TEXTURE SPACE §

by

Rick Gurnsey ‡ and David J. Fleet †

‡ Department of Psychology,
Concordia University, Montréal, Québec.

† Department of Computing and Information Science
Queen's University, Kingston, Ontario.

* To whom correspondence should be addressed

Rick Gurnsey
Department of Psychology,
Concordia University
7141 Sherbrooke Street West,
Montréal, Québec, Canada.
H4B 1R6

phone 514-848-2243
fax 514-848-4545
email gurnsey@vax2.concordia.ca

§ This research was supported by NSERC and FCAR Research Grants to Rick Gurnsey and David Fleet, and an Alfred P. Sloan Research Fellowship to David Fleet. We thank Michael von Grünau, Frédéric Poirier, Marta Iordanova and Cesarea Novielli for their comments on this work. We also thank Peter April for programming support.

ABSTRACT

Similarity judgments from three subjects were obtained for twenty artificial textures comprising filtered noise. Multidimensional scaling (MDS) revealed that three perceptual dimensions explain most of the variance and subjects' solutions are similar. Both individuals' similarity judgments and MDS solutions were highly correlated. A computational model utilizing the energy responses in nine bandpass filters explains an average of 80% of the variability in the original similarity scores of individual subjects. Energy responses are mapped to the perceptual space through a linear transformation that can be decomposed into two components. The first component decorrelates initial filter responses and the second component maps the decorrelated filter responses to a perceptual space. These latter transformations show remarkable agreement between the three subjects.

Studies of visual texture can be motivated from either an ecological or signal processing perspective. The ecological perspective rests on the obvious fact that visual textures are ubiquitous in the natural world; surfaces are rarely composed of materials having uniform reflectance. From this one may conclude that textures should be studied because they provide cues to object identity or that textural discontinuities can provide cues to surface-, depth- or illumination discontinuities. The signal processing perspective views textures as a useful class of stimuli for examining the way in which the visual system encodes distributions of light intensities that are more complex than sine-wave gratings, gabor patches or oriented line segments. Little really hinges on what perspective is taken. The majority of psychophysical texture studies employ artificial textures even though the motivation for such studies may rest on ecological considerations. Our goal in the present paper is to examine the representational system that permits the visual system to make similarity judgments about isolated patches of visual texture. We choose to study artificial textures with controlled spectral characteristics that are free of associations that may undermine our attempt to examine purely visual responses to the textures.

Current theories of the mechanisms subserving texture perception make use of the idea of neural images (Robson, 1980) or filter banks. A neural image represents the retinal image as "seen" through a filter selective for a particular combination of orientation and spatial frequency. Examples of this proposal can be found in Bergen and Landy (1991), Gurnsey, Pearson and Day (1996) and Harvey and Gervais (1981). The simplest version of this "neural image hypothesis" is that each scalar-valued image intensity I(x,y) is transformed into a vector $\mathbf{I}(x,y)$, each component of which represents the average "activity" in a local region around a retinal position (x,y) within a particular neural image. We ask in this paper whether the activities that textures elicit within neural images determine the appearance of textures as revealed by similarity judgments.

The question of perceived similarity is addressed here through multidimensional scaling (MDS), a computational procedure that finds structure in data matrices. Given M objects, subjects may be asked to judge the similarity of the objects in each of the $(M^2 - M)/2$ pairs that can be formed from this set. MDS algorithms (see Schiffman, Reynolds & Young, 1981) attempt to find an arrangement of the M objects in an N-dimensional space that maximizes the negative correlation between the distances that separate objects in this space and the original similarity judgments. If the fit is good between distances in the MDS solution and the original similarities then one would be encouraged to find a theoretical interpretation of the MDS solution. Ideally, one would like to determine the transformation that maps textures from their representations in a physical space (e.g., the Fourier domain) to their representations in the psychological space revealed by MDS. In the case of

texture, the relationships between textures in an MDS solution space might be related to the activities they induce in a set of neural images.

The approach taken here draws parallels between colour vision and texture vision and is inspired by early studies showing that MDS can reveal the mechanisms underlying colour perception. Shepard (1962) demonstrated that similarity judgments (collected by Ekman, 1954) about fourteen monochromatic colour patches ranging in wavelength from 434 nm to 674 nm can be "inverted" through MDS to reveal the internal organization of colour space. Specifically, the MDS analysis revealed that a 2D arrangement of the colour patches explained most of the variance in the original similarity judgments. The recovered 2D solution was essentially the well known colour wheel typically associated with colour opponent mechanisms (Hurvich & Jameson, 1957; DeValois, Smith, Kitai, & Karoly, 1958) and colour naming (Werner & Wooten, 1979). MDS in this case revealed the existence of a representational system for which there is independent evidence. In general, however, MDS is used as an exploratory technique to bootstrap the process of theorizing about mental representations. Richards and Koenderink (1995, p.1323) recently commented that "...texture space, unlike color space, has been extremely resistant to study" and agree that MDS-like scaling techniques may provide useful insights into the nature of texture space (although they prefer an approach different from traditional MDS).

Recently, Rao and Lohse (1996) used MDS in an effort to develop a naming system for visual textures. Such a naming system would be useful for organizing and conveying graphical information (Ware & Knight, 1992). Theoretically, texture naming data might connect to the computations underlying texture perception in the same way colour naming data connect to the opponent theory of colour. Rao and Lohse (1996) had subjects arrange 56 of the Brodatz (1966) textures[1] into groups according to their perceived similarity. From these groups they calculated a similarity measure for each of the 1512 pairings of the 56 stimuli. These similarities (averaged over subjects) were submitted to a non-metric MDS analysis (Kruskal & Wish, 1978) and a three-dimensional solution was accepted. The positions of the stimuli on the MDS solution axes were then related to verbal descriptions of the stimuli that subjects had provided through responses on Lickert scales (see Rao & Lohse, 1996, Figure 9).

Heaps and Handel (1999) conducted experiments similar to those of Rao and Lohse (1996) using natural textures. One of their main conclusions was that perceived similarity may be context dependent and hence the search for a canonical

[1] Richards and Koenderink (1995) also examined the perceptual space of a subset of the Brodatz textures. Their objective was to evaluate their trajectory mapping algorithm as a viable alternative to MDS.

set of dimensions that describe perceptual texture space may be futile. Heaps and Handel make the reasonable point that natural textures afford many bases for similarity judgments. For example, two "visually" similar textures might be judged as dissimilar if an observer's judgments are based on semantic class. Conversely, two visually dissimilar textures might be judged as similar if they are seen as exemplars of the same semantic class. Tactile interpretations (soft, smooth, rough, hard, etc.) of recognizable surfaces (e.g., silk, wood, gravel, marble, etc.) might also compete with visual factors in determining the nature of the similarity judgments that subjects make. It might be argued that these difficulties are due in large part to the use of natural textures for which semantic and material interpretations are available. If the objective is to understand the visual coding mechanisms underlying texture perception then fewer problems of the sort just described might be expected when artificial textures are employed as stimuli.

Several years ago, Harvey and Gervais (1981) used MDS to study the perceived similarities among 30 artificial textures Each texture comprised the same seven, non-harmonically related, vertical sine-wave gratings in cosine phase. The stimuli differed only in the amplitudes of the sinusoidal components which were chosen at random and scaled so that they produced images having the same Michelson contrast $[(L_{max} - L_{min})/(L_{max} + L_{min})]$. In two different experiments Harvey and Gervais (1981) collected similarity measures for each of the $(30^2 - 30)/2 = 435$ pairings of the 30 textures. The similarity judgments were submitted to two MDS analyses (MDSCAL in one case and INDSCAL in a second) both of which revealed that the thirty textures could be arranged in perceptual spaces of three dimensions. That is, the textures could be arranged as points in a three dimensional Euclidian space such that the distances between them were highly negatively correlated with their perceived similarity; textures eliciting high similarity scores were located close to each other in the MDS solution space and textures eliciting low similarity scores were far apart in the MDS solution space.

