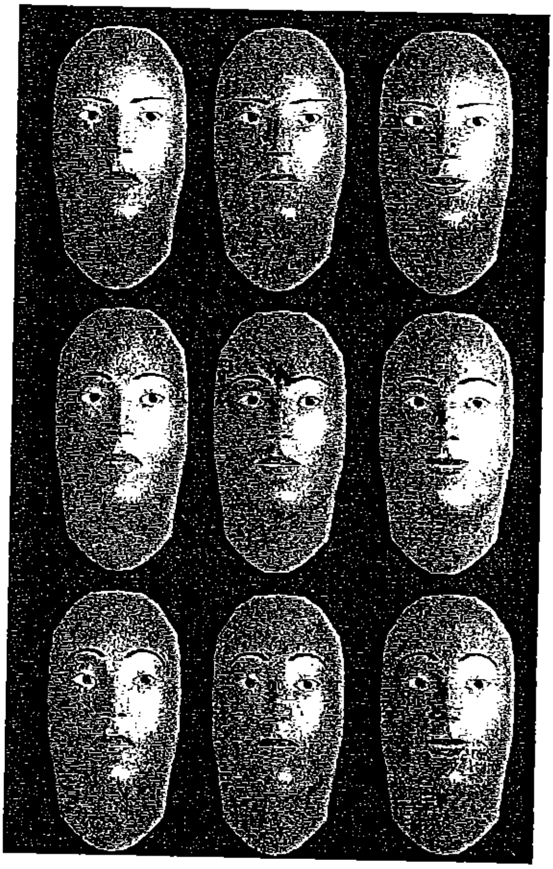

FIG. 8.2. The four faces displaying the maximum feature displacements (at the corners) as well as faces displaying "neutral" displacements. The center face is the "neutral baseline" face, with both mouth and brow displaced at the neutral values. Note that some faces are quite ambiguous and incongruent in their expressions. The unimodal (half-face) conditions displayed only the upper or lower half of the stimulus face.

movement are completely parameterized through a set of about 76 control parameters to permit real-time (30 frames per sec) animation of visible speech and facial expressions. Realistic speech is created by changing the parameters over time according to the overlapping dominance functions of nearby phonemes (Cohen & Massaro, 1993). In this way coarticulation or the influence of neighboring speech segments with each other can be captured in the synthesis (see Massaro, 1998).

P1: FWK Full service
CHAPT-08 LE008/Wenger July 29, 2000 17:39 Char Count= 0

INFORMATION VERSUS INFORMATION PROCESSING

One of the main distinctions throughout this book has been the difference between information and information processing. We believe our model, FLMP, provides a fine-grained analytical tool for separating contributions of information and information processing in facial perception. The FLMP has been used primarily as a model of information processing in the sense that it specifies how sources of information are evaluated, integrated, and selected. Although the sources of information for a given task are specified in the formulation of the model, the exact information value for each source is a free parameter. This allows the FLMP to account for individual differences by the degree to which individuals utilize each source of information. As a result, measures of model performance indicate how well the information processing assumptions of the FLMP describe the observed data independent of the information used by each individual. Information processing is quantitatively specified and the final parameter values offer a theoretical account of the information value for each source for each individual. Thus, the FLMP is not only a method of testing process models, but also can be used to simultaneously explore the nature of the information for a given perceptual task.

Investigations of perception are usually confined to young adults, and data analyses are limited to group averages. The framework of the FLMP, on the other hand, provides a formal analytical method to investigate the behavior of individual participants. As is well known, group results may not represent any of the individuals making up the group. Using the FLMP approach, we can explore individual differences across age groups, gender, races, native languages, and hearing or visual impairments in both information and information processing (Massaro, 1998). One such study, reported in Massaro (1987b), presented preschool children and fourth-grade students with auditory, visual, or bimodal speech. The stimuli varied across a 5 level auditory continuum from /ba/ to /da/, and 2 visual levels, either /ba/ or /da/. This resulted in a 2×5 expanded factorial design in which the children were asked to categorize the stimuli. Model tests of the resulting data showed that the FLMP described the children's performance quite well for both the preschoolers and the fourth graders. Thus, information processing appears to remain constant through development.

What then can account for the consistent finding that overall correct responding is better with increasing age? Analysis of the parameter values was performed to determine the overall influence of each source of information. Analysis of the parameter values is only meaningful if the form of the model is identifiable. In other words, there must be only one possible set of parameter values for the model. In this case, an expanded factorial design was used, ensuring that the parameter values were unique (Crowther, Batchelder, & Hu, 1995; Massaro, 1998).

Given 5 parameters for the auditory source and 2 for the visual source, the amount of influence was taken as the range of parameter values. Because the parameter values varied between 0 and 1, so did the range measure. A large range would indicate a strong effect, whereas a small range would show little effect. Preschool children showed a weaker effect (.483) of the auditory information than fourth graders (.845). Preschool children also showed less of an effect of visual information (.178) than fourth graders (.304). Thus, improvements in performance result from changes in information over the course of development. As children become more experienced with the world around them, they learn what information is useful. How they use this information in speech perception, however, does not appear to change. This research is an example of how the FLMP provides a powerful method for examining individual differences and the basis of those differences in terms of information and information processing.

CHALLENGES FOR THE FLMP

Three long-standing issues debated in the study of facial processing include categorical versus continuous perception, holistic versus analytical processing, and modularity versus general pattern recognition.[5] Several current theoretical and empirical positions in facial perception seem to support categorical information in perception, holistic processing (features are processed dependently), and the hypothesis of modular processes. Because it is assumed in the formulation of the FLMP that information is continuous and features are processed independently, these positions pose a serious challenge for the FLMP. Within the present framework,

[5]The term *pattern recognition* as used here refers to the process of identifying a stimuli and not to the task of old–new recognition (see O'Toole et al., chap. 1, this volume).

perception is viewed as a general process of pattern recognition. This implies that information processing remains constant across perceptual tasks. Thus, modularity or the idea that different perceptual tasks involve different forms of information processing also conflicts with the present framework.

Categorical Perception

Categorical perception has long been an issue in the domain of speech perception but it has more recently come to the forefront in face perception (Beale & Keil, 1995; Cottrell, Dailey, Padgett, & Adolphs, chap. 9, this volume; Etcoff & Magee, 1992). The proposal that features are perceived categorically directly contradicts the FLMP's assumption of continuous information. A categorical model claims that emotion perception is discrete in that gradations of emotion are not easily perceived within an emotion category. Although there is a long history of categorical perception in speech research, theorists currently seem to agree that perceivers have within-category information that is functional in speech perception. This viewpoint did not emerge easily and sometimes a bit of theoretical regression reaches the airwaves. Unfortunately, this progress has not transferred to research on the processing of faces. We review and criticize a few recent experimental claims for categorical perception to set the stage for our research.

Etcoff and Magee (1992) presented faces created by a weighted averaging of line drawings of exemplar faces displaying different emotional expressions. Following the tradition in speech perception studies, they carried out both identification and discrimination tasks. The former requires a categorization, whereas the latter asks for noticing a difference. The identification results showed a systematic change in the identification judgment as the face changed from one emotion category to another. In the ABX discrimination task, three faces were presented, the first two of which differed. The participant was asked to tell which one was identical to the third face. Discrimination performance was better for pairs of faces that tended to be identified as different emotions than for pairs identified as the same emotion. Given that category identity appeared to undermine discrimination, Etcoff and Magee concluded that these facial expressions were perceived categorically because pairs of equally spaced faces along the stimulus continuum did not appear to produce equivalent discrimination differences. Two stimuli within a category were supposedly more poorly discriminated than two stimuli from two different categories.

The emotion results are similar to previous findings of categorical perception in speech but we now have alternative explanations. It is now well known that discrimination tasks underestimate discrimination capacity (Massaro, 1987a). Many discrimination tasks have memory limitations and performance is easily influenced by the participant's use of category labels. The ABX task, for example, makes it difficult to compare the third stimulus X to the first stimulus A. In this task, participants often encode the stimuli categorically and base their discrimination decision on these category labels. Better discrimination for items in different categories than in the same category does not conclusively show that perception is categorical.

More important, categorical perception research does not follow a falsification strategy of inquiry. Given a stimulus continuum between two alternatives, a typical result is that the identification judgments change rather abruptly around the category boundary with changes along the stimulus continuum. Several researchers, like Etcoff and Magee (1992), have interpreted these prototypical findings as evidence for categorical perception. One error in this interpretation, however, is that the dependent measure, proportion of judgments, is being treated as a linear measure of perception. In fact, it has been shown that this type of observed identification function follows directly from continuous perception (Massaro, 1987a, 1987b). Sharp identification boundaries between categories follow naturally from a system with continuous information and a decision criterion (see Massaro, 1987b).

The most direct measure of whether perception is continuous or categorical involves comparing quantitative tests of models that assume either continuous or categorical information (Massaro, 1998; Thomas, 1996). Unfortunately, most categorical theories do not allow compositional determination and are therefore not easily formalized to make testable predictions for this task. For both types of theories, it might be claimed that perception of each face is unique and cannot be predicted from performance on the parts that make it up. On the other hand, there are several other ways categorical perception can be tested. There is a specific categorical model of perception (CMP) in which the participant categorizes information from each feature and responds with the outcome of the categorization of only one of the features with a certain probability, or bias toward that feature. Because this CMP is mathematically equivalent to the SCM, in which the perceiver identifies the stimuli using just a single source of information, a poor fit of the SCM relative to the fit of the FLMP would also provide evidence against this model. Of course, other categorical models are possible and one of these might provide an adequate description of the results. However, the falsification of our categorical model has stood for over two decades and no one has offered a successful alternative categorical model to support the idea of categorical perception.

Modularity

The modularity hypothesis assumes different modes of processing for faces and objects. The FLMP algorithm accounts for the integration of information from different modalities (perception by ear and eye), which challenges the modularity hypothesis in that it attributes differences between the recognition of different modalities or different domains like faces and objects to differences in information. Farah (1995) provided some evidence for a dissociation between face recognition and object recognition. Within our framework, she located this difference at information processing, not information. By different systems for face and object recognition, she meant that "two different systems must: (a) be functionally independent, such that either can operate without the other; (b) be physically distinct; and (c) process information in different ways, so that it (one system) is not merely a physical duplicate of another" (p. 102). Farah's third criterion for different systems is consistent with our belief that previous arguments for modular systems have meant differences in information processing, not simply differences in information.

Farah studied a man, called LH, who was prosopagnosic. People with this neurological disorder have difficulty recognizing the faces of loved ones and well-known celebrities. In one study, he recognized faces and eyeglass frames about equally poorly, whereas normal participants showed a significant 20% advantage for face over eyeglass frames. Furthermore, inverting faces disrupted performance for normal participants somewhat but actually improved performance for LH. These results could have resulted from differences in both information and information processing, or just in information. For example, the loss of configural information for normal participants could account for the poorer performance. For LH, the loss of configural information could have produced better performance by making faces more like nonface objects. One way of testing between these two possibilities would be to utilize the microscope of the expanded factorial design and model testing as in R. Campbell, Zihl, Massaro, Munhall, and Cohen (1997).

Holistic Face Processing

In view of the fact that the FLMP assumes independent features, holistic models of face processing challenge the present framework. It is worthwhile to describe these holistic models and evaluate their conclusions in the context of the FLMP. Holistic processing is a loaded term that is easily criticized but, fortunately, researchers have begun to clarify what they mean by holistic processing (Farah, Wilson, Drain, & Tanaka, 1998). Farah, Tanaka, and Drain (1995) and Carey and Diamond (1994) articulated two different characterizations of holistic processing of the face. The terms *holistic encoding* and *configural encoding* are used to describe these two viewpoints. In holistic encoding, the parts of the face are not separately represented and utilized. Rather, the face is represented as a whole.

As evidence for holistic processing, Tanaka and Farah (1993) found that individual facial features were recognized more easily when displayed as part of a whole face than when displayed in isolation. Whereas recognition of individual features of faces was facilitated by the context of the whole face in normal orientation, recognition was not facilitated in the context of scrambled faces, inverted faces, or houses. In line with these results, Tanaka and Sengco (1997) demonstrated that alterations in facial configurations interfered with the retrieval of facial features, whereas the interference did not appear with inverted faces or nonface stimuli. Moreover, Farah et al., (1998) used a selective attention paradigm and a masking paradigm and compared the perception of faces with the perception of inverted faces, words, and houses. They showed that faces are not only represented more holistically than other stimuli, but also that in immediate perceptual memory and during perception the holistic mode of processing dominated.

These findings and those of Tanaka and Farah (1993) suggested that facial recognition is in some sense a holistic process, differing qualitatively from the recognition of other types of images. They claimed that "the representation of a face used in face recognition is not composed of representations of the face's parts, but more as a whole face" (Tanaka & Farah, 1993, p. 226). In this view, parts of the face are not the atoms of face analysis or representation. This viewpoint is closest to the traditional use of holistic processing in that it bears great similarity to a template matching scheme. According to this viewpoint, the parts of the face would not be as accessible as the complete face.

On the other hand, we must wonder whether the predictions of holistic models are really falsifiable. It seems almost as reasonable to expect the holistic view to predict that the complete face would camouflage one of its parts rather than facilitate its perception. In fact, an advocate of holistic processing in word perception has continuously argued exactly this point (Johnson, 1975; Johnson & Blum, 1988; for an early critique, see Massaro & Klitzke, 1977). Thus it seems that an advocate of this version of holistic perception could have "predicted" either outcome, facilitation or inhibition.

A commendable goal of formalizing models is to prepare them for experimental tests. Unfortunately, we know of no holistic model that can be quantitatively tested against the results. The class of holistic models called holistic encoding would assume that each unique feature combination would create a unique face that could not be predicted from its component features. This formulation captures the idea that somehow the whole is more than some combination of its parts. We are not able to test a specific quantitative formulation of this holistic model because it requires as many free parameters as observed data points. Every face is unique and its identification cannot be predicted on the basis of its components. This model remains untestable until there is some implementation of its principles with fewer free parameters. However, regardless of whether a particular holistic encoding model can be tested, an adequate fit of the FLMP provides evidence against the class of holistic encoding models. If the processing of the whole face is not a function of its component features, then a model assuming that the value of the whole is derived from the values of its parts should fail.

