Steyvers, M., & Busey, T. (2000). Predicting Similarity Ratings to Faces using Physical Descriptions. In M. Wenger, & J. Townsend (Eds.), *Computational, geometric, and process perspectives on facial cognition: Contexts and challenges*. Lawrence Erlbaum Associates.

Predicting Similarity Ratings to Faces using Physical Descriptions

Mark Steyvers & Tom Busey

Indiana University

Email: **msteyver@psych.stanford.edu**

Abstract

A perceptually-grounded, nonmetric feature mapping model is introduced. This model explicitly relates similarity ratings from a facial comparison experiment to various primitive, physically derived facial features such as Gabor jets, principal components and geometric features. In this approach, abstract features are formed that combine and weight information from the primitive facial features. The abstract features form the basis for predicting the similarity ratings for faces. We show how this model extracts abstract "age" and "facial adiposity" features on the basis of all similarity ratings to 50 faces. Whereas traditional multidimensional scaling methods can also uncover important variables for face perception, this model has the additional advantage of making explicit how to compute these variables from primitive facial features. Another advantage of this approach is that the featural descriptions can be used in a generalization test to predict similarity ratings to new faces. We show how this generalization test enables us to constrain various parameters of the model such as the dimensionality of the representation.

Introduction

A major goal in the field of face perception is to determine appropriate representations and processes operating on these representations. Faces are enormously, and perhaps infinitely complex (Townsend, Solomon, & Spencer-Smith, this volume). By the same token, they all share a recognizable shape and configuration: for example, the nose is always between the mouth and the eyes. Although faces consist of a high number of dimensions, the representation of faces may be thought of as a compression or mapping of the featural dimensions into a lower-dimensional space by either ignoring some dimensions or reducing the redundancies among dimensions. Face perception may be thought of as a process by which the physical features of faces are combined in order to support recognition or categorization tasks. To capture the representations that are used in face perception, researchers have adopted one of two major approaches.

The purely psychological and top-down approach

In the purely psychological approach based on multidimensional representations (e.g. Ashby, 1992; Nosofsky 1986, 1991, 1992), a face is represented abstractly as a point in a multidimensional space (Valentine, 1991a, b; this volume). The positions of the points can be derived from data from various psychological tasks with Scaling techniques such as Multidimensional Scaling (MDS) (Kruskal 1964a,b; Shepard 1962a,b, 1974, 1980; Torgeson 1952). In nonmetric MDS, the goal is to find a configuration of points in some multidimensional space such that the interpoint distances are monotonically related to the experimentally obtained dissimilarities. The dissimilarities can be derived from similarity judgements, dissimilarity judgments, confusion matrices, reaction times from discrimination

experiments, correlation coefficients or any other measure of pairwise proximity. In metric MDS, the goal is to find a configuration of points and an appropriate function that transforms the interpoint distances such that the transformed distances match the experimental dissimilarities exactly. In the Appendix, we give a short introduction to nonmetric MDS.

Several researchers using MDS analyses on faces (Busey, this volume, 1998; Johnston, Milne & Williams, 1997; Shepard, Ellis, & Davies, 1977) have developed multidimensional "face-space" representations: the faces are located in a multidimensional space such that similar faces are located in similar regions and that the pairwise distances between the face-locations reflect their perceived similarity. Busey (this volume; 1998) has applied MDS to similarity ratings on all pairs of a set of 100 faces. Based on a six dimensional configuration, the dimensions were interpreted as age, race, facial adiposity, facial hair, aspect ratio of head and color of facial hair. The goal of Busey's work was to predict recognition performance with various computational models that took the configuration of points as a basis for representing the faces.

The resulting MDS solutions for the configuration of points in low-dimensional spaces can give valuable insights about the way faces are perceived, and sometimes forms a useful basis for modeling performance in recognition and/or categorization tasks. Although the resultant dimensions are sometimes given a featural interpretation, this approach explicitly ignores the physical representation of the features comprising the faces. In this purely top-down approach, the multidimensional representations are sometimes difficult to relate back to the physical stimulus.

The purely computational and bottom-up approach

In the purely computational and bottom-up approach (e.g. Hancock, Bruce, & Burton, 1998; O'Toole, Abdi, Deffenbacher & Valentin, 1993; Wiskott, Fellous, Kruger, von der Malsburg, 1997; Yuille, 1991), a face is represented by a collection of features that are explicitly derived from a 2D image that is analogous to the retinal image of the face. For example, a face can be described by the distance between the eyes, the color and texture of the skin or by other features that can be extracted by computational methods.

One method is principal component analysis (e.g. O'Toole et al., 1993; Turk & Pentland, 1991) where the face images are projected onto the eigenvectors (principal components) that capture the significant global variations in 2D image intensities. In another method, face images are processed by overlapping receptive fields (Edelman & O'Toole, this volume; Lando & Edelman, 1995) or Gabor jets (e.g. Wiskott et al., 1996). The responses of these receptive fields are somewhat insensitive to changes in viewing conditions, and retain the local structure of image intensities. In a somewhat older method, faces are encoded with geometric codes such as the distance between the eyes, nose length and lower face width (Laughery, Rhodes & Batten, 1981; Rhodes, 1988). Typically, these codes are derived manually, but there exist several methods to automatically locate feature landmark points (e.g. Lades, Vorbruggen, Buhmann, Lange, von der Malsburg, Wurtz & Konen, 1993; Lanitis, Taylor, & Cootes, 1995; McKenna, Gong, Wurtz, Tanner & Bannin, 1997; Wiskott et al. 1996; Yuille, 1991) that can provide a basis for these codes. In these geometric codes, subtle information about local skin texture is lost, so that by themselves these codes are probably not rich enough to

distinguish between subtle variations that exist in the population of faces

While many of these proposed featural representations for faces provide very rich sources of information and form the basis for many computer face recognition systems, it is not always obvious which features or combinations of features are useful to model human face perception. We define these approaches to be purely computational and bottom-up because the representational spaces are fixed and are not changed to in order to minimize the difference between the simulated performance and observed performance on some face perception task.

Integrating the top-down and bottom-up approaches

To summarize: in a purely psychological and top-down approach, a face is represented as a point in an abstract psychological space where the dimensions are interpreted so that they are related to the physical appearance of the face. In a purely computational and bottom-up approach, a face is represented as a collection of explicitly derived physical features. The goal of this research is to integrate the bottom-up and top-down face encoding approaches into a single framework that links physical features to an underlying psychological space. We refer to two different kinds of spaces. The first, the concrete feature space consists of the collection of primitive physical features for faces (e.g. distance between eyes, texture of skin). The second, the abstract feature space refers to the psychological space that consists of variables (e.g. age, facial adiposity) that are important for modeling performance on psychological tasks. The abstract feature formation is flexible and depends on what perceptual information can be computed from the concrete features and the data that needs to be explained. The process by which the abstract features are derived from the concrete features is made explicit and is constrained by data from a similarity rating task. We call this the feature mapping approach because the goal is to find a mapping between the concrete features and abstracted features. This approach can tell us what features are most important for predicting psychological similarity.

The Rumelhart and Todd (1992) feature mapping model

This feature mapping model is based on work by Rumelhart and Todd (1992) and Todd and Rumelhart (1992). They proposed a model that is fully connectionist. The essential assumption of this model is that the mapping from the concrete feature space to the psychological space can be learned from an analysis of similarity ratings. In their model, the concrete features feed through a single layer network to a new set of nodes. These nodes contain abstracted featural information and are analogous to the dimensions of a MDS solution. The two objects in a similarity rating task are represented separately by two different sets of abstract feature units. The abstracted features of two objects are then compared by feeding through several additional connectionist layers. These additional layers implement a transformation on the distances between the corresponding abstract feature units to a predicted similarity rating. The differences between the predicted and observed similarity ratings are then used for a backpropagation algorithm to optimize the weights between the concrete feature units and the abstract feature units and the weights in the transformation layers. The Rumelhart and Todd model is a metric version of multidimensional scaling: the predicted and observed similarity ratings should have identical values. The nonmetric feature mapping model proposed in this chapter is a nonmetric extension of the Rumelhart and Todd

model; only the rank order of the predicted and observed similarity ratings is important. Any transformation on the observed data that preserves the rank order will lead to the same results. We will now discuss the relative merits of metric and nonmetric scaling methods.