A critical question concerns the relationship between the positions of the textures in the 3D, MDS solution space and the physical description of the stimuli given by the amplitudes of their sinusoidal components. Harvey and Gervais (1981) modelled the internal representations of their textures using the four channel model of Wilson and Bergen (1979); i.e., each texture elicited responses in four, spatial frequency selective channels. Regression analyses were then performed to find the linear combinations of the filter outputs that best matched the positions of the textures on each of the recovered MDS dimensions. This analysis showed that a high percentage of the variability on the first two dimensions of the MDS solution could be explained by a weighted sum of the activities in the four spatial frequency channels. Therefore, the modelled internal representations of the textures were given by linear combinations of the four filter outputs. A final step in the process,

which we describe below, would be to compare the calculated distances between the modelled representations of textures with the raw similarity scores.

In recent work on texture perception there is an emerging dichotomy between so-called high-level (Roa &Lohse, 1996) or attentive (Grossberg & Williamson, 1999; Heaps & Handel, 1999) texture analysis and low-level or preattentive texture analysis (e.g., Harvey & Gervais, 1981; Landy & Bergen, 1991). The present work takes the latter point of view although we acknowledge that the concepts of high-level vs low-level, or attentive vs preattentive texture analysis may be debated. A more neutral position that obviates debates of this sort focuses on the nature of the computations that lead to particular judgments. We ask what biologically plausible transformation takes stimuli, described in physical terms, into the perceptual space that is revealed by MDS.

The purpose of the present study is three-fold. First, the Harvey and Gervais study is one of rather few to address specifically the internal representation of visual textures (cf., Harvey & Gervais, 1978; Roa & Lohse, 1996; Richards & Koenderink, 1995; Heaps & Handel, 1999). Past studies have tended to focus on texture segmentation (Beck, 1982; Gurnsey & Browse, 1989; Gurnsey & Laundry, 1992; Julesz, 1981; Landy & Bergen, 1991 Malik & Perona, 1990; Rubenstein & Sagi, 1990; Voorhees & Poggio, 1988). Studies of segmentation typically focus on the mechanisms that limit the discriminability of two spatially adjacent textures (e.g., Gurnsey & Browse, 1987). The results of such studies are often interpreted in terms of mechanisms that respond to discontinuities within neural images. Fewer studies have examined the perceived similarity (or dissimilarity) of spatially (or temporally) separated textures (Harvey & Gervais, 1978, 1983; Rao & Lohse, 1996; Richards & Koenderink, 1995; Heaps & Handel, 1999). Therefore, it is important to examine the issue of texture representation in contexts other than the texture segmentation task.

Second, the Harvey and Gervais (1981) study provides a very interesting framework within which to advance our understanding of the internal representation of textures. We wish to reexamine their study to determine if their results can be replicated and whether they generalize to different stimuli having the same kind of spatial frequency structure but which are, at the same time, somewhat more in line with the intuitive notion of texture. Whereas Harvey and Gervais used textures comprising seven vertical sine waves, our textures comprise six, narrow bands of 2D noise. As well, Harvey and Gervais sampled the seven dimensional stimulus space randomly, whereas we used a more systematic sampling strategy.

Third, to calculate the response of each channel of the Wilson and Bergen (1979) model to a given stimulus, Harvey and Gervais weighted the amplitude of each sine

wave by the filter's transfer function and summed the results, rather than by applying the filter directly to the image and measuring its energy output. Although this method of analysis may be appropriate for stimuli comprising relatively few sine waves, it is not straightforward for more general stimuli. An important component of the present study is to determine whether the responses of filters applied the images themselves give rise to similar results.

EXPERIMENT 1

METHOD

Subjects

Three subjects with normal or corrected to normal vision participated as subjects. The subjects included the two authors and a third subject who was naive to the purpose of the experiment.

Apparatus.

All aspects of stimulus presentation and data collection were under the control of a Macintosh PowerPC 7100/180 equipped with a 17 inch multi-scan colour monitor. The monitor was set to have a screen resolution of 640 by 480 pixels and the colour lookup table was calibrated to be linear.

Stimuli

Twenty textures were created for use in the experiment. Each texture was created by adding together three band-pass images. Each of the band-pass images resulted from filtering a single image comprising Gaussian noise. The noise was filtered with six isotropic, narrowband filters having non-harmonically related centre frequencies; specifically, 4, 6.52, 10.68, 17.44, 28.48 and 46.52 cycles per patch or 0.53, 0.86, 1.41, 2.30, 3.76 and 6.14 cycles per degree. The six bandpass images were equated for energy; i.e., sum of squared amplitudes. Each bandpass image was then windowed with a circular window, smoothed along the edges.

Taking six items (i.e., bandpass images) three at a time results in 20 combinations. Therefore, twenty different textures were created each comprising three narrow frequency bands. These are shown in Figure 1. A principal components analysis of the frequency components composing the stimuli yielded 5 orthogonal dimensions each explaining 20% of the variability within the set. Therefore, in purely physical terms, five dimensions are required to explain the variability among the textures. All stimuli were normalized by the minimum and

