Modeling Human Emotion Perception

Menna El-Shaer

1 Experiment 1

1.1 Methods

1.1.1 Participants

Thirteen human subjects (four females, nine males) with mean age of 22 years and standard deviation of 2 with normal or corrected-to-normal vision were drawn from the population of students at The Ohio State University. They were seated in front of a personal computer with a 100Hz, 21" CRT monitor. The distance between the eye and the monitor screen was measured to be approximately 50 cm.

1.1.2 Stimuli

Three hundred and fifty grayscale face images were used, consisting of six facial expressions of emotion (happiness, surprise, anger, sadness, fear, and disgust) plus neutral from a total of 50 people. These images were selected from the Database of Facial Expressions of Emotion. Original image sizes of the faces in the database are 1,000 x 750 pixels. To produce the stimuli, we cropped the faces and the image resolution was reduced to about 340 x 500 pixels. We scaled the images back using bilinear interpolation, which preserves most of the spatial frequency components of the image to 300 x 300 pixels to provide a visual angle of 9.2°

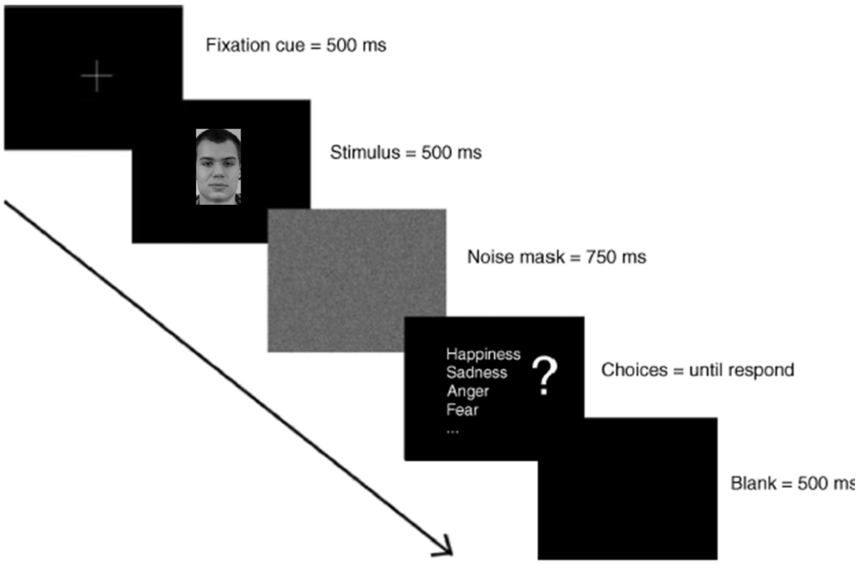

Figure 1: Figure 1: Stimulus timeline. A white fixation cross in black background is shown for 500 ms. Then, a stimulus image was shown for 500ms, followed by a random noise mask for 750 ms. A 7AFC task was used. After the subjects response, the screen went blank for 500 ms and the process was repeated.

1.1.3 Design and Procedure

Subjects observed face images of the seven facial expressions of the 50 identities and they were asked to judge the emotion perceived in each image. Each subject was shown each expression image once with a random order. They were first shown an image of a white fixation cross on a black background for 500 ms prior to the stimulus whose display duration was also 500 ms, followed by a random noise mask shown for 750 ms. A 7AFC (Alternative Forced Choice) task was used, where subjects had to select one of the six emotion labels or neutral. After the subjects response, the screen went black for 500 ms before starting the process again with a different stimulus image. Figure 1 illustrates a typical presentation timeline. The entire experiment lasted about 30 minutes.

1.1.4 Data Analysis

We calculated the accuracy of emotion perception for each emotion and subject. The accuracies were calculated as the number of recorded correct responses for each emotion and subject (N) divided by the total number of trials which represent the number of stimuli images shown (n).

$$Accuracy = N/n \tag{1}$$

1.2 Results

Table 1 shows the probabilities of correct emotion perception of 13 subjects calculated for the six described emotions and neutral expressions of 50 different facial identities.

Emotion	Mean	Standard Deviation
Anger	0.5615	0.2858
Disgust	0.6015	0.2533
Fear	0.5862	0.1965
Happy	0.9862	0.0299
Neutral	0.8846	0.1753
Sad	0.6615	0.2853
Surprise	0.8446	0.1663

Table 1: Table entries show the mean and standard deviations of the correct perception accuracies for all emotion expressions observed.

2 Experiment 2

2.1 Methods

To define the dimensions and the form of the face space, we used norm-based space representations (Neth & Martinez, 2009), where variations in the shape of the facial expression are represented with reference to the norm (statistical mean) face. To compute the mean face, we used the x and y coordinates of 78 landmark points on the face, which were manually selected, to describe its shape. First, we computed the Procrustes mean of 220 different face identities taken from the Facial Expressions Database, and applied Principal Component Analysis (Jolliffe, 2002) to obtain the feature vectors most descriptive of those 220 faces (i.e., accounting for the most to least variance). As a result, the dimensions of the problem were reduced from 78x2 (we used 78 landmarks to define the face, each landmark was determined by its x and y coordinates in the face space) to principal components that constituted 90% of the variance in the face shape data. Taken together, the Procrustes mean and those principal components helped define the computational face space in which we intended to test our proposed models. We defined the resulting principal components as the shape vectors ($\mathbf{x}$) defined as:

$$\mathbf{X}_k = [x_{1k}, x_{2k}, x_{3k}, ..., x_{nk}]^T$$

where n is the number of face identities used in Experiment 1 and $k = 1, ..., K$, where K is the total number of principal components computed for a specific emotion category. Second, we applied Linear Regression Analysis (Cohen, 2003) and Goodness of