Nonmetric vs. metric scaling methods

In nonmetric scaling methods, the goal is to reproduce the monotonic relationships in the proximity matrix obtained from a psychological task. In metric multidimensional scaling, one needs psychological estimates of the metric distances between stimuli. This involves an extra stage of computation in which the interpoint distances are transformed into (for example) expected similarity judgements, same-different judgements or reaction times.

When the experiment is designed such that participants only perform ordinal comparisons between pairs of stimuli (e.g. which of the two pairs of faces is more similar?), then a nonmetric method might be the preferred method to analyze the data. From a theoretical viewpoint, one might prefer the metric method over the nonmetric method since the metric method is more constrained and gives more falsifiable models of the data. From a practical viewpoint, one might prefer the nonmetric method over the metric method. In a metric method, in addition to estimating stimulus coordinates (or weights between the concrete and abstract features in the Rumelhart and Todd model), extra parameters need to be estimated for the transformation stage. This means that the optimization problem for finding good solutions with a metric method is more complex. When a bad solution is obtained with a metric method, it is could be because a bad assumption is made in the transformation stage or because the optimization algorithm suffers from the problem of local minima. Therefore, it is possible that for a given proximity matrix, a nonmetric method results in a reasonable solution whereas a metric method cannot find any reasonable solution. In our research, we chose the nonmetric method to simplify the optimization problem so that good solutions would be more likely than with a metric method.

The Nonmetric Feature Mapping Model for faces

In the feature mapping model, the features comprising each face can be thought of as points in a multidimensional feature space. By feature mapping, the points of the concrete feature space map to points in a lower-dimensional abstract feature space. The exact nature of this mapping is determined by a set of weights. With certain weights, it is possible that the redundancy in the concrete feature set is removed and that useful regularities are retained. Based on a distance function of the differences in this lower dimensional space, the model produces a predicted (dis)similarity rating to the two stimuli that can be compared to the actual (dis)similarity rating. The difference between the predicted and actual similarity ratings can then be used to optimize the weights that determine the nature of the feature mapping. Once the mapping parameters are optimized, the faces have fixed coordinates in the feature abstraction space. We will now summarize the advantages of this approach over the psychological and computational approaches to representations for faces.

Advantages of the feature mapping approach

In MDS, the location of a face is determined by a set of coordinates, or parameters, that are estimated by methods described in the Appendix. When new faces are

introduced, MDS must estimate a new set of parameters in order to determine the face locations. It is therefore not clear how MDS can predict similarity ratings to new faces without introducing new parameters. The first advantage of our feature mapping approach is the possibility of testing its generalization performance without introducing new parameters or changing the parameters. Once all parameters are optimized with respect to some set of stimuli, it is possible to predict the similarity ratings to stimuli that have not been presented before to the model using the same parameter settings. The two sets of features describing a pair of new stimuli are first mapped to points in the abstract feature space. The predicted similarity rating is then some distance function of the points in the abstract feature space. The possibility of assessing the generalization performance is of major importance because it provides a strong test of the feature mapping approach. This technique grounds the representation in the physical stimulus and therefore can make a priori predictions[1].

The dimensions resulting from MDS are constrained by the proximity data obtained from participants. The proximity data in turn is constrained by the processes underlying face perception. In the feature mapping method, the abstracted features (dimensions in MDS) that are formed are influenced by two sources of information. The first source of information is the proximity data from participants which depends on the perceptual processes underlying face perception. The second source of information is provided by the concrete features that can be extracted from images of faces by computational means. Both sources of information will constrain the development of abstracted features to those features that can be specified by computational means and that can predict the proximity data. Therefore, a feature mapping solution might predict the proximity data worse than a MDS solution (given the same number of dimensions) when the chosen set of concrete features does not explain all the variability in the data. However, the dimensions that are developed are computationally specified whereas in the MDS solutions, it is not a a priori guaranteed that the resulting dimensions can be computationally tied to the perceptual information available in face images.

The model

In Figure 1, a schematic overview of the nonmetric feature mapping model is shown. The model takes as input the featural descriptions to a pair of faces. Geometric distances, principal component coefficients and/or Gabor jets were used as featural descriptions; details about these featural descriptions are given in a later section. With the features of a face as input, the model first extracts the relevant features of these faces by mapping from the large concrete feature space to the small abstract feature space. This is done separately for each face of a pair in a similarity rating experiment. This part of the model is connectionist: the input features activations are fed through a fully connected single layer connectionist network with sigmoidal output units. There are many fewer output nodes than input features so the network will typically abstract from the featural information. We will refer to these output units as the abstract feature units. Each abstract feature unit is a sigmoid function of a weighted linear combination of the input features. The matrix W contains all the weights for each input unit to each abstract feature unit. The weight matrix W contains all the parameters of this model.

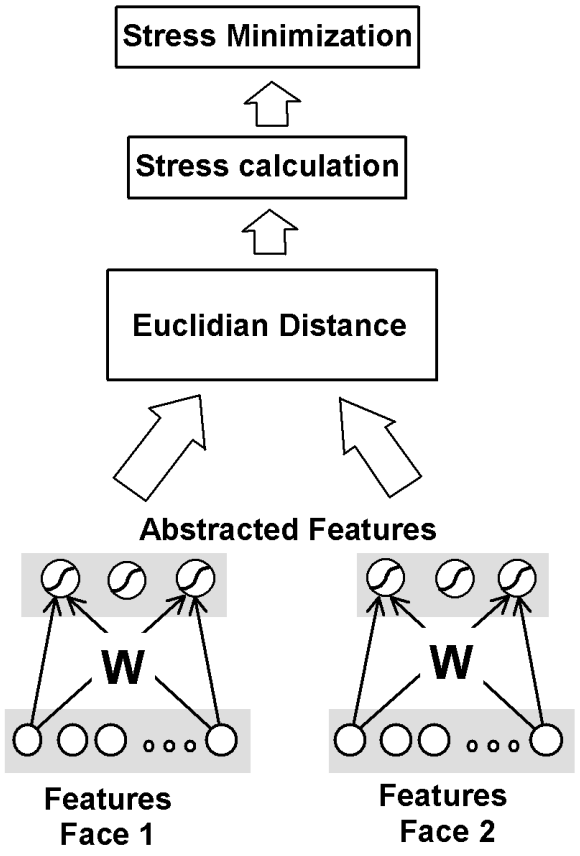


Figure 1. Schematic overview of the nonmetric feature mapping model.

Based on the activations of the abstract feature units, the model can make a prediction about the dissimilarity between two faces. This part of the model is based on nonmetric multidimensional scaling techniques. The distance between the two abstract feature vectors is calculated by a Euclidian distance metric: the bigger the distance between the two vectors, the more dissimilar the two faces are predicted to be. The goal of the model is to have a monotonic correspondence between the Euclidian distances and the observed dissimilarities obtained from the similarity rating task.

This goal is achieved by an optimization algorithm, that operates in two alternating phases (see Appendix). We use stress to compute how much the relationship between the Euclidian distances and observed dissimilarities deviates from monotonicity. The stress is computed in the first optimization stage, in which a monotonic regression (Kruskal, 1964a,b) is performed on the Euclidian distances and observed dissimilarities. In the monotonic regression analysis, target Euclidian distances are computed that lead to a perfect monotonic relationship with the observed dissimilarities and minimize the stress for the given actual Euclidian distances. Now, stress can be calculated and there are also target distances available that can be used in the second optimization phase. In this second phase, the weights in matrix W can be adjusted such that the newly calculated Euclidian distances in abstract feature space correspond more closely to the target distances. We use standard optimization techniques to adjust the weight matrix W to optimize stress (not backpropagation). This cycle of monotonic regression and weight adjustment is repeated until the stress cannot be further improved.

