

Perceptual recognition of facial affect: Cross-cultural comparisons

DOMINIC W. MASSARO and JOHN W. ELLISON
University of California, Santa Cruz, California

Previous research has shown that the perception of affect in faces is well described by the fuzzy logical model of perception (FLMP). In this study, we asked whether the processes involved in recognition depended on the race/culture of the face and/or of the perceiver. A computer-generated face was used to manipulate two features of facial affect: brow deflection and mouth deflection. An expanded-factorial design was used, with four levels of brow deflection crossed with four levels of mouth deflection, as well as their corresponding half-face conditions. Participants identified these faces as either happy or angry. Japanese and U.S. students were tested on faces from these two countries that were texture-mapped onto the animated face. The FLMP gave the best description of performance for both groups and for both types of faces. These findings challenge previous claims of holistic perception, categorical perception, and additive feature integration.

A research project currently under way in our laboratory is to develop a synthetic speech system using both the face and the voice. The primary goal has been to duplicate the movements involved in the production of spoken language (Massaro & Cohen, 1994). In addition, it is possible to texture-map a specific face onto the animated face to give the illusion of an identifiable person who is talking. It quickly became apparent that an important aspect of a talker is his or her facial expression. As participants in communication, we take for granted our ability to convey and perceive facial expressions of emotional affect. To simulate expression in our talking head, it was first necessary to identify features of the face that would communicate specific emotions. We accordingly adapted our paradigm for inquiry to this study. This approach allows the investigation not only of the cues that people use to recognize facial expression, but also of how they are processed to achieve perceptual recognition (Massaro, 1987).

Given our talking head (Cohen & Massaro, 1993, 1994), it was possible to use a set of stimuli that would be standardized and replicable, as well as controllable over a wide range of feature dimensions. The features of this synthetic face are independently controllable, fully quantifiable, and easily replicable. The fact that individual features can be varied within the face without the "cutting and splicing" used by other researchers simplifies the process of displaying features in a controllable manner, while

simultaneously maintaining a highly realistic facial image. Ambiguous, contradictory, and partial feature presentations can be tested very easily, meeting one of the requirements of our paradigm for inquiry. We used an expanded-factorial design, as illustrated in Figure 1. The advantages of the expanded-factorial design are that single features as well as all feature combinations are tested. This design provides a strong test of models of perceptual recognition and judgment (Massaro & Cohen, 1990).

In an initial study, we examined the perceptual recognition of facial affect (Ellison & Massaro, in press). We carried out two experiments using an expanded-factorial design, with five levels of brow deflection crossed with five levels of mouth deflection. The single-factor conditions presented just the upper half of the face or the lower half. These so-called half-face conditions provide corresponding single-factor conditions that are important to include with the factorial conditions. Thus, there was a total stimulus set of 35 faces—25 whole, 5 upper-half, and 5 lower-half faces. In one experiment, we used a two-alternative forced choice between Happy and Angry, and in another, we used nine rating steps from Happy to Angry. The results indicated that participants evaluate and integrate information from both features to perceive affective expressions. Both choice probabilities and ratings showed that the influence of one feature was greater to the extent that the other feature was ambiguous. The fuzzy logical model of perception (FLMP) fit the judgments from both experiments significantly better than an additive model. Given the good fit of the FLMP with its assumptions of continuous and independent features, the research by Ellison and Massaro questions previous claims of categorical and holistic perception of affect.

The goal of the present study was to extend this paradigm to animated faces that give the appearance of a real person. With this technique, we propose to test whether the race of the face and the race/culture of the perceiver

This research was supported in part by research Grant 2 R01 DC 00236-13A1 from the National Institute on Deafness and Other Communication Disorders, National Institutes of Health, the National Science Foundation (BNS 8812728), and the University of California, Santa Cruz. The authors thank Michael M. Cohen for programming help and Doug Hintzman for insightful editorial guidance. Correspondence should be addressed to D. W. Massaro, Department of Psychology, University of California, Santa Cruz, CA 95064 (e-mail: massaro@fuzzy.ucsc.edu).

		Mouth				
		Down	2	3	Up	None
Eyebrow	Down					
	2					
	3					
	Up					
	None					

Figure 1. Expansion of a typical factorial design to include upper face and lower face conditions presented alone. The four levels along the eyebrow and mouth continua represent deflections of the displayed feature along four physical steps from downward displacement to upward displacement.

influence the processes involved in this domain of pattern recognition. Before describing the experiment and the theoretical questions, we review relevant research and formulate the models to be tested. Several recent issues of interest have concerned the recognition and identification of emotional expressions. 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. They interpreted these results in terms of holistic processing in which the parts of a face are not explicitly represented. Recognition would necessarily involve matching the input to a whole face rather than being mediated by its component parts. As described by Ellison and Massaro (in press), however, the advantage of recognizing a feature in a face might be accounted for more parsimoniously by assuming that the other features make an independent contribution to perception. In this case, a positive influence of contextual features also applies to facial affect, as it does in other situations such as the perception of letters in words (Massaro & Cohen, 1994). In addition, there could be information from spatial relations among the parts of the face, which could give an advantage to the whole face relative to a single part. A functional role for spatial relations does not make recognition holistic, because the parts of the face would still have to be represented. Therefore, an advantage for recognizing a feature in a face does not necessarily imply holistic processing as defined by Tanaka and Farah.

Etcoff and Magee (1992) created several sets of face continua that switched between two different emotions. On the basis of a correspondence between their results and traditional results in speech perception, they concluded that most emotions were perceived categorically rather than continuously. We now know, however, that these traditional tasks do not provide an unambiguous measure of the degree to which perception is categorical or continuous (Massaro, 1987, chap. 4). When categorical and continuous models are directly compared, the

continuous model always provides the better description in both speech and emotion (Ellison & Massaro, in press; Massaro & Egan, 1996; Massaro, 1987).

Huber and Lenz (1993) studied the ability of pigeons to learn to discriminate among schematic human faces varying along four different features. They concluded that the pigeons achieved discrimination by adding the information from the different features. However, other theories of feature combination were not tested. Ellison and Massaro (in press), on the other hand, found that the multiplicative combination of the FLMP gave a significantly better description of facial affect recognition than an additive model. We now formulate the FLMP and some alternative models.

