Some Observations on the Effects of Slant and Texture Type on Slant-From-Texture

Pedro Rosas[1,*], Felix A. Wichmann[1,2], & Johan Wagemans[1]

[1]Department of Psychology, University of Leuven,
Tiensestraat 102, B-3000 Leuven

[2]Max Planck Institute for Biological Cybernetics,
Spemannstr. 38, D-72076 Tübingen

[*]corresponding author: Pedro Rosas, email: pedro.rosas@psy.kuleuven.ac.be

Keywords: Texture, slant, slant-from-texture.

Abstract

We measure the performance of five subjects in a 2-AFC slant-discrimination task for differently textured planes. As textures we used uniform lattices, randomly displaced lattices, circles (polka dots), Voronoi tessellations, plaids, 1/f noise, "coherent" noise and a leopard skin-like texture. Our results show: (1) Improving performance with larger slants for all textures, (2) and some cases of "non-symmetrical" performance around a particular orientation. (3) For orientations sufficiently slanted, the different textures do not elicit major differences in performance, (4) while for orientations closer to the vertical plane there are marked differences among them. (5) These differences allow a rank-order of textures to be formed according to their "helpfulness"– that is, how easy the discrimination task is when a particular texture is mapped on the plane. Polka dots tend to allow the best slant discrimination performance, noise patterns the worst. Two additional experiments were conducted to test the generality of the obtained rank-order. First, the tilt of the planes was rotated by ninety degrees. Second, the task was changed to a slant report task via probe adjustment. The results of both control experiments confirmed the texture rank-order previously obtained. We then test a number of spatial-frequency-based slant-from-texture models and discuss their shortcomings in explaining our rank-order. Finally, we comment on the importance of these results for depth perception research in general, and in particular the implications our results have for studies of cue combination (sensor fusion) using texture as one of the cues involved.

1 Introduction

That we are able to perform visually guided tasks in space such as grasping objects, demonstrates that information about the 3–dimensional (3D) layout of the world must be obtainable from vision alone. This field of study is frequently denoted as depth perception and it is assumed that our depth percepts are based on multiple sources of information, in psychophysics often labelled cues, such as disparity, motion parallax, or occlusion, to name but a few. Another example of such a depth cue is texture. Gibson prominently stressed that this cue might be a powerful source of information about (3D) surface layout. He thought of texture "as the structure of a surface, as distinguished from the structure of the substance underlying the surface" (Gibson, 1979, p. 25). He studied the contribution of texture to depth using a slant perception task in which subjects reported the perceived orientation of a slanted plane (without depicting the edges of the surface) with a palm board (Gibson, 1950b). In this setting subjects systematically underestimated slant, and this effect was stronger for what he called "irregular" textures. Since then, the slanted-plane perception task, and variations of it, has become a widely used paradigm for studying the contribution of texture to depth perception (e.g. Cutting & Millard, 1984; Knill, 1998a; Passmore & Johnston, 1995; Rosenholtz & Malik, 1997).

In contrast to other cues, a critical difficulty arises when studying the texture cue to depth: Whilst it is possible to establish a direct parametric relation with a physical variable for some other depth cues such as disparity, motion–in–depth or motion parallax, this is not true for texture. Partly this is because the very concept of "texture" itself is not clear, or at least, elusive to define (Tuceryan & Jain, 1998). According to Gibson the contribution of texture to depth perception is through "texture gradients" (Gibson, 1950a), the amount of variation of a certain parameter in the visual field. On a textured surface receding in depth the amount of variation (the gradient) from point to point in the texture is directly related to the distance from the viewer. The visual system is presumed to perceive depth by detecting this variation. Gibson mentions the "gradient of density of texture", associated to a change in the retinal image of a textured longitudinal surface from coarse to fine. Other gradients, subsequently elaborated on by Purdy

(1958), are gradients of size, gradients of compression (also known as foreshortening), and gradients of convergence (or linear perspective). Gibson's and most current research's use of texture implies that textures are composed of a homogeneous distribution of basic texture elements, frequently termed texels, on the surface. Depth perception is thought to be accomplished by isolating the texels and computing their deformation on different parts of the surface as a measure of the gradients. Cutting and Millard (1984), for example, study the perception of "surface flatness" -and also "curvature"- with stimuli depicting linear perspective, density and compression gradients using regular and irregular octagons as texels to build regular and irregular textures. By independently manipulating these gradients in the octagons present in their stimuli, they conclude that perspective is the most important cue for flatness judgements, whereas compression is most influential for curvature judgements. Similarly, Todd and Akerstrom (1987) manipulated rectangular elements as texels in a shape-from-texture study, finding that the compression gradient is not sufficient to account for the perception of 3D shape. More recently, Knill (1998a) compared human discrimination of slant with ideal observers constructed by modelling three types of gradients: position (related to density), scaling and foreshortening. He concluded that foreshortening is most important for such task. In addition, he reported that discrimination performance improves with increasing slant, and that only marginal differences in performance exist between two types of textures (ellipses and Voronoi tessellations). One should note, however, that the latter statement is based not only on the comparison of only two textures, but also on a single slant only (65 degrees).

Other researchers characterize the textures gradients without explicit identification of texels, which is not trivial for a wide range of textures (Gårding & Lindeberg, 1996; Malik & Rosenholtz, 1997; Li & Zaidi, 2000). Bajcsy and Lieberman (1976) suggested a spatial-frequency-based analysis of slant-from-texture by measuring the change in "Fourier (texture) descriptors" across a "window" on images of outdoor scenes. Malik and Rosenholtz (1994, 1997) developed a shape-from-texture algorithm that measures differences between a pair of nearby image patches in the Fourier domain and called this measure affine texture distortions. They argue that this measure contains enough information to estimate not only slant but also surface curvature. Rosenholtz and Malik (1997) advocate this as a model of human depth-

perception. A final example is the model by Sakai and Finkel (1997) where humans are thought to be assessing slant by monitoring a simpler characterization of the spectrum of the image.

A different approach to slant-from-texture can be found in Witkin (1981), who instead of the homogeneity assumption underlying Gibson's approach, assumes a "directional isotropy" for textures. In this scheme any directional bias in the image is assumed to stem from the orientation of the plane and is consequently used to estimate slant. Thus within Witkin's framework a single patch of texture is sufficient to estimate slant via its deviation from isotropy, whereas slant estimation following the homogeneity assumption relies on comparing the differences between two patches of texture in different parts of the visual field. Note, however, that Witkin's approach does not work for "directional textures" i.e. anisotropic textures (Malik & Rosenholtz, 1997). Blake and Marinos (1990) extended Witkin's approach using an iterative algorithm to extract violations of the isotropy assumption. Experimentally, Rosenholtz and Malik (1997) did not, however, find evidence to reject the use of either Witkin's isotropy assumption or the homogeneity assumption in slant perception.

The previous discussion points to the lack of an obvious parametric description of the texture cue. This situation not only affects the psychophysical research on depth perception, but also the machine vision field in which the choice of a representation of texture is critical for the success of shape-from-texture algorithms (Super & Bovik, 1995). Here, we present a study that does not assume a particular description of the texture cue. Instead, we took an exploratory approach by testing a wide number of types of textures in the slant perception paradigm. By taking such variety of texture instances and a broad range of slants, we observed large differences in human slant-from-texture discrimination. In the future, these data can be used in two ways. First, it can help in the aforementioned debate about texture "descriptors" as it provides a database of psychophysical results for a large number of textures, both with and without obvious texels. Second, it provides a texture performance-database for further depth-related experiments in cue combination research circumventing the problem of the lack of an explicit parametrical texture manipulation. We will return to this idea in the general discussion.