The feature mapping approach is similar to MDS in many respects. In MDS, the stimuli are represented as coordinates in a multidimensional space. In our model, the stimuli can be represented as points in the abstract feature space. As in nonmetric MDS, we use nonmetric methods to evaluate how well the predicted distances correspond to the observed dissimilarities. There are also crucial differences between our approach and MDS. Whereas in MDS, the stress is used to optimize the stimulus coordinates in the multidimensional space, we optimize the mapping from the concrete feature space to the abstract feature space. In MDS, predicting dissimilarities for new pairs of stimuli is impossible without introducing new parameters; the new stimuli are new points in the multidimensional space and therefore new parameters to be estimated. In other words, MDS cannot give parameter-free

predictions for judgements of similarity between new pairs of stimuli. In our model, parameter-free predictions can be made in a straightforward manner. First, the feature mapping matrix W is optimized for the featural descriptions and participants' similarity ratings for one set of faces. Then, holding this feature mapping matrix constant, we can map the concrete features of the new faces to the abstract feature space. Second, we calculate the distances between the new faces in this abstract feature space and compare these to the observed dissimilarities. Another way of describing the difference between MDS and the feature mapping approach is that in the latter approach, the number of parameters does not scale with the number of stimuli under consideration. Since the number of parameters only scales with the number of concrete and abstract features, it is possible that a solution can be achieved with much fewer parameters than with MDS.

In MDS, there is the problem of deciding on the appropriate number of dimensions (Shepard, 1974). Similarly, in our model we face the problem of deciding on the appropriate number of abstract feature nodes. Too many or too few nodes might lead to results that do not generalize well. With the feature mapping approach, a solution to the dimensionality problem is possible. The number of dimensions is determined by the model which generalizes best to similarity ratings of novel pairs of faces. We expect that this generalization test provides strong constraints on the model.

The similarity rating data for the model

All the simulations in this research were based on data from a similarity rating experiment performed by Arici and Busey (in progress). In their experiment, 238 participants rated the similarity of pairs of faces that were presented simultaneously on a computer screen. The faces came from a database of 60 faces. All faces were male, bald faces displayed frontally with similar lighting conditions (see Figure 8 to get an impression of the set of faces). Of a total of 1770 possible pairs of similarity ratings, each of the participants gave ratings to a subset of 177 pairs (this works out to give about 25 ratings for each pair). The individual subject data was first transformed to z-scores by subtracting the subject's mean rating and dividing by the standard deviation. The transformed scores were then averaged over participants.

Table 1

Results of Applying Nonmetric MDS to Similarity Rating Data Set

Dimensionality	S	R_s
1	0.353	0.750
2	0.219	0.857
3	0.159	0.899
4	0.122	0.925
5	0.101	0.939
6	0.086	0.950
7	0.077	0.957
8	0.069	0.961

Note. S=Stress; Rs =Spearman rank order correlation

We applied nonmetric MDS analyses on the proximity matrix so that the results can be compared to the results with the feature mapping model. In Table 1, the stress values are reported as a function of dimensionality. As expected, the stress measure decreases for an increasing number of dimensions. Also reported is the Spearman rank order correlation coefficients (R_s) between the Euclidian distances and the observed dissimilarities. We included this measure here and in the results section so that additional information was available to evaluate the performance of the feature mapping model relative to nonmetric MDS.

Physical features for faces

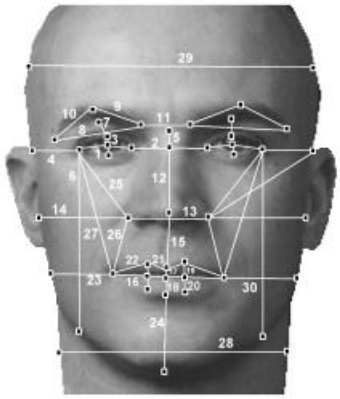
The input to the feature mapping model is a description of the two faces by a set of features. These features were provided by the methods that are outlined below. We simulated the feature model with sets of features from each of these methods separately and in certain combinations. All feature values were standardized to z-scores by subtracting the mean feature value (over all training set faces) and dividing by the standard deviation. The images of the faces were identical to the images used in the similarity rating experiment except for a correction that we performed such that for all images, the center of the bridge of the nose fell on the same image location.

Geometric information

Each face was described by a set of 30 distances as shown in Figure 2. We have used almost the same geometric distances between landmark points on the face as Rhodes (1988). For distances that could be measured in the left or right half of the face (such as the eye-width), we took the average of the distances. In this work, we derived the landmark points manually[2].

Principal component analysis

Principal component analysis (PCA) can be performed on the gray level intensities of the pixels for all the faces under consideration (O'Toole et al. '93; Turk & Pentland, 1991). In this analysis, each face is described as a vector containing the light intensities of all the pixels. Each principal component is then an eigenvector of the covariance matrix of the face vectors. The principal components are ordered by how much of the variance of the covariance matrix is explained. The first/last principal components capture the most/least of the variance of the covariance matrix. Each face can be described as a set of coefficients that expresses a linear combination of principal components. It is possible to reconstruct the face image using only a few coefficients corresponding to the first principal components. In our research with 60 faces, the analysis yields 60 principal components of which we display the first 8 in Figure 3. We performed simulations with using the first 10, 20 or 40 coefficients as featural descriptions.



1	eye width	16	mouth width
2	eye separation	17	top lip thickness
3	eye height	18	bottom lip thickness
4	eye face position	19	top lip thickness at bow
5	eye height inner	20	bottom lip thickness at bow
6	eye height outer	21	center to bow
7	eye eyebrow distance	22	bow to edge
8	eye brow width	23	horizontal mouth - face position
9	eye brow width inner part	24	vertical mouth - face position
10	eye brow width outer part	25	eye nose relation
11	eye brow separation	26	nose mouth relation
12	nose length	27	mouth eye relation
13	nose width	28	neck width
14	horizontal nose - face position	29	upper face length
15	vertical nose - face position	30	lower face width

Figure 2. The 30 geometric distances based on Rhodes (1988).

Figure 3. The first 10 principal components for the 60 faces in the simulations.

Gabor jets, evenly spaced

Gabor filters can perform a local oriented spatial frequency analysis (Daugman 1980, 1985). The filters are operating on

overlapping receptive fields as shown in Figure 4. The origin of each receptive field is positioned in a regular grid. We have used 14x14 receptive fields as shown in the figure. On each receptive field, a Gabor Jet is operating that consists of 12 Gabor functions; these functions were factorially varied over 4 orientations (0, 45, 90, and 135 degrees) and 3 spatial frequencies. Each Gabor function measures the similarity between the intensity profile of the local image patch and the oriented sinusoidal plane associated with the Gabor function. We have used phase shifted pairs of filters to remove (some of) the sensitivity to spatial shifts of image texture. Each Gabor Jet outputs a vector with 12 entries. For the purposes of this model, we averaged the Gabor Jets over the different locations. This results in a loss of local information but retains the overall orientation and spatial frequency information in the image. In this way, the number of features is constrained to 12. In the Discussion, we mention methods to retain the spatial information of the local spatial frequency analysis while at the same time constrain the potentially large number of parameters of the feature mapping model that is given such a representation.

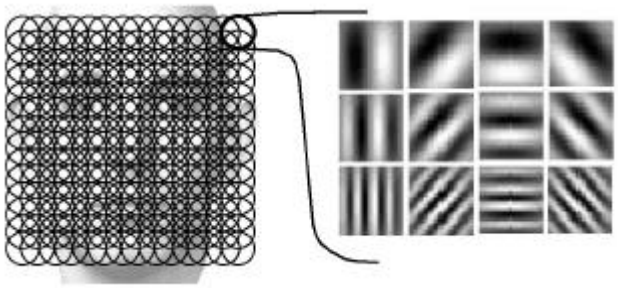


Figure 4. Illustration of the 14x14 receptive fields placed over a face with 2D Gabor functions. The Gabor functions were factorially varied over 4 orientations (0, 45, 90 and 135 degrees) and 3 spatial frequencies. The 12 Gabor functions operating on one receptive field are referred to as Gabor jets.