Fuzzy Logical Model of Perception

In the FLMP, it is assumed that there are three stages or operations between some stimulus event and a response (Figure 2). Specifically, with regard to the present experiments, the FLMP predicts that participants evaluate expressions of emotional affect according to information arriving from multiple sources of information (i.e., individual facial features). In a test of the model, stimulus variation was made along several dimensions (Figure 3). A feature from the upper half of the face and one from the bottom half were chosen to be varied systematically, and independently of each other. The information from each source was evaluated according to degree of match to a prototype or degree of support for a particular affect. This information was then integrated according to a multiplicative combination to determine how representative the stimulus was of a particular affect class. The decision

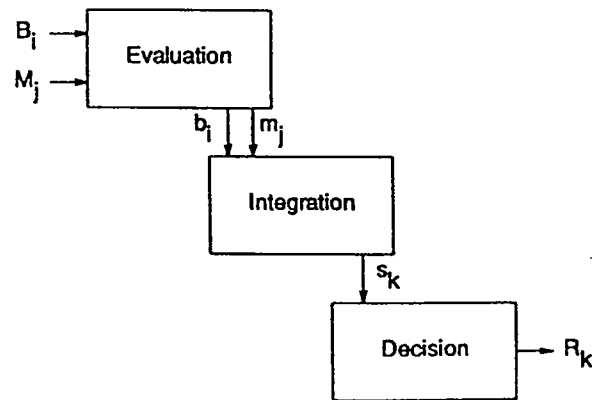


Figure 2. Schematic representation of the three stages involved in perceptual recognition. The three stages are shown to proceed left to right in time to illustrate their necessarily successive but overlapping processing. The sources of information are represented by uppercase letters. Eyebrow deflection is represented by B_i and mouth deflection by M_j . The evaluation process transforms these sources of information into psychological values (indicated by lowercase letters b_i and m_j). These sources are then integrated to give an overall degree of support, s_k , for a given affect alternative k . The decision operation maps the outputs of integration 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.

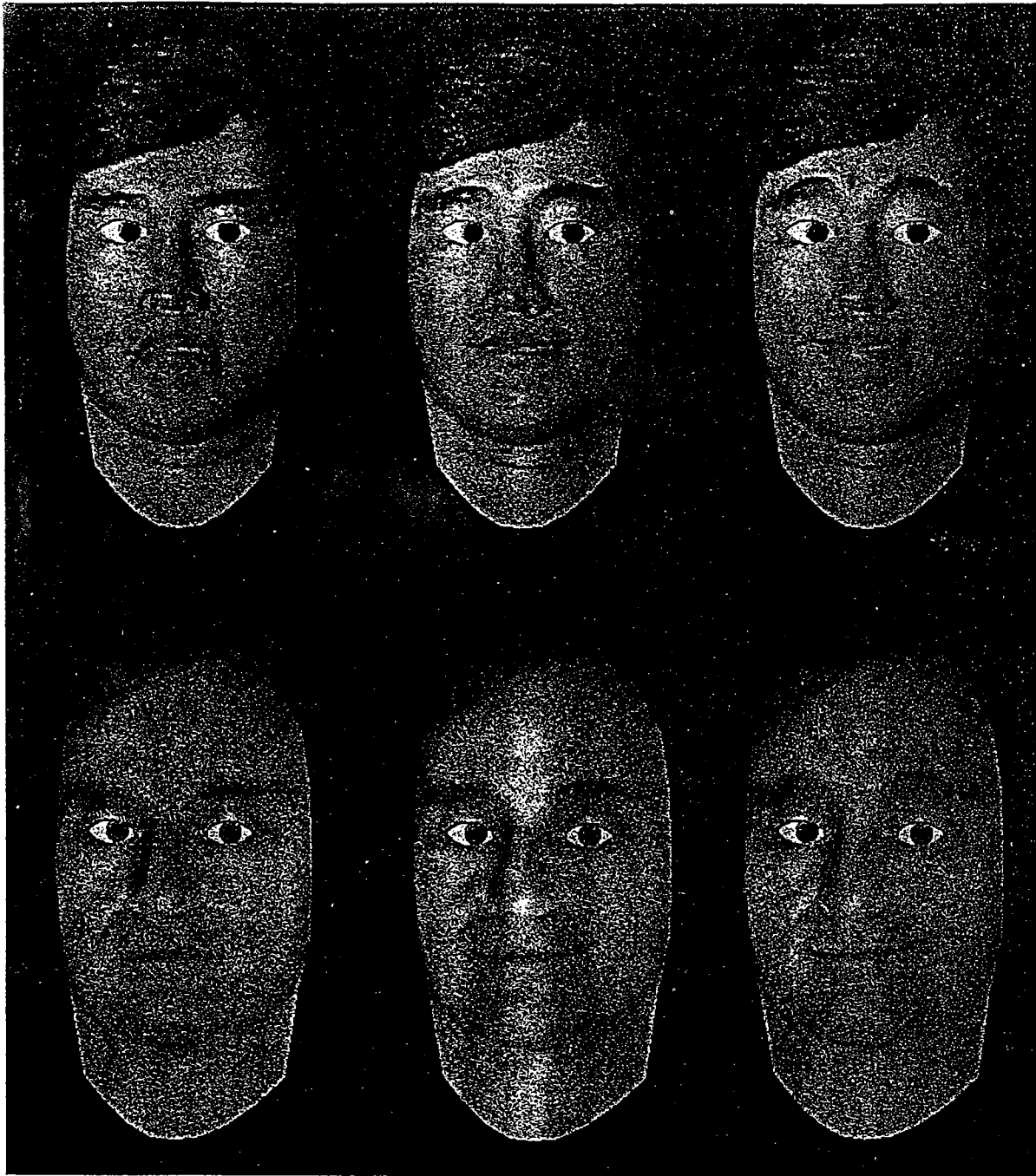


Figure 3. The Japanese and American face types displaying the maximum deflections of the brow and mouth features (at the corners of the figure) as well as faces displaying relatively deflections of each of the features.

process determined the relative goodness-of-match of the stimulus with each prototype and the appropriate response was made.

In the present experiment, the brow in the upper face was varied independently of the mouth in the lower face. To make the brow change from happy to angry, it was varied from elevated and arched to a pulling together of the inner corners and lowering of the eyebrows. The mouth

in the lower half of the face was varied from curled up at the corners for a prototypically happy affect to a flattening of the corners. These features were generated on the basis of features displayed in exemplar photographs in Ekman and Friesen (1975). Although highly arched eyebrows may not strictly be a necessary component of a classic "smile" (Duchenne de Boulogne, 1862/1990; Ekman & Friesen, 1975), this feature is more representative

of happy than of angry expressions. Similarly, the flattening of the corners of the mouth is more representative of an angry than a happy expression. Four levels of brow deflection and four levels of mouth deflection were factorially combined. In addition, trials were presented using only the top half or bottom half of the face; eight half-face conditions were presented, as prescribed by an expanded factorial design.