2 Texture Synthesis

2.1 Overview

For the present study we used a number of textures typically used in the psychophysical literature and some examples from the computer graphics field, such that they potentially could span very different accuracies in depth perception tasks. We did not use natural images because of the difficulty producing arbitrarily slanted samples: In natural images other potential sources of information, not traditionally considered as "texture", are present and entangled with texture itself: shading and specularity, for example, which in the literature are taken as cues themselves. The presence of these cues in some textures illustrates the problem of trying to understand texture in isolation from understanding surface characteristics. Koenderink and van Doorn (1996), for instance, distinguish different forms of shading. The term "smooth shading" is reserved for shading due to the global shape of an object, whereas "illuminance texture" is associated with the surface characteristics. Thus, simple mapping techniques for creating slanted versions of a natural sample are insufficient. In the words of Dana, Van Ginneken, Nayar, and Koenderink (1999, p. 15): "2D texture mapping (...) cannot account for the variations in texture appearance due to local shading, foreshortening, shadowing, occlusion and interreflections". In principle, it may be possible to use a texture synthesis algorithm for replicating a natural sample in different slants but in our opinion the currently available models do not offer completely satisfactory results. A possible solution is, of course, not to perform a mapping at all but to capture natural images at the required slants. Indeed, Dana et al. (1999) built one such database, which is publicly available. Unfortunately, a pilot experiment with these images showed that the size of samples and number of slant levels available was insufficient for our requirements. Therefore, we restricted ourselves to synthetic textures, albeit including some rather natural looking ones (e.g. fractal noise, Perlin noise, leopard-skin).

2.2 Texture Family Descriptions

In the following, we describe the instances of textures used, organizing them in "families" according to the algorithm employed for synthesizing them.

2.2.1 Randomly Displaced Lattice[1].

A plane filled with uniformly placed dots –a lattice- is the simplest and most regular texture of our set. Uniformity here means that the distance between each pair of dots on the plane is the same; therefore, the "cycles of change in brightness" are uniform and the texture is "regular" in the Gibsonian sense (Gibson, 1950b). This regularity makes lattices an ideal stimulus to study perceptual grouping (see Kubovy & Wagemans, 1995; Kubovy, Holcombe, & Wagemans, 1998).

The uniform lattice (dots p0) is computed by dividing a plane in a number of equally separated imaginary columns and rows. At the intersections of these lines a "point" or centroid is placed. Around these centroids a small disc is drawn (4 pixels radius, subtending approximately 0.14 degrees of visual angle), with an inter-centroid distance of 27 pixels (approximately 0.97 degrees of visual angle). We also created less regular versions of the lattice by randomly displacing the evenly placed centroids in the horizontal and vertical direction. We used a normally distributed pseudo-random number generator with zero mean and variance set to either five (dots p5) or ten pixels (dots p10)[2]. This type of procedure was used by Wichmann and Henning (1998) to create semi-irregular stimuli whose global statistics (density, number of discs, mean luminance etc.) are, on average, equal to those of the uniform lattice (dots p0).

---

[1] We borrow a term denoting a similar process implemented in the software Stochastic Geometry (Stoyan, 1992).

[2] The same algorithm could be used to generate "grids" with different displacements by connecting the centroids with lines, and a "checkerboard" pattern -and its displaced versions- by filling the polygons defined by the grid with two different pixel values.

2.2.2 Hard Core Point Process Textures.

A Poisson point process is one where the points are placed randomly and without any spatial variation of the process (Stoyan, Kendall, & Mecke, 1995). A more constrained version of this method is denominated as a "hard core" random point process, when the centroids are initially randomly placed, as in case of the Poisson process, but "forbidden to lie closer together than a certain minimum distance"(Stoyan et al., 1995, p. 162). That is, it is possible to specify a radius of inhibition around each centroid in which no other centroid is positioned. Rosenholtz and Malik (1997) used this procedure to create Voronoi lattices. (Note that other algorithms exist to create Voronoi textures; for example, Knill, 1998b, used one based on a reaction-diffusion model.)

For our textures, we follow the hard core process and created a set of centroids on a plane of 4250x4250 pixels with an increasing inhibition process using the Stochastic Geometry software (Stoyan, 1992), settling on an inhibition radius of approximately 26 pixels. Approximately 3500 centroids were positioned on the surface, and they were used for two different textures. First, a Voronoi tessellation, defined as the set of polygons around the centroids such that all the points inside one polygon lie closer to the respective centroid than to other centroids. Second, by filling circles of 24 pixels radius around the centroids, a polka dot-like pattern (with non-overlapping circles) was obtained[3].

2.2.3 Sinusoidal Plaids.

A number of textures can be constructed by adding sine gratings with different frequencies and orientations. This type of stimuli is typically used in low-level vision studies, but Li and Zaidi (2000) used this type of texture in a series of studies linking the spectral composition of textures and the veridical perception of corrugated surfaces. As an example of this texture family

---

[3] It is straightforward to define a texture formed by ellipses in the same manner. In this case, the major axis of each ellipse must be smaller than the radius of inhibition to avoid overlapping. In addition, the orientation of the ellipses can be manipulated to obtain an isotropic texture (random orientations) or an anisotropic one (setting an orientation bias).

a plaid was created by adding two 2-D sinusoidal functions with the same frequency (approximately 9.4 cps/0.86 cpd) oriented orthogonally to each other.

2.2.4 1 Over F Noise or Natural Noise.

An image whose power spectrum as a function of frequency follows $1/f^{\beta}$ with β between 2 and 4 is sometimes termed a fractal image (Knill, Field, & Kersten, 1990). It has been empirically shown that natural images tend to have a power spectral function that approximately follows such power law with β equal to 2 (Field, 1987). To create a texture with a power spectrum similar to that of natural images we thus generated a random noise texture with 1/f amplitude spectrum[4]. This was done by first filling a plane with "white noise" (normally distributed). The white noise was then "coloured" in the Fourier domain by scaling its amplitude spectra with a 1/f shaped cone. Finally, the texture sample was obtained by computing the inverse Fourier transform.

2.2.5 Perlin Noise or Coherent Noise.

Perlin (1985) introduced an algorithm for creating what he termed "coherent noise". In Perlin's terminology coherent means that the pixel values on the image change smoothly from one point to another, not showing the discontinuities typically obtained from purely random noise processes. Perlin's algorithm takes the positions of the points in space into account during the noise generation, that is, the smoothness does not result from filtering white or binary noise. Different parameterizations of the algorithm create diverse flat textures. We used a parameterization yielding a cloudy-looking texture which in the following we call Perlin noise.

2.2.6 Diffusion.

Turk (1991) introduced a texture synthesis algorithm using the model by Alan Turing for the propagation of chemicals in morphogenesis (Turing, 1952). The algorithm describes biological pattern formations by means of a system of coupled differential equations modeling the diffusion of chemicals at certain rates and the reactions between them. Different textures can

---

[4] A power spectrum of $1/f^2$ corresponds to an amplitude spectrum 1/f.

be generated by manipulating the parameters of the model, such as the diffusion rates and the initial concentration of chemicals. In addition to Turing's basic algorithm, Turk introduced a "cascade process" in which one chemical is "frozen" after some diffusion has taken place, but other processes continue thereafter. Using such a process for a system of two chemicals, we generated a leopard-skin-like texture.

2.3 Texture Mapping

In order to prevent subjects from using a particular feature of an instance of one of our textures to do the task, we generated seven instances of each texture type. (In the case of those textures without a random component by construction, uniform lattice and plaid, a random displacement and a change of phase were introduced, respectively, to create different instances of those patterns.)

Using the texture mapping algorithm by Heckbert (1989) we mapped the texture patches onto planes with inclination between 5 and 80 degrees slant, in single degree steps. The rendering was done using perspective projection for a viewing distance of 60 cm with respect to the center of the image (i.e., the viewing distance during experiments). The final experimental stimulus was a piece of 300x300 pixels cut from the central part of the mapped image, which corresponded to roughly 11.3x11.3 cm on the screen. We used a Gaussian kernel to avoid (or reduce) aliasing as much as possible but without excessively blurring the textures. The amount of filtering (number of times the kernel was applied) was determined ad hoc for each texture. For example, no filtering was applied to 1/f noise textures, while the Voronoi lattice required anti-aliasing treatment because of the "jaggies" which would otherwise form on the lines being projected onto a plane. Figure 1 depicts examples of all the textures used in our study.