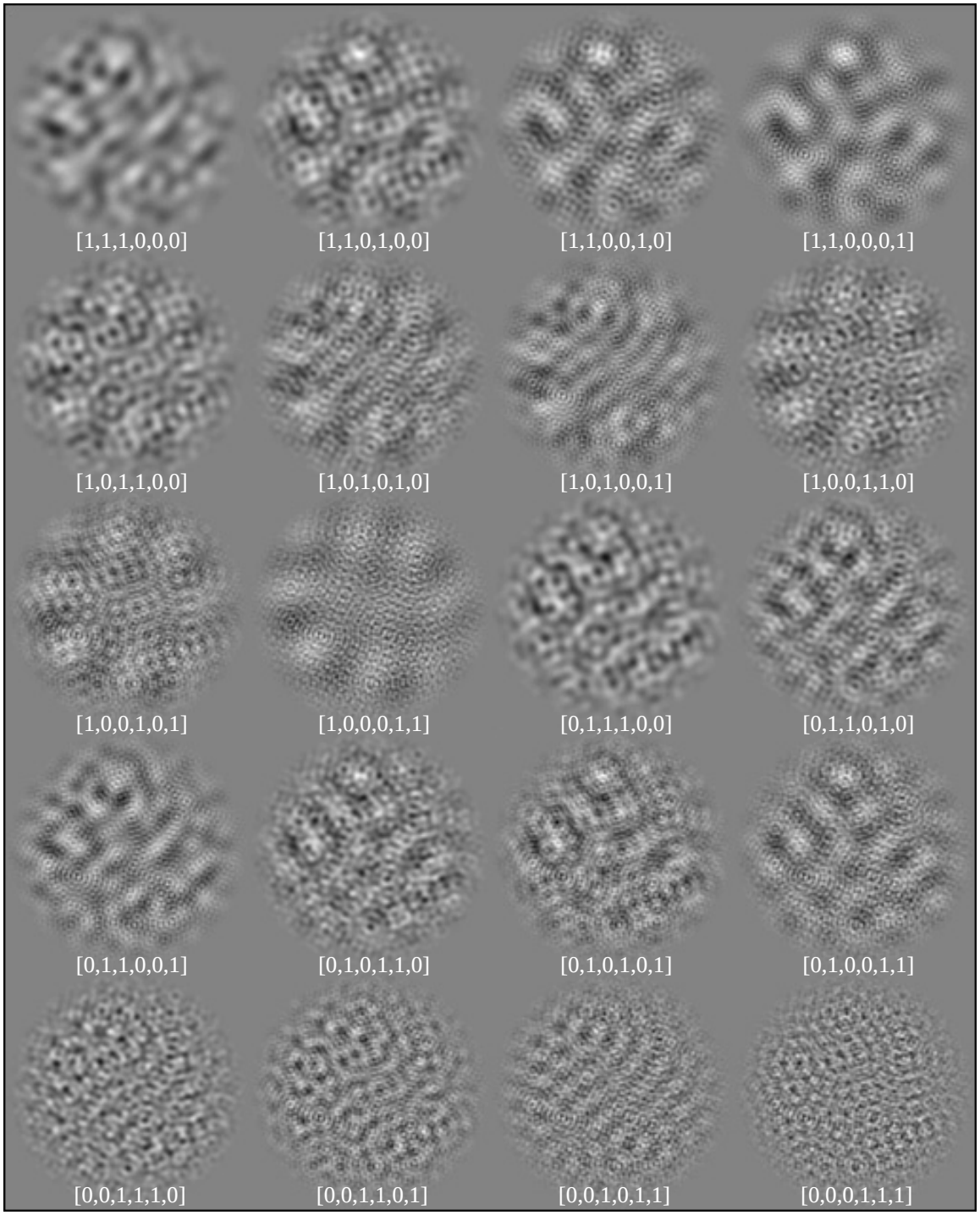

Figure 1. Examples of the twenty stimuli used in the experiment. The six-tuple under each patch indicates the components that compose the patch. The left most element of the six-tuple represents the lowest frequency component and the right-most the highest frequency component. A 1 indicates the component is present and a 0 indicates that the component is absent.

maximum intensities over the entire set of 20 images. Maximum and minimum screen intensities were 87.0 cd/m^2 and 0.4 cd/m^2 respectively and the average Michelson contrast was 81% [although it is not clear that Michelson contrast is a meaningful measure of the contrast in complex images (Peli, 1990)].

Procedure

The procedure followed the method of triads. Three textures were presented on each trial. At a viewing distance of 57 cm, each patch subtended 7.5° visual angle. The three patches were centred on an imaginary circle also having a radius of 7.5° visual angle. Subjects were asked to indicate which two patches appeared the most similar and which two were the least similar. By pressing a predetermined key a small black bar could be moved to connect different pairs of textures. When the bar connected the two most similar textures the "s" key was pressed and when it connected the two least similar textures the "d" key was pressed. All possible triples of textures were presented (1140 triplets) so that each pair of textures appeared 18 times. Each time two textures were judged most similar a counter for that pair (which had been initialized to 0) was incremented by 2. When a pair was judged least similar the counter remained unchanged and for the remaining pair the counter was incremented by 1. Given that each pair occurred 18 times in the course of the experiment the maximum possible similarity score was 2 * 18 = 36 and the minimum score was 0. Trials were run in blocks of 100 and the whole experiment took about two hours to complete.

RESULTS

The similarity judgments (henceforth, *the data*) showed good inter-subject agreement. Each subject produced $(20^2 - 20)/2 = 190$ similarity scores. The squared correlation coefficients (r^2) between subjects' similarity judgments were .682 for subjects RG and FP, .70 for DF and FP and .778 for DF and RG. An obvious first question is whether the data can be explained in terms of the simple correlation (r) between the binary six-tuples representing the stimuli (see the six-tuples associated with each texture patch in Figure 1). There are 190 such correlation-coefficients and 190 similarity scores. For RG, DF and FP the correlation-coefficients explained 25%, 25%, and 27% of the variability in the similarities scores respectively. Therefore, the simple correlations between between pairs of binary vectors do not explain the data. A similar analysis was conducted for cross correlations between the actual texture patches. For all subjects the cross correlations between the stimuli accounted for less than 1% of the variability in the similarity scores.

To understand better the structure of each subject's data, each similarity matrix was submitted to a non-metric MDS analysis using Kruskal's method (SYSTAT, v5.2) and solutions with 1 to 5 dimensions were obtained. As the number of dimensions in the solution increased from 1 to 5, the average stress values decreased (0.235, 0.143, 0.073, 0.049, 0.027) and the average explained variability (r^2) increased (0.65, 0.79, 0.89, 0.92, 0.95). The three dimensional solutions were selected for further analysis because they accounted for 89% of the variance in the original similarity

matrices (on average) and the addition of further dimensions did not improve upon this greatly.

Several aspects of the MDS solutions deserve comment before moving on to further analyses. Table 1 shows the correlation coefficients between all possible pairs of MDS solution vectors (three vectors for each of the three subjects) obtained in the present experiment. The absolute values of nine coefficients are very high (mean = 0.94, in bold text) and 27 are close to zero (mean = 0.08, in plain text). The first column (RG_1) in table 1 indicates that the positions of the twenty textures on the first dimension of RG's MDS solution correlate very highly with the positions of the twenty textures on the first dimension of DF's and FP's MDS solutions. The fourth column of table 1 (DF_1) indicates that the positions of the twenty textures on the first dimension of DF's MDS solution correlates very highly with the positions of the twenty textures on the first dimension of FP's MDS solution. These results indicate that for all three subjects the first dimension of their MDS solutions order the stimuli identically. For DF and FP the positions of the the stimuli on the second and third dimensions (columns DF_2 and DF_3) are also highly correlated, indicating that for these two subjects the second and third dimensions of their MDS solutions order the stimuli identically. (Negative correlations indicate that the *relative* orders of stimuli are very high but their absolute order are reversed). The situation is somewhat different for RG. The positions of the stimuli on his second dimension correlate highly with the DF and FP's third dimension and RG's third dimension correlates highly with DF and FP's second dimension. From these observations it is clear that the three dimensional solutions for the three subjects relate in essentially the same way to the the stimuli but the dimensions do not present themselves in the same order in the three solutions. In subsequent discussion and analysis we flip the order of RG's second and third dimensions to make his solution congruent with those of DF and FP. As well, the signs of the dimensions have been flipped where

	RG_1	RG_2	RG_3	DF_1	DF_2	DF_3	FP_1	FP_2	FP_3
RG_1									
RG_2	-0.04								
RG_3	0.02	0.06							
DF_1	**0.98**	-0.22	-0.06						
DF_2	0.13	0.15	**0.82**	0.06					
DF_3	-0.22	**-0.90**	0.06	-0.05	0.10				
FP_1	**0.97**	-0.17	-0.01	**0.98**	-0.01	-0.16			
FP_2	-0.09	-0.12	**-0.94**	-0.01	**-0.96**	-0.08	0.01		
FP_3	-0.10	**-0.97**	0.02	0.07	-0.05	**0.96**	0.00	0.02	

Table 1. Correlations between the three dimensions of the three subjects solutions.

necessary to make the solutions congruent. These alterations of the MDS solutions do not affect distances in the solutions and hence do not affect the fit of the MDS solutions to the data[2].

The three dimensional solutions (modified as just described) are presented as stereograms in Figure 2. In each stereo pair of Figure 2 there is a coherence to the texture space such that neighbouring texture patches appear more similar than remote texture patches. The dimension depicted on the x-axis (left to right) seems to distinguish stimuli containing predominantly high frequencies from those containing predominantly low frequencies. This might correspond to a verbal label having something to do with coarseness. However, one would be hard pressed to provide verbal characterizations of the other two dimensions; y- and z-axes. Thus, although there is a visual coherence to the three texture spaces this does not seem to correspond directly to a set of verbal labels (cf. Rao & Loshe, 1996).

ANALYSES

What transformation of the stimuli produces the three dimensional perceptual spaces revealed by the MDS analyses? We begin with the possibility that, for each subject, a simple linear transformation of the six dimensional "binary amplitude space" takes it into the three dimensional perceptual space revealed by MDS. If such a transformation exists we may ask if there is a simple and systematic relationship between the transformations derived for each subject. The method of analysis developed to explore these two questions will also be applied to a more plausible internal representation of the stimuli; specifically, the outputs of frequency selective filters.