We assume that participants generate prototypes corresponding to happy and angry affects. The prototype corresponding to a happy face would include an implicit description of both the eyebrow and mouth features—Happy Face: eyebrows elevated and arched; mouth corners curled up. The prototype corresponding to an angry face, on the other hand, might consist of the following description: Angry Face: eyebrows pulled inward and flattened; mouth corners flattened.

No other sources of information contributing to facial affect were listed in these prototype descriptions because we assumed that they are not being influenced systematically by the independent variables of brow and mouth.

At the feature evaluation stage, each physical feature is transformed to psychological values indicating the degree of support for the different alternatives. As can be seen in Figure 2, these psychological values are represented in the model equations in lowercase letters. If B_i represents the brow information, then B_i would be transformed to b_i , the degree to which brow deflection supports the alternative, Happy (H). With just two alternatives, H and Angry (A), we can make the simplifying assumption that the degree to which brow supports the alternative A is $1 - b_i$ (Massaro & Friedman, 1990). Feature evaluation would occur analogously for the mouth, M_j . The mouth information (M_j) would be transformed to m_j to give support for the alternative Happy. The support for Angry would be $1 - m_j$.

Feature integration consists of a multiplicative combination of feature values supporting a given alternative. If b_i and m_j are the values supporting the alternative H , then the total support, $M(H)$, for the alternative H would be given by the product of b_i and m_j :

$$M(H) = b_i m_j. \quad (1)$$

Similarly, the total support, $M(A)$, for the alternative A would be given by the product of $1 - b_i$ and $1 - m_j$:

$$M(A) = (1 - b_i)(1 - m_j). \quad (2)$$

The third operation is decision, which uses a relative goodness rule (Massaro & Friedman, 1990) to give the relative degree of support for each of the test alternatives. In the two-alternative choice task,

$$P(H | B_i, M_j) = \frac{M(H)}{M(H) + M(A)}, \quad (3)$$

where $P(H | B_i, M_j)$ is the predicted probability of a Happy choice given Stimulus B_i, M_j . The FLMP requires four free parameters for the four levels of brow deflection and four for the four levels of mouth.

Additive Model of Perception

In the FLMP, a multiplicative combination of truth values representing the different sources of information is assumed. Other types of models have been proposed to explain performance in various other domains of pattern recognition. In an additive model of perception (AMP) (Cutting, Bruno, Brady, & Moore, 1992; Huber & Lenz, 1993), it is assumed that the sources of information available to the participant are added rather than multiplied, as in the FLMP. Adding the values at integration with a relative goodness rule at decision reduces to an averaging model (Massaro, 1987, chap. 7). In addition, this model can be made more general by allowing one featural dimension to have more influence than the other. Predictions of this weighted averaging model are given by

$$P(H | B_i, M_j) = wB_i + (1 - w)M_j, \quad (4)$$

where w is the weight given to brow and $(1 - w)$ is the weight given to mouth.

Analogous to the FLMP, the AMP requires four free parameters for the four levels of brow and four for the four levels of mouth. An additional free parameter is also necessary to accommodate the weight term. (We do not allow a bias term in the FLMP because the weight given a dimension is completely determined by its degree of support.) It should be stressed that the AMP is mathematically equivalent to a single-channel model in which the participant attends to information from just one feature on a particular trial (Cohen & Massaro, 1992). The AMP is also equivalent to a categorical model (Massaro, 1987, chap. 5), in which the participant categorizes information from each feature and responds with the outcome of the categorization of only one of the features (i.e., the brow) with a certain probability (or bias).

Holistic Model

Unfortunately, there is no straightforward formalization of a holistic model. The FLMP and AMP are capable of predicting performance as a function of the parts of the face. Predictions cannot be made for versions of a holistic model, however, in which either the parts are not represented or else they interact in such a manner that the outcome of their interaction cannot be predicted. In principle, this type of model requires a unique parameter for each unique experimental condition—that is, each variation of the face is considered to be unique. Thus, we cannot formulate a meaningful model to be tested. Given this holistic viewpoint, however, there is no reason to believe that any model formulated in terms of the parts of the face will give an adequate account of performance. This expectation is particularly true for results from an expanded factorial design in which half-face conditions are tested along with the whole faces. The same holistic model should not be able to describe both the half-face and whole-face conditions. Because of theoretical parsimony, a successful featural model would challenge this holistic approach.

A persistent issue in the study of human facial expression is the extent to which it is universal or culturally dependent (Ekman, 1992, 1993; Fridlund, 1994; Russell, 1994). Although this is a highly controversial question, we limit our investigation to the question of the recognition of facial expression. Furthermore, we are primarily concerned with whether the processes involved in the recognition of facial expression differ cross-culturally. More specifically, we ask whether the processes differ for recognition of a face from a different race relative to the same race, and whether persons from different cultures use different processes in recognition. We believe that this question can be answered because our study is carried out in the framework of the FLMP. The FLMP model allows us to distinguish between information and information processing. Within the context of the present experiment, information can be considered to be equal to the outputs of the evaluation process—the b_i and m_j of Equation 1. Information processing, on the other hand, refers to the nature of the integration and decision operations. On the one hand, we can ask (1) whether the same information is used in the recognition of the two different faces and/or (2) whether this information is the same across cultures. In addition, we can assess (1) whether the same processes occur in the recognition of two different faces and/or (2) whether information processing is consistent across cultures.

In a classic study, Ekman and Friesen (1975) asked college students in the United States and Japan to view stress-inducing films either alone or in view of a research assistant. When alone, the Japanese and American students appeared to make identical facial expressions. When in view of another person, however, the Japanese appeared to mask their facial expression more than the Americans. Although these results are highly controversial (see Fridlund, 1994), they provide a starting point for the present study. If indeed there are cultural differences in the display of facial expression, we would also expect to find cultural differences in the information value of the features used for recognition. If the Japanese tend to attenuate their facial expression, for example, then we might expect that Japanese students would tend to judge a Japanese facial feature as having a more extreme emotion than the equivalent American face. On the other hand, Japanese students might tend to be less extreme in their judgments overall relative to the American students if they tend to attenuate their own overt interpretations of emotion. Furthermore, in Ekman's (1972) neurocultural model, culture-specific learned habits can alter the appearance of the face linked with a specific emotional expression. In this case, the cues that perceivers use to recognize facial expression in one culture might differ somewhat from those used in another. All of these potential differences would be differences in information and not information processing.

Independently of these possibilities, however, the theoretical framework of model testing allows us to ask whether the processes involved in the recognition of fa-

cial expression differ in the two cultures. Given the distinction between information and information processing, it is possible that the features used to categorize facial expression might vary somewhat across cultures. Although the cues might be different, information processing might be invariant across culture, and consistent with the prediction of the FLMP. On the other hand, it is possible that the FLMP is capable of describing the information processing of one culture, but not another. Given the ubiquitous findings of support for the FLMP in a variety of pattern recognition tasks across development and aging, different languages, and content domains, we predicted that the information processing of both the Japanese and American students would be well described by the FLMP. This approach is analogous to previous cross-linguistic experiments comparing speech perception processes of speakers of different languages (Massaro, Cohen, & Smeele, 1995).