3 Experiment 1: Slant discrimination

3.1 Experimental Setup

The experimental setup was carefully designed to avoid, as much as possible, a cue conflict situation arising due to the physical flatness of the screen: The monitor, a Sony GDM-F500R, was located behind a black wooden plate with a circular aperture of 10 cm diameter, subtending approximately 10 deg of visual angle at the subjects' eye. The monitor was completely covered with the exception of the viewing aperture. A viewing tube of approximately 53 cm length and 31.5 cm diameter was set on the front of the wooden plate. At the viewing tube's end opposite to the monitor, a head and chin rest was located, i.e. roughly 60 cm away from the monitor. This device was aligned such that the subject's opened eye (monocular view) was looking roughly at the centre of the viewing aperture. The setup itself was inside an area enclosed by black curtains and subjects did not see the casing of the monitor, or the computer driving it, at any time during experimentation. The monitor was carefully calibrated to correct its nonlinear gamma function and CRT-induced geometric distortions, using a Photometrics SenSys digital camera employing the method described by Wichmann (1999). The computer driving the monitor and the experiment was a Macintosh PowerMac G4. The contrast and brightness settings of the display were adjusted such that maximum luminance level was 114 cd/m^2, and the minimum 7 cd/m^2.

3.2 Methods

Five subjects participated in standard temporal two-alternative-forced-choice (2AFC) experiments. Subjects were presented with textured planes at physically different levels of slant and they had to indicate which of the two images appeared more slanted in depth. Each texture was tested independently, that is, during a single trial both planes had the same texture. The order in which the subjects were tested with the different textures was different for each subject. All subjects had normal or corrected to normal vision. Two were completely naive to the purposes of the experiment, one had a general notion of the research and two (PR, JW) were authors.

At the beginning of each trial a red fixation cross appeared for 800 ms (the inter-trial interval) in the centre of the screen. The first-interval image's contrast was ramped-in (50ms) stayed at maximum contrast for 200ms, and then was ramped-out (50ms). During the inter-stimulus interval (400 ms) the screen was set to the mean luminance. Then the second-interval commenced using the same fade in-maximum contrast-fade out sequence as used for the first image (50ms + 200ms +50ms). Finally, subjects had to respond within a 750 ms response interval. No performance feedback was provided but tones marked the beginnings of the first and second interval[5].

The display of images, timing and answer collection was controlled using the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997).

Four slant levels were chosen as standards as a compromise between the total amount of trials to be collected and sampling of the possible range of slants sufficiently finely: 26 degrees, 37 degrees, 53 degrees and 66 degrees away from vertical, i.e. 66 degrees is very slanted (near horizontal), 26 degrees has little slant (near vertical).

A combination of adaptive and constant stimuli procedures was used to collect data. The adaptive procedure was used to obtain a crude first estimation of the psychometric function. From this estimate some critical stimuli values were extracted to carry out a constant stimuli procedure. Finally, psychometric functions were fit to the pooled data obtained from both procedures. For the adaptive procedure the implementation of QUEST (Watson & Pelli, 1983) in the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997) was employed. For each psychometric function, three interleaved QUEST procedures were run to estimate the 60, 82 and 90% correct levels. Two psychometric functions were estimated at each standard for each texture: one for smaller slants relative to the standard, and one for bigger slants. Thus, in total, 24 randomly interleaved QUEST procedures were run simultaneously for each texture (4 standards x 2

[5] The sounds were used to avoid possible confusions between the first and second image because of the fast pace of the experiment. One of two possible configurations was applied to each subject: low tune signalling first image and high tune for second image, or vice versa.

psychometric functions x 3 thresholds). At the beginning of each trial, one of these structures was randomly and independently selected to determine the next standard and comparison stimulus level. The total number of trials for the adaptive procedure part was determined by the stop criterion of the adaptive procedure, which we set to 50 trials for each threshold estimation; hence in total 1200 trials were run for each texture and each subject during the adaptive part of the experiment. Using the Psignifit Toolbox, which implements the constrained maximum-likelihood method described by Wichmann and Hill (2001a, 2001b), with a logistic function as underlying shape, four stimulus levels were estimated from the adaptive procedure data for each of the 24 functions: 63, 75, 90 and 95% correct. At these stimulus levels 50 additional trials were run using the method of constant stimuli[6]. Thus, the constant stimuli part comprised 1600 additional trials for each texture (4 standards x 2 psychometric functions x 4 stimuli levels x 50 trials). The final estimation of the psychometric functions reported in the results sections and shown in the figures, was made with the fused data collected in both procedures, normally representing 350 trials for each psychometric function (see note 6). All fits were done with the Psignifit Toolbox, and forced to cross chance performance (50% correct in 2AFC) at the slant level of the standard. To avoid bias resulting from level-independent lapses during the maximum-likelihood estimation of the psychometric functions, the lapse rate was not assumed to be fixed (and equal to 0), but rather an extra (constrained) parameter in the fit (for details see Wichmann & Hill, 2001a). As a result, the maximum performance, or upper asymptote, of the psychometric functions is not necessarily 1.0 but between 0.95 and 1.0, given the constraints used.

3.3 Results

Differences among psychometric functions reflect the differential difficulties among experimental conditions. In our particular case, the psychometric functions are associated with a certain texture and represent the probability of detecting which of two planes is more slanted.

[6] In some exceptional cases, when performance higher than 80% correct was unattainable, only three stimulus levels were tested in the constant stimuli part.

Thus, a shallow psychometric function for a particular texture indicates that with this texture slant discrimination is difficult. Or, expressed the other way around, the steepness of the psychometric function is associated with the "helpfulness" of the texture in slant-from-texture.

In Figure 2 the estimated psychometric functions for the textures tested are depicted for every subject. Performance is plotted as a function of the slant on linear coordinates. Different symbols and colours code the various textures (see figure legend). Each row is associated with the data collected from a single subject. On every row each subplot contains the psychometric functions around one standard, which is marked by the segmented line: the leftmost subplot corresponds to the comparisons with the least slanted standard (26 degrees), the second is associated with 37 degrees, then 53 and 66 degrees slant, respectively. The curves depict the estimated psychometric functions for discriminating the standard against planes with smaller slants (left side of the standard) and bigger slants (right side of the standard). Error bars were computed by a parametric bootstrap procedure (Wichmann & Hill, 2001b) and correspond to 68% confidence intervals (SD). The psychometric functions indicate that discrimination performance is not equal at every standard: The task was more difficult for the less slanted standard, yielding shallower psychometric functions for all textures. The difficulty decreased as the slant of the planes increased, reflected by steeper psychometric functions, though the change in performance is clearly not equal for all textures. We observe an asymmetry of the psychometric functions around the standards with steeper slopes to the right side of each standard. For large slants (66 degrees) the textures are extremely similar in their helpfulness for discrimination: the psychometric functions are virtually identical. However, differences become apparent as the standard gets closer to the fronto-parallel plane. For example, for subject BW at standard 66 degrees the discrimination is the same for Perlin noise (gray triangles pointing downwards) and circles (gray circles) as texture, while at 37 degrees and 26 degrees there are very marked differences between the two. We will further comment on this issue when discussing the texture rank order to be extracted from our data.

3.4 Statistical analysis: effects of slant and texture.

Here we use statistical techniques to assess the effects of slant and texture on the psychometric functions. In particular, we study how slant and texture type affect the slope of the psychometric functions at 75% correct performance.

We proceed as follows to test the putative parallelism of a set of psychometric functions, modifying the technique of Wichmann (1999): First, all the data (conditions) which should be compared to each other are fused and shifted, and the best common slope (at 75% correct) for the fused data is computed. Then, new psychometric functions are fit to the individual data sets fixing the slope (at 75% correct) of the resulting function to that of the common slope. The goodness-of-fit of such models is used as indicator of parallelism among the psychometric functions: A failure of the common-slope-model represents evidence against the parallel hypothesis, because a single slope does not explain the data (fitted psychometric functions). The goodness-of-fit will be assessed by two means: by computing the deviance of the model and by studying the correlation between the deviance residuals and the model predictions. The deviance, or log-likelihood ratio statistic, represents how a model deviates from a "full" or saturated model[7], larger values of it indicating a poor fit. To determine the critical region for interpreting a particular value of deviance, we generate the deviance distribution using Monte Carlo techniques and performing a single-sided test with 5% significant level (deviance takes only positive values). Analogously, we compare the observed correlation between the deviance residuals and the model predictions with the Monte Carlo-based estimated distribution for the correlation. A double-sided test with 5% significant level was performed in this case.

3.4.1 Effects of slant

[7] The "full" or saturated model is the extreme one containing as many parameters as data points, leaving no random component. Of course, its generalization ability is severely limited and, as a model, it is uninformative, but it is useful as a baseline for comparison.

