CONNECTIONIST APPROACHES TO VISUALLY-BASED FACIAL FEATURE EXTRACTION

ALICE J. O'TOOLE AND HERVÉ ABDI

ABSTRACT. We examine here some properties of a connectionist autoassociative matrix for storing, in a parallel and distributed fashion, face stimuli that are coded as simple patterns of spatially varying light intensities. First, we find that the opposition of positive and negative point contributions for nearly all the eigenvectors forms head/hair shapes, often containing the positions and shapes of eyes. Second, the opposition of positive and negative points that contribute strongly to the determination of the first eigenvector appear to separate male and female head/hair shapes. We find also that pixel positions that contribute strongly to the eigenvectors generally form spatially contiguous groups in the face pattern, often form face/head shapes, and occasionally consist of points that form a hairstyle. The results are discussed in terms of previous results indicating the salience of these 'features' for discrimination and identification, and in terms of Bruce & Young's (1986) visually-derived semantic code.

Research on the topic of face processing has expanded rapidly in recent years and now encompasses a variety of disciplines that span the cognitive sciences. Within these disciplines, an equally diverse range of face processing tasks has been studied, including recognition, classification into physical and semantic categories, naming, recall involving descriptions of features, and matching of faces and facial features. The diversity of these tasks challenges a wide range of the aspects of the human face processing system. While much early research on face processing was not theoretically motivated, a variety of strong and relatively comprehensive theories of face processing have since emerged (Hay & Young 1982, Ellis 1986, Rhodes 1985, and Bruce & Young 1986) and currently form a framework within which most present research on the topic is conducted.

Despite the expansion of research on these diverse aspects of face processing, relatively little theoretical and empirical work has concentrated on

Correspondence should be sent to: Hervé Abdi, School of Brain and Behavioral Sciences, The University of Texas at Dallas, MS:Gr4.1, Richardson, TX 75083-0688, E-mail: herve@utdallas.edu, home page: http://www.utdallas.edu/~herve
Reference: O'Toole A.J., Abdi, H. (1989). Connectionist approaches to visually-based facial feature extraction. In Tiberghien G. (Ed.) *Advances in Cognitive Sciences*. London: Wiley. pp 123–140.

understanding the properties of faces as spatial patterns,[1] Faces are complex and highly similar patterns, with a nose, two eyes, and mouth appearing on all faces in roughly the same configuration. As such, the visual information that may be used to recognize, discriminate, categorize, and match faces is likely to involve quite subtle aspects of the spatial pattern.

While it is clear that psychological data (summarized and extended by Bruce, 1979) indicate that for familiar faces a purely visual representation is likely to fall far short of capturing the information that humans use for face processing, the properties of faces as spatial patterns are nonetheless worth considering for a number of reasons. Primarily, there is a great deal of psychological data indicating that people are able to extract meaningful information including sex, age, and race, from unfamiliar faces. Further, people are also able (or at least quite willing!) to make arbitrary categorizations of unfamiliar faces into semantic categories such as occupations and even personality types (Bower & Karlin 1972, Klatsky et al. 1984, Abdi 1986). These analyses are likely to involve a true combination of visual information and semantic associations that have previously been made to similar familiar faces. The meaning, in perceptual or quantitative terms, of similar, even for the relatively simple categorizations of sex and age, is a question that has not often been addressed in these studies despite its inherent theoretical interest[2].

1. Computational Issues

The issues surrounding visual representation and processing of complex images have been addressed in recent years generally within the context of artificial intelligence. For many psychologists, the goal of applying such strategies to perceptual and cognitive tasks has been to model human information processing. Within the domain of face processing, a number of researchers have begun recently to consider some of these issues (see Bruce & Burton, in press, for a thorough review), and have modelled a variety of face processing tasks. One major difficulty with these face processing models, and in fact nearly all computational image models, is the lack of an adequate definition and representation of the features that make up the image. Clearly, however, the adequacy of these features is strongly dependent on the image processing task to be performed (i.e., recognition, categorization, or the extraction of some specific useful piece of information, such as the expression on a face), as well as on the criteria with which the model is to be evaluated. This computational observation is quite in tune with present

[1]Spatial patterns is meant here in a sense independent of visual processing. Clearly, Sergent (1982) has pioneered work considering faces as visual stimuli in a general processing sense.

[2]The studies by Pittenger & Shaw (1975a, b) are notable exceptions for the perception of age.

theories of face recognition that posit several relatively separate or at least independent face codes that subserve different processes.

2. Rationale

The purpose of the present paper is to explore the usefulness of connectionist or neural-network models in conjunction with a relatively simple spatial coding, for quantifying a type of facial feature set that has not traditionally been considered in models of face processing. We hope to show some interesting relationships between these kinds of features and those that have been discussed in a number of studies and reviews as salient features for recognition and similarity judgements of unfamiliar faces (Davis et al. 1979, Shepherd et al. 1981). Further, we will speculate a little about how one may reliably separate faces into some simple visually-derived semantic categories using information in the system.

Before proceeding, it is perhaps worth putting the present work in perspective. The data presented here are clearly exploratory in nature. For the present, we tend to view the methodology used here more as a tool-box full of new ways to think about and to quantify the visual information in faces than as a model of face processing. The usefulness of these data for predicting psychological phenomena awaits much empirical work. For the present, we limit ourselves to pointing out the ways in which these analyses may give insight into some presently existing psychological data.