The second characterization of holistic processing, called configural encoding, refers to the possibility that the spatial relations among the parts of the face are more influential than the parts themselves. The parts are represented but it is the relations among the parts that are critical for analysis. This interpretation of holistic processing is also consistent with Tanaka and Farah's (1993) finding that individual facial features were more easily recognized when part of the complete face than when presented alone. According to this view, the complete face would provide spatial relations that would not be available in a part of the face presented in isolation. We have no objection to this possibility. In the framework of the FLMP, a relation between two parts of the face could function as an additional source of information. The configural feature would be encoded and evaluated independently like the isolated features. Then, all features both configural and isolated would be combined multiplicatively.

Unfortunately the hypothesis of configural encoding is also difficult to test. The nature of the relation between features has not been specified as of yet. Is this relation the spatial distance between features, the relative positions of the features, or the angle of a straight line connecting them? In terms of our typical factorial design, we manipulate two factors independently along a continuum. If every combination of these two factors forms a different relational feature, then we have as many parameters as data points. Clearly, this is an untestable model.

Although these issues have been extensively addressed in the facial perception literature, the FLMP approach provides a level of specification sufficient to critically test between the competing positions. This framework, therefore, offers the potential to falsify alternative explanations and advance our understanding of facial perception. Unfortunately, theories that are not specific enough to be formalized cannot be tested against the FLMP and cannot be quantitatively falsified. Despite this, we attempt to formulate and test a holistic model in the facial identity section later. This formulation requires additional assumptions about independence and dependence at various stages of processing. In addition to this test of competing models, however, the fit of the FLMP alone still allows us to determine roughly how well a model that assumes analytic processing predicts the data. This measure of fit can be evaluated using either previous fits to other data sets as a rule of thumb or the benchmarking procedure described earlier.

FACIAL AND OTHER CUES TO EMOTION

There is no doubt that the production of facial expressions is an effective means of communicating emotion. Darwin (1872) argued that facial expressions have their origins in basic acts of self-preservation common to human beings and other animals, and those acts were related to the emotional states now conveyed by the descendent expression.

We recognize and characterize facial expressions of emotion in other humans with a high degree of accuracy and consistency (Collier, 1985; Ekman, 1993; Ekman & Friesen, 1975; Ekman, Friesen, & Ellsworth, 1972). The face is not unique in this regard, in that we are also tuned to various nonfacial displays of emotional arousal. Hand and body gestures are well-known communicators of affective states (Archer & Silver, 1991). Even other species produce and respond to visible displays of emotion. Parakeets, for example, are sensitive to the size of the iris (Brown & Dooling, 1993). This cue is only one of several that parakeets use to signal relevant information, such as sex, age, and emotional arousal. These cues were shown to be highly functional because they were discriminated more quickly than other nonfunctional features.

Varying Ambiguity in the Identification of Emotion

Baldi, our talking head, makes possible a set of quite realistic faces for research that are standardized and replicable, as well as controllable over a wide range of feature dimensions. Displays of ambiguous or contradictory features or partial face presentations can be made more easily than with previous types of facial stimuli (see Fig. 8.2). Thus, it quickly became apparent that we could initiate a cottage industry in the study of facial and vocal cues to emotion. There was no shortage of literature on facial cues to emotion but we found a tremendous void in the domain of vocal cues. We learned that Baldi had to be given increased resolution in certain parts of the face, as well as additional controls over these parts.

We use the expanded factorial design to study the pattern recognition of emotion (Ellison & Massaro, 1997). The affective categories happy and angry were chosen because they represent two of the basic categories of emotion. Of course, happy and angry faces are not discrete, nonoverlapping emotional displays, but a face can vary in the degree to which it represents one emotion as opposed to the other. To implement the expanded factorial design, it was necessary to choose two features to vary systematically to create a range of emotions between happy and angry.

We chose two features that seem to differ somewhat in happy and angry faces. The features varied were brow displacement (BD) and mouth corner displacement (MD). As can be seen in Fig. 8.2, BD was varied from slightly elevated and arched for a prototypically happy emotion to fully depressed and flattened for a prototypically angry emotion. MD was varied from fully curled up at corners for a prototypically happy emotion to fully curled down at corners for a prototypically angry emotion. An important criterion for manipulating two features is that they can be varied independently of one another. Thus, varying one cue in the upper face and one cue in the lower face was an ideal solution. Furthermore, there appear to be motor neurons from the neocortical motor strip in which the upper and lower face are served by different neurons (Fridlund, 1994). Five levels of the upper face conditions and 5 levels of the lower face conditions were factorially combined, along with the 10 half-face conditions presenting the upper face or lower face alone. The feature values were obtained by comparison to features displayed in exemplar photographs in Ekman and Friesen (1975).

These two features are neither necessary nor sufficient for happy or angry faces, but they are simply correlated with these emotion categories. Like other categories, emotion categories are fuzzy in that no set of necessary and sufficient features characterizes a particular emotion. Even for natural faces, there is some controversy concerning the degree to which observers can accurately categorize different emotions. As concluded by Fridlund (1994), there is no evidence for the claim that a given facial expression is unambiguously linked with a single emotion category. In addition, several other features are also correlated with these affective categories. For example, there is a tendency for a tightening around the eyes and a lifting of the cheeks in spontaneous smiling (Allen & Atkinson, 1981; Duchenne de Boulogne, 1990; Ekman et al., 1981). This is another example of the one-to-many and many-to-one relation between sensory cues and perceived attributes. We limited our study to just two features to keep the number of unique faces reasonably small and the number of test observations relatively large. Our task was a two-alternative forced choice between HAPPY and ANGRY. There were 35 different test faces. Participants were not shown any exemplar faces, nor were they given any feedback. After 10 practice trials, each stimulus face was randomly presented 16 times to each of 26 participants for identification.

The points in Fig. 8.3 give the observed average results as a function of the mouth and brow variables. The left panel shows performance when just the upper half of the face was presented. Changes in the displacement of the brow were effective in changing the identified emotion in the expected direction. Similarly, the lower half of the face influenced the number of "happy" judgments in the anticipated way. The steeper curve for the mouth variable illustrates that it was somewhat more influential than the brow variable. The middle panel gives the factorial combination of the two halves of the face. As can be seen in the figure, each of the two variables continued to be influential even when paired with the other variable.

The average results show most conclusively how two sources of information are more informative than just one. The probability of a happy judgment was about .80 when just the most upward deflection of the brow was presented and was about .88 for the most upward deflection of the mouth. However, when the two features were presented together in the whole face, the probability of a happy judgment was near 1. An analogous result was found for the most downward deflection of these two

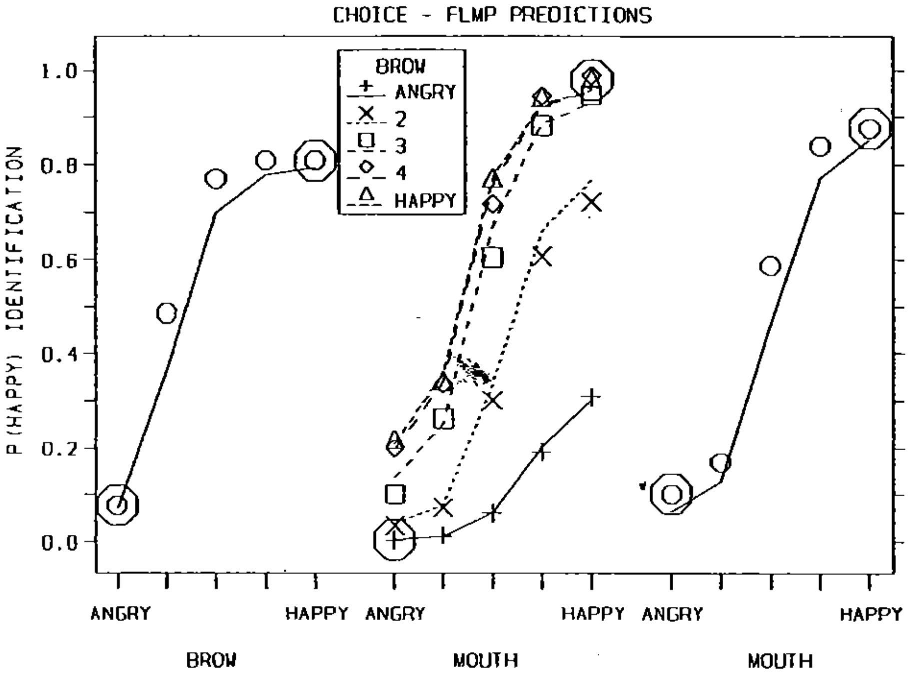

FIG. 8.3. Predicted (lines) and observed (points) proportion of happy judgments as a function of the levels of the brow and mouth variables. The left panel shows performance for just the upper half of the face and the right panel for just the lower half. The middle panel gives performance for the factorial combination of the two halves. Average results across 26 participants are shown. The circled points show the superadditivity predicted by the FLMP (from Ellison & Massaro, 1997, Experiment 1).

variables. These superadditive outcomes are consistent with our general view of pattern recognition. We now derive the predictions of the FLMP to test the model quantitatively against all of the results.

Implementation of the FLMP

In our implementation of the FLMP for emotion perception, participants are assumed to have prototypes corresponding to happy and angry faces. A happy face is characterized by the eyebrows slightly elevated and arched and the mouth corners fully curled up. An angry face is represented as having the eyebrows fully depressed and flattened and the mouth corners fully curled down. Of course, there are other sources of information described in the prototypes, but these do not require our attention because they should not be influenced systematically by the two independent variables.

Feature integration consists of a multiplicative combination of the feature values supporting a given alternative. Thus, support for each alternative is:

$$S(H) = b_i m_j$$
$$S(A) = (1 - b_i)(1 - m_j)$$

The probability of an H response is then:

$$P(H|B_i M_j) = \frac{S(H)}{S(H) + S(A)}$$

where $P(H|B_i M_j)$ is the predicted choice given stimulus B_i and M_j.

As in the case of bimodal speech, the FLMP requires 10 free parameters for the 5 levels of BD and the 5 levels of MD. These 10 parameters are used to calculate the percentage correct identification in all 35 conditions. In the two-choice identification task, the FLMP's RMSDs for individual participants ranged between .047 to .128 with an average RMSD of .082. As can be seen in the figure, the FLMP gives a good account of the results. When both the brow and the mouth are deflected upward, the face is perceived to be happy. The reader might have also noticed that only half of the American football is present in the factorial part of the design. This simply means that the mouth variable did not give enough support for happy to dominate the judgments at the right side of the factorial plots. Thus, the brow variable did not provide unambiguous support for the happy emotion. The half of a football is consistent with the asymmetry of the parameter values. Our justification for interpreting the parameter values is based on the use of an expanded factorial design (Massaro, 1998). The average parameter values for the brow variable were .046, .349, .711, .788, and .804 as this variable was changed from angry to happy. The analogous values for the mouth variable were .051, .107, .479, .823, and .881. In both cases, the parameter values are more extreme at the angry than at the happy end of the continuum. In this case, a downward deflection of the brow will carry more influence than the upward deflection of the mouth, and analogously for the reverse pairing.

Reaction times (RTs) of the identification judgments were also analyzed. The RTs of the identification judgments can be used to test the FLMP's prediction that RT should increase to the extent the facial information is

PI: FWK Full service
CHAPT-08 LE008/Wenger July 29, 2000 17:39 Char Count= 0

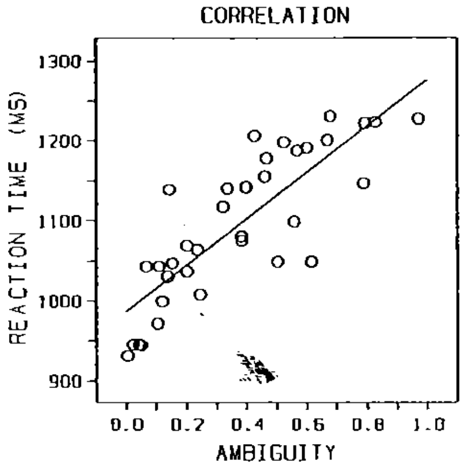

FIG. 8.4. Reaction times (RT) averaged across all participants and plotted as a function of ambiguity (A given by Equation 7) for each of the 35 conditions of the expanded factorial design (from Ellison & Massaro, 1997, Experiment 1).

ambiguous. Ambiguity is defined as the extent to which the probability of a judgment, in this case $P(Happy)$, approaches 0.5.

$$A = 1 - 2(|0.5 - P(Happy)|) \tag{7}$$

Thus, ambiguity varies between 0.0 when $P(Happy)$ is 0 or 1, and 1.0 when $P(Happy)$ is 0.5. An RT averaged across all participants was computed for each of the 35 stimulus conditions and correlated with the A values computed from the average results of the identification task. Figure 8.4 shows the strong positive (.83) correlation between this measure of ambiguity and RT.

Ambiguity predicts RT for both the bimodal and unimodal conditions. For the unimodal condition when a half-face is made more ambiguous, then its identification RT increases. For the factorial conditions, RTs appear to increase to the extent the two half-faces are both ambiguous or when they conflict with one another (e.g., a "happy" brow and an "angry" mouth, which creates an ambiguous stimulus).

Given this simple relation between identification judgments and RTs, the preceding ambiguity equation could be incorporated into the formulation of the FLMP so that the FLMP could predict RTs about as well as identifications. This would not require the use of identification judgments for $P(Happy)$ as this term could be replaced by $P(H\,|\,B_i M_j)$ shown previously

P1: FWK Full service
CHAPT-08 LE008/Wenger July 29, 2000 17:39 Char Count= 0

and in Equation 4. Thus, the assumptions of the FLMP could potentially be shown to hold for RT data as well as identification judgments.