In this case, the parallelism to test corresponds to the psychometric functions at different standards for a given texture: If there were no effect of slant, the slope would be constant across different standards. Thus, the data across the four standards for every texture is fused after shifting the stimulus axis such that the trials are distributed around an arbitrary standard level. The test indicates that for 40 out of the 41 cases corresponding to the 164 psychometric functions representing the functions at the "left side" of standards (smaller slants psychometric functions), the common slope refit of individual data sets failed. In the exceptional case (subject BW, texture 1/f noise), though, the common slope model showed a lack of fit to the fused data, rendering the common slope value questionable. For the bigger slants psychometric functions (right side of standards) in 35 out of 41 cases the common slope refit of individual data sets fail. The exceptions in this case are: subject HZ textures Voronoi (but also in this case the common slope model failed) and dots p5, subject JS texture 1/f noise (again, the common slope model failed), subject JW texture Perlin noise, subject PR textures 1/f noise and dots p0.

To summarize, parallelism across slants can be rejected in more than 90% of the conditions tested, i.e. there is a significant effect of slant on the steepness of the psychometric function.

In Figure 3 we illustrate two cases for the change in slope at different standards; the slope at 75% correct of bigger slants psychometric functions (the right side of the standard) is depicted against the standards tested. The common slope is indicated as a horizontal line, error bars correspond to 68% confidence intervals. The plot on the left (subject JW texture dots p0) shows a clear case of the slant affecting performance since the common slope value falls outside the confidence intervals of the four individual slopes. The case plotted on the right (subject HZ dots p0) shows a less evident change of slope: The common slope value falls within the 68% confidence intervals of three individual values, failing only for the slope at the least slanted plane[8].

[8] Indeed, if visual inspection were done with 95% confidence intervals for this type of cases, parallelism would not be rejected. However, refitting of the individual data sets with the slope set to the common value results in failure in most of these cases, reflecting the lower power of a test based on overlapping of confidence intervals.

Further, the asymmetry around the standards mentioned in section 3.3 can be assessed by testing the parallelism between the two psychometric functions for a particular texture around a standard: A symmetric pair would have the same (absolute) slope[9]. In this case, the refitting tests of the 164 pairs of psychometric functions result in 78 cases in which the slope at 75% correct of the psychometric functions at the right side of the standard (discrimination against more slanted planes) is larger than the slope of psychometric functions at the left side of the standard (discrimination against less slanted planes), 77 cases with no statistical difference and 9 cases with the opposite asymmetry (of these, 2 cases could be dismissed because of the bad fit due to poor performance). The 78 cases with the right-side-steeper asymmetry, representing 48% of the results, are consistent with the effect of slant on performance because the slope of the psychometric function is higher for comparisons of the standard against more slanted planes. They are distributed as follows: 18 cases around standard 26 degrees slant, 27 cases around standard 37 degrees slant, 23 cases around standard 53 degrees slant and 10 cases around standard 66 degrees slant.

3.4.2 Effects of texture

For studying the effects of texture on the slope of the psychometric function the hypothesis to test is the parallelism among psychometric functions at a certain standard for all textures: If there were no effect of texture, the slope would be the same across the textures tested.

In 37 cases of the 40 tests, the common slope model did not fit the data for all textures, indicating a significant texture effect on the slant-from-texture discrimination. The three exceptions were observed for large slants: two subjects (BW and JW) at standard 66 degrees slant for psychometric functions of smaller slants (left side of the standard) and one subject (JW) at standard 53 degrees for psychometric functions at the side of bigger slants. Of course, the failure of the refitted models does not mean that any subset of the textures at that standard does

[9] The stimulus axis for one set of data is flipped and then the data sets for left and right side are shifted such that they can be fused into one set.

not have a common slope, but that at least one texture differs significantly from the others. For example, taking the psychometric functions of smaller slants: At standard 66 degrees if the textures 1/f noise and leopard are not considered for subject HZ's data, the common slope model fits well the sets of the remaining patterns. The same occurs if textures circles and dotp0 are removed for subject JS's data, and textures 1/f noise, Perlin noise and circles for subject PR. However, at lower slants the influences of textures on discrimination performance is much more pronounced and as a consequence a subset with a common slope is smaller. For example, at standard 37 degrees for subject BW and JW the refitting with a common slope works only for one pair of textures (circles and dots p0 for BW and leopard and dots p5 for JW).

Figure 4 depicts an example corresponding to (absolute values of) slopes of smaller slants psychometric functions at 75% correct for one subject (PR) at two standards, with the horizontal line indicating the common slope computed for the fused data. On the left, for the standard 66 degrees slant the common slope value falls within the 68% confidence intervals for six out of eight texture's slope whereas on the right, for standard 37 degrees slant the psychometric function slopes are very different. For the case on the left a common-slope model can be refitted successfully up to a subset of five textures, while for the case on the right only one pair of textures with common slope can be extracted from these data (This is based on statistical analysis, not on visual inspection. See note 8).

3.5 Analysis: rank-order

Given the observed differences among textures in our slant-from-texture experiment it is possible to generate a rank-order of the textures according to their "helpfulness" in slant-from-texture. To compare the textures across the range of tested slants we need to quantify helpfulness, taking both the slopes for lower- and higher standards into account as well as the relative shifts of the psychometric functions ("thresholds"). To accomplish this we integrate the area between the lower- and higher psychometric functions between the two performance levels

60 and 80% correct[10]. The procedure is exemplified in Figure 5. Lower values of what we simply denote "area" indicate a more helpful texture.

In Table 1 the computed area values for all textures and subjects are indicated, including the 68% confidence limits extracted from the psychometric functions confidence intervals. To facilitate the comparison of these data among subjects, in Table 2 the values are normalized according to the maximum area value observed for each subject in all conditions of texture and slant.

If the helpfulness is rather constant for a certain texture in the range of slants, a plot with the computed areas is similar to a horizontal line, while a steeper curve indicates a worsening helpfulness. In Figure 6 the areas of the four textures eliciting a similar rank-order among subjects are plotted. The error bars are based on the 68% confidence intervals for the fit of each psychometric function (left side and right side) at the performance limits for the area. They are not displayed, though, when the performance was so poor that the estimated value to render performance 80% correct was lower than 0 degrees: texture 1/f noise at standard 26 degrees for subjects BW, HZ and JS, and at standard 37 degrees for subject JS; texture Perlin noise at 26 degrees for subject JW and JS.

The plots again show the effect of slant rendering textures less helpful as the area increases with orientations getting closer to the vertical plane. It is also possible to see differences among the textures at some standards: The most helpful texture being circles, characterized by the flattest area function across standards, then followed by leopard, whose area is similar to circles for very slanted planes but bigger at the other standards. The third texture in this ranking is Perlin noise and finally 1/f noise having in general the biggest area of the four textures (While circles is not the best texture for all the subjects, 1/f noise was the worst texture for all subjects). In the next two sections, we will report additional experiments conducted to test the generality of this rank-order.

---

[10] The performance levels for this computation were arbitrary chosen within a certain range. Note that we tried a number of different combinations of such limits without affecting the rank order reported here.

4 Experiment 2: slanted “wall” discrimination

The stimuli used in the previous experiment depicted a surface similar to a ground plane. Then, the improvement in slant discrimination as the slant increases could be explained by a specialized mechanism “tuned” for surfaces “to walk on” (Gibson, 1979). To test this, we tilted our stimuli by ninety degrees to depict a surface similar to a slanted wall whose farthest point was on the left side of the screen, and asked observers to discriminate slant. In this case, slant is the angle between the surface and a vertically oriented plane orthogonal to the line of sight. A major difference in performance would be evidence for a mechanism more sensitive to “floors” or “surfaces to walk on”.

4.1 Setup and methods

The same setup and methods were used as the previous experiment. The number of texture types and standards tested was reduced because we did not want to exhaustively characterize observer’s performance but only ascertain whether there were important differences to the results previously obtained. Two subjects (who had already participated in the previous experiment) were tested with stimuli depicting two textures, one classified as very helpful (circles) and the other very poor in helpfulness (Perlin noise). Two standards were tested in one case (26 and 66 degrees slant for subject JS), and three in the other (26, 53 and 66 degrees slant for subject PR). The subjects had to indicate which of the two images presented on every trial was the more slanted.