Let $\mathbf{a}_j$ be a six element column vector of zeros and ones describing the frequency components of the jth stimulus. Let $\mathbf{A}$ be a matrix whose jth row is $\mathbf{a}_j'$, where $\mathbf{a}_j'$ is the transpose of $\mathbf{a}_j$. Let $\mathbf{y}_k$ (for k {1, 2, 3}) be vectors representing the coordinate locations of the 20 stimuli along each of the three MDS solution axes. We wish to determine the best linear combination of the six dimensional stimulus vectors, $\mathbf{a}_j$, that predicts their positions along the three MDS axes, $\mathbf{y}_k$. For the kth coordinate axis, this amounts to finding the vector $\mathbf{x}_k$ that minimizes:

$$|| \mathbf{y}_k - \mathbf{A}\,\mathbf{x}_k ||^2 \qquad [1]$$

[2] MDS solutions have no inherently "correct" orientation because it is only the distances between points in the solution that are relevant to the fit to the data. Thus, it may seem odd that the solutions for all three subjects require no rotation to bring them into alignment. In part this may be because for all three subjects the first dimension accounts for a very large proportion of the variance in the data and hence tends to anchor the solutions.

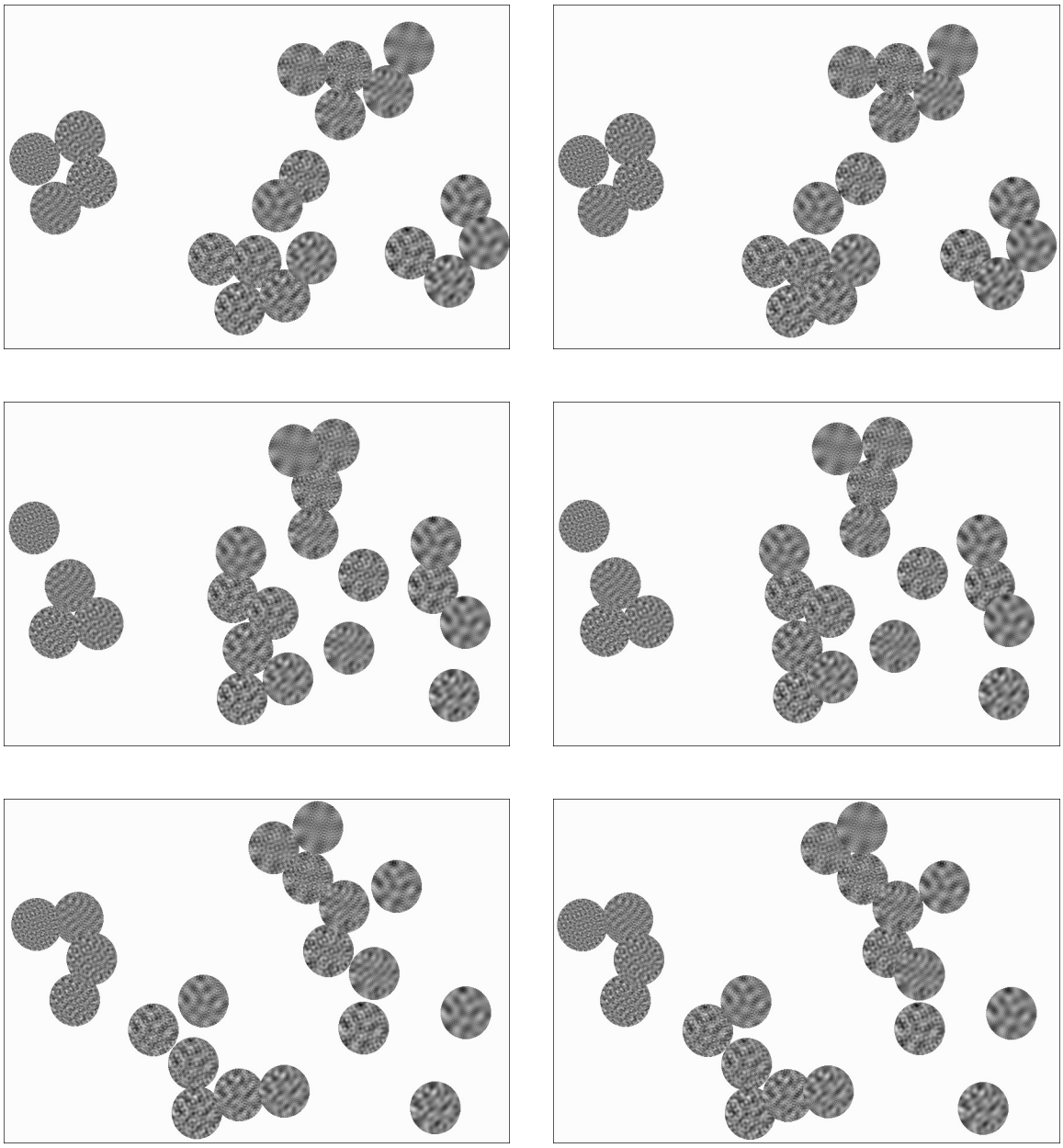

Figure 2. Stereo-pairs depicting the three-dimensional, MDS solutions for each of the three subjects (top, RG; middle, DF; bottom, FP). These three dimensional solutions accounted for 90% of the variability in the original similarity matrices. Please note that these depictions deviate slightly from the actual solutions. This was done to ensure minimal overlap between patches that are close together in the solution space.

For notational convenience we let $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3]$ and let $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3]$. We can now solve for the columns of $\mathbf{X}$ simultaneously by minimizing

$$||\mathbf{Y} - \mathbf{A}\,\mathbf{X}||^2. \qquad [2]$$

Let $\widehat{\mathbf{X}}_S$ denote the least squares estimate of the transformation from stimuli to MDS coordinate positions for subject *s*. Accordingly, let $\widehat{\mathbf{Y}}_S = \mathbf{A}\widehat{\mathbf{X}}_S$ denote the predicted

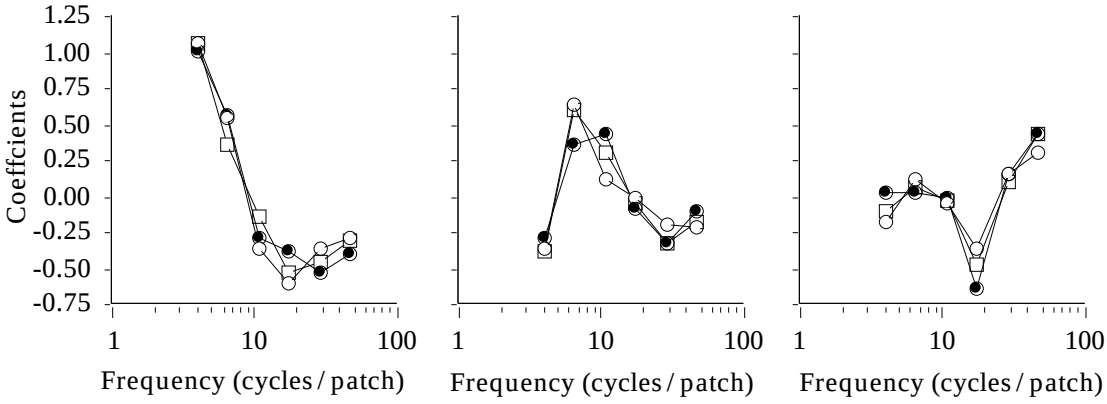

Figure 3. A depiction of the coefficients that map textures to from their frequency components to their internal representations. Each panel shows the coefficients for three subjects. The inner product of the coefficients and the frequency components in a texture map that texture to its position on one of the MDS solution axes. DF open circles, RG filled circles, FP open squares.

MDS coordinates for subject *s* given this transformation.