3. Neural networks—Applications to visual data

As a brief but necessary digression, neural network or connectionist models make use of the potential of large arrays of single units to compute in a parallel and distributed fashion. Parallel refers to the fact that large numbers of computations are independent of one another and therefore may occur (potentially at least) simultaneously. Distributed refers to two aspects of the models. First, the representation of a single stimulus is not contained in the activation of a single unit, but rather is distributed throughout the memory in a non-local sense. Second, the units of the model play a role in the representation of the entire stimulus set and are not dedicated to a single stimulus[3].

Connectionist models come in all shapes and sizes, with learning (adaptive) and preset (a priori) constraint versions, linear and non-linear neural activation functions, fully and partially connected units, and a variety of other parameters that affect performance in various (and generally unanalyzable!) ways. The unanalyzability is not meant here in a bad sense, but may, in fact, constitute much of the current intrigue the models hold, not

---

[3]The system presented in Aleksander (1983) and Stonham (1986) uses distributive coding in the first but not the second sense.

only for psychologists, but for engineers, neuroscientists, and physicists as well. Despite the variety of models and algorithms of connectionist approaches, what ties these strategies together as a group is their parallel and distributed nature, and the distinctive physiological flavor of their computations.

As a cautionary note, it is perhaps worth keeping in mind that despite the strong popularity of connectionist models in the last few years, these models are not the long awaited cure-all for many long-standing and presently intractable problems, especially in visual science. For example, virtually nothing is known about how constantly changing images on the two retinae (due to an observer and a world in motion and also to the intelligent and efficient movements of the eyes), are turned into meaningful objects and events in the brain. The perceptual world is at least four-dimensional with useful and meaningful information being extracted constantly from the x, y, z, and time dimensions. The present simulations, and all computational models of face processing that we know of, are limited sadly to the first two of these four dimensions. Thus, one point to bear in mind is simply that however far connectionist models go in expanding our understanding of some old problems, they are not likely to go far in solving the problems as a whole until more psychological and neurophysiological data contribute to a better understanding of these fundamental issues.

Despite the pessimistic nature of such a warning, one cannot help but be guardedly excited about the potential of neural network approaches to offer some intriguing new insights that were not available with previous ways of conceptualizing models of face processing. As Anderson & Rosenfeld (1988) have pointed out, much of this contrast lies in the fact that neural network models are computing memories, not computers. The present model deals almost exclusively with the properties of a network as a memory, rather than as a dynamic system implementing the coordination of processes (a framework that has been more the tradition in models of face processing). As such, a connectionist system such as the one used here attempts to find statistical regularities in the patterns that are inherent in the distributed nature of the memory.

4. PREVIOUS WORK

Network models have been applied previously to the task of recognition of spatially transformed faces (Millward & O'Toole 1986, O'Toole et al. in press). This work built on a demonstration by Kohonen (1977) that illustrated the usefulness of a simple linear associator for storing and retrieving images in a parallel and distributed system. Briefly, the above studies showed that the performance of a simple associative model of face recognition, using a low-level spatial coding, was able to mimic the pattern of recognition transfer errors found with human subjects performing a recognition transfer task for spatially filtered faces. While the conclusions that one may draw from models of this sort are quite limited, perhaps the major point of the study was simply that the extraction and coding of high-level features, as has traditionally been defined, was not necessary to parallel the performance of the observers on the task.

One problem in using models such as those employed in O'Toole et al. (in press), is simply that it is often quite difficult to analyze the data-base properties of the model in terms useful or at least familiar to psychologists. In statistical terms, however, Sirovitch & Kirby (1987) have provided a thorough analysis of the properties of covariance matrix created from face images cropped to include only the eyes and nose. The goal of their work was to provide a low-dimensional optimized representation effaces from a given stimulus set using principal components analysis techniques. Sirovitch & Kirby (1987) show that faces stored in such a matrix can be reconstructed to quite a recognizable form (to within 3% error) using roughly 40 parameters and the corresponding 40 eigenvectors of the matrix. The parameters correspond to the weighting coefficients applied to the eigenvectors to recreate a given stored face from a linear combination of the eigenvectors. While the work of Sirovitch & Kirby (1987) is presented in a somewhat different light, it is clear that the covariance matrix they use is equivalent to a linear auto-associative connection matrix.

For the psychological properties of these systems, Abdi (1987, 1988) has gone some way toward clarifying the issues. The purpose of Abdi's work was to make clear and expand the relationship between a multivariate statistical approach and neural-modelling approach. One psychologically interesting aspect of an auto-associative matrix created using faces is that some of its eigenvectors are strikingly face-like. This is demonstrated both in Abdi (1988) and Sirovitch & Kirby (1987). Abdi (1988) has related these eigen- vectors to prototypes of faces. Thus, he found in a matrix composed of both male and female faces, eigenvectors corresponding to feminine faces, and One eigenvector corresponding to a distinctly masculine face.

It is here that we should note the gradual blurring of terminology in psychology. Sirovitch & Kirby's (1987) use of eigenvectors is clearly as features from which a face is built. While these features are more configural than most researchers in psychology are used to talking about (i.e., eyes, nose, mouth, etc.), they are quite a valid conceptualization of a feature. The major difference between these features and more local ones is simply the semming lack of psychological reality of the eigenvectors that are not all face-like or do not contain distinguishable sub-parts of the face. Nonetheless, these features are often quite necessary for "putting the face back together again." In this sense, such an interpretation is rather at odds with traditional psychological ideas that features are something that can be "pointed out," "described," or at least "distinguishable" in a face.