4.2 Results and Discussion

The previously observed pattern of results is also present in the data collected in this experiment: At very slanted planes the discrimination task is easier, performance is not symmetric around the standard, and psychometric functions for each texture have similar slopes at very slanted planes, while there are bigger differences among them for smaller slants.

Figure 7 shows a direct comparison of slant discrimination of ground-like surfaces and wall-like surfaces. We compared the results by inspecting the overlapping error bars at 75% correct of the psychometric functions. From these 20 comparisons (2 textures x 2 standards x 2 psychometric functions for subject JS and 2 textures x 3 standards x 2 psychometric functions for subject PR) 55% show equal performance in both experiments; 30% show better performance in the ground-like surfaces and 15% better performance in wall-like surfaces. These results, albeit they cannot entirely dismiss a specialized mechanism tuned to floor-like surfaces, indicate that such a mechanism only plays a minor, if any, role in the results obtained in our previous experiment.

5 Experiment 3: probe adjustment

To have another test of the generality of the discrimination-based rank-order of textures we used an adjustment task to collect the slant perceptions and build a rank-order. In this experiment, the subjects had to adjust a probe for reporting the perceived orientation of the plane.

5.1 Setup and Methods

The same physical setup was used as in previous experiments. An additional circular aperture of approximately 5 cm diameter was cut on the cardboard covering the monitor screen, below the stimulus aperture. Through this opening, the subjects could see the probe (a movable green line) plotted over a reference axis. If we imagine the screen lying on an X-Y plane (Z then being the depth), the probe represented the view in a Y-Z plane. Figure 8 depicts the situation. Subjects did not report any problems performing this viewpoint change to do the task. Four subjects that have participated in Experiment 1 participated in this new experiment.

The experiment was self-paced as the subjects could see the stimulus while adjusting the probe using the keyboard. The probe's movement was restricted to between 0 and 90 degrees

slant. After aligning the probe to the desired slant, the subjects pressed a key to report their setting and the next stimulus appeared. No feedback was provided.

The stimuli were the same as in Experiment 1, except that displaced versions of the dot lattice were discarded because in the discrimination task no systematic difference across all subjects was found, leaving seven textures for testing. Again, four slants were chosen as standards to collect subjects' estimates. Taking into account the poor performance in Experiment 1 at small slants, the selected orientations for testing were closer to the ground-plane: 37, 53, 66 and 73 degrees slant. Subjects had to complete 40 judgments of each target slant. All textures were randomly interleaved during the experiment.

5.2 Results and Discussion

Figure 9 shows the results for each subject for the four textures eliciting the rank-order extracted from Experiment 1: circles, leopard, Perlin noise and 1/f noise. On each subplot, filled symbols indicate the mean response of the 40 settings and bars indicate the sample standard deviation; the dotted lines indicate the "true" slants tested.

A possible measure of the reliability of a texture for this task is the inverse of the sample standard deviation. This is because the variability of the subjects' settings should be related to how reliably the texture conveys slant. However, the data for the noises (1/f noise and Perlin noise) have some peculiarity that troubles this criterion as a unit to find a rank order comparable to the helpfulness order previously found. For example, for subject JS the sample standard deviations for 1/f noise are rather small with the single exception of the most slanted plane, while the means indicate that JS mainly perceived a zero-slanted plane for those orientations. Then, a rank-order based on the sample standard deviation would classify this texture as helpful, although it is not depicting much slant until surfaces are close to the ground plane.

An alternative criterion is to build a rank-order on the correlation coefficient between the raw true orientations depicted and the raw reported judgments. In particular, such criterion would penalize a texture yielding smaller changes in the judgments than the true orientations (flatter judgments). A rank-order based on this rule is comparable to the rank-order based on the

discrimination task because a texture rendering constant judgments with the probe would also make the slant discrimination more difficult.

In Table 3 we show the correlation coefficients between the effective slants tested (37, 53, 66 and 73 degrees slant) and the judgments by means of the probe for the textures on the ordered set. The rank-order based on this task is coherent with the 2AFC-based ranking for three subjects, while for the fourth subject (JS) there is one case of reversal (Perlin noise better than leopard).

6 Models of Slant-From-Texture

In computer vision there are many types of texture analysis algorithms, statistical or geometrical in nature, as well as model-based methods (Tuceryan & Jain, 1998). Frequently they are attractive and elegant for the particular practical problems at hand in the computer science community, but, as we have already discussed in the introduction, they have difficulty as models of human texture analysis as they typically rely on extraction of texels, or "texture-base-units" which are typically absent in natural textures.

Thus here we concentrate on Fourier-based or spectral models of slant-from-texture. Such models are more attractive for our purposes because, first, they do not rely on texel or geometry extraction and should thus be, in principle, applicable to all types of textures, even to our stochastic ones. Second, starting with the seminal work of Campbell and Robson (1968), there is overwhelming evidence suggesting that the early visual system performs a (local) frequency analysis on the input image, some form of wavelet decomposition (for a comprehensive review see the book by Graham, 1989).

Several models exist in vision science and in computer vision that use a spectral-based characterization of textures. Some examples are: Turner (1986) for texture discrimination, Bergen and Landy (1991), Jain and Farrokhnia (1991) and Jain and Bhattacharjee (1992) for texture segmentation, and Greenspan (1996) for texture recognition. For shape from texture, Bajcsy and Lieberman (1976) first suggested to estimate slant-from-texture by local spatial-frequency-based analysis, and ever since then various algorithms have been proposed for this problem using the spectral components of texture (e.g. Clerc & Mallat, 2002; Kanatani, 1984; Krumm & Shafer, 1992; Malik & Rosenholtz, 1997; Ribeiro & Hancock, 2000a, 2000b; Sakai & Finkel, 1995, 1997; Super & Bovik, 1995; Turner, Bajscy, & Gerstein, 1989; Turner, Gerstein, & Bajcsy, 1991; Turner, Salganicoff, Gerstein, & Bajcsy, 1989).

Turner, Gerstein, and Bajcsy (1991) explicitly address the slant underestimation effect reported by Gibson (1950b), especially observed for what Gibson called "irregular" textures. In

this model sets of Gabor filters in quadrature are used to extract the "Gabor-energy" at different positions of the image. Turner et al (1991) proposed two slant estimation algorithms using the two-dimensional distributions of filters outputs (Gabor-energy). In the first case the slant is estimated by minimizing an error measure between the actual output of the filter bank and the predicted output for a texture with known frequency components (Turner, Bajcsy, & Gerstein, 1989). Given that this is very unlikely as a strategy for the human visual system as it would need to know the power spectrum of the (non-slanted) texture, Turner proposed a second method in which the output of the filter bank is compared to "templates" stored in the system, representing different slants (Turner, Salganicoff, Gerstein, & Bajcsy, 1989). Even if we are willing to accept the presence of such templates in the human visual system, it is not clear how to derive quantitative or even qualitative predictions from Turner's models. Turner argues in general terms that textures whose spectra contain isolated peaks would lead human subjects to perceive more slant than textures with more broadband frequency components: "Although Gibson described the textures for which less accuracy was attained as 'irregular', from the standpoint of the model presented in this paper, the textures are better termed 'broad spectrum' or 'multipeaked spectrum' " (Turner et al., 1991, p. 223). To illustrate his argument, he takes images from the Brodatz album (Brodatz, 1966), and present the bricks image D95 as a regular texture, and calf skin D06 as irregular texture.

In Figure 10 we show horizontal and vertical slices through the Fourier amplitude spectrum of these textures and also of the four textures from our rank-order. By visual inspection, the spectrum of 1/f noise seems similar to that of calf skin and would classify as irregular texture. Although the spectra of circles and leopard textures do have more peaks than 1/f noise, they are not as "isolated" as the bricks image. The spectra of dots and plaid show isolated peaks, making them similar to the bricks images. According to Turner's model, plaid and dots should thus be better textures to convey slant than circles and leopard, contrary to our empirical findings reported earlier. Thus a simple texture characterization in terms of the "peakedness" of the amplitude spectrum is insufficient to explain human slant-from-texture perception.

Sakai and Finkel (1997) proposed a method that uses a different characterization of the spectra, and one that performs an analysis within textures themselves, that is, their model is image based and does not require any template stored in the system. Similar in spirit to Turner, Sakai and Finkel suggest that for perceiving depth the visual system tracks the peak frequency when the spectrum contains high peaks. In the absence of strong and isolated peaks, the system is presumed to track the mean frequency or Average Peak Frequency (APF) of a texture patch in different regions of the image.