Figure 3 provides a visualization of the columns of $\hat{\mathbf{X}}_S$ that map the spatial frequency components of each stimulus to its position on the three solution axes, for each of the three subjects. There is remarkable agreement in the form of the coefficients. For the left panel the averaged r^2 between coefficient vectors is 0.96, for the centre panel it is 0.89, and for the right panel it is 0.83. The left panel shows that predicted position on the first dimension of the MDS solution can be obtained from a weighted sum of the frequency components that pits high frequencies against low frequencies. That is, a stimulus will map to one end of the first dimension if it contains predominantly low frequencies and to the other end if it contains predominantly high frequencies. The centre panel shows that the second dimension tends to contrast the second and third frequency components in the stimulus (6.54 and 10.68 cycles per patch) with the remaining frequency components, although almost zero weight is given to the fourth component (17.44 cycles per patch). The right panel shows that the third dimension contrasts the fourth component with the fifth and sixth components.

Using $\hat{\mathbf{Y}}_S$ as a model of the internal representation of the textures for subject *s*, the distance between all texture pairs within the model can be computed and compared with the original similarity judgments. Correlations between model distances and similarities yielded r^2s of 0.75, 0.75, and 0.73 for subjects RG, DF and FP respectively, thus accounting for an average of 74% of the variability in the original similarity scores. Because of the high negative correlation between the model distances and similarity scores, the models may be taken as reasonable hypotheses about the internal representation of textures in the current sample. The models fall short, however, of the MDS solutions themselves, which accounted for an average of 89% of the variability in the original similarity scores. On the other hand, the

models correlate as well with the similarity scores as subjects correlate with each other. Thus, each MDS solution may involve a certain amount of idiosyncratic variance that is not well captured by this version of the simple linear model.

The preceding analysis represented each stimulus as a binary code that indicated which frequency components were present in the stimulus (see Figure 1). If the visual system consisted of a large number of very narrowly tuned filters, then the preceding analysis might indicate how the outputs of those filters are combined to define the perceptual space containing the textures that we have examined. However, it is clear that the visual system employs a relatively small set of broadly tuned filters rather than many tightly tuned filters.

At this point, we turn our attention to a more realistic computational account of the similarity data. In particular, we ask if the energy responses of band-pass filters can be mapped to the 3D MDS solutions. Energy responses were computed by convolving each stimulus with a filter then taking the square root of the sum of squared responses. The filters were band-pass, with Gaussian amplitude spectra on a log frequency axis, i.e.,

$$f(d) = e^{\frac{-[\log(cf) - \log(d)]^2}{\sigma}} \qquad [4]$$

where d is distance from the origin of the 2D Fourier transform, $cf = N/\omega$ is the centre frequency of the filter (N is the size in pixels of the window containing the texture patch and ω is the wavelength in pixels) and σ is the bandwidth of the filter. These are lognormal filters with a constant octave bandwidth; on a linear frequency axis the bandwidth increases with frequency. Obviously our ability to fit the data will be affected by the number of available filters, but the number of filters required is unclear. Therefore, eight filters banks were created. The filters in each bank had "centre-wavelengths" ranging from 2^{min} to 2^{max} pixels ($min = 1$, $max = 5$). The step size from min to max was s^{-1} with s in [0.5, 1.0, 1.5, ... , 4]. The bandwidths of all filters in a bank were set to $\sigma = s^{-1}$; that is, as the number of filters increased their bandwidths narrowed proportionately.

In each analysis, **A** (see equation 1) was a 20 by n array of energy responses; $n = s * (max - min) + 1$. The analysis was conducted exactly as with the binary coded frequencies except that each row of **A** corresponded to energy responses from a set of n filters. Three related questions were addressed. First, is it possible to combine the energy responses to account for a significant amount of the variability in the original similarity data. This can be answered by determining the correlation, and hence r^2, between the original similarities and the model distances. (Recall that the r^2s for the models shown in Figure 3 averaged 0.74.) Second, are the

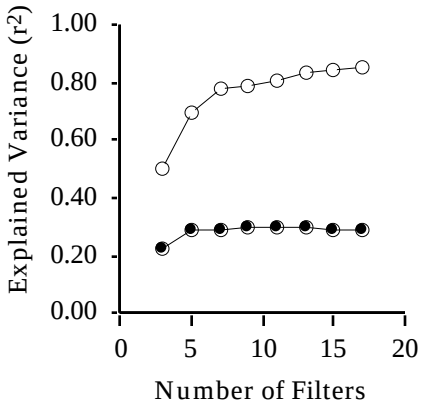

Figure 4. Summary of changes in average explained variance (squared correlation between subject similarity judgments and some function of the filter outputs) as a function of the number of bandpass filters. Open circles depict variance explained by the model (averaged over three subjects). Closed circles depict variance explained by the cross-correlation of filter responses.

transformations that account for the original data similar for each of the three subjects. Recall that for the coefficients in Figure 2, the averaged r^2 values were 0.96, 0.83, and 0.89 for panels 1 to 3 respectively. Third, are there simple characterizations of these transformations as there appear to be for the coefficients shown in Figure 3? The answer to this question requires visual inspection.

Figure 4 plots the average agreement (r^2) between the original similarity data and model distances as a function of the number of filters in the filter bank. Agreement improves as the number of filters increases and begins to asymptote when there are seven filters in the bank. With seven filters the average r^2 was 0.78 and for 17 filters it was 0.85. These correlations represent improvements over the 74% explained variance when the analysis was applied to binary coded frequency components. These results are very encouraging because they indicate that a simple linear combination of energy responses provides an excellent account of the original similarity data. That is, the original similarity judgments can be almost entirely explained by the pattern of activity that the textures elicit within a set of neural images.

Figure 5 depicts the coefficients that map energy responses (in this case, 9 filters) to each of the MDS solution axes. (Although we have shown just one instance of the obtained coefficients, the characteristics that we discuss are independent of the number of filters employed.) In answer to the second question, the coefficients show little similarity across subjects in any of the three panels. In answer to the third question, there is no simple characterization of these transformations as there was for the coefficients shown in Figure 3. Therefore, the good news is that these coefficients do as good a job of explaining the the original similarity data as do the MDS solutions themselves. The bad news is that there is nothing pretty about the

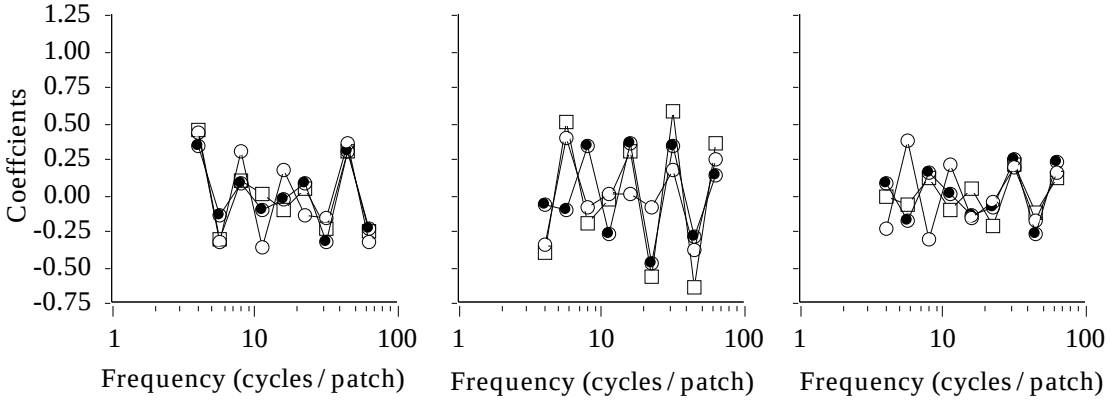

Figure 5. A depiction of the coefficients that map textures to from 9 responses to their positions on the MDS solution axes. Each panel shows the coefficients for three subjects. DF open circles, RG filled circles, FP open squares.

coefficients.