From a neural modelling perspective, conceptualization of the eigenvectors of a covariance matrix as features, is not new. Anderson et al. (1977) proposed that since the eigenvalues are a measure of the importance for discriminating between the items learned, the eigenvectors corresponding to the largest eigenvalues may be thought of as distinctive features. Further, Anderson & Mozer (1981) expand these ideas and propose the following:

> If we were to call the eigenvectors with large positive eigenvalues features, then we would have a system that has some similarities to what feature analysis is supposed to do but where features are now represented by patterns of activity rather than by selective neurons (p. 221–222).

They call these types of features macrofeatures and contrast them to micro features, which correspond to something like the properties for which individual neurons display selectivity (e.g., oriented line segments for some neurons in visual cortex). Anderson & Mozer (1981) claim that the macrofeatures are what perception and feature analysis actually use and are the psychologically salient entities. They present a series of simulations of English letter categorization coding the letters as spatially varying patterns. Anderson & Mozer (1981) present the five eigenvectors with the largest eigenvalues and show that there is no easily discernible relationship between the eigenvectors and the types of features that are typically seen in intui- tively derived feature sets for letters. They do, however, show one eigenvec- tor that is intuitively interpretable as a relatively specific local feature detector. However, there is an important difference in the letter stimuli used by Anderson & Mozer (1981) and the face stimuli used here. This is simply that faces have a strong inherent configural structure, whereas no such obvious structure exists for letter stimuli. Nonetheless, Anderson & Mozer (1981) point out that interpretable or not, macrofeatures are completely defined in a formal sense.

The second rather different characterization of the eigenvectors as prototypes (Abdi 1988) is also quite a reasonable interpretation of at least the face-like eigenvectors. The psychological validity of such an interpretation could (and, one hopes, will soon!) be tested in some traditional prototype recognition tasks that test the distinguishability of faces based on their similarity to the (or some?) prototype(s) defined as such[4]. The relationship between these face-like eigenvectors and the non-face-like eigenvectors is

[4]The question of whether there are one or more prototypes is an interesting one, but is beyond the scope of the present paper.

not easily interpretable in either a traditional feature or prototype context. The terminology used here is somewhat arbitrary in the sense that "prototype" is a psychological term that tends to be defined somewhat recursively by the performance of subjects rather than by any physical criteria. Further, in the domain of face processing, researchers do not always agree on its use or meaning.

An interesting point about the two ways of conceptualizing the eigenvectors is that it mixes traditional psychological ideas about features and prototypes in a usual manner. Thus, we have in the eigenvectors both the feature components needed to build many different kinds of faces, as well as components that have properties which themselves may be related to prototypes (since some of the eigenvectors are themselves members of the set of all faces). We hope to offer an interesting third dimension to this mixing of terminologies.

5. DEFINING THE MODEL

5.1. **Stimuli.** We used 28 faces from a relatively homogenous pool of both female (ra=16) and male (n==12) college students aged approximately 18 to 22 years[5]. The facts were not carefully aligned, as they were in Sirovitch & Kirby (1987). Our reason for not aligning the faces is that there is no simple way to do this, given a set of relatively naturally posed photographs. Further, we hoped to preserve some of the subtle difference in poses of the subjects in the photographs. Also, we used complete faces, including the hair and lower face features rather than a sub-section of the face as was used in Sirovitch & Kirby (1987), Millward & O'Toole et al. (in press).

Faces in the present model are denned as state vectors $\mathbf{f}_1, \ldots, \mathbf{f}_j, \ldots, \mathbf{f}_J$, where J is the number of faces. These vectors consist of the concatenated rows of pixel intensities extracted from digitized photographs of faces. While from the point of view of the network, the particular coding used (i.e., pixel intensities) is a somewhat arbitrary decision, it is not without impor- tance. This is a consequence of the memory-driven nature of the system. We have chosen to use a pixel representation for three reasons. First, while some psychological data (Millward & O'Toole 1987, O'Toole et al. in press) indicate that a pixel representation produced approximately equivalent performance to a system sensitive only to line information in the faces, there are other data that indicate that zero-crossings may not capture important information for face processing. In particular, Galper & Hochberg (1972) showed a recognition performance deficit when faces are presented in the photographic negative, despite the preservation of line information in faces transformed in this way. By using pixel information, we hoped to preserve

[5]Clearly more faces and, more importantly, more diverse faces would have been better. The limited number used here is due to some technical problems.

as much information as possible. Second, as Bruce (1988) points out, the importance of information that clarifies the three-dimensional structure of a face should not be underestimated. Shading and stereo information are the two obvious sources for such information. Thus, while primal sketch type representations do not contain shading information, they assume stereo information for the creation of the 3-D representation. Given that our model does not include stereo information [6]. it is perhaps wise to include the shading information contained in the pixel representations. Third, from a practical point of view, use of pixels provides a visible representation, and thus the performance of the model is easily accessible.