The APF is measured by first extracting the local spatial frequency of the images with a bank of oriented filters implemented by Difference of Gaussian (DoG) or Gabor patches. The image is convolved with the filters, and then the output of identically tuned filters is spatially pooled within a small neighbourhood by taking the maximum response of the units. In effect this removes the sensitivity of the DoG or Gabor filters to local phase to mimic the receptive field profiles of complex cells. The frequency tuning of the filter with the highest response determines the peak frequency at every point, and finally the APF is taken as a Gaussian average of the peaks over a larger neighborhood[11]. Sakai and Finkel argue that the larger the changes of the APF within an image, the stronger the perception of slant. Clearly, if true then for our experiments better performance should be associated with larger changes in the APF between the top (far) and the bottom (near) of our textured planes.

Figure 11 shows the APF of the textures used in our experiments on the vertical axis as a function of position along the texture on the horizontal axis. Each subplot contains the analysis for the four slants used (color coded) and represents the average of the APF results across the different instances used for every texture type in our experiments. The APF values are displayed normalized according to the maximum APF observed across the slants tested for every texture.

---

[11] Following Sakai and Finkel, the Gabor bank we used contained frequencies between approximately 0.5 and 9 cpd. Because our stimuli had zero tilt, we used a single orientation for the filters. The pooling of convolution results was done over a radius of four pixels, and the Gaussian average on a radius of 25 pixels.

In general the changes of APF are larger for the more slanted planes, as they should be. In addition, there are differences among the APFs that are consistent with the rank-order found: texture circles spans bigger changes than Perlin noise, and Perlin noise also spans bigger changes than 1/f noise. However, the APF obtained for slanted planes mapped with leopard would suggest that the task is easier with this texture than it is with circles. Also, the APF changes for the texture plaid and uniform dots are much larger than for the rest of patterns, with the sole exception of leopard. This would predict a high performance with such patterns which we did not observe consistently across subjects.

7 Conclusions and future reseach

By means of an extensive psychophysical study the helpfulness of different types of textures in a slant-discrimination task of slanted planes has been characterized at different slant levels. The results show that the slant discrimination of textured planes is affected by both the slant level and the texture mapped on the surface.

The first effect is such that the more slanted the surface, the easier the discrimination becomes. This observation has already been described by Knill (1998a), though here we reported the effect for a wider number of patterns including textures that are not composed nor can be approximated by texels. The effect of slant was further discussed in terms of how it changes the slopes of the psychometric functions. If there were no slant effect, we would have obtained a set of parallel psychometric functions at the different standards tested such that a rigid shift on the stimulus axis would suffice to predict performance at any slant. Now, although it is rather intuitive to describe performance as a function of the surface slant, as we have done it here, a linear relation between such axis and the underlying mechanism for slant discrimination cannot be assumed a priori. The change in the information relevant for the task between, say, 10 and 20 degrees slant is not necessarily the same as the change between 60 and 70 degrees slant. If one could find a transformation of the stimulus axis that produced parallel psychometric functions, the function on the horizontal axis would thus represent a dimension linearly related to the information subjects are using to perform the task. We transformed our stimulus axis according

to functions related to the geometry of the problem (cosine of slant as descriptor of "compression", change of the visible area of the surface, etc.) trying to get more insight in the mechanisms underlying the slant discrimination, without finding a single transformation that works in all cases. Indeed, Levine (1970) has proven that a set of curves cannot be rendered parallel if they are crossed. For our case, this implies that there is *no single* transformation of the stimulus axis for creating a parallel psychometric family because in a range of the axis they are parallel (similar slope and they touch/cross each other) while in another they are not. This result implies that subjects were using different mechanisms, strategies or cues even for the same texture at different slant levels, or at least that they were using a complex cue whose reliability itself may not be a function of slant.

In addition, the number of patterns and the range of slants used in this study allowed us to show that texture type also affects performance. Knill (1998a) reported only marginal differences between a Voronoi lattice and an ellipses-based texture, but the comparison was limited to a single slant rather close to the ground plane (65 degrees slant). Here we observed that there are differences between texture patterns and that such differences are clearer when the surface slant is closer to the vertical plane. By measuring the change of the area enclosed by the psychometric functions yielded by each texture as helpfulness measurement, we could detect systematic differences among four textures across the subjects. This allowed us to build a rank-order that also held in an experiment using the method of probe adjustment.

As an attempt to explain our psychophysical results we studied the spatial-frequency components of our stimuli, in particular computing the Average Peak Frequency proposed by Sakai and Finkel (1997). The results suggest that the spatial-frequency characterization is part of the underlying mechanism yielding the observed rank-order. However, such model (and also Turner et al., 1991) would predict a high position in such ranking for textures with simple and strongly peaked spectra, such as the plaid texture, which we did not observe consistently across subjects. This suggests that other mechanisms are likely involved in this task. In addition, our results may point to the importance of the issue of how the filters properties (number, frequency

range, etc) are set—as argued by Greenspan (1996) for texture recognition[12]—and how the features are extracted from the output of the filter bank (see Kruizinga, Petkov, & Grigorescu, 1999, in the context of texture discrimination). Given that we observe systematic differences among subjects for some of the textures, a human slant-from-texture model needs to allow some critical parameters to be flexible enough to accommodate these differences within a modeling framework. Perhaps attentional factors are involved in that some subjects attend more to certain parts of the spectrum for some textures than others.

Given our results we propose two lines for further research. First, different hypotheses characterizing the information in texture related to depth can be tested against our data. Any model of slant from texture should be coherent with the differences among the types of texture we have observed, and should explain them. The other way around, a study of structural differences of the textures in this ordered set can motivate a model, for example, by extending the spatial-frequency analysis outlined here. We are currently studying the same set of textures mapped onto elliptical cylinders. So far, our results indicate differences with the rank-order reported here in a discrimination task and a probe adjustment task. A spatial-frequency comparison of the textures mapped to these different geometries shows relevant changes in the APF that again explains at least partially the differences in rank-order between planes and cylinders (Rosas, Schepers, Wichmann, & Wagemans, 2002).

Second, we argue that the helpfulness order is closely related to a reliability order. Recently, many studies have proposed and tested reliability-based depth-cue integration mechanisms (e.g. Ernst & Banks, 2002; Landy, Maloney, Johnston, & Young, 1995), the general idea being that the vision system should combine different cues by applying different weights to them according to the reliability of each cue. If the textures we have ranked here were presented with other cues in a depth-perception study, it would be possible to test such an integration mechanism based on reliability: the change in the stimulus from a helpful texture to a less helpful texture type should be reflected in the integration of cues, penalizing the less helpful

[12] This leads to the idea of defining the filters by an unsupervised learning algorithm for texture recognition.

texture. We are currently pursuing this line of research combining texture with the (object) motion cue (Rosas & Wagemans, 2003), and texture with haptic information (Rosas, Wichmann, Ernst & Wagemans, 2003; Rosas, Wagemans, Ernst & Wichmann, 2004).

8 References

Bajcsy, R., & Lieberman, L. (1976). Texture gradient as a depth cue. *Computer Graphics and Image Processing, 5*(1), 52-67.

Bergen, J. R., & Landy, M. S. (1991). Computational modeling of visual texture segregation. In Landy, M.S., & Movshon, J.A. (Ed.), *Computational models of visual processing* (pp. 253-271). Cambridge, MA: The MIT Press.

Blake, A., & Marinos, C. (1990). Shape from texture: Estimation, isotropy and moments. *Artificial Intelligence, 45*, 323-80.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision. 10*(4), 433-436.

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

Campbell, F. W., & Robson, J. G. (1968). Application of Fourier analysis to the visibility of gratings. *Journal of Physiology (London), 197*(3), 551-66.

Clerc, M., & Mallat, S. (2002). The texture gradient equation for recovering shape from texture. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 24*(4), 536-549.

Cutting, J. E., & Millard, R. T. (1984). Three gradients and the perception of flat and curved surfaces. *Journal of Experimental Psychology: General, 113*(2), 198-216.

Dana K. J., Van Ginneken B., Nayar S. K., & Koenderink J. J. (1999). Reflectance and texture of real-world surfaces. *ACM Transactions on Graphics, 18*(1), 1-34.