Gabor jets, matched locations

In this representation, we apply the same Gabor Jet analysis as outlined in the previous section but the origins of Gabor Jets are now positioned at 48 feature landmark positions on the face as shown in Figure 5. We average over location so that this also results in a description of the face in terms of 12 features. The positioning of the Gabor Jets at the feature landmark points ensures that the Gabor Jets for any pair of faces are aligned so that they analyze corresponding local regions. The feature landmark points correspond to the landmark points that were used to measure the geometric distances. This provides a rationale for combining the geometric distance information with the Gabor Jet information. In this research, the placement of the landmark points was done manually.

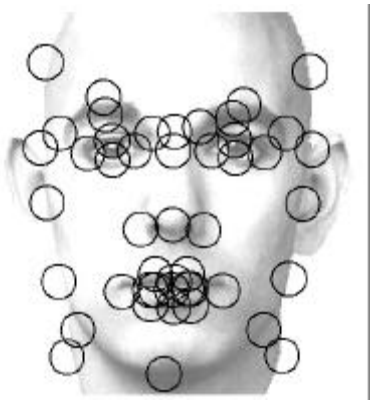


Figure 5. The Gabor Jets placed over feature landmark points of the face. These points coincide with the points that were used to calculate the geometric distances (Figure 2).

Simulation methods

In our simulations we used the similarity rating data set for 60 faces. The averaged similarity data (over all 238 participants) for 50 randomly chosen faces was used optimize the feature mapping matrix W. We will refer to this set as the training set. The number of similarity ratings in the training set is 1225. The averaged

Table 2

Results of Applying the Nonmetric Feature Mapping Model for Training and Test Sets

	No Geometric Information						Geometric information combined with 2D intensity model					
			Training Set		Test Set				Training Set		Test Set	
2D Intensity Model	*N*	*K*	R_s	*S*	R_s	*S*	*N*	*K*	R_s	*S*	R_s	*S*
None							30	4	.805	.177	.775	.153
PCA (10 components)	10	4	.476	.248	.451	.272	40	2	.791	.250	.752	.242
PCA (20 components)	20	4	.667	.210	.647	.217	50	4	.900	.130	.780	.140
PCA (40 components)	40	4	.889	.141	.780	.193	70	2	.843	.224	.734	.243
Gabor Jets, Evenly Spaced	12	2	.501	.291	.146	.310	42	3	.851	.186	.857	.153
Gabor Jets, Matched Locations	12	4	.869	.171	.598	.261	42	4	.888	.144	.791	.182

Note. N=number of concrete features; K=number of feature abstraction units; Rs=Spearman rank order correlation; S=Stress

data (of the same 238 participants) for the remaining 10 faces was used as the test set to measure the generalization performance. The test set consisted of 45 similarity ratings. We did not use the part of the dataset that consisted of similarity ratings to pairs of training and test set faces (500 ratings): the generalization performance is measured in terms of its ability to predict similarity ratings to pairs of new faces.

Two measures of performance were used: the stress and the Spearman rank order correlation coefficient for the Euclidian distances and observed dissimilarities. We ran simulations using the features from each feature description method separately and we also simulated some combinations of methods. For each set of features, we ran different simulations in which we varied the number of abstract feature nodes between 1 and 7.

Results

The results of the simulations are summarized in Table 2. We show the results for the geometric distance method and the PCA and Gabor jet methods by themselves and in combination with the geometric distances. N is the number of features used in the description of the faces. The measures of fit were the Spearman's rank order correlation coefficient R_s and Kruskal stress; these are shown for both training and test sets. In the table, we show the number of feature abstraction units, K, that resulted in the best generalization to proximity data of new faces as measured by R_s (we varied K between 1 and 7). Two general patterns of results can be observed. First, each feature method by itself gives comparable results for both training and test sets: the geometric distances, PCA (40 components) and Gabor jet method (matched locations) by themselves result in similar rank order correlation and stress values. Second, the Gabor jet feature sets

(evenly spaced and matched locations) lead to much better results when they are augmented by the geometric distances.

In Figure 6, we show the effect of varying the number of feature abstraction units on the stress and rank order correlation for two feature methods: (a) the evenly spaced Gabor jet method with geometric distances, and (b) the PCA (10 coefficients) method with geometric distances. It can observed that overall, the stress decreases for both training and test set when the number of feature abstraction units increases. The rank order correlation curve for the test set "bends over" at some number of feature abstraction units. Applying a generalization-test criterion, this is the number of feature abstraction units that we choose because this leads to the best generalization to new faces.

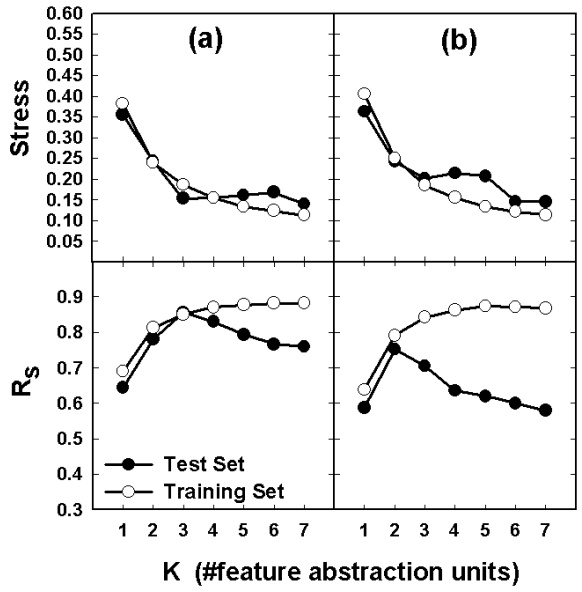


Figure 6. The effect of varying the number of feature abstraction units on Kruskal stress and Spearman rank order correlation R_s for two feature description methods: (a) the evenly spaced Gabor jet model plus geometric distances and (b) the PCA (10 coefficients) method plus geometric distances.

In Figure 7, the Shepard diagram is shown for one specific simulation result: the evenly spaced Gabor jets method with 3 feature abstraction units. In the diagram, the Euclidian distances vs. observed dissimilarities are plotted for both training and test sets. Also plotted is the monotonic regression line for the training set. It can be observed that a reasonable monotonic correspondence exists between the Euclidian distances and observed dissimilarities. The feature mapping model with this feature description method produces rankings for the dissimilarity of pairs of old and new faces that are similar to the average rankings that participants give.

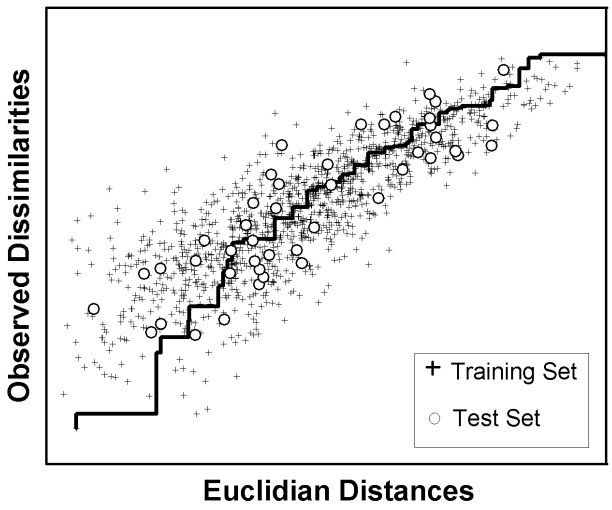


Figure 7. The results of the feature mapping model using the Gabor jets, evenly spaced with three feature abstraction units. In this Shepard diagram, the observed dissimilarities are plotted against the Euclidian distances of the model. The results are shown for the dissimilarities for pairs of faces from the training set and the test set. The results of a monotonic regression on the training set dissimilarities is indicated by the solid line.