5.2. **Procedure.** To store the faces, an autoassociative model was created using the following rule:

$$\mathbf{A} = \sum_j \mathbf{f}_j \mathbf{f}_j^T \tag{1}$$

where, $\mathbf{A}$ is the memory matrix, $\mathbf{f}_j$, is the j-th input vector of pixel intensities.

The relationship between this type of matrix to the matrix used in some previous work (Millward & O'Toole 1986, O'Toole et al.) in which an error correction rule was used to learn the faces as follows. The Widrow-Hoff (Duda & Hart 1973) learning rule changes the memory matrix weights in the following manner:

$$\delta\mathbf{A} = \alpha\left(\mathbf{f}_j - \mathbf{A}\mathbf{f}_j\right)\mathbf{f}_j^T \ , \tag{2}$$

where α is a learning constant. This weight-changing procedure is applied iteratively until some learning criterion is reached.

For the purposes of eigenvector extraction, the above corrected matrix at the point of convergence (i.e. its Penrose pseudo-inverse, Green & Carroll 1976) may equally well have been created without the learning rule simply as associative matrix. The effect of the learning rule is not to change the eigenvectors, but rather to equalize their importance by equalizing the eigenvalues. The learning procedure has the effect of greatly improving the system estimations of the faces, and is therefore very important for recognition tasks. However, given that we are presently concerned with the eigenvectors, an equivalent way of proceeding is to use the uncorrected autoassociative matrix. A second way to define this matrix, and one more intuitively in tune with our later analyses, is as follows. If $\mathbf{X}$ is a matrix containing the original face vectors as columns $1, dots, J$, then an associative matrix memory $\mathbf{M}$ can be made as:

$$\mathbf{M} = \mathbf{X}\mathbf{X}^T \ . \tag{3}$$

[6]Bruce (1988) argues that shading information is likely to be more important for face processing than stereo information.

It is clear that $\mathbf{M}$ will be a matrix of rank R, where $R \leq J$, assuming that there are more pixel elements than face stimuli. Thus, the matrix $\mathbf{X}$ will have two sets of R eigenvectors corresponding to the non-zero eigenvalues. One set of these eigenvectors will correspond to the rows of $\mathbf{X}$ (i.e., the pixels) and can be extracted from the $I \times I$ matrix $\mathbf{X}\mathbf{X}^T$, and the other set will correspond to the columns of $\mathbf{X}$ (i.e., the faces) and can be extracted from the $J \times J$ matrix $\mathbf{X}^T\mathbf{X}$.

For the present, we will consider the eigenvectors corresponding to the pixels. Each of the I pixels is a point in the space defined by the eigenvectors. It is an interesting psychological point to recall that each of these pixel values as it varies across the J faces is differentially important for determining each of the eigenvectors. This importance or "contribution" of any given pixel $\mathbf{p}$ to any given eigenvector $\mathbf{e}$ can be found by taking the squared length of the projection of the vector $\mathbf{p}$ along the eigenvector $\mathbf{e}$.

$$\|\mathrm{proj}_{\mathbf{e}}\mathbf{p}\| = \left(\frac{\mathbf{p}^T\mathbf{e}}{\|\mathbf{e}\|}\right)^2 . \tag{4}$$

Each contribution is given the sign of the projection used to compute it. In these differential contributions we can see the importance of individual pixels for determining the eigenvectors. Further, since the faces have been coded in graphic format, we can display these important pixels/areas of the face for each eigenvector. There are a number of ways to implement this selection process. Perhaps the simplest way to do this is just to display points that exceed the mean point contribution for a given face by the standard deviation of the point contributions for a given face weighted by some arbitrary amount. Since there are at present no compelling reasons to complicate this process, we have chosen this simple scheme.

A second interesting psychological point to recall is that the contributions of the pixel values to each of the eigenvectors may be either positive or negative. In simple terms, the intuitive meaning of this positive to negative contrast is that it may serve to oppose different areas of the face in a way that may be useful for discriminating groups from one another.

Before presenting the data, it is worth exploring some of the possible ways in which this analysis might turn out, and further, speculating a little about what the different results might mean. The first, and perhaps least intuitively interesting possibility, would be that the important points for the eigenvectors will be non-contiguous and randomly scattered around the face. What this would mean is that the model was not sensitive to the inherent structure of a face pattern per se.

A second general possibility is that the important pixels will be found in a (some) contiguous group(s). This possibility presents the further possibility of either local feature contiguity (e.g., the region around the nose may be seen to be strongly important for some eigenvector), or some type of global configural continuity. Global configurational information may either be contiguity of points that may span the face in some way (e.g., an outline of a face shape), or perhaps the combination of two or more local areas of the face (e.g., the eyes and chin region).

Finally, none of the above possibilities for any given eigenvector precludes the existence of the other possibilities for other eigenvectors, or even for the same eigenvector. Thus, it is possible to have all or any combination of the following occurring for the different eigenvectors, and also to have mixing within an eigenvector of the various patterns of results. It is clear that this analysis does not limit itself to considering 'features' of a certain predefined type, for example, local versus configural, but rather will naturally make use of the local and/or configural structure of the information.

6. Results

The presentation of the results is somewhat difficult in that they are clearly of a qualitative and not quantitative nature. Thus, we try here to make some general statements about the data and then to provide some examples of the interesting phenomena. First, we have found, almost without exception, that the important points for the eigenvectors are contiguous in the face. Second, we have found that by displaying the points contributing positively to the eigenvectors versus those contributing negatively, something like a head shape, often including the positions and the shapes of the eyes, emerges. This appears often in a form that separates the hair, almost like a wig, from the internal features and the background of the picture. Third, in displaying the strongly important points, we have found a variety of interesting configurations, sometimes interpretable and sometimes not.