The simplicity and regularity of the coefficients in Figure 3 provided a promising picture of how the spatial frequency content of the texture patches may relate to their perceptual similarities. These coefficients might be seen as analogous to the opponent colour mechanisms underlying colour appearance and colour naming. This simplicity and regularity is lost however, when textures are described more realistically in terms of energy responses (see Figure 5). The most inelegant feature of the coefficients in Figure 5 is their high frequency oscillations from positive to negative. It is particularly odd that these oscillations produce a high degree of *correlation* in the coefficients that map energy responses to the *uncorrelated* MDS solution axes (see Table 1). These correlated oscillations suggest that the coefficients are performing, in part, a function that is independent of the particular dimension of the MDS solution to which they map energy responses.

One possibility is that the coefficients shown in Figure 5 represent a transformation that both decorrelates filter responses and explains the structure of perceptual texture similarities as arranged in MDS space. Because the filters are broadly tuned, they overlap in frequency space, and hence their responses contain uninformative sources of correlated variation. Such unwanted sources of variance can be eliminated by a linear transformation of the filter responses that decorrelates their outputs while retaining that structure in the responses owing to the stimuli. This kind of transformation whitens the filter responses to uncorrelated white noise.

To create an appropriate decorrelator, we first find the correlation matrix, **C**, of the responses. In particular, the components of **C** are given by

$$\mathbf{C}_{ij} = \mathbf{F}_i' * \mathbf{F}_j \quad [5]$$

where $\mathbf{F}_i$ and $\mathbf{F}_j$ denote vectors containing the coefficients of the discrete-time Fourier transforms of the filters' impulse responses. If **v** is a vector of filter responses resulting from the application of the filters to white noise, then $\mathbf{C} = E[\mathbf{v}\ \mathbf{v}']$ is their correlation matrix, where E[.] denotes mathematical expectation. To decorrelate the filter responses one need only apply a linear transformation given by $\mathbf{Q}^{-1}$, where $\mathbf{C} = \mathbf{Q}\ \mathbf{Q}^{-1}$ is called the Cholesky decomposition of **C**. One can show that if **v** is mean zero with correlation matrix **C**, then $\mathbf{Q}^{-1}\ \mathbf{v}$ has a correlation matrix equal to the identity matrix. The resulting transformation looks very much like a local form of linear inhibition between filters with nearby frequency tuning.

Proceeding with the analysis as above, we were interested in the structure of the transformation that mapped the decorrelated filter responses on to the MDS coordinate positions. As above, this is obtained as a least squares solution to

$$||\ \mathbf{Y} - \mathbf{A}\ \mathbf{Q}^{-1}\ \mathbf{X}\ ||^2. \quad [6]$$

Figure 6 shows the coefficients of $\hat{\mathbf{X}}_S$ obtained using equation 6 for each of the three subjects. In this case **A** represents the responses from a bank of nine filters. The obtained coefficients explain an average of 80% of the variability in the original similarity data, which is an improvement over the model shown in Figure 3. The most important point to note is that the coefficients in Figure 6 have lost their high frequency oscillations and are now quite similar in form to those in Figure 3. As in Figure 3 the coefficients for each of the three subjects are very similar and they have the same simple structure. We may conclude that the coefficients in Figure 5 combine two processes, one that decorrelates filter responses and another that maps

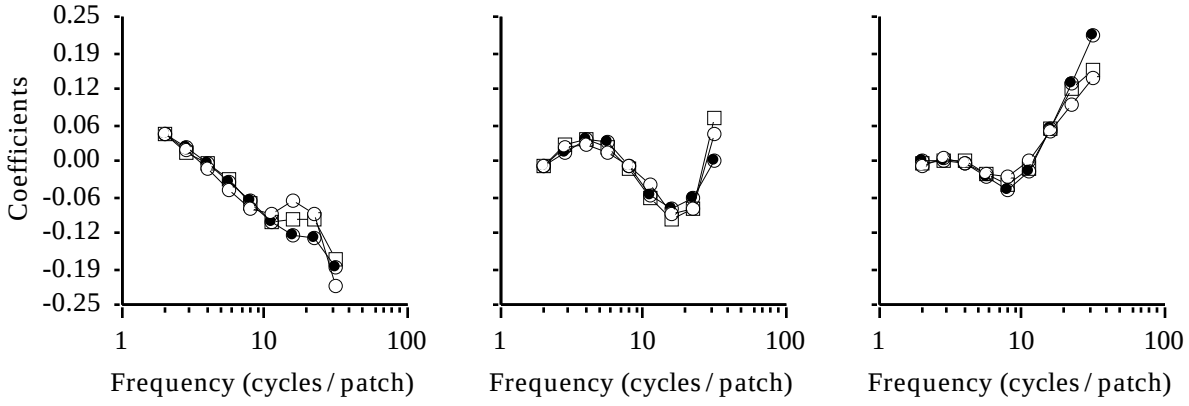

Figure 6. A depiction of the coefficients that map textures to from 9 decorrelated filter responses to their positions on the MDS solution axes. Each panel shows the coefficients for three subjects. DF open circles, RG filled circles, FP open squares.

the unique contribution of each filter to the MDS solution axes.

We have chosen to show the coefficients that map nine decorrelated filter responses in Figure 6 in part because they correspond approximately to the asymptotic level of explained variance (as shown in Figure 4) and partly because more than nine filters becomes biologically implausible. If five or seven filters are employed then coefficients explain less of the original variance in the similarity scores but they have the same general form shown in Figure 6. If eleven to seventeen filters are employed then more variance is explained in the similarity scores. In these cases, when $\widehat{\mathbf{X}}_S$ is derived from the raw energy responses the oscillations are extremely pronounced and when $\widehat{\mathbf{X}}_S$ is derived from the decorrelated energy responses the oscillations are removed (as in Figure 6) but the pattern of coefficients become increasingly dominated by the responses of high frequency filters. In this case much more weight is given to the high frequency filters thus flattening the remaining coefficients. One reason for this is that there is relatively low gain on the high frequency filters so that more weight must be given to them to map to the MDS axes. The coefficients can be made more regular by normalizing the filter responses on a per stimulus basis, and by changing filter bandwidths. These operations reduce the high frequency coefficients but the patterns of coefficients become idiosyncratic as the number of filters increases.

DISCUSSION

In the present experiment subjects' similarity judgments are highly correlated and their 3D MDS solutions (which were very similar) explained most of the variability in their data. If texture were not encoded by some basic visual process then we might have expected subjects' judgments to be idiosyncratic and their MDS solutions to be unrelated. In fact, subjects seem to rely on similar internal representations to judge the similarity of texture patches and this situation is a precondition to searching for simple architectures that are candidates for this internal representation. The finding that three dimensions are sufficient to explain the majority of variability in the data agrees with the results of Harvey and Gervais (1981) even though different stimuli and a different sampling of the stimulus space were employed. Our computational analysis extends that of Harvey and Gervais and shows that a sequence of biologically plausible transformations provides a coherent account of how subjects judge the similarity of textures under conditions in which edge-based strategies are impossible (Gurnsey & Laundry, 1992; Wolfson & Landy, 1995; Graham, 1991) and no obvious verbal labels are available (cf. Rao & Lohse, 1996; Heaps & Handle, 1999). We have concluded that a linear transformation of decorrelated filter responses provides a reasonable hypothesis about the basis of texture space. We next address the plausibility of this proposal.

Number of filters. Figure 4 (unfilled circles) shows that increasing the number of neural images provides the conditions for increasingly better fits to the similarity data. Although increasing the number of filters might seem to be an obvious way to improve the fit, such an expectation requires accepting the assumption that filters are the basis for similarity judgments in the first place. Figure 5 shows that model fits tend to asymptote at about 7 to 9 filters. This is more filters than are usually assumed to subserve spatial vision (Wilson, McFarlane & Phillips, 1983) but does not seem to be an unreasonable number. Thus we may conclude that seven to nine filters provide a reasonable basis for the present texture space.