Fit statistics to find the best fitting function of each model to determine which model of perception, whether it is the categorical or the continuous best fits the behavioral data. We used the least squares approach to do the fitting analysis, where we tried to find the best fitting function that minimizes the sum of squared residual errors (E) defined as:

$$E = \sum_{i=1}^{n} (Y_i - y_i)^2 \tag{2}$$

,where Y represents the correct perception accuracies from Experiment 1.

$$y = c_1 \mathbf{X} + c_2 \tag{3}$$

is the regression function that maps the shape vectors ($\mathbf{X}$) to the behavioral accuracies (y). In matrix form, $\mathbf{c} = [c_1, c_2]^T$ defines the unknown coefficients (c_1) and (c_2) that can be calculated using:

$$\mathbf{c} = (A^T A)^{-1} A^T y \tag{4}$$

where A is an nx2 matrix consisting of $\mathbf{x}$ as the first column and *ones* as the second column.

After obtaining the best model parameters by Regression Analysis, we applied Leave-one-out Cross-validation Analysis (Picard, 1984) to test the reliability of our model.

2.2 Results

We calculated the training errors or Root Mean Square Errors ($RMSE$) of the linear model as:

$$RMSE = \sqrt{avg(Y - y)^2} \tag{5}$$

where Y are the actual perception accuracy values, while y are the predicted values of the perception accuracies by either the Categorical or the Continuous model, and Coefficients of Determination (R^2) as:

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (Y_i - y_i)^2}{\sum_{i=1}^{n} (Y_i - \bar{Y})^2} \tag{6}$$

where n is the number of face identities, which in linear regression statistics is the same as the square of the Pearson Correlation Coefficient (r) between the predictors, i.e. shape vectors, and the response variable, i.e. the perception accuracy in our case.

Tables 2 and 3 show $RMSE$ and r values for the Categorical and Continuous models respectively.

Emotion	$RMSE$	r
Anger	0.2463	0.4922
Disgust	0.2031	0.5864
Fear	0.1581	0.5830
Happy	0.0255	0.5047
Neutral	0.1396	0.5940
Sad	0.2395	0.5297
Surprise	0.1386	0.5393

Table 2: Table entries show the Linear Regression $RMSE$ and correlation coefficient values computed by the Categorical model.

Emotion	$RMSE$	r
Anger	0.2060	0.6854
Disgust	0.2101	0.5459
Fear	0.1472	0.6541
Happy	0.0228	0.6361
Neutral	0.1507	0.4957
Sad	0.2262	0.5988
Surprise	0.1364	0.5603

Table 3: Table entries show the Linear Regression $RMSE$ and correlation coefficient values computed by the Continuous model.

Leave-one-out Cross-validation was performed by removing one face identity at a time, repeating the computational analysis and using the removed face identity to test our models by plugging in the shape of that identity in equation (3) and calculating the perception accuracy (y) for the model. We then calculated the RMSE using equation (5). We repeated this for all 50 identities and calculated the mean $RMSE$ and standard deviations. Tables 4 and 5 show the computed mean $RMSE$ and standard deviation results for the Categorical and Continuous models respectively.

Emotion	Mean	Standard Deviation
Anger	0.3386	0.0880
Disgust	0.3094	0.0834
Fear	0.2332	0.0485
Happy	0.0370	0.0125
Neutral	0.2138	0.0759
Sad	0.3431	0.0953
Surprise	0.1958	0.0614

Table 4: Table entries show the mean Linear Regression $RMSE$ and standard deviations for all emotion expressions, after using one identity at a time to test the Categorical model.

Emotion	Mean	Standard Deviation
Anger	0.3450	0.1039
Disgust	0.2937	0.0561
Fear	0.2374	0.0520
Happy	0.0361	0.0086
Neutral	0.2129	0.0667
Sad	0.3422	0.0835
Surprise	0.1869	0.0317

Table 5: Table entries show the mean Linear Regression $RMSE$ and standard deviations for all emotion expressions, after using one identity at a time to test the Continuous model.

3 References

The Resolution of Facial Expressions of Emotion, Shichuan Du, Aleix M. Martinez, Journal of Vision, 11(13):24, 113, 2011

Emotion perception in emotionless face images suggests a norm-based representation, Donald Neth, Aleix M. Martinez, Journal of Vision, January 8, 2009 9(1): 5; doi:10.1167/9.1.5

Procrustes Methods in the Statistical Analysis of Shape, Colin Goodall, Journal of the Royal Statistical Society. Series B (Methodological), Vol. 53, No. 2, 1991, pp. 285-339

Jolliffe, I. T., 2002, Principal Component Analysis, Second ed. Springer Series in Statistics. New York: Springer-Verlag New York

Cohen, J., Cohen P., West, S.G., & Aiken, L.S., 2003. Applied multiple regression/correlation analysis for the behavioral sciences. (2nd edition) Hillsdale, NJ: Lawrence Erlbaum Associates

Cross-Validation of Regression Models, Picard, Richard; Cook, Dennis, Journal of the American Statistical Association, 1984, 79 (387): 575583