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature, 415*(6870), 429-433.

Field, D. J. (1987). Relations between the statistics of natural images and the response properties of cortical cells. *Journal of the Optical Society of America A, 4*(12), 2379-94.

Gårding, J., & Lindeberg, T. (1996). Direct computation of shape cues using scale-adapted spatial derivative operators. *International Journal of Computer Vision, 17*(2), 163-91.

Gibson, J. J. (1950a). *The perception of the visual world*. Boston: Houghton Mifflin.

Gibson, J. J. (1950b). The perception of visual surfaces. *American Journal of Psychology, 63*, 367-84.

Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston: Houghton Mifflin Company.

Graham, N. (1989). *Visual pattern analyzers*. New York: Oxford University Press.

Greenspan, H. (1996). Non-parametric texture learning. In Nayar, S.K., & Poggio T. (Eds), *Early Visual Learning* (pp. 299-328). New York: Oxford University Press.

Heckbert, P. S. (1989). *Fundamentals of texture mapping and image warping*. Master's Thesis. University of California, Berkeley.

Jain, A. K., & Bhattacharjee, S. (1992). Texture segmentation using Gabor filters for automatic texture processing. *Machine Vision and Applications, 5*, 169-184.

Jain, A. K., & Farrokhnia, F. (1991). Unsupervised texture segmentation using Gabor filters. *Pattern Recognition, 24*(12), 1167-1186.

Kanatani, K. (1984). Detection of surface orientation and motion from texture by a stereological technique. *Artificial Intelligence, 23*, 213-237.

Knill, D. C. (1998a). Discrimination of planar surface slant from texture: human and ideal observers compared. *Vision Research, 38*(11), 1683-711.

Knill, D. C. (1998b). Surface orientation from texture: ideal observers, generic observers and the information content of texture cues. *Vision Research, 38*(11), 1655-82.

Knill, D. C., Field, D., & Kersten, D. (1990). Human discrimination of fractal images. *Journal of the Optical Society of America A, 7*(6), 1113-1123.

Koenderink, J. J., & van Doorn, A. J. (1996). Illuminance texture due to surface mesostructure . *Journal of the Optical Society of America A, 13*(3), 452-463.

Kruizinga, P., Petkov, N., & Grigorescu, S. E. (1999). Comparison of texture features based on Gabor filters. *Proceedings of the 10th International Conference on Image Analysis and Processing* (pp. 142-147).

Krumm, J., & Shafer, S. A. (1992). Shape from periodic texture using the spectrogram. *Proceedings of the 1992 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '92)* (pp. 284-289).

Kubovy, M., Holcombe, A. O., & Wagemans, J. (1998). On the lawfulness of grouping by proximity. *Cognitive Psychology, 35*, 71-98.

Kubovy, M., & Wagemans, J. (1995). Grouping by proximity and multistability in dot lattices: A quantitative Gestalt theory. *Psychological-Science, 6*(4), 225-234.

Landy, M. S., Maloney, L. T., Johnston, E. B., & Young, M. (1995). Measurement and modeling of depth cue combination: in defense of weak fusion. *Vision Research, 35*(3), 389-412.

Levine, M. V. (1970). Transformations that render curves parallel. *Journal of Mathematical Psychology, 7*(3), 410-443.

Li, A., & Zaidi, Q. (2000). Perception of three-dimensional shape from texture is based on patterns of oriented energy. *Vision Research, 40*(2), 217-42.

Malik, J., & Rosenholtz, R. (1994). A computational model for shape from texture. In Bock, G.R., & Goode, J.A. (Eds), *Higher-Order Processing in the Visual System* (pp. 272-286). Chichester: John Wiley & Sons.

Malik, J., & Rosenholtz, R. (1997). Computing local surface orientation and shape from texture for curved surfaces. *International Journal of Computer Vision, 23*(2), 149-68.

Passmore, P. J., & Johnston, A. (1995). Human discrimination of surface slant in fractal and related textured images. *Spatial Vision, 9*(1), 151-161 .

Pelli, D. G. (1997). The Video toolbox software for visual psychophysics: Transforming numbers into movies. *Spatial-Vision. 1997, 10*(4), 437-442.

Perlin, K. (1985). An image synthesizer. *Computer Graphics, 19*(3), 287--296.

Purdy, W. C. (1958). The hypothesis of psychophysical correspondence in space perception. *Dissertation Abstracts, 19*, 1454-1455 .

Ribeiro, E., & Hancock, E. R. (2000a). Estimating the 3D orientation of texture planes using local spectral analysis . *Image and Vision Computing , 18*(8), 619-631.

Ribeiro, E., & Hancock, E. R. (2000b). Improved orientation estimation for texture planes using multiple vanishing points . *Pattern Recognition , 33*(10), 1599-1610.

Rosas, P., Schepers, J., Wichmann, F. A., & Wagemans, J. (2002). Surface-slant and surface-curvature from texture. *Perception, 31, S27a.*

Rosas, P. & Wagemans, J. (2003). Combination of texture and object motion in slant discrimination. Journal of Vision 3(9), 842a, http://journalofvision.org/3/9/842/, doi:10.1167/3.9.842.

Rosas, P., Wichmann, F. A., Ernst, M. O. & Wagemans, J. (2003). Texture and haptic cues in slant discrimination: Measuring the effect of texture type on cue combination [Abstract]. Journal of Vision, in press.

Rosas, P., Wagemans, J., Ernst, M. O. and Wichmann, F. A.(2004). Texture and haptic cues in slant discrimination: Reliability-based cue weighting without statistically optimal cue combination. Manuscript in preparation.

Rosenholtz, R., & Malik, J. (1997). Surface orientation from texture: isotropy or homogeneity (or both)? *Vision Research, 37*(16), 2283-93.

Sakai, K., & Finkel, L. H. (1995). Characterization of the Spatial Frequency Spectrum in Perception of Shape-from-Texture . *Journal of the Optical Society of America A, 12*(6), 1208-1224.

Sakai, K., & Finkel, L. H. (1997). Spatial-frequency analysis in the perception of perspective depth. *Network: Computation in Neural Systems , 8*, 335-352.

Stoyan, D. (1992). Stochastic geometry version 4.1.

Stoyan, D., Kendall, W. S., & Mecke, J. (1995). *Stochastic Geometry and its Applications*. Chichester: John Wiley & Sons.

Super, B. J., & Bovik, A. C. (1995). Planar surface orientation from texture spatial frequencies. *Pattern Recognition , 28*(5), 729-743.

Todd, J. T., & Akerstrom, R. A. (1987). Perception of three-dimensional form from patterns of optical texture. *Journal of Experimental Psychology: Human Perception and Performance, 13*(2), 242-255.

Tuceryan, M., & Jain, A. K. (1998). Texture Analysis. In Chen, C. H., Pau, L. F., & Wang, P. S. P. (Eds), *The Handbook of Pattern Recognition and Computer Vision* (2nd ed., pp. 207-248). River Edge, NJ : World Scientific Publishing Co.

Turing, A. (1952). The chemical basis of morphogenesis. *Philosophical Transactions of The Royal Society B, 237*, 37-72.

Turk, G. (1991). Generating textures on arbitrary surfaces using reaction-diffusion. *Computer Graphics (SIGGRAPH '91 Proceedings)* (pp. 289-298).

Turner, M. R. (1986). Texture discrimination by Gabor functions. *Biological Cybernetics, 55*, 71-82.

Turner, M. R., Bajcsy, R., & Gerstein, G. L. (1989). Estimation of textured surface inclination by parallel local spectral analysis. (Report No. MS-CIS-89-42 GRASP LAB 184). Department of Computer and Information Science School of Engineering and Applied Sciences University of Pennsylvania.

Turner, M. R., Gerstein, G. L., & Bajcsy, R. (1991). Underestimation of visual texture slant by human observers: a model. *Biological Cybernetics, 65*(4), 215-226.

Turner, M. R., Salganicoff, M., Gerstein, G. L., & Bajcsy, R. (1989). Receptive fields for the determination of textured surface inclination. (Report No. MS-CIS-89-48 GRASP LAB 187). Department of Computer and Information Science School of Engineering and Applied Sciences University of Pennsylvania.

Watson, A. B., & Pelli, D. G. (1983). QUEST: A Bayesian adaptive psychometric method. *Perception & Psychophysics, 33*(2), 113-20.