Rating Judgments

Rating judgments also provide a valuable dependent measure of pattern recognition. Ellison and Massaro (1997) also obtained rating judgments. The procedure was identical to the identification task except that the 22 participants received instructions to rate the emotion on a scale from 1 to 9. Figure 8.5 shows the ratings averaged across the participants, along with the predictions of the FLMP. The independent variables influenced performance in the same manner as in the two-choice task.

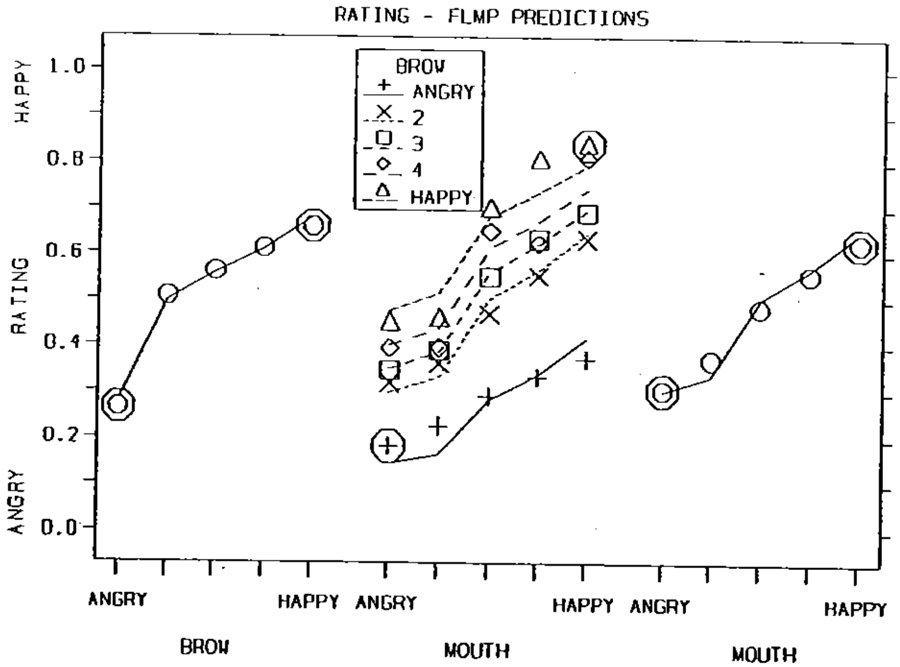

FIG. 8.5. Predicted (lines) and observed (points) rating of happy judgments as a function of the brow and mouth conditions. The left panel shows performance for just the upper half of the face and the right panel for just the lower half. The middle panel gives performance for the factorial combination of the two halves. Predictions are for the FLMP. The circled points illustrate the superadditivity predicted by the FLMP (from Ellison & Massaro, 1997, Experiment 2).

Although we show just the average results, the ratings for individual participants follow the predictions of the FLMP. We have circled 6 points in Fig. 8.5 to illustrate that the superadditivity predicted by the FLMP holds for rating judgments as well as for identification judgments. We circled points that supported the same alternative. As can be seen in the figure, the rating judgment given the two sources of information combined is more extreme than the judgment given either one of these sources presented alone. These results are strong evidence against a WTAV or SCM in which the rating for two sources of information cannot be more extreme than the rating for either source presented alone. Consistent with this observation, the model tests for the rating judgments gave the same conclusions as for the identification judgments. The RMSD for the FLMP fit to average rating data was .047 compared to an RMSD of .076 for the fit of the WTAV and SCM.

FACE IDENTITY

The recognition and identification of emotional expressions has usually been isolated in the literature from the processing of facial identity. One testable assumption, however, is that the identity of faces is derived from the features that make them up in the same manner that the expression of a face is computed from facial features. In other words, the difference between the two domains may be one of different subsets of features rather than information processing. Although it is necessarily the case that the features for facial expression differ from those for facial identity, the processing involved in these two domains could be identical. Previous findings of dissociations between emotion and identity, such as segregated processing in the brain (e.g., Sergent, Ohta, MacDonald, & Zuck, 1994), might reflect only differences in information.

Notwithstanding the large number of faces potentially stored in memory and the high degree of similarity among faces, a known face is easily identified in about half a second. Bahrick, Bahrick, and Wittlinger (1975) found above 90% recognition of yearbook photos of schoolmates, independent of class sizes between 90 and 800, and independent of time from graduation between 3 months and 35 years. We continue to learn new faces with ease. People can successfully encode large numbers of new faces from photographs inspected only briefly (5 sec each) and subsequently pick these from distracters at recognition rates of over 90% (Carey, 1996).

Besides this enormous capability to learn and remember faces, another striking aspect of face processing is its robustness under certain manipulations. For example, a face can successfully be identified even after changes in expression, illumination, or when distorted as in caricature (Ekman, 1973; Rhodes, Brennan, & Carey, 1987; Troje & Bülthoff, 1996). Research has also confirmed that identification is almost unaffected by a change of viewpoint (Valentin, Abdi, Edelman, & Posamentier, chap. 11, this volume). Whereas identification performance was unaffected by moderate transformations from full frontal face to three-fourths view between presentation and test (Davies, Ellis, & Shepard, 1978; Patterson & Baddely, 1977) performance was somewhat decreased when the face was in profile (Galper & Hochberg, 1971). The best transfer of performance is observed between faces taken from mirror symmetric views (Troje & Bülthoff, 1997).

Distinctive Faces

By studying the role of distinctive features we are concerned with information in the face and facial representation. Bruce (1988) defined a distinctive face as one whose visual appearance is relatively unusual compared with the set of faces under consideration. A great deal of research has shown that distinctive faces are especially easy to identify. This is true for the recognition of familiar faces (Valentine & Bruce, 1986) and for recognizing previously presented unfamiliar faces (Light, Kayra-Stuart, & Hollander, 1979). One explanation of the advantage in memory of distinctive faces is that they are encoded on distinctive properties or features (Shepard, Gibling, & Ellis, 1991). This interpretation is consistent with the creation of caricatures. A cartoonist exaggerates the distinctive features of a face while preserving the typical ones. Valentine (1991) argued that a common adaptive mechanism may underlie distinctiveness effects in face recognition (see Busey, chap. 5, this volume; Valentine, chap. 3, this volume). According to A. W. Ellis (1992), it is at least theoretically conceivable that at some point in our evolutionary history selection pressures favored the rapid recognition of a face as belonging to a member of a group different from one's own. In monkeys, apes, or early hominids, the mechanisms proposed to underlie distinctiveness effects would also lead to the rapid identification of an individual as belonging to a species, subspecies, or group with different facial characteristics. Circumstances can be imagined in which that might have been adaptive.

Usually, organisms do not identify faces but other organisms. It is only reasonable that other features not on the face might contribute to the identification. We identify faceless friends over the telephone and an acquaintance even when his face has undergone an extreme change such as the shaving of a beard. We might also easily identify someone because of her distinctive hairstyle or walk. On the other hand, features not on the face such as a hat, headband, or glasses could disrupt face identification. As anecdotally described by Young and Bruce (1991), Little Red Riding Hood mistook a wolf for her grandmother. Although she could see that the wolf's eyes, nose, and teeth were larger than her grandmother's, she failed to identify the wolf because she was influenced by the hat the wolf was wearing, as well the context of him being in her grandmother's bed. Research has shown that children around the age of 6 years confuse faces because of an exclusive focus on such paraphernalia (Diamond & Carey, 1977). More generally, research in developmental psychobiology has confirmed that the young are more greatly influenced by salient or intense contextual cues than are adults (Kraebel, Vizvary, Heron, & Spear, 1998). For prosopagnosic patients who are unable to identify familiar faces, paraphernalia accompanying a face often are the only means to identify a face. Successful identification is usually achieved by relying on clothes or voice (H. D. Ellis & Young, 1989).

As can be seen in our short review, one line of research on face identity focuses on the features that are used in the processing of face identification. In terms of our approach of pattern recognition we name this focus the perspective of information. Face identification can be understood as a pattern recognition situation that provides multiple sources of information, including distinctive and nondistinctive features of a face, situational context, or other nonfacial features of the organism.

The Role of Experience and Development

The influence of experience on information or on modes of processing can be studied at least from two perspectives. On the one hand, there is the perspective that focuses on the comparison between the processing modes in domains differing in the amount of experience (e.g., comparing face processing to the processing of nonface objects, see Diamond & Carey, 1986; Tanaka & Gauthier, 1997). On the other hand, there is the equally important perspective of perceptual development that examines the influence of increasing age. In general, perceptual development brings about a gain in experience and perceptual learning. In what follows, we explain this perspective in greater detail.

Perceptual development research has been dominated by the long-standing idea of a developmental shift from holistic modes of processing in young children (around age 4–7) to analytic modes of processing in older children and adults (Kemler Nelson, 1989; Shepp, 1978; Smith & Kemler, 1977; Werner, 1957).[6] Although this view of developmental shift has been challenged and modified, subsequent studies did not question the assumption of an adult as an analytic processor, but simply the proposal that children are typically holistic processors. Examining children's modes of processing in more detail, one line of research argued that the diagnosis of holistic processing in children derives from an inattention to individual differences. Consistent with our value of analyzing individual participants, analyzing individual data in the context of commonly used tasks like the restricted classification task paradigm or a concept-learning task proved to be very informative. It was found that children had a strong bias to use just a single dimension to make their judgments (e.g., Schwarzer, 1997; Thompson, 1994; Ward, Vela, & Hass, 1990; Wilkening & Lange, 1989).

Another line of research explored the nature of perception in preschool children during the earliest moments of visual processing (Thompson & Massaro, 1989). The goal was, in contrast to the mentioned restricted classification or concept-learning tasks, to investigate perceptual processing while minimizing decision processes. The children's judgments were best described by the predictions of the FLMP. Like adults, children evaluate features independently and combine them during an integration operation. The multiplicative algorithm described how children integrate the features better than the additive integration rule. These results question the belief that children's processing is mainly holistic.

By analogy to the developmental studies concerning holistic processing in children, the goal of the following experiments on face perception was to investigate in greater detail the putative holistic face processing in adults (see introduction, this chapter). Can this conclusion about holistic face processing in adults be maintained even if individual data are analyzed and methods for examining early feature processing are used? Examining the processes of face identification in the context of mathematical model testing, especially in the context of the FLMP, can answer this question.

As already noted, if face identification performance can be explained by the predictions of the FLMP, the underlying processing sharply contrasts with holistic processing in terms of a holistic encoding of faces. For holistic face processing, according to this definition, processing of the complete face cannot be reduced to processing of separated facial parts. If the FLMP does indeed describe the processes involved in facial identification,

[6]It should be noted that this body of research focused exclusively on nonfacial visual stimuli.

adult face processing could be characterized as analytical in the sense that separate facial features are taken into account. This conclusion is in line with the developmental perspective of an adult as an analytic processor.

However, in theory the same conclusion could even be drawn if nonintegrative models like the SCM fit the observed data. This is because the SCM specifies independent evaluation of single features which is the antithesis of holistic models. The SCM is a nonintegrative model that assumes that only one of multiple inputs is used. In contrast, the FLMP is an integrative model that proposes a multiplicative combination of several features. Thus, examining the fit of the FLMP in comparison to the SCM answers the question of analytic integrative or analytic nonintegrative processing. Additionally, the general question of analytic or holistic processing is addressed by comparing the fit of the FLMP to a holistic model (HM).

Empirical Studies on Face Identity

Analogous to the studies on facial emotion mentioned earlier, we used the expanded factorial design to study the processing of face identity. The stimulus faces were generated using a database of three-dimensional head models from the Max Planck Institute of Biological Cybernetics in Tüebingen, Germany.[7] The head models did not contain distinctive features such as glasses, beards, or earrings. The hair had been removed digitally, because the scanning technique had problems digitizing the hair (for details, see Troje & Bülthoff, 1996). The basis of the faces used in our experiments was one synthetic face, namely the average face of the database. Unfortunately, Baldi could not be used because this research was initiated outside of our lab, however, we look forward to using Baldi in future face identity work.

Constructing the Facial Stimuli

Because in theory a face provides a multitude of dimensions we varied—using Rhodes' (1988) terminology—those first-order facial features that were typically used by participants. These features characterize (a) the appearance of the eyes and eyebrows, and (b) the mouth of the faces. Thus, as in the experiments on facial emotion, we varied one feature (eyes) in the upper part of the synthetic face and one feature (mouth) in the lower part of this face and, again, could vary the upper and lower part of the face

[7] We thank Nikolaus Troje for constructing and providing the faces.

PI: FWK Full service
CHAPT-08 LE008/Wenger July 29, 2000 17:39 Char Count= 0

independently of one another. Using the method of the correspondence-based representation of faces developed by Vetter and Troje (1997)—which allowed for the construction of continua along facial features—we created 5 levels for both the upper and lower part of the face. The upper face conditions comprised variations of the eyes and eyebrows as well as variations of the height of the forehead. On the other hand, the lower face conditions consisted of variations of the mouth and chin. Combining the 5 levels of the upper face conditions and the 5 levels of the lower face conditions using the expanded factorial design (5×5 plus the 10 half-face conditions presenting the upper face and lower face alone) resulted in 35 stimulus faces.

To use the 35 faces in the context of a face identification task we defined two prototypical faces. These faces contained the extreme levels on both features, eyes and mouth. One prototype had a long forehead, narrow eyebrows, and a wide mouth (prototype with Level 1 for the eyes and Level 5 for the mouth) and the contrasting prototype had a short forehead, wider eyebrows, and a small mouth (prototype with Level 5 for the eyes and Level 1 for the mouth). We named these prototypes Bob (5,1) and John (1,5).

After being familiarized with the faces of Bob and John, the participants' task was to identify each of the 35 stimulus faces (each was presented 16 times in random order). To minimize the effect of memory, we fastened pictures of Bob and John beside the response buttons. To make sure that the participants did not use elaborate problem-solving strategies to give their identification response, we displayed the stimulus faces only for 500 msec each.