METHOD

Subjects

The Japanese participants were 16 native Japanese students who were attending a 3-week English language course at the University of California, Santa Cruz (UCSC). The students ranged in age from 19 to 20 and were all right-handed. There were 14 males and 2 females in the group. Although the students understood some English, the experimental instructions were written in Japanese Kanji and were explained in Japanese by a Japanese interpreter. The United States (hereafter "American") participants were 14 regular UCSC students who participated as part of their required undergraduate coursework. The students ranged in age from 18 to 41 and were all right-handed. There were 6 males and 8 females in the group. All participants were naive to the procedures and hypotheses of the study, and had not seen the synthetic face stimuli prior to the experimental sessions.

Apparatus

The stimuli used in these experiments were generated by facial synthesis software utilizing a parametrically controlled polygon topology synthesis technique with texture-mapped skin surfaces and ray-tracing lighting simulation. Because the skin surface texture is chosen arbitrarily, we can use digitized images of actual persons, wrapping the features closely around the polygons of the face frame. This capability, combined with the facial motion simulation attained by varying polygon vertex angles and edge lengths, allows the program to simulate motion of features in realistic face images in a quantifiable and reproducible manner. A complete description of the software of the face and its use in experimentation is given by Cohen and Massaro (1994).

Twenty-four face stimuli were constructed to portray affective expressions that varied along two feature dimensions in an expanded-factorial design. The features manipulated in the stimulus set were brow and mouth characteristics. These two features were orthogonally varied in the factorial condition along continua of four steps. In addition, upper-half and lower-half faces were presented for brow-only and mouth-only conditions at the same four steps. The features were determined as described in the description of the FLMP and were chosen because they were consistently identified with Happy and Angry affects in previous experiments (Ellison & Massaro, in press). Each face began at a neutral expression (not included as a level in the design), moving to the target feature value(s) during the first 600 msec and remaining on, without motion, for an additional 400 msec. A feature value was changed from neutral to the target

value according to a negatively accelerated exponential function. The rate parameter of this function was selected to provide a smoothly visible and plausible rate of feature deflection.

These 24 face stimuli were implemented in the stimulus set by using digitized images of an adult male American face and an adult male Japanese face for the texture-mapped skin. These face images were retouched slightly prior to the mapping process to facilitate the assignment of features to the wire-frame polygons of the synthetic face. The underlying wire-frame structure, however, was the same across the two face types for a given level of deflection of the features. Figure 3 depicts some of the faces at their target values for the Japanese and American faces. This stimulus set, then, incorporated 24 faces—16 in the whole-face condition and 4 in each half-face condition, for each of two face types, for a total of 48 unique faces.

The experiment was controlled by a Silicon Graphics 4D/Crimson VGX workstation running under the IRIX operating system. The stimulus faces were presented in color to each participant on a 12-in. (diagonal) NEC Model C12-202A color monitor. Responses were collected on TVI 950 VDT terminals and their associated keyboards.

Procedure

Each participant was tested in one of four separate sound-attenuated rooms. A two-alternative forced choice (2AFC) response paradigm was used. Participants were required to respond to each stimulus face with either "Happy" or "Angry" by pressing a correspondingly labeled key on either the left or right edge of the keyboard. The keys were labeled in Kanji for the Japanese, and the position of the keys was counterbalanced across blocks of participants. We did not show exemplar faces of the two alternative responses or give feedback during the experiment. The experiment began with a short title sequence. The faces were displayed for 1,000 msec each, with a 100-msec 1000-Hz orienting beep sounding 700 msec prior to

stimulus onset. The stimulus faces (Figure 3) were sized to fill the vertical dimension of the 12-in. monitor screens (16 cm high) and were viewed at a distance of about 40 cm. Thus the face subtended a visual angle of about 31° horizontally and 43° vertically. No visual fixation point was provided. Because the control program collected all 4 participants' responses before displaying the next face in the stimulus set, there was a short but variable time between stimulus trials, on the order of 3 to 4 sec. Each experimental session included 10 practice trials and 288 experimental trials; the experimental trials were selected randomly without replacement within each block of 48 trials. Each participant was tested in two sessions, separated by a 5-min rest period, for a total of 12 observations for each of the 48 unique faces.

RESULTS

Given the 2AFC task, the probability of identifying the face as happy, $P(\text{Happy})$, was sufficient to represent choice performance. Figure 4 gives the average results for the Japanese participants identifying the Japanese and American faces. As can be seen in the figure, the results are very similar for the two different faces. There were significant main effects of both brow and mouth and their significant interaction (all $ps < .001$). The probability of a happy identification increased with upward brow deflection and with upward mouth corner deflection. The statistically significant interaction indicated that the influence of one variable was larger to the extent the other variable was ambiguous. It was also the case that brow deflection was generally less influential than mouth deflection. As