Both old and new faces can be represented as points in the multidimensional feature abstraction space. We can plot the faces in this space to visually inspect the variation of faces along each feature abstraction dimension. We can also visually inspect if the test set faces are mapped to locations where there are similar training set faces. This is illustrated in Figure 8a for the previously mentioned simulation result. The faces are shown in the 2D space determined by the activations of two of the total three feature abstraction units. The test faces

are indicated by a surrounding square box. From visual inspection, it seems reasonable to interpret the dimensions corresponding to feature abstraction units #1 and #2 as "age" and "facial adiposity" respectively: from left to right, the faces get older and from bottom to top, the faces get wider and pudgier. From visual inspection, the placement of the new faces makes intuitive sense: the older, pudgier faces are placed in the upper right corner of the space and the younger, skinnier faces are placed in the lower left corner. The third unit (not shown in the figure) did not lend itself to such an interpretation. The third unit might explain part of the variance of both training and test set by clustering similar faces in similar regions so that the faces are locally but not globally ordered along this third dimension. If the number of feature abstraction units is restricted to two, then with this feature method, age and facial adiposity are again our interpretations of the two units. We report the case of three units here because that case leads to the best generalization to new faces. In Figure 8b, the same plot is shown for the simulation results with the PCA method with 10 components combined with geometric distances using two feature abstraction units. The two feature abstraction units seem to be picking up on similar information as in the previous simulation result: the first and second unit can be interpreted as "age" and "facial adiposity" respectively.

To test our subjective interpretations of the representational role of the feature abstraction units, we correlated the activity of the feature abstraction units with participants' age and facial adiposity ratings. We obtained age and facial adiposity ratings for all 60 faces from a group of 12 participants that was not involved in the similarity rating experiment. Each feature abstraction unit for the two previously mentioned simulation results was correlated with the averaged age and adiposity ratings. The results are summarized in Table 3. For the purpose of comparison, the table also shows the correlations for the dimensions of a three dimensional nonmetric MDS solution for all 60 faces.

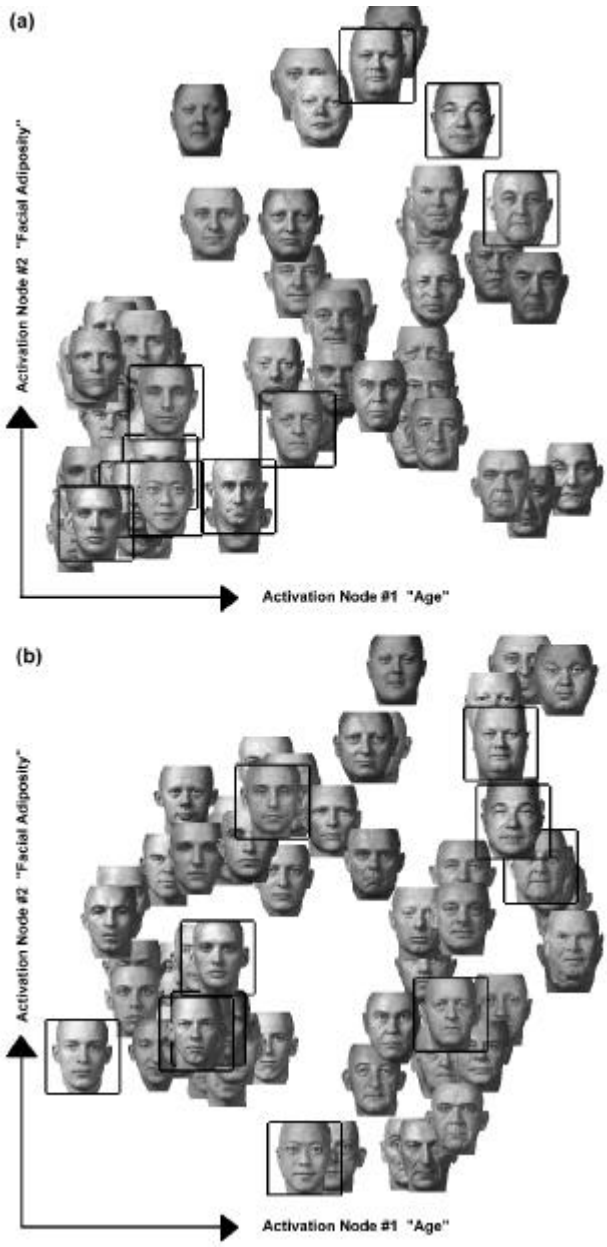


Figure 8. The training and test set faces (marked by a box) plotted in the space determined by the activations of the first two abstract feature units for (a) the evenly spaced Gabor jet method with geometric distances using three abstraction units (only the first two are shown here) and (b) the PCA method with 10 components with geometric distances and two feature abstraction nodes.

In the top part of the table indicated by "Unrotated Dimensions," the feature-mapping and MDS representations are left unchanged. Our subjective interpretations of the

ıble 3

ɔrrelations of the Unrotated and Rotated Solutions of Two Feature Mapping Models and a MDS lution with age and facial adiposity ratings

	Gabor jets, evenly spaced and geometric distances		PCA (10 components) and geometric distances		MDS	
Unrotated Dimensions	Age	Adiposity	Age	Adiposity	Age	Adiposity
1	.791	.325	.866	.783	.434	-.364
2	.659	.901	-.020	.679	.810	.903
3	-.005	-.552			.062	.020
Rotated Dimensions						
1	.889	.444	.909	.435	.909	.435
2	.487	.943	.344	.889	.465	.947
3	.039	.065			.280	.324

representational role of the two feature abstraction dimensions are confirmed: the first unit correlates highly with participants' age ratings and to a lesser degree with participants' adiposity ratings. For the second unit, the opposite is true: it correlates highly with adiposity and to a lesser degree with age. In the bottom part, the representations are rotated such that each of the units/dimensions correlates maximally with one of the observed rating (age or adiposity) and minimally with the other. This method leaves all pairwise distances unchanged[3]. The rotated solutions show similar correlations with participants' ratings for both the feature mapping representations and the MDS analysis.

One of the advantages of the feature mapping approach is the possibility of inspecting the feature mapping; this can give insight into how the abstract features are "built" from the individual concrete features. Each feature abstraction unit is a sigmoidal function of a weighted combination of input features. In Figure 9, the weights are shown for the previously mentioned Gabor jet simulation (left) and PCA simulation (right). In these bubble plots, positive/negative weights are shown as filled/unfilled circles respectively. The size of the circles reflects the absolute magnitude of the weights. Each column displays the weights to one of the feature abstraction units. In Figure 8 and Table 3, it was established that it is reasonable to interpret the first and second unit as representing age and adiposity respectively.

When we inspect the weights for the geometric distances in the left and right plots, there are various patterns to be discerned. For example, the "age" unit decreases its activity for larger eye-widths (distance #1), larger distances between the nose and the side of the face (distance #14) and larger distances between the nose and the chin (distance #15). Its activity is increased for larger distances between the tip of the nose and corner of the mouth (distance #26) and for larger lower face widths (distance #30). These dependencies make some sense when one considers the changes in the facial structure for aging faces. The "adiposity" unit's activation is increased by larger distances between the mouth and the side of the face (distance #23) and larger

lower face widths (distance #30). Overall, the weights to the Gabor jets in the simulation result shown on the left are more difficult to understand. However, the large positive weight from the "age" unit to the Gabor function that picks up diagonally oriented features (from the lower left to upper right in a local patch of the image) does make sense: the faces were lit mostly from the right side of the face such that the cheek wrinkle was more visible in the left side of the face where it runs diagonally from the left side of the mouth to the left side of the nose.