Figure 8.6 (see points) shows the mean probabilities of identifying the faces as Bob's face as a function of the levels of the mouth and eyes variables. The left panel shows the identification for just the lower half of the face and the right panel for just the upper half. As can be seen, the steeper curve for the eyes (upper) variable illustrates that it was somewhat more influential than the mouth (lower) variable. However, both half-face conditions were effective in changing the identification from Bob to John. In the whole-face conditions, people's identification was mostly influenced by the upper part of the face. The influence of the mouth was much less than the influence of the eyes. Thus, the lower part of the face was very informative in the half-face conditions, but not in the whole-face condition.

Of primary interest in this analysis is which model could best describe our results. Because the purpose of our experiments was to examine the questions of analytic or holistic processing as well as integrative or nonintegrative processing, we compared our results with the fit of the following models. As described earlier, both the fit of the FLMP and SCM assume

P1: FWK Full service
CHAPT-08 LE008/Wenger July 29, 2000 17:39 Char Count= 0

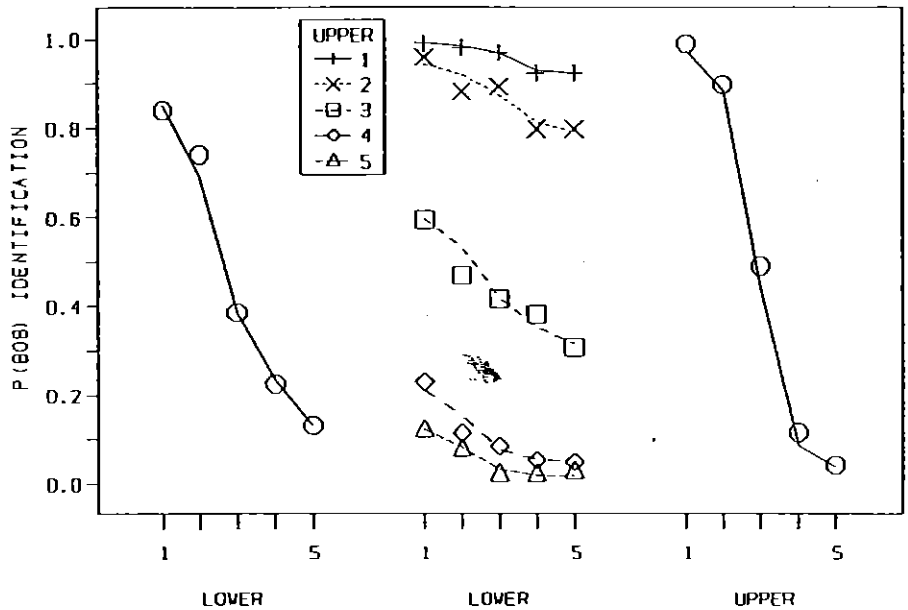

FIG. 8.6. Predicted (lines) and observed (points) proportion of Bob identifications as a function of the upper and lower face manipulations. The left panel shows the identification for just the lower half of the face, the right panel for just the upper half, and the middle panel for the whole-face condition.

analytic processing in the context of independent evaluation of the facial attributes. The HM is a nonanalytic model in which dependence is assumed at evaluation. If the FLMP gives a better fit of the data than the SCM, we can argue that the facial features were in fact integrated. Also, if the FLMP gives a better fit than the HM we can conclude that informational dependence occurs after evaluation.

In view of the fact that the influence of the lower part of the face was less than the upper part, we included a weighted FLMP (wFLMP) in addition to the simple FLMP. In the wFLMP the contribution of the lower face is attenuated by some proportion in the whole-face condition relative to the half-face condition. This model might be better able to describe the results than the simple FLMP because the informativeness of the upper and lower part of the face is relative when they are presented together. Thus, informativeness as given by the parameter values of the model changes from the half-face to whole-face conditions. Because we used the parameters fit to the half-face condition to predict the whole face condition some information context bias or weight should be added. This bias was added by reducing the influence of the upper face parameters for predicting only the whole-face condition:

$$f_i(b) = w f_i(u) + (1 - w).5 \tag{8}$$

where $f_i(b)$ is the feature value in whole-face conditions, $f_i(u)$ is the feature value in half-face conditions, and i tracks the level of the feature. The w is a free parameter indicating the relative amount of influence on trials of the whole-face conditions. Given the 5×5 expanded factorial design and the simple FLMP, 10 free parameters are necessary to fit our model to the 35 conditions: 5 parameters for each level of eye variations and 5 for mouth variations. In the wFLMP, given the additional weight parameter, 11 free parameters are necessary.

In our formulation of the SCM, we assume that only one of the inputs, the upper or lower part of the face, is functional in whole-face conditions. Thus, this model predicts that processing is nonintegrative and self-terminates when information from either part of the face is sensed (Townsend & Nozawa, 1995). The SCM represents the extreme of the analytic position because it allows for individual elements of the face to be sufficient for identification even in the presence of other elements or features. The logic of the SCM is as follows. The information on the upper part of the face is selected with some bias probability p, and the lower part information of the face with bias $1 - p$. For a given whole face condition the upper face information will be identified as Bob with probability u_i and the lower face information with probability l_j. Thus the predicted probability of the identification of Bob given the ith level of the upper face information, U_i, and the jth level of the lower face information, L_j, is:

$$P(Bob \mid U_i L_j) = p u_i + (1 - p) l_j \tag{9}$$

This equation for the SCM predicts the probability of identifying Bob for each of the 35 conditions in our expanded factorial experiment. Because the 35 equations have 5 different values of u_i and 5 different values of l_j and we also do not know the value of p in the whole face conditions, 11 free parameters are necessary: the value of p, the 5 u_i values, and the 5 l_j values.

Our primary assumption in creating the HM was that the subjective value of facial attributes is interactive at the lowest levels of processing. In other words, a nonanalytic model cannot allow for independence at the evaluation stage. This leads us to our secondary assumption, which is how we define dependence at evaluation. It is reasonable that as one feature becomes more salient it will influence the subjective value of the other feature to a greater extent. This can be captured quantitatively by a multiplicative rule:

$$c_{ij} = u_i l_j$$

where c_{ij} is the feature resulting from the integration of upper u_i and lower l_j parts of the face. For the HM we also assume that integration and decision processes are the same as the FLMP. Thus, the probability of a Bob response is:

$$P(Bob \mid U_i L_j) = \frac{c_{ij}}{c_{ij} + (1 - c_{ij})} \tag{10}$$

The results confirmed our expectations that the wFLMP is better able to describe the results than the simple FLMP. In comparison to the simple FLMP, the wFLMP yielded the best fit to observed data. Whereas the RMSDs of the wFLMP ranged between .0293 and .0895 with an average of .0611, the RSMDs of the simple FLMP were significantly higher, .0548 to .1257, with an average of .0807. The fit of the wFLMP model was also better than the SCM. The RMSDs of the SCM ranged from .0293 to .0998, with an average .0737. Finally, the HM fit the observed data worst of all, with RMSDs between .1326 and .2700 and an average of .1887.

Thus, the better fit of the wFLMP and FLMP lends support to the view that face processing, even for face identification, is analytic at evaluation but requires integration before an identification decision is made. Further, these results support the assumptions of the FLMP that independent facial features are multiplicatively combined and that decision is determined by the relative support of all alternatives. Unfortunately, we cannot claim to have falsified holistic encoding due to the many assumptions used in constructing the HM. Given the particularly poor fit of the HM, however, theories of holistic encoding must be questioned.

Given these results, the question arises of why face identification was more influenced by the eyes than by the mouth. Are these findings only observable in the context of identifying faces with very short presentation times that possibly induce an incomplete visual exploration of the whole face? Or is it just the case that the variations of the eyes are more informative than the mouth variations? As noted, the variations of the eyes also included variations of the eyebrows and the forehead, whereas the variations of the mouth consisted only of width variations. To answer this question, we

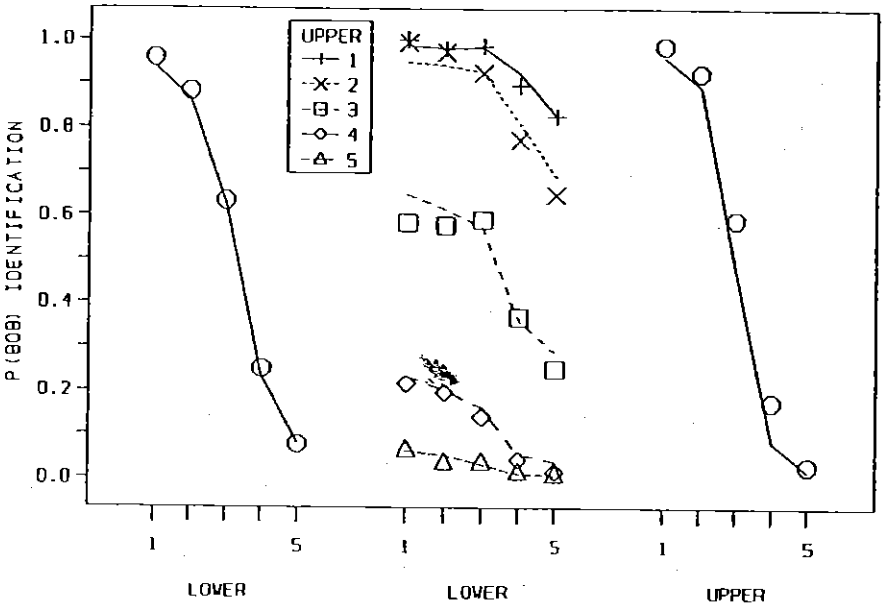

FIG. 8.7. Predicted (lines) and observed (points) proportion of Bob identifications as a function of the upper and lower face manipulations for the increased presentation time of 1,000 msec. The left panel shows the identification for just the upper half of the face, the right panel for just the lower half, and the middle panel for the whole-face condition.

increased the presentation time of the faces to 1,000 msec in a second experiment. The results showed that again—although to a weaker degree—the lower part of the face was less influential than the upper part of the face in the whole-face condition (see Fig. 8.7). Thus incomplete visual exploration does not seem to be the reason for the main influence of the eyes in face identification. Rather, the larger influence of the eyes could be due to the fact that the eye variations are more informative than the variations of the mouth. Possibly, the fact that more information changed in the upper part of the face makes it easier to discriminate the variations of the eyes than those of the mouth. Participants will usually be more influenced by features that are easy to discriminate than by features that are more difficult to discriminate (Garner & Felfoldy, 1970).

Moreover, the model fits replicated the first results in that the wFLMP showed a better description of the data than the simple FLMP. RMSDs for the wFLMP ranged between .0353 to .1246 with an average of .0667, whereas the simple FLMP RMSDs ranged from .0381 to .1498 with an average of .0891. Further, the FLMP and wFLMP gave a better fit to the data than the SCM and HM. SCM RMSDs ranged from .0511 to .1601 with an average of .0960, and the HM RMSDs ranged from .0859 to .2221 with an average of .1636. The good fit of both of the FLMPs and the poor fit of the SCM and HM agrees with the previously mentioned results in that analytic processing as well as the multiplicative information integration is central in face processing.

In sum, our studies proved successful in addressing the question of how facial features are evaluated and integrated to achieve the identification of faces. Despite the stronger influence of the eyes, both features, eyes and mouth, were effective in changing the identification from John to Bob. These processes were well described by the biased wFLMP and FLMP relative to the poorer description of the HM and the SCM. Given that the good fit of the FLMP challenges the issue of holistic face processing in terms of holistic encoding, our results on face processing are in agreement with the general proposal of an adult as an analytic processor. Beyond that, our results underscore that the separated facial features were not processed in isolation but were integrated multiplicatively in the process of face identification.

FACIAL SPEECH

Speech perception has long been dominated by the study of how individuals hear sounds and interpret this information as language. Focusing on the auditory input, however, little attention has been given to the visible nature of speech in face-to-face communication. The human face is a rich source of information for a large variety of tasks. Faces not only convey person identity but also give cues to gender, emotional states, direction of attention as well as speech. It is common to associate visible speech perception or lipreading as it is sometimes only learned by those with hearing impairments. Of course, research does show that visible speech is a useful source of information for the hearing-impaired (Massaro & Cohen, 1999; Walden, Prosek, Montgomery, Scherr, & Jones, 1977). However, research also indicates that untrained normal hearing individuals use visible speech to recognize words (Massaro, Cohen, & Gesi, 1993), consonants (C. S. Campbell & Massaro, 1997), and vowels (Jackson, Montgomery, & Binnie, 1976; Montgomery & Jackson, 1983). In fact, the use of visible speech cues is so natural and automatic for normal hearing individuals that it is difficult to ignore. The McGurk effect shows that visible speech that conflicts with auditory information can still influence perceptual judgments (McGurk & MacDonald, 1976). For example, combining the visual sentence *My gag kok me koo grive* with the auditory sentence *My bab pop me poo brive* gives the impression that the speaker is saying *My dad taught me to drive* (Massaro, 1987b).

Information in Speechreading

One of the most obvious sources of information for visible speech perception is the mouth of the speaker. Early views of speech training for the deaf proposed that attention should be focused exclusively on the lips. Thus, visible speech perception came to be called lipreading. In fact Summerfield (1979) provided some support for this by showing that identification improved 31% when lips alone were added to distorted auditory speech. Summerfield further proposed that the lips could be analyzed into three functional features: lip occlusion, horizontal lip extension, and oral area. However, additional work has shown that features functional in visible speech perception come from areas other than the lips. Such features include the jaw bone and skin (Benoit, Guiard-Marigny, Le Goff, & Adjoudani, 1996), cheek movement and jaw rotation (Erber, 1974), tongue movement (Bunger, 1952), and teeth visibility (McGrath, 1985). In our framework, these observations reflect the multiple sources of information or features that are available in visible speech perception. Because speech information comes from various areas of the face, visible speech perception is now more appropriately called speechreading.