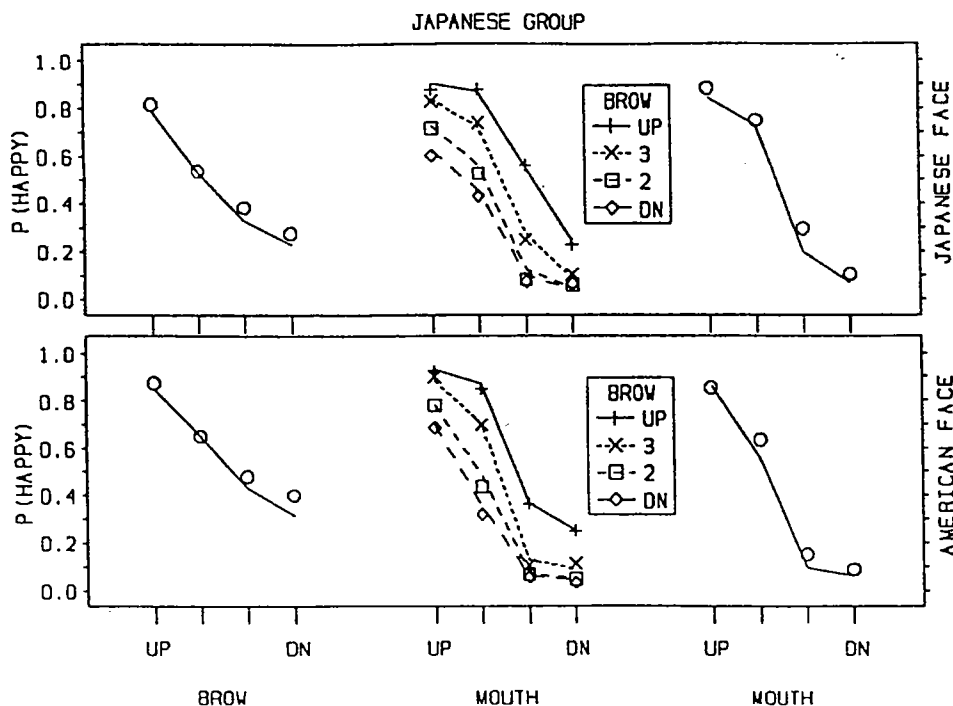


Figure 4. Predicted (lines) and observed (points) $P(\text{Happy})$ judgments of the Japanese group as a function of the brow and mouth conditions, for both the Japanese face (top panel) and the American face (bottom panel). The left and right portions of the graph give the single-factor half-face conditions, whereas the middle portion gives the factorial full-face conditions. Predictions are for the fuzzy logical model of perception.

can be seen in the right and left panels of Figure 4, the change in $P(\text{Happy})$ was less in the half-face brow than in the half-face mouth condition.

Figure 5 gives the average results for the American participants identifying the Japanese and American faces. These results are similar to those of the Japanese participants shown in Figure 4. At this coarse level of analysis, it appears that there were no large ethnic or cultural differences. We also use the model analysis to provide a more rigorous test of these influences.

Test of the Models

The models were fit separately to each of the individual participants and to the mean participant computed by averaging the results across participants. Model fitting was accomplished through the use of the STEPIT subroutine (Chandler, 1969), which estimates the parameters of the model to generate predicted performance. The subroutine changes the model parameters until values are found to minimize the differences between predicted and observed performance. This routine minimizes the root mean square deviation (RMSD) between an individual participant's observed and predicted choice probabilities. The value of the RMSD gives the goodness of fit of a model, with lower RMSDs indicating better fits.

Given the 4×4 expanded factorial design, eight free parameters are necessary to fit the FLMP to the 24 conditions for a given face (Japanese or American): four pa-

rameters for each level of brow deflection and four for mouth deflection. The parameter values represent the degree to which these features match those in the happy prototype. The FLMP predictions of the average results are given in Figures 4 and 5.

The AMP was also fit to the data using the same procedure. Given the additional weight parameter in the AMP, nine free parameters are necessary. Table 1 gives the average RMSD values as a function of group, model, and face. An analysis of variance was carried out on the RMSD values with group, model, face, and participant as factors. The RMSDs for the FLMP were significantly lower than those for the AMP [$F(1,20) = 37.42, MS_e = .032, p < .001$]. Thus, the FLMP yielded a significantly better description of the results. The only other significant effect was participant group [$F(1,20) = 6.59, MS_e = .0015, p = .015$]. Both models gave a better description of the American than the Japanese subjects. However, as will be seen in a benchmark measure of performance, this difference reflects slightly less extreme response proportions (and therefore larger expected variance) for the Japanese relative to the American participants (Table 1). Thus, the statistically significant effect of group is not psychologically meaningful.

The two models were also fit to the average results averaged across participants within a group. The FLMP fit of the mean participant gave an RMSD of .0320, whereas the fit of the AMP to the mean participant data was .0796. For the American group, the FLMP fit of the mean par-

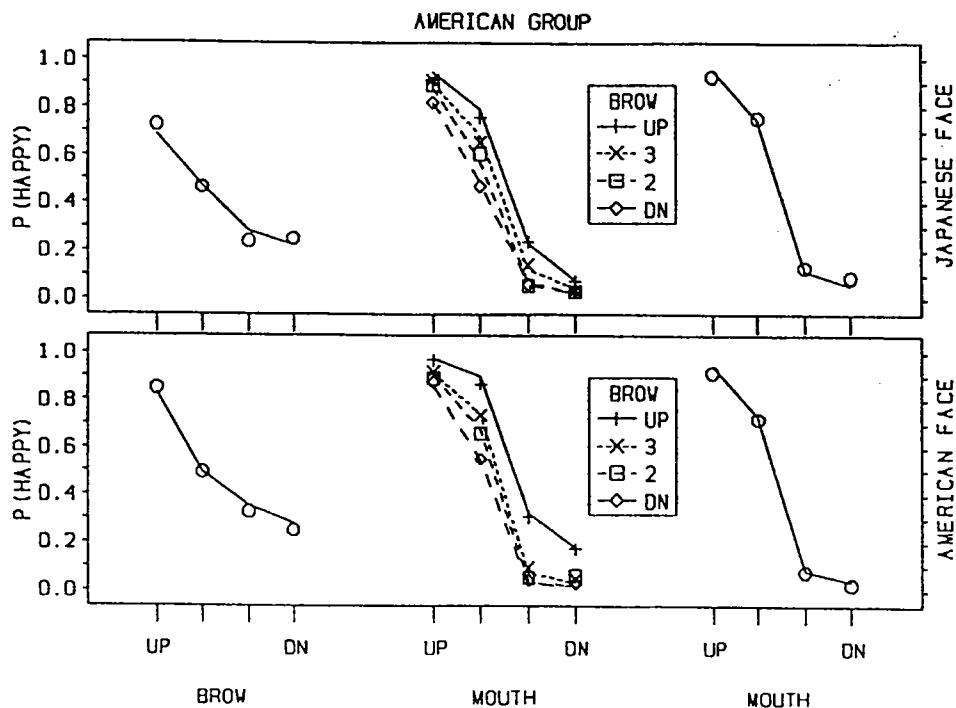


Figure 5. Predicted (lines) and observed (points) $P(\text{Happy})$ judgments of the American group as a function of the brow and mouth conditions, for both the Japanese face (top panel) and the American face (bottom panel). The left and right portions of the graph give the single-factor half-face conditions, whereas the middle portion gives the factorial full-face conditions. Predictions are for the fuzzy logical model of perception.

Table 1
Average Observed RMSD and Adjusted Benchmark RMSD(ab) of the Individual Participant Fits by the FLMP and AMP for the American and Japanese Faces and for the American and Japanese Groups

Face	Model	Fit	Group	
			American	Japanese
American	FLMP	RMSD	.0661	.0879
		RMSD(ab)	.0551	.0611
	AMP	RMSD	.0989	.1241
		RMSD(ab)	.0517	.0572
Japanese	FLMP	RMSD	.0726	.0784
		RMSD(ab)	.0577	.0616
	AMP	RMSD	.0927	.1183
		RMSD(ab)	.0541	.0578

Note—FLMP, fuzzy logical perception model; RMSD, root mean square deviation; AMP, additive model of perception.

participant gave an RMSD of .0262, whereas the fit of the AMP to the mean participant data was .0437.