For the first point, the fact that the important points tend strongly to be contiguous in the face indicates that the eigenvectors are capturing some of the structure of the face pattern. For the display of pixels contributing positively and negatively to the eigenvectors, the oppositions for the first 16 eigenvectors are illustrated in Fig. 1 and may be characterized mostly as head and/or hair shapes. In a few cases, some configural information about the positions of the nose and mouth is also seen. Some of the positive negative point separations may be further characterized as distinctly female (e.g., 1, 4, 7), whereas others appear to be distinctly male (3, 11, 14).

For the display of strongly important points for the eigenvectors, we will start by using the first eigenvector as an example. In Fig.2a we display the points that exceed the mean plus the standard deviation of the point contributions of the first eigenvector by a factor of 1.2. Then, a head/hair shape with relatively long curly (or at least fluffy) hair appears. This head shape was relatively typical of the women in our sample. On the other hand, in Fig. 2b we display the pixels that contribute negatively to the eigenvector

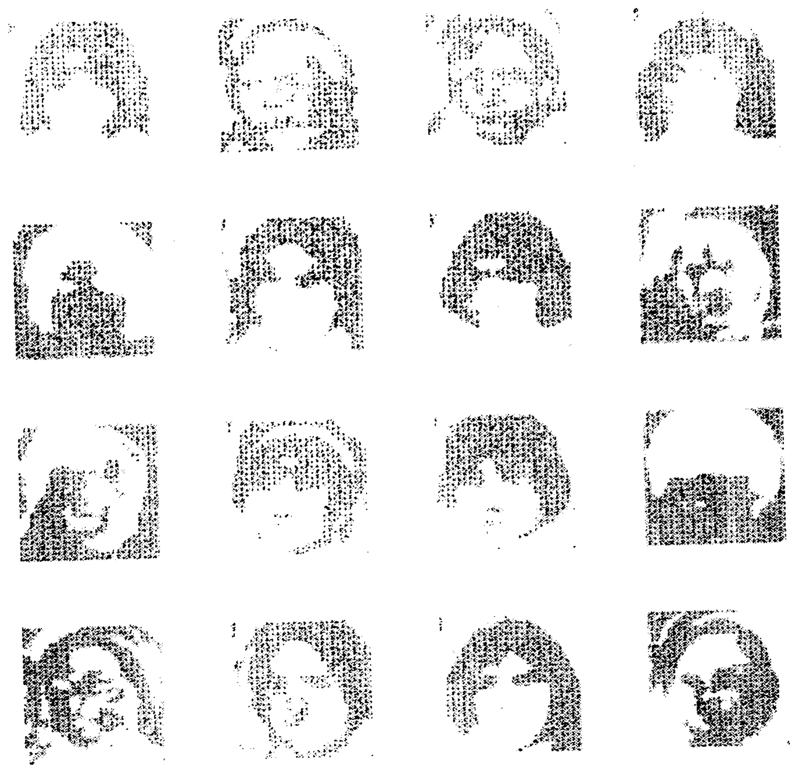

FIGURE 1. Points contributing positively (white) and negatively (black) to the first 16 eigenvectors. These generally form head, hair, and face shapes sometimes contain- ing the shapes and positions of the eyes. The faces are numbered $1 \ldots 4,\ 5 \ldots 8$, etc., left to right.

and that had contributions less than the mean contribution minus the standard deviation of contributions weighted by a factor of 1.2. What appears is a male, short-cropped hair shape that was relatively typical of the men in our sample. Finally, the two are shown superimposed in Fig. 2c.

The display of male versus female-like head/hair shapes in the important point contributions indicates that this eigenvector may be capturing the features that contrast the male and female faces in our sample. This interpretation was supported by a correspondence analysis (Benzecri, 1973) on the pixel representations of the faces. Briefly, a correspondence analysis optimizes the representation of distances between faces in a low dimensional space. We have applied this analysis to the faces represented as pixels. Thus, the distance relationships that emerge are based only on the spatially varying patterns. The resultant first axis of this analysis does quite a good job of separating the male and female faces in the sample. We found that the axis placed all 12 of the male faces on one side of the axis, and 14 of the 16 female faces on the other side of the axis. The two female faces that were placed on the male side of the axis consisted of one very short-haired woman and a second woman whose hair was tightly pulled back off her face. This is an indication that the model may have been using hair/head shape to separate the males and females, and not some more subtle feature such as the delicacy of inner features.

A second interesting point about the important pixels, even when the opposition between positive and negatively contributing pixels does not lend itself to easy interpretation, is that they often form hair styles that appear almost as wigs. Several examples are displayed in Fig. 3. This is an indication of the salience of these features for the statistical separation of groups of faces in the set.

7. Discussion

A number of interesting and potentially useful 'features' have emerged that reflect the spatial structure of the stimuli. These are face shape (sometimes including the eye shapes and positions), hair style, and some more complex configural type features. We recall once again the exploratory nature of the data, and thus remind the reader that in this section we limit ourselves to offering some speculations about the potential relationship between these data and some previous psychological data. Further, we limit ourselves to addressing only unfamiliar face recognition.