Energy. Recently, Heeger and Bergen (1995; Bergen, 1994) have suggested the distribution of responses within neural images characterize the appearance of many textures. They have provided impressive illustrations that textures are often indistinguishable when forced to have the same distribution of responses in a set of neural images. In some sense the two textures are metameric (Richards, 1979). Thus in many cases the first order statistics (in Julesz's sense) within neural images determine the appearance of textures. This might be seen as an alternative to the present proposal where energy responses are taken as the determinants of perceived similarity. However, we will show that the two proposals are not different because energy is the only important feature of the distribution of responses within a neural image.

The band pass filters used here are zero-mean and therefore the expected response within all neural images will also be zero. Energy is the sum of squared responses within a neural image and is therefore a measure of the variance within the neural image. Each panel of Figure 7 shows the distribution of responses within one neural image for each of the 20 stimuli in Figure 1. The mean (μ) and standard deviation (σ) within a neural image were computed and all responses exceeding $\pm 3\sigma$ were eliminated. Of course μ was close to 0 in all cases and σ depended on the match of the filter to the frequency content of the display. Filter responses within $\pm 3\sigma$ of the mean were then normalized to integer values in the range -31 to 31 and frequency counts made. The effect of these trivial computations is to normalize the distributions with respect to energy. In other words, energy differences have been factored out of the response distributions. Figure 7 shows that when energy differences are factored out the distributions become essentially independent of filter and image. In fact, a scaled Gaussian distribution with a μ of 0 and σ of 10 explains 99% of the variability in the means of each of the 9 panels of figure 6. We conclude that the energy within a neural image is probably the most important feature of the response distribution and hence a reasonable statistic upon which to build and account of texture space.

Interpreting the Regression Coefficients. If one accepts that energy responses in

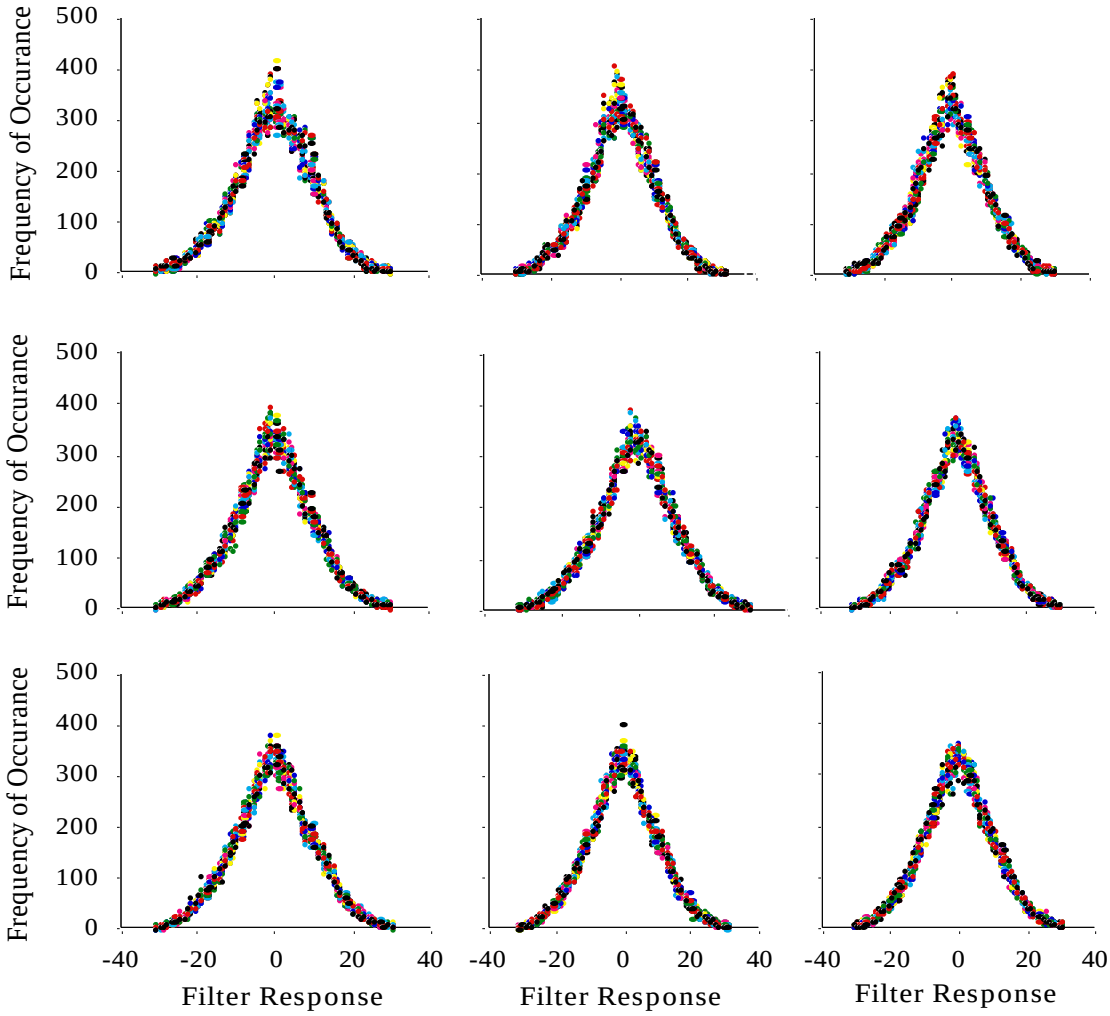

Figure 7. Normalized response distributions. Each panel plots the distribution of responses of a single filter applied to each of the twenty images shown in Figure 1. Only responses within three standard deviations of the mean have been retained. Those responses remaining have been normalized to the range -31 to 31, thus factoring out energy differences between the distributions. As a result, the response distributions are now independent of filter and image.

seven to nine bandpass filters provides a reasonable foundation on which to build a theory of texture perception, one still needs to justify why a linear combinations of energy responses should be the determinant of texture perception. The main response is that both the MDS solutions and the regression coefficients define spatial representations of texture and do an excellent job of explaining subjects' similarity judgments. The issue here is that many factors that can influence the pattern of coefficients obtained in the regression analysis. These factors include the nonlinearity used in computing "energy," the number of filters, the filter bandwidths, the filter sensitivities and whether or not the filter responses are decorrelated. There are undoubtedly other factors that would affect the coefficients. Therefore, while one may accept that the linear model is useful and appropriate, any particular model is not unique and independent arguments must be made to support it.

The present results indicate that the most important dimension (dimension 1; see Figures 3 and 6) in the MDS solutions contrast high vs low frequency components in the stimulus. This result is in agreement with the results of Harvey and Gervais (1981). This contrast also appears to emerge in the Rao and Lohse (1996) paper; their third dimension seems to pick out a coarse vs fine distinction between stimuli. All of these data are consistent with the intuition that the central tendency in the frequency content of textures should have a strong influence on perceived similarity. In this sense, the coefficients in Figures 3 and 6 have some empirical and intuitive support. The third dimension in the present solution appears to correspond to the second dimension in the Harvey and Gervais (1981, Figure 3) study in that it contrast mid-range frequencies against high and low frequencies (although the correspondence is not exact and analytical methods are different). As mentioned earlier, it is difficult to establish a correspondence between the second and third dimensions of our solutions with the solutions of Rao and Lohse (1996) and Heaps and Handel (1999) because in these latter cases the dimensions in question were derived from verbal labels that did not constrain the present results[3].

In addition to finding some empirical and intuitive support, we can bootstrap support for the current regression weights through consistency arguments. Figure 3 shows that there is strong agreement across all three subjects about the mapping from the spatial frequency content of the stimuli to their positions in the MDS solution space. The obtained regression coefficients define a functional or abstract correspondence between the physical descriptions of the stimuli and their psychological structure as revealed by MDS. When the responses to the stimuli are more realistic (i.e., the outputs of bandpass filters) we find that the coefficients that map from response space to MDS space loose their coherence (agreement between subjects). If we attach importance to the pattern of weights found in the “ideal case” (i.e., Figure 3) then these may act as a reference point or constraint when making choices about the structure of the model. The use of energy and a decorrelation matrix are guided by this idea. That is, these choices lead to a mapping function (Figure 6) that is in good agreement with the “idealized” situation described in Figure 3.