The advantage of the FLMP did not interact with participant group or face type. These results speak to the question of information processing differences as a function of participant's ethnic identity and culture and that of the face being judged. Given that the goodness-of-fit of the FLMP and its advantage over the AMP did not vary significantly with ethnic group or face type, we can conclude that information processing is similar and follows the form of the FLMP.

Given the good fit of the FLMP, the parameter values can be used to evaluate the influence of the brow and mouth features. To assess the magnitude of the effects of the brow and mouth features, the parameter values were used to compute the range between the first and fourth levels of each factor. The larger the difference between the parameter values at these two levels, the larger the influence of that factor. This difference is between 0 and 1, assuming that the influence of a factor will always be monotonic in the appropriate direction. Figures 6 and 7 plot these parameter values for the two faces and the two groups of participants. In both figures, brow effect is plotted against mouth effect for each participant and each face type. The figures indicate that most participants in both groups were more influenced by changes in the mouth than in the brow. However, large individual differences in parameter values are observed within each group. Thus, there were clearly information differences within a group.

To address the issue of information differences between the American and Japanese students, the parameter values of the FLMP fits were submitted to an ANOVA, with mouth, brow, group, and face as factors. Figure 8 gives the average brow and mouth parameters for the two faces and two groups. As can be seen in the figure, no major differences were found. As mentioned in the analysis of choice responses, it can be seen that the mouth was more influential than the brow. Although a small effect, one consistent finding was that the brow of the American face was evaluated as happier overall than that of the Jap-

anese face. Finally, the middle two levels of the mouth parameter were seen as happier for the Japanese face than for the American face for the Japanese participants (all $ps < .05$).

Although the FLMP provides a significantly better fit than the AMP, it is valuable to determine how good the fit is in an absolute sense. A benchmark measure has been developed to provide this index of goodness-of-fit of a model (Massaro & Cohen, 1993). Even if a model is perfectly correct, we cannot expect it to fit results perfectly. A probabilistic prediction is necessarily associated with some variability. This variability depends on the number of observations and the response probability. Thus, we can expect some mismatch between the predicted and observed values, even if the model is correct. In the present task with two response alternatives, the observed variability should be equal to that expected from binomial variability. It is possible to determine the expected binomial variability as a function of the observed response probabilities and the number of observations at each experimental condition. We can therefore ask if the fit of a model is poorer than this expected variability.

The standard deviation of a binomial distribution (with two outcomes) is equal to the square root of its binomial variance:

$$\sigma = \sqrt{\frac{pq}{N}}, \quad (5)$$

where p is the probability of one outcome (response), q the probability of the other ($q = 1 - p$), and N is the number of observations.

The benchmark RMSD is determined by computing the binomial variance for each of the 24 experimental conditions, averaging these 24 values, and finally taking the

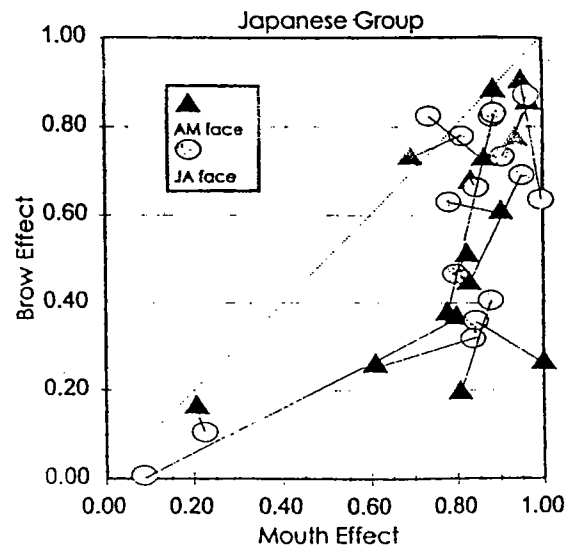


Figure 6. Brow effect and mouth effect for the Japanese group. Lines connect individual participant's values for the American and Japanese faces.

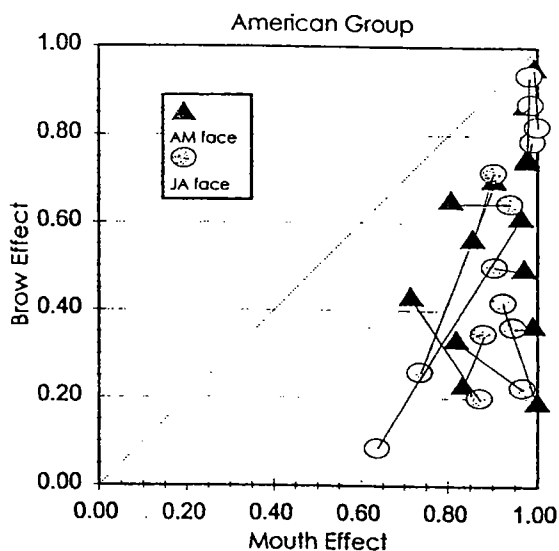


Figure 7. Brow effect and mouth effect for the American group. Lines connect individual participant's values for the American and Japanese faces.

square root. If $RMSD(b)$ is defined as the benchmark $RMSD$, then

$$RMSD(b) = \sqrt{\frac{\sum_{i=1}^k (pq/N)}{k}}, \quad (6)$$

where k is the number of experimental conditions (24 for each face in the present study).

One possible limitation with $RMSD(b)$ as a benchmark is the fact that the models have free parameters, whereas this benchmark does not. Each free parameter in a model could allow one experimental condition to be predicted exactly. If each free parameter is used to predict one of the experimental conditions exactly, and the model makes correct predictions of the other conditions without any free parameters, the expected $RMSD$ for the predictions would necessarily be less than the $RMSD(b)$ given by Equation 6. If there are as many free parameters as experimental conditions, then the $RMSD$ would be zero. Accordingly, it is necessary to adjust the $RMSD(b)$ on the basis of the number of free parameters used by a given model. Following the logic of statistics, this adjusted benchmark, $RMSD(ab)$, is given by

$$RMSD(ab) = \frac{(k-f)}{k} \times RMSD(b), \quad (7)$$

where f is the number of free parameters of the model, k is the number of experimental conditions being predicted, and $RMSD(b)$ is given by Equation 6. With no free parameters, the benchmark would reduce to Equation 6. With k parameters, the adjusted benchmark would be zero. In general, the $RMSD(ab)$ values provide a benchmark to evaluate the observed goodness of fit of a model.