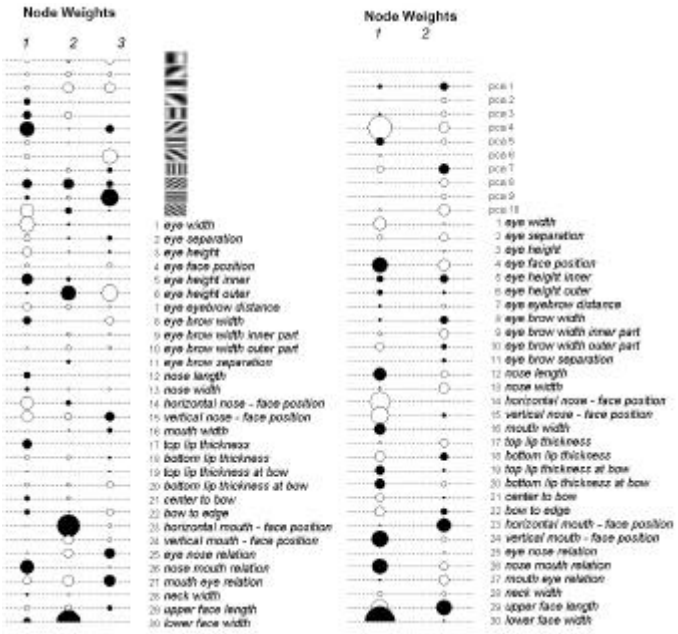


Figure 9. A pictorial presentation of the feature mapping developed in two simulations. On the left, the weights are shown for the evenly spaced Gabor jets with geometric distances simulation using three feature abstraction units. On the right, the weights are shown for the PCA method (10 components) with geometric distances using two feature abstraction units. Positive/negative weights are shown as filled/unfilled circles respectively. The size of the circles reflects the absolute magnitude of the weights.

Obviously, this is not a good way to pick up on the age aspect of a face in general. The nature of the feature mapping is dependent on the perceptual processes giving rise to the proximity data and the information contained in the set of features provided by computational methods. Not surprisingly, this analysis suggests that the amount of wrinkling is a feature that is used by participants when making proximity judgements. More dedicated computational methods can analyze the amount of wrinkling in a way that is invariant over lighting conditions, facial pose etc.

Discussion

We have shown that the feature mapping approach can predict the similarity ratings to both old and new faces with a reasonable degree of accuracy. Applying the PCA and Gabor jet method combined with geometrical codes has led to similar results. For both methods, age and facial adiposity were the two abstract featural dimensions that were extracted. When inspecting the feature mapping that was developed with these methods, we were able to get a better understanding of what combinations of features are successful in predicting the structure of the similarity data. Interestingly, the feature mapping method already resulted in good generalization performance when used with geometric distances only. Adding the textural information provided by Gabor jets or principal components led in some cases to better and in some cases to worse generalization performance. These results suggest that the geometric distances by themselves contain important information to explain the similarity ratings between pairs of faces for the set of faces we used. It is possible that with a more heterogeneous set of faces, the feature mapping model would show a larger benefit for adding textural informa tion. When the feature mapping method was provided with a combination of geometric distances and Gabor jets, the resulting abstract feature units "fused" the information from both sources of information to create the "age" and "facial adiposity" units. This coincides with experimental results by Mark, Pittenger, Hines, Carello, Shaw, and Todd (1980) that show that participants use both

shape information (which could be directly based on geometric distances) and texture information (based on the Gabor jets for example) in a complex interrelated way to determine the age of faces.

One weakness of the simulations reported here, is that we averaged the outputs of the Gabor filters over different locations so that spatial information is lost. It is possible that a subtle texture variation at the eyes is important for making similarity judgements, but that the feature mapping model could not pick up on this information since it was only provided with averaged textural information. We could have opted for a separate weight for every different filter at every position but this would have led to an excessive number of parameters. We see two solutions to this problem of reducing the input dimensionality while retaining spatial information of faces. First, PCA can be applied to the filter outputs at different locations (e.g. Dailey & Cottrell, 1998). In this way, a large input dimensionality can be reduced to a much smaller one if there is redundancy between the filter outputs. A second solution is to keep a separate weight for every filter at every location but to impose a spatial smoothness constraint on their outputs. For example, nearby filters are likely to capture the same kind of information so that the smoothness constraint will force the weights to have similar values. Regularization theory (Tikhonov & Arsenin, 1977) or hyper basis functions (Poggio & Girosi, 1990) can be applied to impose smoothness constrains so that the effective number of parameters is much smaller than the number of filter weights. In future research, we hope to explore both of these techniques.

To summarize, we believe the inclusion of physical descriptions of visual objects is important in an analysis of behavioral data for these objects. By mapping from the physical descriptions through a series of stages to the behavioral data, it is possible to apply this mapping to novel objects and measure the generalization performance. The generalization test is a strong test to select among various candidate models and can put strong constraints on those models.

Related Research

Rhodes (1988) regressed a large set of physical measurements, ratios of measurements and participants' ratings about the appearance of 41 faces against the dimensions of a multidimensional scaling solution for these faces. Two MDS dimensions correlated well with the ratings for age and weight ratings respectively. Many physical measurements relating to the appearance of the eyes and nose were also found to correlate with the MDS dimensions. Her research shares important similarities with ours. We also find that age and weight (facial adiposity) to be important variables in participants' similarity judgements. Also, the regression method can be viewed as an alternative to the feature mapping method. First, a MDS solution is obtained for a set of faces. Then, as in Rhodes' work, a multiple regression analysis is performed on each dimension so that each dimension is a linear combination of physical measurements. Then, holding the regression parameters constant, similarities for novel faces can be computed by using the measurements for these novel faces and predicting the coordinates for these novel faces. Therefore, as in our research, this method is grounded in the physical representation of a face and allows generalization to novel faces. There are two important differences between this regression and the feature mapping method. First, the feature mapping method is an integrated approach that was especially designed to linking physical measurements with proximity data whereas in the regression method, purely psychological scaling techniques are combined with multiple regression techniques. Second, as stated before, the dimensions resulting from MDS are only

constrained by the proximity data and not in any way by what perceptual information can be compututionally extracted from the face images. This can result in important differences between the dimensions developed by MDS and by the feature mapping method. In feature mapping, the dimensions developed depend on how predictive the information contained in the concrete features is for the structure in the proximity data. So, if the shape of the eyes is a major factor in the similarity judgements, it is only possible for the feature mapping model to discover this when there are primitive features that can capture the shape of the eye. Therefore, it is possible that the feature mapping model does not develop certain dimensions. On the other hand, for the dimensions that are developed, the model has developed an explicit computational procedure to compute the dimensions.

Hancock et al. (1998) have reported results of predicting the similarity ratings to faces on the basis of various principal component decompositions and Gabor jet models. In one method, as a measure of proximity, they calculated the Euclidian distance between the principal component coefficients of pairs of faces. In another method, they calculated an alignment penalty for two faces based on the graph-matching system by Wiskott et al. (1997). In this system, a face is analyzed by Gabor jets that are placed on the vertices of a graph. When presented with two faces, the goal of the system is to align the vertices of the graphs representing each face such that the differences in Gabor jet outputs at corresponding vertices and the differences between corresponding distances is minimized: the alignment penalty expresses the degree of mismatch between two faces by weighting both kinds of distances. Based on these two methods, Hancock et al. reported very low rank order correlations between the Euclidian distances/ alignment penalties and human proximity data. One possible reason why they obtained poor fits is that they did not find the relevant combinations of features to predict the proximity data: all features were weighted equally in the comparison between two faces. In contrast, in the feature mapping approach, features are weighted according to well they predict similarity ratings. Biederman and Kalocsai (1997) also tested the system of Wiskott et al. (1997). As opposed to the results obtained by Hancock et al., they refer to an unpublished study that shows that the model's similarity ratings were strongly correlated with human performance in a same-different judgment task.

The advantage of the approach by Hancock et al. and Biederman et al. is that they tested a feature method that automatically places Gabor jets on corresponding positions of faces. In our research, we have manually placed the feature landmark points on the faces. In future research, we hope to test automatic methods (Lades et al. 1993; Lanitis et al. 1995; McKenna et al. 1997; Wiskott et al. 1996; Yuille, 1991) to position Gabor jets on faces.