The Multidimensional Fuzzy Logical Model

The psychophysical study of features has typically taken the approach of manipulating one feature of interest and holding all other information constant. Changes in correct identification indicate how functional that feature is for the experimental task. Although this method has yielded a large body of empirical data, problems can arise trying to manipulate more complex features such as those in the mouth region of a speaker's face. As an alternative, we proposed an extension of the traditional FLMP called the multidimensional fuzzy logical model (MD-FLMP). The MD-FLMP allows one to specify the information or features within the formulation of the model itself (C. S. Campbell & Massaro, 1997). In other words, if each speech token is a point in a multidimensional space, one can specify the feature or dimension axes and the vector direction of each token

TABLE 8.1
Six Visible Features From C. S. Campbell and Massaro (1997)

	Viseme								
Feature	*/ba/*	*/va/*	*/tha/*	*/da/*	*/za/*	*/la/*	*/ra/*	*/ĵa/*	*/wa/*
Duration	–	–	–	–	+	+	+	+	+
Tongue-Tip movement	–	–	+	+	+	+	–	–	–
Lip rounding	–	–	–	–	–	–	+	+	+
Mouth narrowing	–	–	–	–	–	–	–	–	+
Dental adduction	–	–	+	–	+	–	–	+	–
Lower-Lip tuck	–	+	–	–	–	–	–	–	–

(R. N. Shepard, 1980). For example, Table 8.1 shows that seven feature axes are hypothesized with two directions per axis, + or –. A + indicates the presence of that feature for a token prototype whereas a – indicates the absence. As shown in Table 8.1, /tha/ and /da/ both share tongue-tip movement and thus are hypothesized to be similar. In other words, the prototypes for /tha/ and /da/ are assumed to be close in multidimensional space. Exactly how close is not specified. The length or magnitude of the prototype vector is parameterized usually with one parameter per feature axis. However, many other possibilities exist. The parameters are iteratively adjusted between 0 and 1 until the RMSD between observed and predicted data has been minimized. By examining the parameter values we can determine how each feature axis contributed to the overall fit of the model. A parameter value near .5 indicates a vector with no magnitude in either direction and thus no real contribution to the fit of the model. Alternatively, a value near 1.0 indicates a strong contribution and a value near 0 indicates a contribution in the opposite direction as that hypothesized; the + should have been a –. The observed data to which the MD-FLMP is fit is the confusion matrix among all speech tokens. It is assumed that the more confused two tokens are, the more similar or closer they are in multidimensional space (Luce, 1963). This similarity can be specified simply by having tokens share one or more features in common.

The MD-FLMP has many benefits. As mentioned already, the MD-FLMP does not require any specific manipulation, unlike the traditional

FLMP, which requires a factorial or expanded factorial design.[8] Thus, the MD-FLMP can be used to explore information in more ecologically valid situations. Second, data from older experiments can be modeled or remodeled for the purposes of meta-analysis or to test some new hypothesis. Rather than the time-consuming process of redesigning and running experiments, new assumptions can be quickly formulated in the model and fit to preexisting data. Third, features are hypothesized a priori or before model fitting. Unlike multidimensional scaling and parallel distributed processing models with hidden units, this allows for a stronger understanding of the connection between the physical dimensions of the stimuli and the psychological dimensions given by the feature axes. Finally, it provides a substantial decrease in the number of free parameters required to fit the model (Massaro & Cohen, 1999). Whereas the ratio of parameters to data points is 1 to 3 or 1 to 4 for the traditional FLMP, the ratio is around 1 to 13 for the MD-FLMP.

Model Formulation

Similar to the traditional FLMP, all the information processing assumptions of evaluation, integration, and decision are also given in the formulation of the MD-FLMP. Each feature is an independent continuous source of information that is evaluated against prototypes in memory:

$$f_x = g(F_x)$$

Each feature, f_x, is a function of the value of the feature, F_x, in the stimulus. The result of this function is a fuzzy truth value between 0 and 1. The support for the response given the stimulus is defined by the similarity of the stimulus and response in terms of the number of features they share. In other words, the support for the response k and stimulus j given feature x is f if they share this feature and $(1 - f)$ if they do not:

$$S(kj \mid x) = \begin{cases} f_x & \text{if} \quad f_{xj} = f_{xk}, \\ (1 - f_x) & \text{if} \quad f_{xj} \neq f_{xk} \end{cases} \tag{11}$$

Because $S(kj \mid x) = S(jk \mid x)$ then the model makes the same predictions

[8]The only design requirement of the MD-FLMP is that some measure of psychological similarity among response alternatives be obtained. For identification and categorization tasks this measure would probably be response confusions. However, for same–different tasks this would be the proportion of "same" responses for all pairs of alternatives.

for each cell above and below the diagonal. In other words, this formulation assumes that the confusion matrix is symmetrical about the diagonal.[9]

The support for each response alternative is calculated by combining the support from each feature. All the sources of information are integrated according to a multiplicative operation in the MD-FLMP. As with the traditional FLMP, other integration assumptions may be formulated and tested. The general form of the equation for the support of a response k given a stimulus j is:

$$S(k \mid j) = \prod_{x=1}^{n} S(kj \mid x) \tag{12}$$

where n is the number of features. Looking at Table 8.1, the support for a /tha/ response given a /da/ stimulus would be a function of the match of five features and the mismatch of one feature (dental adduction):

$$S(/tha/ \mid /da/) = f_d f_t f_r f_n (1 - f_a) f_l$$

The symmetrical predictions of the model mean that $S(/tha/ \mid /da/) = S(/da/ \mid /tha/)$.

Finally, the decision operations of assessment and response selection is made according to the RGR. The probability of response given the stimulus for each cell of the confusion matrix is:

$$P(k \mid j) = \frac{S(k \mid j)}{\sum_{k=1}^{m} S(k \mid j)} \tag{13}$$

where m is the number of response alternatives.

In what follows, we demonstrate the use of the MD-FLMP in three experiments to explore issues of information and information processing in speechreading. We show how assumptions about information or the features used in speechreading can be easily formalized and how alternative theories can be tested and falsified. The MD-FLMP also allows us to discover how information changes across experimental manipulations such as stimulus degradations, natural and synthetic speech, and change of viewpoint.

[9] In terms of R. N. Shepard's (1980) work, the support $S(kj)$ for alternative k given stimulus j can be thought of as the distance d_{kj} between category k and j in multidimensional space. If we assume symmetry then the distance from k to j is the same as the distance from j to k or $d_{kj} = d_{jk}$ (R. N. Shepard, 1980, Equation 2b).

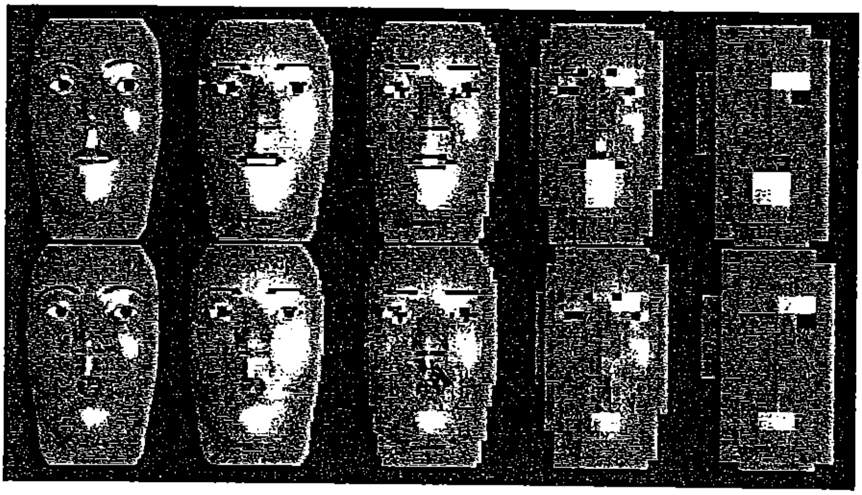

FIG. 8.8. The synthetic talking head, Baldi, from 145 cycles per face at far left through 4 cycles per face at far right. Upper panels show articulation of /va/, whereas the lower panels show /wa/. Taken from C. S. Campbell and Massaro (1997).

Degrading the Face Stimuli

The MD-FLMP was first used to test a set of six visible facial features hypothesized to be functional in speechreading (C. S. Campbell & Massaro, 1997). In this first experiment we created a nine-alternative forced-choice consonant–viseme categorization experiment with five levels of spatial degradation (see Fig. 8.8 for examples of /va/ and /wa/). The visemes (/ba/, /va/, /tha/, /da/, /ra/, /la/, /za/, /zha/, and /wa/) presented were representative of all nine consonant viseme classes in English (Walden et al., 1977). Viseme classes are groups of phonemes that are not visually distinct. For example, /ba/ and /pa/ are in the bilabial viseme class because they are difficult to distinguish using only visual information. Viseme identification was used due to the ease of this task by normal hearing, untrained participants. The results showed that accuracy was fairly resistant to degradation caused by quantization but confusions among viseme classes increased as the amount of quantization increased. The MD-FLMP model was constructed with six visible features (see Table 8.1) serving as sources of information to predict these confusions. The model fits showed that the six visible features predicted the pattern of confusions quite well. Analysis of parameter values indicated that these features were either highly or moderately functional for visible speech perception. Additionally, a multiplicative

P1: FWK Full service
CHAPT-08 LE008/Wenger July 29, 2000 17:39 Char Count= 0

feature integration model fit the observed data better than an additive integration model. The first experiment replicated and extended these findings to cover a different range of spatial degradation. Essentially, features functional in speechreading should generalize to new participants and similar stimuli. To assess the adequacy of the six visible features, we decided to test them against a competing set of features. Similar to C. S. Campbell and Massaro (1997), participants were presented with all nine consonant visemes at five levels of spatial quantization and asked to categorize each token. The only difference was that the levels of spatial quantization were changed to include 145, 32, 18, 10, and 7 cycles per face. Cycles per face is the number of pixels across the face at eye level divided by two. The results replicated the C. S. Campbell and Massaro (1997) experiment as shown in Fig. 8.9. Spatial quantization had a strong influence reducing accuracy from 66% in the undegraded condition to 34% at 7 cycles per

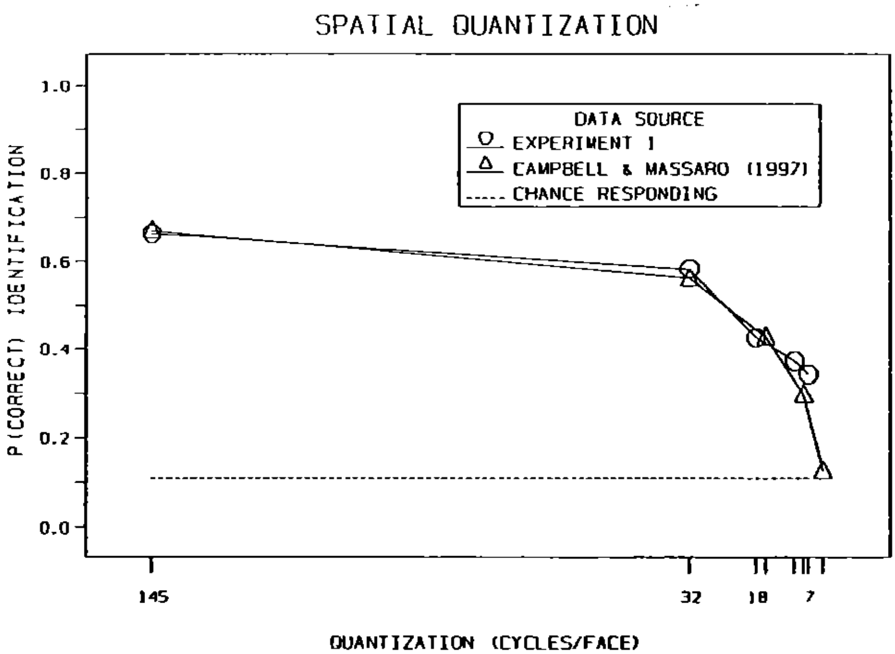

FIG. 8.9. Mean percentage correct viseme identifications across levels of spatial quantization. The current experiment (circles) was measured at 145, 32, 18, 10, and 7 cycles per face, whereas C. S. Campbell and Massaro (1997) (triangles) was measured at 145, 32, 16, 8, and 4 cycles per face. Results are consistent across experiments.

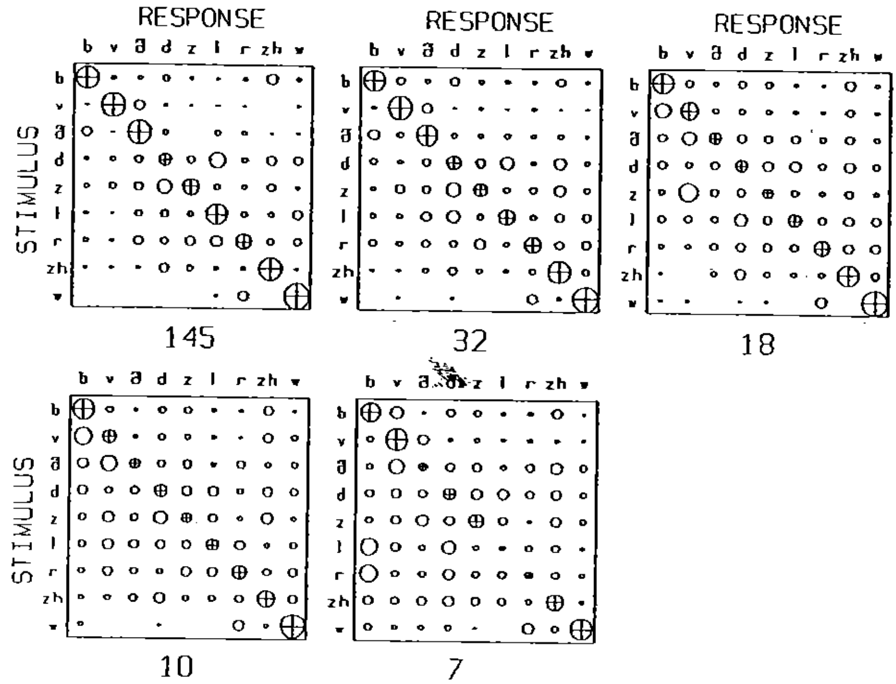

FIG. 8.10. Confusion matrices for each level of spatial quantization. The area of each circle indicates the mean proportion of responses given that stimulus. The circles on the diagonal with crosses indicate the proportion of correct responses, and the off-diagonal circles show confusions (errors).

face. Figure 8.10 gives the average confusion matrices for each level of quantization. The diagonal line of circles gives the proportion of correct responses, and the off-diagonal circles indicate confusions. At 145 cycles perface, we see that /da/ is confused with /la/ much of the time, whereas /wa/ is seldom confused with any other viseme. As the quantization increases, the proportion of correct responses on the diagonal decreases and the proportion of confusions increases.