An $RMSD(ab)$ was computed for each participant for each face. Table 1 gives the $RMSD(ab)$ values for the two models. An analysis of variance was carried out comparing the $RMSD$ and $RMSD(ab)$ values, with group and face as additional factors. For the FLMP, the $RMSD(ab)$ values averaged .0590, which was significantly lower than the $RMSD$ of .0760 [$F(1,28) = 17.65, p < .001$]. Thus, the fit of the FLMP is somewhat worse than the corresponding benchmark. For the AMP, the $RMSD(ab)$ values averaged .0552, which was significantly lower than the $RMSD$ of .1094 [$F(1,28) = 149.03, p < .001$].

Finally, we can test the FLMP's prediction that reaction time (RT) should increase to the extent that the facial information is ambiguous. In the framework of the FLMP, a more ambiguous face will require more time for the decision process (see Massaro & Cohen, 1994). Decision time appears to be a roughly linear function of the relative goodness-of-match given by the relative goodness rule. In this case, choice RT should be a linear function of ambiguity, when ambiguity is defined as the extent to which $P(\text{Happy})$ approaches .5:

$$A = .5 - |P(\text{Happy}) - .5|, \quad (8)$$

where $|x|$ is the absolute value of x . In this case, ambiguity varies between 0 when $P(\text{Happy})$ is 0 or 1, and .5 when $P(\text{Happy})$ is .5. An RT was computed for each of the 24 stimulus conditions at each of the two faces and correlated with the A values. Figure 9 shows the strong relationship between this measure of ambiguity and choice RT for these 48 conditions. There was a strong positive correlation between A and RT ($r = .846, p < .001$). This result is consistent with the FLMP assumption that it is the overall ambiguity of the stimulus event that is important for processing time.

DISCUSSION

The present experiments proved successful in addressing the issue of how two features of facial affect are evaluated and integrated to achieve perceptual recognition. Both brow deflection and mouth deflection were effective in changing the judgment from happy to angry. In addition, the influence of one of these features was larger to the extent that the other feature was ambiguous. These results were well described by the FLMP, a model that has proven successful in a wide variety of domains. Given that the FLMP assumes independent features, its good fit to experimental results challenges holistic models (Tanaka & Farah, 1993). The FLMP also assumes continuous features, which poses a problem for theories of categorical perception of affect (Etcoff & Magee, 1992). Finally, the multiplicative integration of the FLMP, as well as the poor fit of the AMP, weakens additive models of feature integration (Huber & Lenz, 1993).

The RTs of the judgments also increased to the degree that the test face displayed an ambiguous expression. In accord with the predictions of the FLMP, participants took longer to respond when the two features were dis-

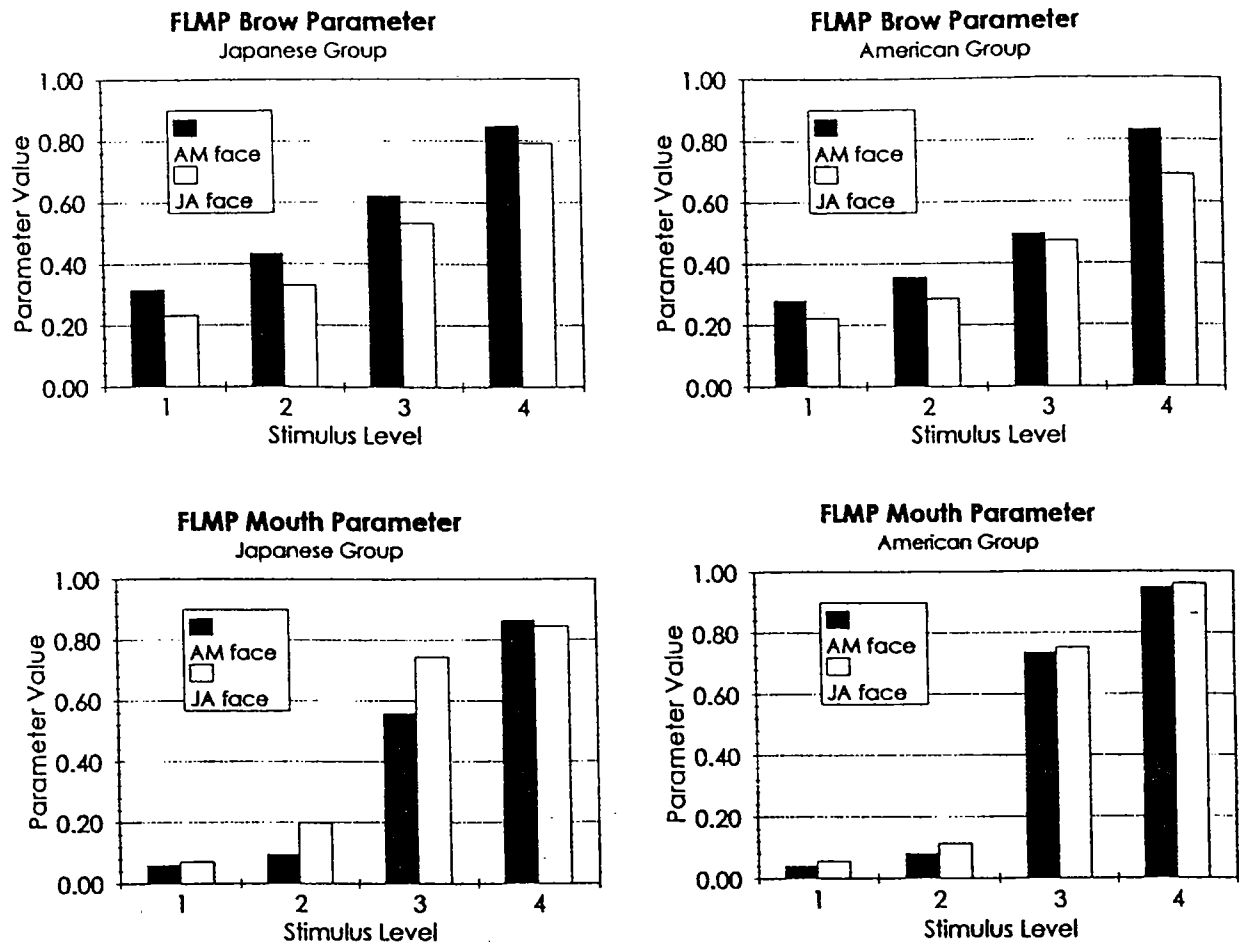


Figure 8. Average parameter values for the brow and mouth features for the Japanese and American participants and the Japanese and American faces.

played in conflicting directions, when the two features were only slightly deflected (ambiguous), or when one of the features was missing and the other was ambiguous. The fastest responses were when features were congruent and unambiguous.