Cutzu and Edelman (1998) obtained human proximity data for complex 3D objects. These artificially generated objects were fully specified by a set of parameters (defining the shape of the parts, their orientations, and relative positions). They generated different sets of objects of which they knew the exact configuration in parameter space: for example, they generated stimuli whose parameters formed a triangle, square, cross or a star in the parameter space. The question of interest was whether the structure in the artificially generated objects was reflected in the human proximity data. They performed MDS analyses on the proximity data and then calculated the degree of match with the underlying triangle, square, cross or star configuration. The degree of match was determined by a Procrustes transformation that allows scaling, rotation, reflection and translation of a configuration to fit another

configuration. They found similarities between the structure of the objects in parameter space and the structure of the human proximity data to these objects. In one simulation, they tested the ability of simple receptive field activities to these objects to model the human proximity data. They found a poor correspondence between the structure in human proximity data and the structure in proximities of the simple receptive field activities to these objects; the receptive field activities tended to be more sensitive to the orientation than the identity of an object. Cutzu and Edelman have focused on low-dimensional representations for complex visual objects. Similarly, in our research, we have focussed on a relatively few number of feature abstraction units to capture the proximity structure of faces. They also indirectly compared the structure of human proximity data with the structure of the physically described objects with the Procrustes method. In contrast, in the feature mapping model, we can directly compare the model's similarity rating with the human similarity rating.

New directions for the MDS approach

The traditonal approach in MDS has been to ignore any quantitative information about the stimuli of interest. The idea is that useful information about the perception of for example faces can be extracted without any explicit reference to the features that are relevant for face perception. In fact, Roger Shepard (1980), one of the key researchers developing the MDS approach stated that "This purely psychological approach has advantages over a psychophysical one in the case of complex naturalistic stimuli such as faces, for which the relevant physical dimensions are as yet poorly characterized …" (p. 390). Since 1980, there has been a lot of progress in developing featural representations for faces. We do not claim that we now know exactly what features are used in face perception. Instead, we believe that with methods similar to the feature mapping approach, it is possible to make more precise what features and combinations of features are useful in modeling face perception.

Acknowledgements

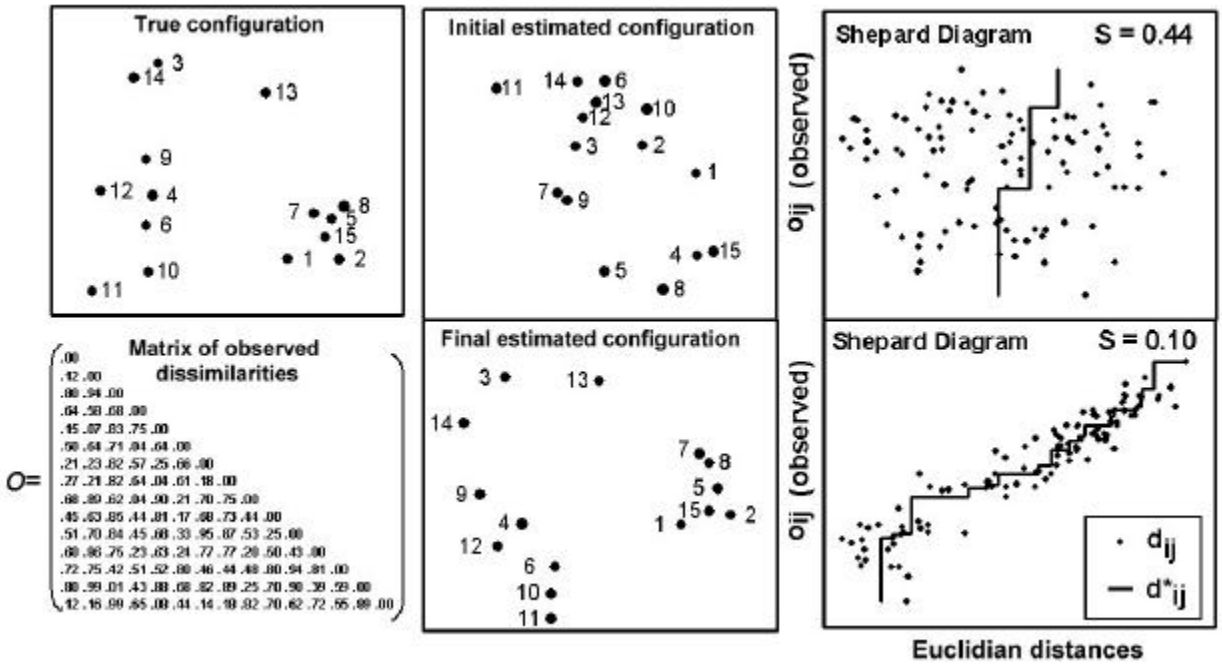
The authors wish to acknowledge the insights gained from suggestions given by Peter Todd, Peter Grünwald, Matthew Dailey, Gillian Rhodes, Michael Wenger, Rob Nosofsky, Rob Goldstone and Richard Shiffrin.

Appendix

Nonmetric multidimensional scaling (MDS)

The goal of nonmetric MDS is to find a configuration of points in some multidimensional space such that the interpoint distances are monotonically related to the experimentally obtained dissimilarities. We will illustrate the nonmetric MDS approach through an example. This example will serve to introduce the concepts behind the approach that are useful for later sections, and is not meant as an in depth introduction (see Shepard, 1962a,b; Kruskal, 1964a,b; Schiffman, Reynolds & Young, 1981 for details).

Let us assume that we actually know for 15 items in a two dimensional psychological space the coordinates for each item along each dimension. Let K represent the number of dimensions, and x_{ai} the coordinates for each item *i* along dimension *a*. Let us assume that all coordinates fall in the range [0 1]. In the upper left panel of Figure A1, we show the configuration of these points in the two dimensional space. To simulate the noisy data from an experiment, we generate the observed dissimilarities o_{ij} between item *i* and *j* with:



$$o_{ij} = \left[\sum_{a=1..K} \left| x_{ai} - x_{aj} \right|^{r} \right]^{1/r} + \mathrm{n}(\mu;\sigma) \qquad \text{(A1)}$$

The first term represents the Minkowski distance metric. We set *r* to 2 which gives rise to a Euclidian distances (r=1 leads to city block distances). The second term represents normally distributed noise with mean μ and standard deviation σ (in this example μ=0, σ=0.08). In the lower-left panel is shown the data matrix of observed dissimilarities. When given just this data matrix, the goal of non-metric MDS is to find points x'_{ai} whose interpoint distances, d_{ij} are *monotonically* related to the observed dissimilarities o_{ij}.

As a first attempt to retrieve the original configuration of points, we start with a random two dimensional configuration for x'_{ai} (upper-middle panel) and take the Minkowski distance metric with r=2 to generate the pairwise Euclidian distances d_{ij}. In the upper right panel, we show a Shepard diagram that relates the distances d_{ij} to the observed dissimilarities o_{ij}. It can be observed that for this initial configuration, there is no significant relationship between the distances d_{ij} and the observed dissimilarities o_{ij}. Kruskal (1964a,b) proposed a measure for the deviation from monotonicity between the distances d_{ij} and the observed dissimilarities o_{ij} called the stress function:

$$S = \sqrt{\frac{\sum_{ij} \left(d_{ij} - d^{*}_{ij} \right)^2}{\sum_{ij} d_{ij}^{\,2}}} \qquad \text{(A2)}$$

Note that the observed dissimilarities o_{ij} do not appear in this formula. Instead, the discrepancy between the predicted distances d_{ij} and the target distances d^{*}_{ij} are measured. The d^{*}_{ij} can be found by monotonic regression (Kruskal, 1964). In the Shepard diagram, instead of

showing individual points for the target distances d_{ij}^{*} , we connected them by a solid line. The target distances d_{ij}^{*} represent the distances that lead to a perfect monotonic relationship to o_{ij} (as can be seen by the solid line) that minimize the stress function for the given d_{ij} . The Kruskal stress measure is a lack of fit measure: when S equals 0, there is a perfect monotonic relationship between the distances d_{ij} and the observed dissimilarities o_{ij} . The goal in nonmetric MDS is to find the configuration of points that gives the minimum stress value. Kruskal (1964a,b) and Takane and Young (1977) have proposed optimization algorithms for this problem. These algorithms minimize stress in two alternating phases. In the optimal scaling phase, a monotonic regression analysis finds the target distances d_{ij}^{*} for fixed d_{ij} (therefore fixed x'_{ai}) such that the stress is minimized. In the second phase, the coordinates x'_{ai} are optimized to bring the distances d_{ij} closer to the target distances d_{ij}^{*} (these are held constant in this phase) in order to minimize stress. This optimization is continued until stress cannot be improved further. In the lower middle panel is shown the configuration x'_{ai} after optimization. The stress has been reduced to 0.10 (compared to 0.44 for the random start configuration). In the lower right panel is shown the Shepard diagram that shows a reasonable degree of monotonic correspondence between the distances and observed dissimilarities. It can be observed that the final configuration is similar to the true original configuration (since the observed dissimilarities were based on noisy samples of the true distances, the retrieved configuration cannot be expected to be exactly the same).