One interesting aspect of the kinds of 'features' that we have found is that they recall some work reviewed by Shepherd et al. (1981) on cue saliency in face recognition, as well as some work by Davies et al. (1979) on similarity effects in face recognition.

To begin,. Shepherd et al. (1981) reviewed a number of studies that addressed one of the following questions:

- Which parts of the face are the most important for enabling a subject to identify a face as one he has seen one a previous occasion?
- What parts of the face are most important in judging that two faces are similar?

The kinds of studies reviewed involved tasks including fragmentation of the face for recognition, alteration of features, face description, and multidimensional scaling. Shepherd et al. (1981) present a very thorough review of the literature up to that time, and conclude that

> Probably the most consistent finding has been that hair is the most important single feature for the purposes of these studies (p. 130).

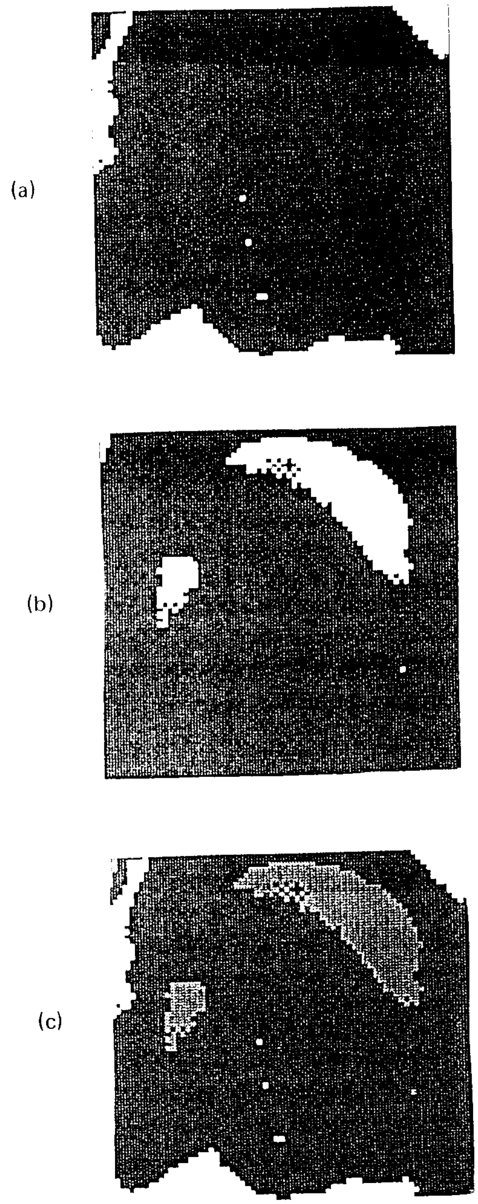

FIGURE 2. **(a)** The important positive point contributions for the first eigenvector in white. These form the outline of a relatively female head with thick hair. **(b)** The important negative point contributions (in white) for the first eigenvector. These separate a relatively male head with short-cropped hair from the background. **(c)** The super-position of Fig. 2a and 2b.

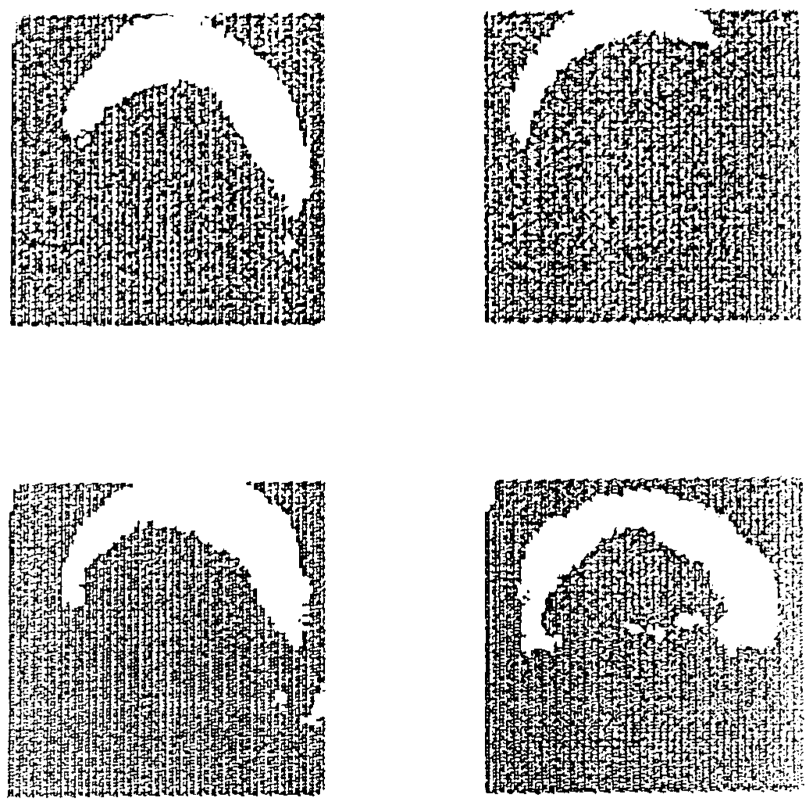

FIGURE 3. **(a)** Fig. 3 The strongly important points (in white) for the fifth, twelfth, nineteenth, and twenty-third eigenvectors.