Similar to the model tests of C. S. Campbell and Massaro (1997), three types of MD-FLMP models (simple, full, and weighted) were fit to the confusion matrices. The five levels of quantization provided five confusion matrices of data to be predicted. Each matrix had 81 cells (9 stimuli × 9 responses) for a total of 405 data points (see Fig. 8.10). The simple model contained only one free parameter for each feature resulting in 6 parameters to predict 405 data points. Because each parameter can assume

TABLE 8.2
Weighted Model Parameter Values for Six Visible Features

	Visible Features					
Parameters	*Duration*	*Tongue-Tip*	*Rounding*	*Narrowing*	*Adduction*	*L-Lip Tuck*
Feature values	0.706	0.780	0.666	0.999	0.800	0.922
	Spatial Quantization (cycles per face)					
	145	*32*	*18*	*10*	*7*	
Weight values	0.997	0.949	0.734	0.622	0.565	

only one value, then the simple model cannot account for any change in feature information due to degradation. As expected, the simple model gave a somewhat poor description of the confusion matrices with an average RMSD of .1324. To account for quantization, a full model was created with 5 parameters for each feature (30 parameters total). These 5 parameters allowed the feature values to change as a function of quantization. This full model resulted in a significantly smaller RMSD of .1167.[10]

If we assume that all six features are degraded the same proportion by quantization, we can replace the five parameters with a single weighting parameter. Similar to the weighted FLMP used previously (see Equation 8), the weight forces the value of the feature parameter to .5 (not informative) as its values decreases. The weighted model gave an average RMSD = .1251, which was halfway between the simple and full model fits. Thus, it improved the fit by half the amount possible while saving nearly two thirds of the parameters. Table 8.2 shows the average parameter values for the weighted model. Mouth narrowing and lower-lip tuck appear to be highly informative, whereas duration and rounding are not quite as functional. As expected, the weight values decrease as the degradation increases.

[10]The models were fit to each participant individually and all tests of significant differences between models were performed using a t test on these data.

Overall, the model fits of this experiment replicate those performed in C. S. Campbell and Massaro (1997). Additionally, the parameter values were similar across experiments, showing that the six visible features generalize across different levels of degradation and different participants. Our model tests thus far have assumed that information is processed according to the constraints of the FLMP. However, it is possible that information is integrated additively instead of multiplicatively (Anderson, 1981). To test this we created simple, full, and weighted versions of a multidimensional additive model of perception (MD-AMP). Consistent with C. S. Campbell and Massaro (1997), the resulting model fits show that the MD-FLMP predicted the data much better than the MD-AMP for all three models. The MD-AMP had a mean RMSD of .179 for each model compared to much lower RMSDs for the MD-FLMP.

Although we have been testing various theories of processing throughout this chapter, we can also test competing theories of information using the MD-FLMP. Miller and Nicely (1955) used a set of five linguistic features to describe the confusions among English consonants in audible speech. Table 8.3 gives the assignment of these linguistic features to the nine visemes. For voicing, nasal, and fricative, each feature can be either present or notpresent. Place, however, has three levels indicated with 1 for front, 2 for middle, and 3 for back, indicating where the vocal tract is constricted. The linguistic feature set was fit with all three FLMP models and these results were compared to the same models fit with the visible feature set. Results show that the original visible feature set described the data much better than the linguistic features. For the full model, the RMSD for the

TABLE 8.2
Five Linguistic Features From Miller and Nicely (1955)

	Viseme								
Feature	*/ba/*	*/va/*	*/tha/*	*/da/*	*/za/*	*/la/*	*/ra/*	*/ĵa/*	*/wa/*
Voicing	+	+	+	+	+	+	+	+	+
Nasal	–	–	–	–	–	–	–	–	–
Fricative	–	+	+	–	+	–	–	+	–
Place	1	1	1	2	2	2	3	2	1
Duration	+	+	+	+	–	–	–	–	–

TABLE 8.3
Full Model Parameter Values for Linguistic Features

	Linguistic Features				
Quantization	*Voicing*	*Nasal*	*Fricative*	*Place*	*Duration*
145	0.504	0.500	0.849	0.791	0.765
32	0.490	0.496	0.784	0.781	0.746
18	0.500	0.500	0.682	0.718	0.671
10	0.508	0.500	0.626	0.702	0.682
7	0.504	0.500	0.705	0.646	0.553

visible features was .111, whereas for the linguistic features it was .151. The difference in RMSDs for the two competing feature sets averaged .034 across simple, full, and weighted models. One reason why the linguistic features did not do as well probably stems from the fact that these features were created to describe audible speech. It has been shown that the functional features for visible speech are complementary to those of audible speech (H. W. Campbell, 1974). For example, voicing is an important feature for audible speech but it has very little value in visible speech. More importantly, however, the poor performance of the linguistic features is due to the fact that they fail to differentiate certain visemes. In Table 8.3, for example, /za/ and /zha/ have the same features as do /va/ and /tha/, respectively. Table 8.4 shows the average parameter values for each linguistic feature and level of quantization (the full model). The parameter values for the voicing and nasal features are all about .5. This indicates that these features were not useful for predicting the pattern of data given by participants. This is consistent with previous research showing that voicing and nasality are not very functional in speechreading (Dowell et al., 1982; Massaro & Cohen, 1999). The remaining features are fairly functional, ranging from .849 to .765. Appropriately, as spatial quantization increases, these parameter values move toward .5. The features become less functional because the stimulus is more degraded. These model tests yielded three main conclusions. First, the six visible features provide a better description of speechreading performance than the linguistic features. Second, the integration of visible speech information is better described by the FLMP than the AMP. Third, this modeling approach is reliable because stable fits were obtained across slightly different experiments.

Natural versus Synthetic Speech

In the second experiment, we aimed to generalize the psychological validity of the six visible features by fitting identifications of a natural speaker instead of Baldi. The motivation for this experiment is that Baldi has been shown to be somewhat less intelligible than natural speech (Massaro, 1998). Additionally, it has been claimed in the domain of auditory speech perception that synthetic speech lacks the informational richness of natural speech (Nusbaum, Dedina, & Pisoni, 1984). This means that information in synthetic speech is not merely a degraded version of natural speech, but provides different cues as well. If synthetic speech is qualitatively different from natural speech, the patterns of confusions for each stimulus should differ. Because the feature model used here is fit to confusion matrices, any difference in the patterns will then be reflected in the overall fit of the model. Thus, if natural speech were qualitatively different from synthetic speech we would expect either a better or poorer fit of the six visible features to data from a natural speaker. The parameter values for each feature provide an additional metric by which to evaluate the informativeness of features for natural versus synthetic speech. Differences in these values provide useful information for guiding improvements of our synthetic speech. Similar to the previous experiment, participants were presented with all nine consonant–vowel visemes at five levels of spatial degradation. The natural speaker was an adult male taken from laserdisc (Disc II) of the Johns Hopkins lipreading corpus (Bernstein & Eberhardt, 1986).

Overall, our results show that accuracy was higher for natural speech than synthetic speech. Accuracy was 76% in the undegraded condition versus 66% in C. S. Campbell and Massaro (1997) using synthetic speech, a difference of 10%. Figure 8.11 shows that, similar to the first experiment, performance from both natural and synthetic speech takes the form of a positively decelerated function and thus, was robust to the influence of spatial quantization. The six visible features were fit to the confusion matrices produced by speechreading the natural speaker. Only the full model and weighted model were tested because the simple model fails to describe confusions across levels of quantization. The results of the model tests confirmed that the fits using synthetic speech and natural speech were all about the same. For the full model, the mean RMSD of the six visible

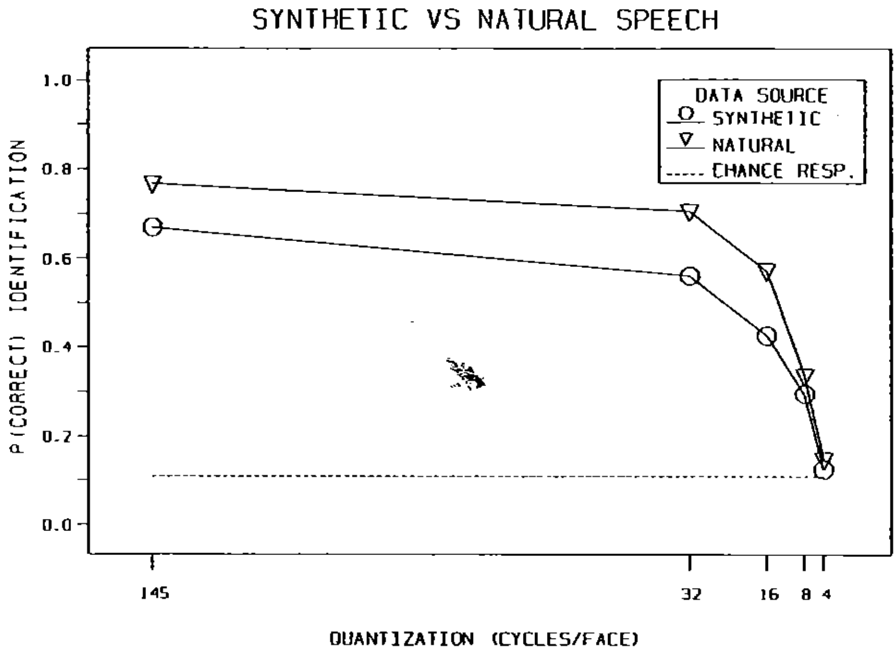

FIG. 8.11. Mean percentage correct identifications across levels of spatial quantization for natural (triangles) and synthetic (circles) speech.

features was .117 for synthetic speech and .125 for natural speech. For the weighted model, the mean RMSD was .125 for synthetic and .112 for natural speech.

Additional model tests were performed to examine whether the information processing assumptions of the FLMP hold for natural speech in the same manner as synthetic speech. Because the FLMP has been fit mainly to data from synthetic speech, it might be claimed that this model cannot be generalized to natural speech. Clearly, however, the FLMP adequately fit the confusions from natural speech for both the full model and the weighted model. The RMSDs of the AMP were worse than the FLMP at .191 for the full model and .191 for the weighted model. Thus, consistent with previous research in bimodal speech perception (Massaro, 1987b) the FLMP generalizes quite well from synthetic to natural speech. Although the model fits for natural and synthetic speech are nearly the same, analysis of the feature parameter values can give fine-grained information about the intelligibility of natural and synthetic speech. Differences in feature parameter values allow for specific recommendations on improving our synthetic speech.

TABLE 8.4
Parameter Values for Natural and Synthetic Speech

	Visible Features					
Stimuli	*Duration*	*Tongue-Tip*	*Rounding*	*Narrowing*	*Adduction*	*L-Lip Tuck*
Natural	0.783	0.921	0.858	0.818	0.942	0.995
Synthetic	0.706	0.780	0.666	0.999	0.800	0.922

Table 8.5 shows a comparison of average parameter values for each of the six visible features. The parameter value for duration was somewhat higher for natural than for synthetic speech, and the value for tongue-tip movement was much higher. This indicates that minor improvements need to be performed for viseme duration and relatively greater adjustments will be required for tongue-tip movement. Thus, the tongue is an area of ongoing work for improvements in our synthetic speech (see Cohen, Beskow, & Massaro, 1998). The rounding feature has a much lower parameter value for synthetic than natural speech, indicating that rounding also needs to be improved. As noted earlier, the viseme class /r/ tends to be less intelligible for synthetic than natural speech (Massaro, 1998). The lips will need to be adjusted to produce a more realistic rounding movement. The features adduction and lower-lip tuck are also somewhat lower for synthetic than natural speech. The visibility of the teeth and the lower-lip movement need to be improved. One way to improve adduction would be to simply increase the whiteness of the teeth relative to the lip coloration. Of course, this may occur automatically if a light source is positioned to shine directly into the mouth. Unlike the other features, the parameter value for narrowing is higher for synthetic than natural speech. Thus, narrowing shows that synthetic features can actually be made more informative than features in natural speech. Our model fits indicate that the six visible features generalize quite well from synthetic to natural speech. The features do not merely describe the perception of visible speech from synthetic talking heads but speechreading in normal face-to-face communication as well. The MD-FLMP modeling approach allows us not only to compare information assumptions across stimuli but to also make more specific comparisons by examining parameter values. In this sense, the MD-FLMP acts as a diagnostic tool for evaluating synthetic stimuli.

Speechreading From Different Views

A great deal of facial perception research has presented faces only in the frontal view under optimal conditions of lighting and distance. Faces are viewed from a variety of angles and distances in more typical situations. Research in face recognition has shown that performance is little influenced by faces rotated 45 degrees in depth (Davies et al., 1978; Hill, Schyns, & Akamatsu, 1997; Patterson & Baddely, 1977). Somewhat greater losses in face recognition performance occur when the face is in profile (Galper & Hochberg, 1971). Highly robust recognition across viewpoints has also been shown in speechreading. Visible vowel recognition is almost completely unaffected by head rotations in depth of 0 and 90 degrees (Wozniak & Jackson, 1979) and 45 degrees (Neely, 1956). Similar results were found for speechreading words (Ijsseldijk, 1992) and sentences (Bauman & Hambrecht, 1995). Visible consonants are also fairly robust to rotations in depth of 45 degrees but performance tends to decline more rapidly at 90 degrees (C. S. Campbell & Massaro, 1998).

As speechreading has been shown to be robust across viewpoints, it is reasonable to assume that the features functional in frontal views are also functional in profile. Therefore, the six visible features should also generalize from frontal to profile viewpoints. To test this model, the six visible features were fit to confusion matrices generated from speechreading Baldi in frontal and profile views at five levels of quantization. The visemes and levels of quantization were the same as those tested in the previous experiments.