To simplify the example, we started initially with a random configuration for the items. A random configuration is not a good configuration to start with and might take a long time to converge. It is better to perform Torgeson scaling, based on a theorem by Young-Householder (see Schiffman, Reynolds & Young, 1981; Torgeson, 1952 for details), to get an good initial configuration for the stress minimization algorithm.

References

Ashby, F.G., (1992). *Multidimensional models of Perception and Cognition*. Erlbaum, Hillsdale, NJ.

Barsalou, L. (in press). Perceptual symbol systems. *Behavioral Brain Sciences*.

Biederman, I., & Kalocsai, P. (1997). Neurocomputational bases of object and face recognition. *Philosophical Transactions of the Royal Society London: Biological Sciences, 352*, 1203-1219.

Busey, T. (1998). Physical and psychological representations of faces: Evidence from morphing. *Psychological Science, 9*, 476-482.

Cutzu, F., & Edelman, S. (1998). Representation of object similarity in human vision: psychophysics and a computational model. *Vision Research, 38*, 2229-2257.

Dailey, M.N. & Cottrell, G.W. (1998). Task and spatial frequency effects on face specialization. In: *Advances in Neural Information Processing Systems, 10*, Cambridge, MA: MIT Press, pp. 17-23.

Daugman, J.G. (1985). Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *Journal of the Optical Society of America A, 2*, 1160-1169.

Fiser, J., Biederman, I., & Cooper, E.E. (1996). To what extent can matching algorithms based on direct outputs of spatial filters

account for human object recognition? *Spatial Vision, 10*, 237-271.

Hancock, P.J.B., Bruce, V., & Burton, M.A. (1998). A comparison of two computer-based face identification systems with human perceptions of faces. *Vision Research, 38*, 2277-2288.

Harnad, S. (1990). The symbol grounding problem. *Physica D*, 42, 335-346.

Johnston, R.A., Milne, A.B., & Williams, C. (1997). Do distinctive faces come from outer space? An investigation of the status of a multidimensional face-space. *Visual Cognition, 4*, 59-67.

Lando, M., & Edelman, S. (1995). Receptive field spaces and class-based generalization from a single view in face recognition. *Network*, 6:551-576.

Kruskal, J.B. (1964a). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrica*, 29, 1-27.

Kruskal, J.B. (1964b). Multidimensional scaling: A numerical method. *Psychometrica*, 29, 115-129.

Lades, M., Vorbruggen, J.C., Buhmann, J., Lange, J., von der Malsburg, C., Wurtz, R.P., & Konen, W. (1993). Distortion invariant object recognition in the dynamic link architecture. *IEEE Transactions on computers*, 42, 300-311.

Lanitis, A., Taylor, C.J., & Cootes, T.F. (1995). An Automatic Face Identification System Using Flexible Appearance Models. *Image and Vision Computing*, 13, 393-401.

Laugherey, K., Rhodes, B. & Batten, G. (1981). Computer-guided recognition and retrieval of facial images. In G. Davies, H. Ellis & J. Shepard (Eds.), *Perceiving and remembering faces* (pp. 251-271). London, UK: Academic Press.

Mark, L.S., Pittenger, J.B., Hines, H., Carello, C., Shaw, R.E., Todd, J.T. (1980). Wrinkling and head shape as coordinated sources of age-level information. *Perception & Psychophysics*, 27, 117-124.

McKenna, S.J., Gong, S., Wurtz, R.P., Tanner, J., & Banin, D. (1997). Tracking facial feature points with Gabor Wavelets and shape models. *Proceedings of the 1st International Conference on Audio-Video- based Biometric Person Authentication.*

Nosofsky, R.M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General*, 115, 39-57.

Nosofsky, R.M. (1991). Tests of an exemplar model for relating perceptual classification and recognition memory. *Journal of Experimental Psychology: Human Perception and Performance, 17*, 3-27.

Nosofsky, R.M. (1992). Exemplar-based approach to relating categorization, identification, and recognition. In F.G. Ashby (Ed.), *Multidimensional Models of Perception and Cognition*. Hillsdale, NJ: Lawrence Erlbaum Associates.

O'Toole, A.J., Abdi, H., Deffenbacher, K.A., Valentin, D. (1993). Low-dimensional representation of faces in higher dimensions of the face space. *Journal of the Optical Society of America A, 10*, 405-411.

Poggio, T., & Girosi, F. (1990). Regularization algorithms for learning that are equivalent to multilayer networks. *Science, 247*, 978-982.

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

Rumelhart, D.E., & Todd, P.M. (1992). Learning and connectionist representations. In D. Meyers and S. Kornblum (Eds.), *Attention and Performance*. Cambridge, MA: MIT Press.

Schiffman, S.S., Reynolds, M.L., & Young, F.W. (1981). Introduction to multidimensional scaling; theory, methods, and applications. New York, NY, Academic Press.

Sheperd, J.W., Ellis, H.D., & Davies, G.M. (1977). Perceiving and remembering faces. Technical report to the Home Office under contract POL /73/1675/24/1.

Shepard, R.N. (1962a). The analysis of proximities: Multidimensional scaling with an unknown distance function. I. *Psychometrica*, 27, 125-140.

Shepard, R.N. (1962b). The analysis of proximities: Multidimensional scaling with an unknown distance function. II. *Psychometrica*, 27, 219-246.

Shepard, R. N. (1974). Representation of structure in similarity data: Problems and prospects. *Psychometrika, 39*, 373-421.

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

Takane, Y. & Young, F.W. (1977). Nonmetric individual differences multidimensional scaling: an alternating least squares method with optimal scaling features. *Psychometrica*, 42, 7-64.

Tikhonov, A.N., & Arsenin, V.Y. (1977). *Solutions of ill-posed problems*. Wiley, New York.

Todd, P.M. & Rumelhart, D.E. (1992). Feature abstraction from similarity ratings: a connectionist approach. *Unpublished manuscript*.

Torgeson, W.S. (1952). Multidimensional scaling: I. Theory and method. *Psychometrica*, 17, 401-419.

Turk, M., & Pentland, A. (1991). Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3, 71-86.

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

Valentine, T. (1991b). Representation and process in face recognition. In Watt, R. (Ed.), *Vision and visual dysfunction*. Vol. 14: Pattern recognition in man and machine series editor, J. Cronley-Dillan). London: Macmillan.

Wiskott, L., Fellous, J.M., Krüger, N., & von der Malsburg, C. (1997). Face Recognition by Elastic Bunch Graph Matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):775-779.

Yuille, A.L. (1991). Deformable templates for face recognition. *Journal of Cognitive Neuroscience, 3*, 59-70.

Footnotes

[1]On a more general note, Barsalou (in press) and Harnad (1990) have claimed that any psychological theory that uses only amodal variables (variables of which it is not specified how they relate to a perceptual modality) faces problems because such a theory is too unconstrained and therefore cannot make strong a priori predictions.

[2]The landmark points for the geometric distances were actually never put on the faces directly for the purpose of obtaining these distances. Instead, the distances were computed based on landmark points that initially served the purpose of control points for morphing algorithms. These control points were put on manually by one user. Because only one user provided the control points, we do not have data available to assess the reliability of this procedure (as was done in Rhodes, 1988).

[3]This is true because we use the Euclidian distance metric. With other Minkowski distance metrics, the representation cannot be arbitrarily rotated while preserving all pairwise distances.