They comment also that the salience of hairstyle is quite ironic in that hairstyle is perhaps the most changeable feature in the face. As such, its inherent value as a consistent cue for face identification is clearly limited.

The present analysis has also indicated the statistical importance of the hair shape for accounting for variance in spatially varying face patterns. It is certainly true that logically speaking, hair is not the best cue to rely on for identification purposes; nonetheless, the hair takes up a relatively large area of the face and may be salient simply by virtue of this fact. Further, the studies reviewed by Sheperd et al. (1981) attest to the fact that observers are often unable to ignore this 'relatively uninformative' feature in favor of some other potentially more reliable features (at least for unfamiliar faces).

Shepherd et al. (1981) also note that in a multi-dimensional scaling study (Shepherd et al. 1980), face shape emerged as an important feature. More evidence for this contention can be seen in Davies et al. (1979). They presented a hierarchical cluster analysis of judged similarity of male Caucasian faces, and showed that the clusters related strongly to subject performance on a recognition task. The analysis employed in Davies et al. (1981) revealed four clusters. The largest cluster consisted of "young men with long, thin faces and short, tidy hair." Of the smaller clusters, one consisted of "much older men with square chins, lined complexions, fat faces, and short, tidy, receding, hair." A second cluster consisted of "young men with short, fat faces, curly hair, and narrowed eyes," and a final cluster consisted of "young men with long, oval faces, thick, medium-length hair, and wide eyes." Davies et al. (1979) suggest on the basis of their results, and on the basis of some previous studies, that face shape, age, hairstyle, and eyes are particularly salient features for discriminating faces.

The macrofeatures found here, and their variants made by examining the important pixels, capture a number of important statistical properties of faces that correspond (somewhat more than expected) to these salience studies. Thus, we have found that many of the eigenvectors highlight face, head, or hair shapes. As mentioned previously, it is likely that our macrofeatures have been more interpretable than those of Anderson & Mozer (1981), owing to the difference in the inherent structure of faces versus letter stimuli.

The present data may be related also to Bruce & Young's (1986) suggestion of the existence of one type of code used in the face processing system. To account for the ability (or at least willingness) of people to put unfamiliar faces into categories of male/female, ethnic group, and persona- lity type, etc., Bruce & Young (1986) have proposed the existence of a visually-derived semantic code for faces. This code consists of the information about a face's owner that can be obtained from unfamiliar faces. Further, since the experiments used to support the existence of this code have all been performed by using photographs of faces, it is clearly possible to extract such information from a presentation of the face that presumably does not offer much sophisticated information for forming what Bruce & Young (1986) have called a structural code of the face (i.e., one that is view-independent).

Attributions such as sex, age, and personality types to unfamiliar faces are made on the basis of what the faces look like (Bruce 1988, p. 60). As noted, these judgments are likely to rely on comparisons between the unfamiliar face and other familiar faces that have since been categorized along the relevant dimensions. Present theories of face recognition do not generally propose a form for this similarity comparison. While human observers find it reasonably easy to categorize faces along these visually derived semantic dimensions, from a computational point of view, quantification of the visual information that reliably separates male from female faces, old from young faces (much less the face of an engineer from that of a pop singer), is not all a trivial task. This is especially the case, given the relatively similar nature of all faces.

The present results suggest one possible way to quantify some of the information important for these categorizations. For the categories of sex and age, it is very likely that the statistical structure of the spatially varying face patterns[7] as touched by the present methods, would be sufficient to show reliable separation into these categories. The present data indicate that this is true at least for the category of sex (the age of the faces was not varied).

For the category of age, a careful look at the cluster descriptions of Davies et al. (1979) and at the classic ecologically-motivated studies of Pittenger & Shaw (1975a, b) provide some insights. In the Davies et al. (1979) study, one important feature dividing the young and old men was related to hair shape (receding in the older men). In face, all of their clusters involved complex interactions of head, hair, and face shapes. Pittenger & Shaw (1975a), investigating the physical correlates of the perception of relative age, propose cardioidal strain (a mathematical transformation applied to head shape) as a visual invariant carrying age information. They showed experimentally that observers can and actually do make use of this information in judging relative age of profiles of faces. While these studies were confined to profiles of faces, they indicate the great importance of natural variations of head shapes due to ageing for the perception of relative age. In a follow-up longitudinal study using full-face photographs of chil- dren/adolescents, Pittenger & Shaw (1975b) found that features such as hair style, structural information in the internal configuration of features, and the outline shape of the face also contribute to the perception of age. All of these features are likely to be easily accessible with the present analyses.

8. SUMMARY

The present work suggests one possible way of quantifying spatially varying intensities in an image in a way that may be useful for the categorization of faces into some simple visually-derived semantic categories. The approach taken is consistent with connectionist approaches to memory and with statistical approaches related to principal component analysis. Future work should concentrate on testing the psychological relevance of these types of features for discrimination and categorization tasks.

ACKNOWLEDGMENT

This work was supported by grant BNS85-18675 NSF5-26195 to James A. Anderson. Alice, J. O'Toole was supported during the work by a Chateaubriand Post-Doctoral Fellowship from the French Embassy to the United States.

[7] Assuming reasonable scaling and alignment of the faces.