Relative to frontal view, the profile view reduced accuracy by only 16% in the undistorted condition and 11% at 32 cycles per face. These results support the notion that speechreading is fairly robust to variations in viewpoint. Even in profile, performance was resistant to the effects of quantization (Fig. 8.12). Comparing confusion matrices of frontal and profile views indicates that the patterns of confusions appear to differ widely. In the frontal view, typical confusions are seen between /za/ and /tha/ as well as /da/ and /la/. These same confusions are seen in the profile view but they are not as pronounced. In the profile view, /tha/, /da/, /za/, and /la/ are often confused with /va/. Across all levels of quantization, /ra/, /zha/, and /wa/ in profile are confused with each other. Given these large differences, there may be qualitative differences in the information used across views. Thus, the six visible features may not fit the pattern of confusions in the profile view.

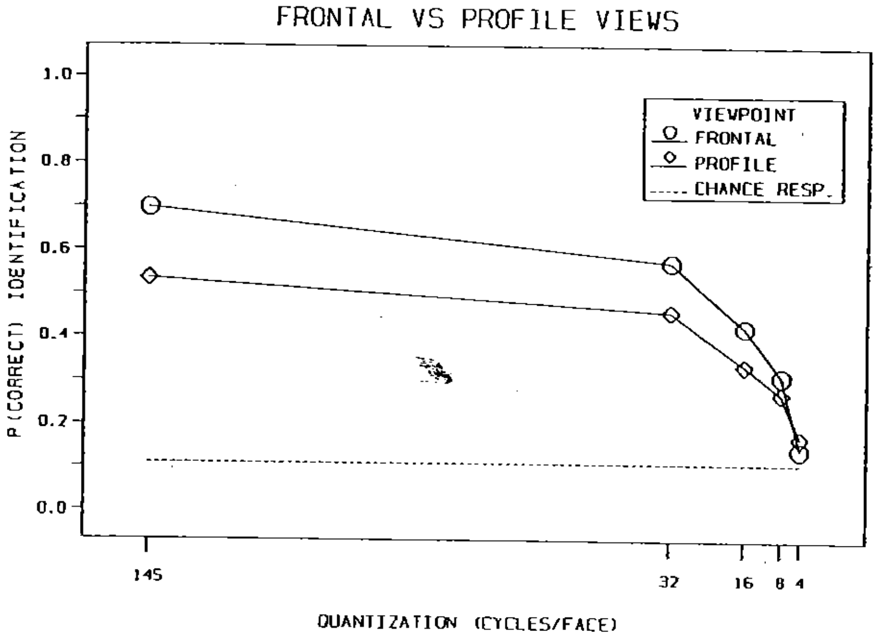

FIG. 8.12. Mean percentage correct identifications across levels of spatial quantization for frontal and profile viewpoints. Circles indicate frontal viewpoint and diamonds indicate profile viewpoint.

Confirming this suspicion, it was found that the six visible features do not fit the data from the profile view as well as the frontal view. The mean RMSD was .125 (full) and .136 (weighted) for the frontal view and .145 (full) and .154 (weighted) for the profile view. This suggests that the features functional in speechreading differ across viewpoints. An analysis of parameter values shows that tongue-tip movement, dental adduction, and to a lesser extent mouth narrowing are not as informative in the profile view. The parameter value for dental adduction is .859 in the frontal view but drops to .679 in profile. Likewise, the parameter value for tongue-tip movement is .855 in the frontal view decreasing to .774 in profile. This makes sense intuitively as well. It is difficult to see the tongue and teeth inside the mouth from a side view.

If the quality of information actually changes, the six features do not simply degrade in profile but, rather, new information must be introduced that was not in the frontal view. To test this we added a feature that is highly visible from the side but not from the front. Two features that are visible in profile include lip protrusion and lip retraction (Bauman & Hambrecht, 1995). Because lip protrusion is already present in the lip rounding feature, lip retraction was added as a seventh feature. Several consonants have lower-lip retraction to restrict the vocal tract toward the front of the mouth. This includes the labialdentals /va/, the interdentals /tha/, and to a lesser extent the alveolars /da/, /za/, and /la/. Because lower-lip retraction appears similar to the lower-lip tuck, many of the labial consonants are confused for /va/ when viewed in profile. Tests of the visible feature set with lip retraction show a significant improvement in fit to the profile view with mean RMSDs of .136 (full) and .151 (weighted). The tests also show no changes for the fits to the frontal view. This confirms that changes in the quality of information occur from frontal to profile views.[11]

Because the type of information used in speechreading changes with profile view, it may be claimed that information processing also changes. To test this, MD-FLMP and MD-AMP versions of the seven visible feature model (including lower-lip retraction) were constructed. Replicating previous model tests in this chapter, the FLMP provided a better fit to the frontal view data than the MD-AMP. The mean RMSDs for the MD-FLMP were .122 (full) and .135 (weighted) versus .185 (full) and .186 (weighted) for the MD-AMP. The MD-FLMP also fit better than the MD-AMP in the profile view. Mean RMDSs for the MD-FLMP were .136 (full) and .151 (weighted) versus .180 (full) and .180 (weighted) for the MD-AMP. Consistent with previous findings, information processing remains the same across perceptual tasks as only the information changes. Overall, these model tests showed that information in speechreading changes in quantity and quality as a function of viewpoint. Decreases in quantity were indicated by reduced speechreading accuracy in profile compared with frontal views. Changes in quality were given by the differences in the pattern of confusions for profile and frontal views combined with the poor fit of the six visible features to the profile data. The possibility that new information is functional in profile speechreading was tested by adding a seventh feature, lower-lip retraction, to the original six feature set. The significantly better fit of the seven-feature model confirmed that changes in information quality occur across viewpoints. Tests for changes in information processing, however, did not show differences.

[11] Similar to the first experiment, the weighted model does not fit the data quite as well as the full model for the profile view. The assumption of the weighted model that the effect of quantization is the same for each feature may not be valid. For example, one feature may be very resistant to quantization, whereas another succumbs more quickly. Further tests of the differential effect of spatial distortion on individual features are needed.

Summary

We have shown in three experiments how the MD-FLMP is a productive framework for exploring facial speech. Using this framework, we were able to show that the six visible feature model was superior to a competing seven linguistic feature model for describing the important information in synthetic and natural speech. However, the six visible features were not sufficient to account for speechreading in profile indicating that the features functional in speechreading change across viewpoints. The addition of the lip retraction feature was an attempt to uncover what new features are important in profile speechreading. This model was also used to explore specific patterns in the observed data through an analysis of the parameter values. Initially parameter values were inspected to find the relative importance of each feature in speechreading. However, this analysis also indicated the importance of each feature for synthetic and natural speech. Finally, these differences were used as a diagnostic and to make recommendations to improve our synthetic speech. It is also possible that this analysis could be used to explore what information good speechreaders use compared to poor speechreaders. This knowledge could then be used to improve speechreading training programs for people with hearing impairments. In addition to the tests of information in face perception, tests of information processing showed that the FLMP provides a better description of speechreading synthetic, natural, and profile speech than an MD-AMP. Overall, the MD-FLMP approach is useful because it allows one to easily formalize and test information and information processing assumptions, falsify alternative information and information processing theories, and evaluate changes in information across stimuli and participants.

GENERAL CONCLUSIONS

We have presented in this chapter an information processing framework for studying face perception and formalized the approach in a mathematical model called the FLMP. The value of this method was tested in the three areas of face perception: emotion identification, face identification, and speechreading. In all three areas, the FLMP predicted performance better than several alternative models such as the SCM and the AMP. The success of the FLMP provides support for the information processing assumptions formulated in the model and casts doubts on current positions in the face perception literature. The CMP, which is mathematically equivalent to the SCM, was falsified for both emotion and face identifiction. This is not surprising given that categorical theories of perception have been systematically falsified in other areas of perception such as speechreading and auditory speech recognition (Massaro, 1987b). Because the FLMP assumes each source of information is independent, the success of this model also provides evidence against holistic theories of face perception. If holistic models were valid, multiple sources of information should be dependent (evaluated as a whole). Our formulation and testing of this notion of holism in the HM, however, showed very poor fits for the face identification experiment.

Contrary to the modularity viewpoint, the FLMP seemed to provide equally good predictions of performance for all three areas of face perception. This suggests that information processing is the same for these areas and thus, face perception in general. To account for task-specific differences in performance we need only look to the information.

The general modeling approach used in our framework has been shown to provide a powerful and flexible method for formulating alternative hypotheses. Different information processing assumptions such as additive integration (AMP) and nonintegration (SCM) were easily formalized and tested. Competing assumptions about the information or features functional in speechreading were also formalized using a new model called the MD-FLMP. Thus, our modeling approach is flexible enough to formalize and test a wide range of hypotheses concerning facial perception. The only requirement for formalizing verbal positions is that they meet some minimal standards of clearness and completeness. Once formalized, however, these hypotheses share a common mathematical language allowing for better analytic comparisons among them and the possibility of prediction testing.

In the formulation of the FLMP, information is free to vary through a set of parameters. This allows one to pullapart issues related to information and issues related to information processing. Thus, our framework provides an analytical method for exploring information and a formal method for testing information processing. Throughout this chapter we have shown how the parameter values give clues to information in a given task. For example, our study of emotion identification showed that the range of parameter values was more extreme for the angry than the happy end of the continuum. This indicated that downward deflection of the mouth and brow were more influential than upward deflection. Additionally, for speechreading, parameter values for dental adduction and tongue-tip movement dropped substantially from frontal to profile views. This indicated that these features were not as functional in profile.

Throughout this chapter we have defined and evaluated an information processing framework and have shown how this method can advance our understanding of face perception. Our framework combines three areas (a formal modeling approach, experimental paradigm, and facial animation technology) to provide a powerful yet flexible tool for inquiry. In the future we hope to see this information processing approach combined with models of encoding, sensory system models, and models of psychological evidence spaces to provide a unified account of face perception.

REFERENCES

Allen, V. L., & Atkinson, M. L. (1981). Identification of spontaneous and deliberate behavior. *Journal of Nonverbal Behavior, 5*, 224–237.

Anderson, N. H. (1962). Application of an additive model to impression formation. *Science, 138*, 817–818.

Anderson, N. H. (1965). Averaging versus adding as a stimulus-combination rule in impression formation. *Journal of Experimental Psychology, 70*(4), 394–400.

Anderson, N. H. (1973). Functional measurement of social desirability. *Sociometry, 36*(1), 89–98.

Anderson, N. H. (1974). Information integration theory: A brief survey. In D. H. Krantz, R. C. Atkinson, R. D. Luce, & P. Suppes (Eds.), *Contemporary developments in mathematical psychology* (Vol. 2, pp. 236–305). San Francisco: Freeman.

Anderson, N. H. (1981). *Foundations of information integration theory.* New York: Academic.

Anderson, N. H. (1982). *Methods of information integration theory.* New York: Academic.

Anderson, N. H. (1996). *A functional theory of cognition.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Archer, D. (Producer), & Silver, J. (Director). (1991). *A world of gestures: Culture and nonverbal communication* [Videorecording]. (Available from University of California Extension Media Center, Berkeley, CA)

Bahrick, H. P., Bahrick, O. O., & Wittlinger, R. P. (1975). Fifty years of memory for names and faces: A cross-sectional approach. *Journal of Experimental Psychology: General, 104*, 54–75.

Bauman, S. L., & Hambrecht, G. (1995). Analysis of view angle used in speechreading training of sentences. *American Journal of Audiology, 4*, 67–70.

Beale, J. M., & Keil, F. C. (1995). Categorical effects in the perception of faces. *Cognition, 57*, 217–239.

Benoit, C., Guiard-Marigny, T., Le Goff, B., & Adjoudani, A. (1996). Which components of the face do humans and machines best speechread? In D. G. Stork & M. E. Hennecke (Eds.), *Speechreading by humans and machines* (pp. 315–328). New York: Springer-Verlag.

Bernstein, L. E., & Eberhardt, S. P. (1986). *Johns Hopkins lipreading corpus videodisk set.* Baltimore: Johns Hopkins University.

Brown, S. D., & Dooling, R. J. (1993). Perception of faces by budgerigars (Melopsittacus undulatus): II. Synthetic models. *Journal of Comparative Psychology, 107*, 48–60.

Bruce, V. (1988). *Recognizing faces.* Hove, UK: Lawrence Erlbaum Associates.

Brunswik, E. (1956). *Perception and the representative design of psychological experiments.* Berkeley: University of California Press.

Bunger, A. M. (1952). Speech reading—Jena method (2nd rev.) Danville, IL: Interstate.

Campbell, C. S., & Massaro, D. W. (1997). Perception of visible speech: Influence of spatial quantization. *Perception, 26*, 627–644.

Campbell, C. S., & Massaro, D. W. (1998). Visible speech perception and robustness in face processing. In J. P. H. Wechsler, V. Bruce, F. Fogelman Soulii, & T. Huang (Eds.), *Face recognition: From theory to applications* (Vol. 163, pp. 391–401). Berlin: Springer-Verlag.

Campbell, H. W. (1974). *Phoneme recognition by ear and by eye: A distinctive feature analysis.* Unpublished doctoral dissertation, University of Nijmegen, Nijmegen, Holland.

Campbell, R., Zihl, J., Massaro, D. W., Munhall, K., & Cohen, M. M. (1997). Speechreading in the akinetopsic patient, L.M. *Brain, 120,* 1793–1803.

Carey, S. (1996). Perceptual classification and expertise. In R. Gelman & T. Kit-Fong Au (Eds.), *Perceptual and cognitive development* (pp. 49–69). San Diego, CA: Academic.

Carey, S., & Diamond, R. (1994). Are faces perceived as configurations more by adults than by children? *Visual Cognition, 1,* 253–274.

Cohen, M. M., Beskow, J., & Massaro, D. W. (1998, December). *Recent developments in facial animation: An inside view.* Paper presented at AVSP '98, (Sydney, Australia).