Fixed Architectures The present conceptualization of texture space is in many ways similar to ideas about colour perception. Although there may be cognitive or interpretational influences on colour perception most theorizing about colour is couched in terms of fixed transformations of the responses three initial filters (e.g.,

[3] It would be interesting to relate the frequency content of the natural textures used by Roa and Lohse (1996) and Heaps and Handel (1999) to the MDS solutions they obtained. It is conceivable that the frequency content of the stimuli in question provides a more coherent explanation of their MDS solutions than do groups of verbal labels.

Wandell, 1995). Our proposal shares much with this latter view. A number of interesting question arise when drawing this comparison. The first is whether our analysis has revealed *the* fixed architecture underlying texture space. Our conclusions must be tempered by a problem that confronts all MDS studies, namely, we have sampled only a small region of the space of visual textures. If the region of texture space sampled been different our solutions may have changed. Therefore, further experiments must be undertaken that expand the domain of sampled textures (e.g., manipulating orientation, phase and range of spatial frequencies) to determine if our current space remains a coherent sub space of a broadened sampling of texture. A positive finding would suggest that we have indeed uncovered a fixed component in the architecture of the visual system.

There are, however, certain results that would be at once consistent with the current model yet require some modifications to it. For example, it is likely that changing viewing distance--and hence the entire spatial frequency range covered by our stimuli--would have no effect on similarity data, MDS solutions or regression analyses. Such a result would be inconsistent with the current proposal because the same pattern of weights would be required for many shifts up and down the frequency axis. On the other hand, this result could be dealt with by a constancy mechanism that applies a fixed *pattern* of weights, normalized to the entire range of frequencies presented in the stimulus set.

Finally, a comment is in order concerning the relationship between the components of the present model and theories of colour. Wandell (1995) suggests that wavelength opponent mechanisms may reflect mechanisms that decorrelate the responses of the three wavelength selective cone types which have overlapping sensitivity functions. The decorrelation matrix $\mathbf{Q}^{-1}$ (eqn. 6) may be seen as performing an analogous function in the case of texture. However, whereas the decorrelated filter output may provide a basis for perceptual colour space, this is not so for texture because the decorrelated filter responses do not map directly to the dimensions of the MDS solution. Rather, we suggest that following the process of decorrelation, there is a linear combination of decorrelated filter responses that maps to perceptual texture space.

REFERENCES

Beck (1982). Texture segmentation. in Beck, J. (Ed.), Perceptual organization and representation. Hillsdale, NJ: Lawrence Erlbaum.

Bergen, J.R. (1994). Texture perception: filters, non-linearities and statistics. Investigative Ophthalmology and Visual Science, 35(4), p.1477.

Bergen, J.R. & Landy, M.S. (1991). Computational modeling of visual texture segregation, in Computational Models of Early Visual Processing, Eds. Michael S. Landy and J. Anthony Movshon. Cambridge MA: MIT Press.

Brodatz, P. (1966). Textures: A photographic album for artists and designers. New York: Dover.

DeValois, R.L., Smith, C.J., Kitai, S.T., & Karoly, A.J. (1958). Responses of single cells in different layers of the primate lateral geniculate nucleus to monochromatic light. Science, 127, 238-239.

Ekman, G. (1954). Dimensions or color vision. Journal of Psychology, 38, 467-474

Graham, N. (1991). Complex Channels, early local nonlinearities and normalization in texture segregation. in Computational Models of Early Visual Processing, Eds. Michael S. Landy and J. Anthony Movshon. Cambridge MA: MIT Press.

Grossberg, S. & Williamson, J.R. (1999). A self-organizing neural system for learning to recognize textured scenes. Vision Research. 39,1385-1406.

Gurnsey, R., Pearson, P., & Day, D. (1996). Texture discrimination along the horizontal meridian: effects of magnification, frequency content and micropattern orientation. Journal of Experimental Psychology: Human Perception and Performance, 22, 738-757.

Gurnsey, R. & Browse, R.A. (1989). Asymmetries in visual texture discrimination, Spatial Vision, 4, 31-44.

Gurnsey, R. & Laundry, D. S .(1992). Texture discrimination with and without abrupt texture gradients. Canadian Journal of Psychology, 46, 306-332.

Harvey, L.O. & Gervais, M.J., (1978). Visual texture perception and Fourier analysis. Perception & Psychophysics, 24, 534-542.

Harvey, L.O. & Gervais, M.J., (1981). Internal representation of visual texture as the basis for the judgment of similarity. Journal of Experimental Psychology: Human Perception and Performance, 7, 741-753.

Heaps, C. & Handel, S. (1999). Similarity and Features of natural textures. Journal of Experimental Psychology:Human Perception and Performance, 25, 299-320.

Heeger, D.J. & Bergen,J.R. (1995). Pyramid Based Texture Analysis/Synthesis. Computer Graphics Proceedings, p. 229-238, 1995.

Hurvich, L. & Jameson, D. (1957). An opponent-process theory of color vision. Psychological Review, 64, 384-404.

Kruskal, J. & Wish, M. (1978). Multidimensional Scaling. (Sage University Paper series on Quantitative Applications in the Social Sciences, 07-11) London: Sage Publications.

Landy, M.S. & Bergen, J.R. (1991). Texture segregation and orientation gradient. Vision Research, 41. 679-691.

Malik, J. & Perona, P. (1990). Preattentive texture discrimination with early vision mechanisms. Journal of the Optical Society of America A7, 923-932.

Peli, E. (1990). Contrast in complex images. Journal of the Optical Society of America A 7, 2032-2039.

Rao, A.R. & Lohse, G.L. (1996). Towards a texture naming system: identifying relevant dimensions of texture. Vision Research, 36, 1649-1669.

Richards, W. (1979). Quantifying sensory channels: generalizing colorimetry to orientation and texture, touch and tones. Sensory Processes, 3, 207-229.

Richards, W. & Koenderink, J. (1995). Trajectory mapping: a new nonmetric scaling technique. Perception, 24, 1315-1331.

Robson, J.G. (1980). Neural Images: The physiological basis of spatial vision. In C.S. Harris (Ed.). Visual Coding and Adaptability. Hillsdale NJ: Erlbaum. (1980)

Rubenstein, B.S. & Sagi, D. (1990). Spatial variability as a limiting factor in texture-discrimination tasks: implications for performance asymmetries. Journal of Optical Society of America A. 7, 1632-1643.

Shepard, R. (1962). The analysis of proximities: multidimensional scaling with an unknown distance function.II. Psychometrika, 27, 219-246.

Voorhees H, & Poggio T. (1988), Computing texture boundaries from images. Nature, 333:364-367.

Wandell, B.A. (1995). Foundations of Vision. Sunderland MA: Sinauer Associates, Inc.

Ware, C. & Knight, W. (1992). Orderable dimensions of visual texture for data display: orientation, size and contrast. In Bauersfeld, P., Bennett, J, & Lynch, G. (Eds) ACM conference on human factors in computing systems (CHI) 203-206.

Werner, J.S. & Wooten, B.R. (1979). Opponent chromatic mechanisms: Relation to photopigments and hue naming. Journal of the Optical Society of America, 69, 422-434.

Wilson, H.R & Bergen, J.R. (1979). A four mechanism model of spatial vision, Vision Research, 19, 19-32.

Wilson, H.R., McFarlane, D.K., & Phillips, G.C. (1983). Spatial frequency tuning or orientation selective units estimated by oblique masking. Vision Research, 23, 873-882.

Wolfson, S.S. & Landy, M.S. (1998). Examining edge- and region-based texture analysis mechanisms. Vision Research, 38, 439-446.