References

Abdi, H. (1986) Faces, prototypes, and additive tree representations. In: H. D. Ellis, M. A. Jeeves, F. Newcombe & A. Young (eds.) *Aspects of Face Processing*. Dordrecht: Martinus Nijhoff.

Abdi, H. (1987). Do we really need a contingency model for concept formation? *British Journal of Psychology 78* 113–125.

Abdi, H. (1988). A generalized approach for connectionist auto-associative memories: interpretation, implication, and illustration for face processing. In: J. Demongeot (ed.), *Artificial Intelligence and Cognitive Sciences.* Manchester: Manchester University Press.

Aleksander, I. (1983). Emergent intelligent properties of progressively structured pattern recognition nets. *Pattern Recognition Letters 1,* 375–384.

Anderson, J. A. & Mozer, M. C. (1981). Categorization and selective neurons. In: G. E. Hinton & J. A. Anderson (eds) *Parallel Models of Associative Memory*. Hillsdale, N. J: Lawerence Eribaum Associates, Inc.

Anderson, J. A. & Rosenfeld, E. (1988). *Neurocomputing: Foundations of Research.* Cambridge: MIT Press.

Anderson, J. A., Silverstein, J. W., Ritz, S. A. & Jones, R. S. (1977). Distinctive features, categorical preception, and probability learning: Some applications of a neural model. *Psychological Review 84* 413–451.

Benzécri, J.-P. (1973). *L'analyse des Données (2 Vol.).* Paris: Dunod.

Bower, G. H. & Karlin, M. B. (1972). Depth of processing pictures of faces and recognition memory. *Journal of Experimental Psychology 103* 751–757.

Bruce, V. (1979). Searching for politicians: an information-processing approach to face recognition. *Quarterly Journal of Psychology, 31* 373–395.

Bruce, V. (1988). *Recognising Faces.* London: Eribaum.

Bruce, V. & Burton, M. (in press). Computer Recognition of faces. In: A. W. Young & H. D., Ellis (eds), *Handbook of Research on Face Processing.* Amsterdam: North Holland.

Bruce, V. & Young, A. W. (1986). Understanding face recognition. *British Journal of Psychology 77* 305–327.

Davies, G. M., Shepherd, J. W. & Ellis, H. D. (1979). Similarity effects in face recognition. *American Journal of Psychology 92* 507–523.

Duda, R. 0. & Hart, P. E. (1973). *Pattern Classification and Scene Analysis,* New York: Wiley.

Ellis, H. D. (1975). Recognizing faces. *British Journal of Psychology 66* 409–426.

Ellis, H. D. (1986) Processes underlying face recognition. In: R. Bruyer (ed.) *The Neuropsychology of Face Perception and Facial Expression.* Hillsdale: Eribaum.

Galper, R. E. & Hochberg, J. (1972) Recognition memory for photographs of faces. *American Journal of Psychology 84* 351–354.

Green, P. E. & Carroll, J. D. (1976). *Mathematical Tools for Applied Multivariate Analysis.* New York: Academic Press.

Hay, D. C. & Young, A. W. (1982). The human face. In: A. W. Ellis (ed), *Normality and Pathology in Cognitive Functions.* New York: Academic Press.

Klatsky, R., Martin, G. L. & Kane, R. A. (1982). Semantic interpretation effects on memory for faces. *Memory & Cognition 10* 195–206.

Kohonen, T. (1977). *Associative Memory—a System Theoretical* Approach. Berlin: Springer-Verlag.

Millward, R. B. & O'Toole, A. J. (1986). Recognition memory transfer between spatially transformed faces. In: H. D. Ellis, M. A. Jeeves, F. Newcombe & A. Young (eds.) *Aspects of Face Processing.* Dordrecht: Martinus Nijhoff.

O'Toole, A. J., Millward, R. B. & Anderson, J. A. (in press). Recognition memory for spatially transformed faces: a physical system approach. *Neural Networks.*

Pittenger, J. P. & Shaw, R. E. (1975a) Aging faces as viscal-elastic events: Implications for a theory of non-rigid shape perception. *Journal of Experimental Psychology: Human Perception and Performance 104* 374–382.

Pittenger, J. P. & Shaw, R. E. (1975b). The perception of relative and absolute age in facial photographs. *Perception & Psychophysics 18* 137–143.

Rhodes, G. (1985). Lateralized processes in face recognition. *The British Journal of Psychology 76* 249–271.

Sergent, J. (1982). The cerebral balance of power: Confrontation or cooperation? *Journal of Experimental Psychology: Human Perception & Performance 8* 254–271.

Shepherd, J. W., Davies, G. & Ellis, H. D. (1981). Studies of cue saliency. In: G. Davies, H. D. Ellis & Shepherd, J. W. (eds). *Perceiving and Remembering Faces.* 105–131.

Shepherd, J. W., Davies, G., Ellis, H. D. & Freeman, J. (1980). *Identification after Delay. Final Report to the Home Office Research Unit for Grant RES 522/4/1.* Department of Psychology, University of Aberdeen.

Sirovitch, L. & Kirby, M. (1987). Low-dimensional procedure for the characterization of human faces. *Journal of the Optical Society of America 3* 519–524.

Stonham, J. (1986). Practical face recognition and verification with WISARD. In: H. Ellis, M. Jeeves, F. Newcombe & A. Young (eds.) *Aspects of Face Processing,* Dordrecht: Martinus Nijhoff.