Cohen, M. M., & Massaro, D. W. (1993). Modeling coarticulation in synthetic visual speech. In N. M. Thalmann & D. Thalmann (Eds.), *Models and techniques in computer animation* (pp. 139–156). Tokyo: Springer-Verlag.

Collier, G. (1985). *Emotional expression.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Crowther, C. S., Batchelder, W. H., & Hu, X. (1995). A measurement-theoretic analysis of the fuzzy logic model of perception. *Psychological Review, 102*(2), 396–408.

Cutting, J. E. (1998). Information from the world around us. In J. Hochberg (Ed), *Handbook of perception and cognition: Perception and cognition at century's end* (2nd ed., pp. 69–93). San Diego, CA: Academic.

Darwin, C. (1872). *The expression of emotion in man and animals.* London: Murray.

Davies, G. M., Ellis, H. D., & Shepard, J. W. (1978). Face recognition accuracy as a function of mode of representation. *Journal of Applied Psychology, 63,* 180–187.

Dewey, J. (1886). *Psychology.* New York: Harper.

DeYoe, E. A., & Van Essen, D. C. (1988). Concurrent processing streams in monkey visual cortex. *Trends in Neurosciences, 11,* 219–226.

Diamond, R., & Carey, S. (1977). Developmental changes in the representation of faces. *Journal of Experimental Child Psychology, 23,* 1–22.

Diamond, R., & Carey, S. (1986). Why faces are and are not special: An effect of expertise. *Journal of Experimental Psychology: General, 115,* 107–117.

Dombi, J. (1982). A general class of fuzzy operators, the DeMorgan class of fuzzy operators and fuzziness measures induced by fuzzy operators. *Fuzzy Sets and Systems, 8,* 149–163.

Dowell, R. C., Martin, L. F. A., Tong, Y. C., Clark, G. M., Seligman, P. M., & Patrick, J. F. (1982). A 12–consonant confusion study on a multiple-channel cochlear implant patient. *Journal of Speech and Hearing Research, 25,* 509–516.

Duchenne de Boulogne, G. B. (1990). *The mechanism of human facial expression.* Cambridge, UK: Cambridge University Press.

Ekman, P. (Ed.). (1973). *Darwin and facial expression: A century of research in review.* San Diego, CA: Academic.

Ekman, P. (1992). *Telling lies: Clues to deceit in the marketplace, politics, and marriage.* New York: Norton.

Ekman, P. (1993). Facial expression and emotion. *American Psychologist, 48,* 384–392.

Ekman, P., & Friesen, W. (1975). *Pictures of facial affect.* Palo Alto, CA: Consulting Psychologists Press.

Ekman, P., Friesen, W., & Ellsworth, P. (1972). *Emotion in the human face: Guidelines for research and an integration of findings.* New York: Pergamon.

Ekman, P., Hager, J. C., & Friesen, W. (1981). The symmetry of emotional and deliberate facial action. *Psychophysiology, 18,* 101–106.

Ellis, A. W. (1992). Cognitive mechanisms of face processing. *Philosophical Transactions of the Royal Society of London, 335,* 113–119.

Ellis, H. D., & Young, A. W. (1989). Are faces special? In A. W. Young & H. D. Ellis (Eds.), *Handbook of research on face processing* (pp. 1–26). Amsterdam: Elsevier.

Ellison, J. W., & Massaro, D. W. (1997). Featural evaluation, integration, and judgement of facial affect. *Journal of Experimental Psychology: Human Perception and Performance, 23*, 213–226.

Erber, N. P. (1974). Effects of angle, distance, and illumination on visual reception of speech by profoundly deaf children. *Journal of Speech and Hearing Research, 17*, 99–112.

Etcoff, H. L., & Magee, J. J. (1992). Categorical perception of facial expressions. *Cognition, 44*, 227–240.

Farah, M. J. (1990). *Visual agnosia: Disorders of object recognition and what they tell us about normal vision*. Cambridge, MA: MIT Press.

Farah, M. J. (1995). Dissociable systems for visual recognition: A cognitive neuropsychology approach. In S. M. Kosslyn & D. N. Osherson (Eds.), *Visual cognition: An invitation to cognitive science, Vol. 2* (2nd ed., pp.101–119). Cambridge, MA: MIT Press.

Farah, M. J., Tanaka, J. W., & Drain, H. M. (1995). What causes the face inversion effect? *Journal of Experimental Psychology: Human Perception and Performance, 21*, 628–634.

Farah, M. J., Wilson, K. D., Drain, M., & Tanaka, J. N. (1998). What is "special" about face perception? *Psychological Review, 105*(3), 482–498.

Fridlund, A. J. (1994). *Human facial expression: An evolutionary view*. San Diego, CA: Academic.

Galper, R. E. & Hochberg, J. (1971). Recognition memory for photographs of faces. *American Journal of Psychology, 84*, 351–354.

Garner. W. R., & Felfoldy, G. L. (1970). Integrality of stimulus dimensions in various types of information processing. *Cognitive Psychology, 1*, 225–241.

Hill, H., Schyns, P. G., & Akamatsu, S. (1997). Information and viewpoint dependence in face recognition. *Cognition, 62*, 201–222.

Ijsseldijk, F. J. (1992). Speechreading performance under different conditions of video image, repetition, and speech rate. *Journal of Speech and Hearing Research, 35*, 466–471.

Jackson, P. A., Montgomery, A. A., & Binnie, C. A. (1976). Perceptual dimensions underlying vowel lipreading performance. *Journal of Speech and Hearing Research, 19*, 796–812.

James, W. (1890). *The principles of psychology*. New York: Holt.

Johnson, N. F. (1975). On the function of letters in word identification: Some data and a preliminary model. *Journal of Verbal Learning and Verbal Behavior, 14*, 17–29.

Johnson, N. F., & Blum, A. J. (1988). When redundancy hurts letter detection: An attempt to define one condition. *Perception & Psychophysics, 43*, 147–155.

Kemler Nelson, D. G. (1989). The nature and occurrence of holistic processing. In B. E. Shepp & S. Ballesteros (Eds.), *Object perception: Structure and process* (pp. 357–386). Hillsdale, NJ: Lawrence Erlbaum Associates.

Kraebel, K. S., Vizvary, L. M., Heron, J. S., & Spear, N. E. (1998). Effect of context salience on heart rate orienting and habituation in preweanling and periadolescent rats. *Behavioral Neuroscience, 112*(5), 1080–1091.

Light, L. L., Kayra-Stuart, F., & Hollander, S. (1979). Recognition memory for typical and unusual faces. *Journal of Experimental Psychology: Human Learning and Memory, 5*, 212–228.

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

Luce, R. D. (1963). Detection and recognition. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology* (Vol. 1, pp. 103–189). New York: Wiley.

Luce, R. D. (1977). The choice axiom after twenty years. *Journal of Mathematical Psychology, 15*, 215–233.

Massaro, D. W. (1975a). *Experimental psychology and information processing*. Chicago: Rand McNally.

Massaro, D. W. (Ed.). (1975b). *Understanding language: An information processing analysis of speech perception, reading, and psycholinguistics*. New York: Academic.

Massaro. D. W. (1987a). Categorical partition: A fuzzy logical model of categorization behavior. In S. Harnad (Ed.), *Categorical perception: The groundwork of cognition* (pp. 254–283). New York: Cambridge University Press.

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

Massaro, D. W. (1989). *Experimental Psychology: An information processing approach.* San Diego, CA: Harcourt Brace Jovanovich.

Massaro, D. W. (1998). *Perceiving talking faces.* Cambridge, MA: MIT Press.

Massaro, D. W., & Cohen, M. M. (1999). Speech perception in perceivers with hearing loss: Synergy of multiple modalities. *Speech, Language and Hearing Research, 42,* 21–41.

Massaro, D. W., Cohen, M. M., & Gesi, A. T. (1993). Long-term training, transfer, and retention in learning to lipread. *Perception & Psychophysics, 53,* 549–562.

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

Massaro, D. W., & Klitzke, D. (1977). Letters are functional in word identification. *Memory & Cognition, 5,* 292–298.

Massaro, D. W., & Oden, G. (1980). Evaluation and integration of acoustic features in speech preparation. *Journal of the Acoustical Society of America, 67,* 996–1013.

McGrath, M. (1985). *An examination of cues for visual and audio-visual speech perception using natural and computer generated faces.* Unpublished doctoral thesis, University of Nottingham, Nottingham, UK.

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

Miller, G. A., & Nicely, P. (1955). An analysis of perceptual confusions among some English consonants. *Journal of the Acoustical Society of America, 27,* 338–352.

Montgomery, A. A., & Jackson, P. L. (1983). Physical characteristic of the lips underlying lipreading performance. *Journal of the Acoustical Society of America, 73,* 2134–2144.

Neely, K. K. (1956). Effect of visual factors on the intelligibility of speech. *Journal of the Acoustical Society of America, 28,* 1275–1277.

Nusbaum, H. C., Dedina, M. J., & Pisoni, D. B. (1984). *Perceptual confusions of consonants in natural and synthetic CV syllables* (Speech Research Laboratory Tech. Note No. 84–02). Bloomington: Indiana University, Speech Research Laboratory.

Patterson, K., & Baddely, A. D. (1977). When face recognition fails. *Journal of Experimental Psychology: Human Learning and Memory, 3,* 406–417.

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

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

Rhodes, G. (1988). Looking at faces: First-order and second-order features as determinants of facial appearance. *Perception, 17,* 43–63.b

Rhodes, G., Brennan, S., & Carey, S. (1987). Identification and ratings of caricatures: Implications for mental representations of faces. *Cognitive Psychology, 19,* 473–497.

Schwarzer, G. (1997). Analytic and holistic modes in the development of melody perception. *Psychology of Music, 25,* 35–56.

Sergent, J., Ohta, S., MacDonald, B., & Zuck, E. (1994). Segregated processing of facial identity and emotion in the human brain: A PET study. *Visual Cognition, 1,* 349–369.

Shepard, J. W., Gibling, F., & Ellis, H. D. (1991). The effects of distinctiveness, presentation time and delay on face recognition. *European Journal of Cognitive Psychology, 3*(1), 137–145.

Shepard, R. N. (1980). Multidimensional scaling, tree-fitting, and clustering. *Science, 210*(4468), 390–398.

Shepp, B. E. (1978). From perceived similarity to dimensional structure. In E. Rosch & B. Lloyd (Eds.), *Cognition and categorization* (pp. 135–167). Hillsdale, NJ: Lawrence Erlbaum Associates.

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,* 279–298.

Summerfield, A. Q. (1979). Use of visual information in phonetic perception. *Phonetica, 36,* 314–331.

Swets, J. A. (1998). Separating discrimination and decision in detection, recognition, and matters of life and death. In D. Scarborough and S. Sternberg (Eds.), *An invitation to cognitive science: Methods, models, and conceptual issues* (Vol. 4, pp. 635–702). Cambridge, MA: MIT Press.

Tanaka, J. W., & Farah, M. J. (1993). Parts and wholes in face recognition. *Quarterly Journal of Experimental Psychology, 46A*, 225–245.

Tanaka, J. W., & Gauthier, I. (1997). Expertise in object and face recognition. In R. L. Goldstone, P. G. Schyns, & D. L. Mendin, (Eds.), *Psychology of learning and motivation series, special volume: Perceptual mechanisms of learning* (Vol. 36, pp. 83–125). San Diego, CA: Academic.

Tanaka, J. W., & Sengco, J. A. (1997). Features and their configuration in face recognition. *Memory & Cognition, 25*(5), 583–592.

Thomas, R. D. (1996). Separability and independence of dimensions within the same-different judgment task. *Journal of Mathematical Psychology, 40*, 318–341.

Thompson, L. A. (1994). Dimensional strategies dominate perceptual classification. *Child Development, 65*, 1627–1645.

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*, 334–362.

Townsend, J. T., & Nozawa, G. (1995). Spatio-temporal properties of elementary perception: An investigation of parallel, serial, and coactive theories. *Journal of Mathematical Psychology, 39*, 321–334.

Troje, N. F., & Bülthoff, H. H. (1996). Face recognition under varying pose: The role of texture and shape. *Vision Research, 36*, 1761–1771.

Troje, N. F., & Bülthoff, H. H. (1997). How is bilateral symmetry of human faces used for recognition of novel views? *Vision Research, 38*, 79–89.

Valentine, T. (1991). A unified account of the effects of distinctiveness, inversion and race in face recognition. Quarterly *Journal of Experimental Psychology, 43A*, 161–204.

Valentine, T., & Bruce, V. (1986). Recognizing familiar faces: The role of distinctiveness and familiarity. *Canadian Journal of Psychology, 40*, 300–305.

Vetter, T., & Troje, N. F. (1997). Separation of texture and shape in images of faces for image encoding and synthesis. *Journal of the Optical Society of America, 14*, 2152–2161.

Walden, B. E., Prosek, R. A., Montgomery, A. A., Scherr, C. K., & Jones, C. J. (1977). Effects of training on the visual recognition of consonants. *Journal of Speech and Hearing Research, 20*, 130–145.

Ward, T. B., Vela, E., & Hass, S. D. (1990). Children and adults learn family-resemblance categories analytically. *Child Development, 61*, 593–605.

Werner, H. (1957). *Comparative psychology of mental development.* New York: International Universities Press.

Wilkening, F., & Lange, K. (1989). When is children's perception holistic? Goals and styles in processing multidimensional stimuli. In T. Globerson & T. Zelnicker (Eds.), *Cognitive style and cognitive development* (pp. 141–171). Norwood, NJ: Ablex.

Wozniak, V., & Jackson, P. (1979). Visual vowel and diphthong perception from two horizontal viewing angles. *Journal of Speech and Hearing Research, 22*, 354–365.

Yager, R. R. (1980). On a general class of fuzzy connectives. *Fuzzy Sets and Systems, 4*, 235–242.

Young, A. W., & Bruce, V. (1991). Perceptual categories and the computation of "grandmother." *European Journal of Cognitive Psychology, 3*, 5–49.

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