The present results in the facial expression domain are highly analogous to those found in speech perception. Multiple sources of information have been shown to be influential in speech perception. In addition, the FLMP gives a good description of performance when multiple speech cues are varied in an expanded factorial design (Massaro et al., 1995). Furthermore, there appear to be information (but not information processing) differences across different individuals, ages, and languages (Massaro, 1987, 1992, 1994; Massaro et al., 1995). In the facial expression domain, we have inferred information differences as a function of both individual and ethnicity. However, information processing was invariant across individuals, ethnicity of the perceiver, and the face being judged.

As with all research, additional empirical evidence is necessary to determine the generality of the conclusions. Two obvious limitations are that our experiments to date

have used just two emotions and have limited judgments to just two possible response alternatives. Our paradigm is easily extended to other emotions and multiple response alternatives, however, and findings in the speech domain give us confidence that the present conclusions go well beyond the present task. We encourage other scientists to use the expanded factorial design to help provide a much-needed database on the perception of facial affect.

In summary, there is a good deal of literature (and debate) regarding whether the cues for facial affect are universal across cultures. The present research goes further by asking whether the information processing of facial cues is universal. We have shown how the distinction between information and information processing can be formalized within the framework of the FLMP. Given this framework, we were able to address not only whether the cues are universal but also whether the cues are combined in the same manner across cultures and for faces of different races. The results indicate that the perceptual processing of facial affect follows well-established principles of pattern recognition, independent of the ethnicity and culture of the perceiver and of the perceived face.

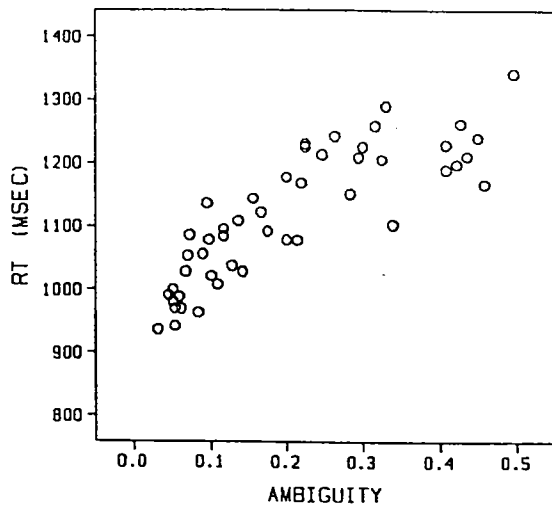


Figure 9. Reaction time (RT) as a function of ambiguity (A) for the 48 conditions (the 24 expanded factorial design conditions \times the two faces). Each point is the average of 15 participants.

REFERENCES

- CHANDLER, J. P. (1969). Subroutine STEPIT—Finds local minima of a smooth function of several parameters. *Behavioral Science*, *14*, 81-82.
- COHEN, M. M., & MASSARO, D. W. (1992). On the similarity of categorization models. In F. G. Ashby (Ed.), *Multidimensional models of perception and cognition* (pp. 395-447). Hillsdale, NJ: Erlbaum.
- 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.
- COHEN, M. M., & MASSARO, D. W. (1994). Development and experimentation with synthetic visible speech. *Behavior Research Methods, Instruments, & Computers*, *26*, 260-265.
- CUTTING, J. E., BRUNO, N., BRADY, N. P., & MOORE, C. (1992). Selectivity, scope, and simplicity of models: A lesson from fitting judgments of perceived depth. *Journal of Experimental Psychology: General*, *121*, 364-381.
- DUCHENNE DE BOULOGNE, G.-B. (1990). *The mechanism of human facial expression*. Cambridge: Cambridge University Press. (Original work published 1862)
- EKMAN, P. (1972). Universals and cultural differences in facial expressions of emotion. In J. Cole (Ed.) *Nebraska Symposium on Motivation*, 1971 (Vol. 19, pp. 207-283). Lincoln: University of Nebraska Press.
- EKMAN, P. (1992). Are there basic emotions? *Psychological Review*, *99*, 550-553.
- 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.
- ELLISON, J. W., & MASSARO, D. W. (in press). Featural recognition and identification of facial affect. *Journal of Experimental Psychology: Human Perception & Performance*.
- ETCOFF, N. L., & MAGEE, J. J. (1992). Categorical perception of facial expressions. *Cognition*, *44*, 227-240.
- FRIDLUND, A. J. (1994). *Human facial expression: An evolutionary view*. San Diego, CA: Academic Press.
- HUBER, L., & LENZ, R. (1993). A test of the linear feature model of polymorphous concept discrimination with pigeons. *Quarterly Journal of Experimental Psychology*, *46B*, 1-18.
- MASSARO, D. W. (1987). *Speech perception by ear and eye: A paradigm for psychological inquiry*. Hillsdale, NJ: Erlbaum.
- MASSARO, D. W. (1992). Broadening the domain of the fuzzy logical model of perception. In H. L. Pick, Jr., P. Van den Broek, & D. C. Knill (Eds.), *Cognition: Conceptual and methodological issues* (pp. 51-84). Washington, DC: American Psychological Association.
- MASSARO, D. W. (1994). Bimodal speech perception across the lifespan. In D. J. Lewkowicz & R. Lickliter (Eds.), *The development of intersensory perception: Comparative perspectives* (pp. 371-399). Hillsdale, NJ: Erlbaum.
- MASSARO, D. W., & COHEN, M. M. (1990). Perception of synthesized audible and visible speech. *Psychological Science*, *1*, 55-63.
- MASSARO, D. W., & COHEN, M. M. (1993). The paradigm and the fuzzy logical model of perception are alive and well. *Journal of Experimental Psychology: General*, *122*, 115-124.
- MASSARO, D. W., & COHEN, M. M. (1994). Visual, orthographic, phonological, and lexical influences in reading. *Journal of Experimental Psychology: Human Perception & Performance*, *20*, 1107-1128.
- MASSARO, D. W., COHEN, M. M., & SMEELE, P. M. T. (1995). Cross-linguistic comparisons in the integration of visual and auditory speech. *Memory & Cognition*, *23*, 113-131.
- MASSARO, D. W., & EGAN, P. B. (1996). Perceiving affect from the voice and the face. *Psychonomic Bulletin & Review*, *3*, 215-221.
- MASSARO, D. W., & FRIEDMAN, D. (1990). Models of integration given multiple sources of information. *Psychological Review*, *97*, 225-252.
- RUSSELL, J. A. (1994). Is there universal recognition of emotion from facial expression? *Psychological Bulletin*, *115*, 102-141.
- TANAKA, J. W., & FARAH, M. J. (1993). Parts and wholes in face recognition. *Quarterly Journal of Experimental Psychology*, *46A*, 225-245.

(Manuscript received June 28, 1995;
revision accepted for publication December 5, 1995.)