Wichmann, F. A. (1999). *Some aspects of modelling human spatial vision: Contrast discrimination.*

Unpublished doctoral dissertation, University of Oxford.

Wichmann, F. A., & Henning, G. B. (1998). No role for motion blur in either motion detection or motion based image segmentation. *Journal-of-the-Optical-Society-of-America-A, 15*(2), 297-306.

Wichmann, F. A., & Hill, N. J. (2001a). The psychometric function: I. Fitting, sampling, and goodness of fit. *Perception & Psychophysics, 63*(8), 1293-1313.

Wichmann, F. A., & Hill, N. J. (2001b). The psychometric function: II. Bootstrap-based confidence intervals and sampling. *Perception & Psychophysics, 63*(8), 1314-1329.

Witkin, A. P. (1981). Recovering surface shape and orientation from texture. *Artificial Intelligence, 17*, 17-45.

9 Figure and Table Captions

Figure 1: Examples of slanted planes with the different textures used in this study. The patterns are, from top to bottom row: dots with zero displacement (uniform lattice), dots with five pixels displacement, dots with ten pixels displacement, circles (polka dots), Voronoi, plaid, 1/f noise, Perlin noise and leopard. On each row the slants depicted are, from left to right: 26 degrees, 37 degrees, 53 degrees and 66 degrees. (Note: Although the borders of the samples are squares in this figure, the stimuli in the experiments were shown within a circular aperture).

Figure 2: The psychometric functions estimated for all the textures observed by the subjects in Experiment 1 showing performance (proportion correct) on the x-axis as a function of the surface slant on the y-axis. Each row corresponds to the data collected for a single subject. On every row the leftmost subplot corresponds to the comparisons with the least slanted standard (26 degrees), the second to a standard of 37 degrees, the third to 53 degrees and the fourth to 66 degrees. The standard levels are indicated by a solid vertical line. To avoid visual clutter due to the large number of data points per texture and the large number of textures per condition, we only show raw data for the second standard for every subject (slant 37 degrees, second column); symbol size is proportional to the number of trials. The symbols shown in the first, third and fourth column do *not* represent empirical data; they are only used to help clarify the figure legend.

Figure 3: Two examples of the results when refitting slopes at 75 percent correct for different standards.

Figure 4: Slopes at 75 percent correct for the textures tested by one subject at two standards.

Figure 5: The area between the psychometric functions for bigger and smaller slants at a certain standard, enclosed by two performance levels, is taken as the measure for helpfulness.

Figure 6: Areas of the four textures eliciting a similar rank-order among subjects. The error bars are based on the 68% confidence intervals for the fit of each psychometric function. They are not displayed, though, when the performance higher than 80% correct was unatainable in the range of slants tested.

Figure 7: Psychometric functions for the discrimination of "ground"-like surfaces (black lines) and "wall" surfaces (gray lines). Two textures and two subjects.

Figure 8: Probe alignment task. The subjects adjusted a movable line drawn over a reference axis to indicate the perceived orientation of the slanted plane depicted in the upper aperture. They had to indicate the angle of the plane from the ground plane as seen from the left side of the surface, depicted by the circular arrow (which was not present during the experiment).

Figure 9: Results of the probe alignment experiment for four textures. The dotted horizontal lines indicate the slants tested. Bars represent the sample standard deviation, full circles the sample mean.

Figure 10: Discrete Fourier Transform (DFT) of textures. First two rows with images from the Brodatz album (bricks D95 and calf skin D06). In rows three to six are the textures of the rank-order found in our experiments: circles, leopard, Perlin noise, and 1/f noise. Rows seven and eight contain textures plaid and uniform dots. On every row the image processed is shown (a

Hanning window was applied to the raw images before performing the DFT) together with the spectrum and slices of the spectrum on the horizontal and vertical axes.

Figure 11: Average Peak Frequencies of textures used in our experiments. On each plot the curves depict textures at different slants (see colour coding), and they are normalized according the maximum APF observed for that plot. By visual inspection, the changes in APF would predict that leopard, plaid and uniform dots should facilitate the most the perception of slant, followed by circles, then Perlin noise and finally 1/f noise.

Table 1: Raw area values for all textures and subjects. Numbers between brackets indicate the 68% confidence limits.

Table 2: Normalized area values for all subjects and textures. The raw values obtained for each subject were normalized by the maximum area value observed for the subject.

Table 3: Correlation coefficients between the raw slants shown and subjects probe adjustments.

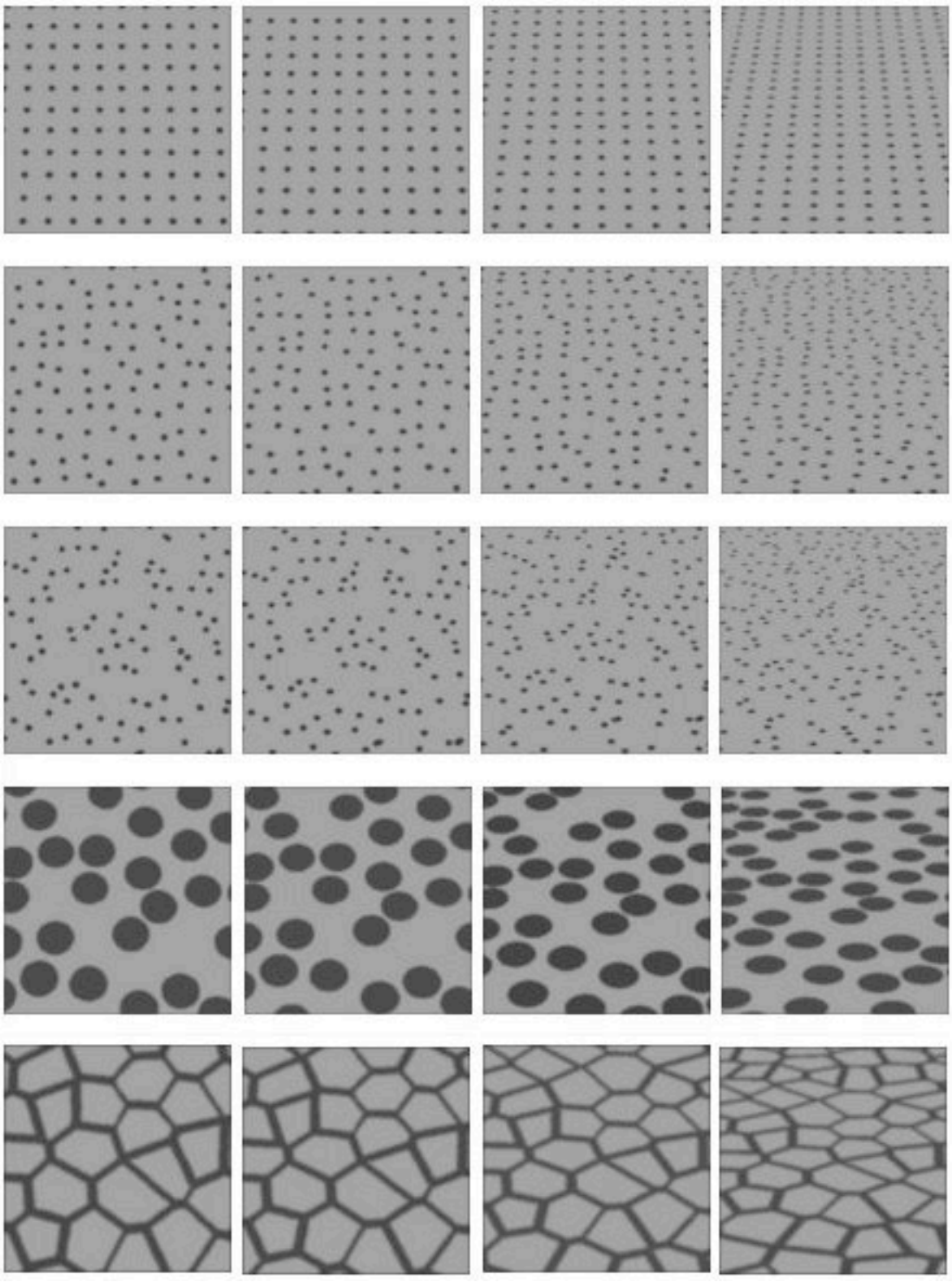

Figure 1

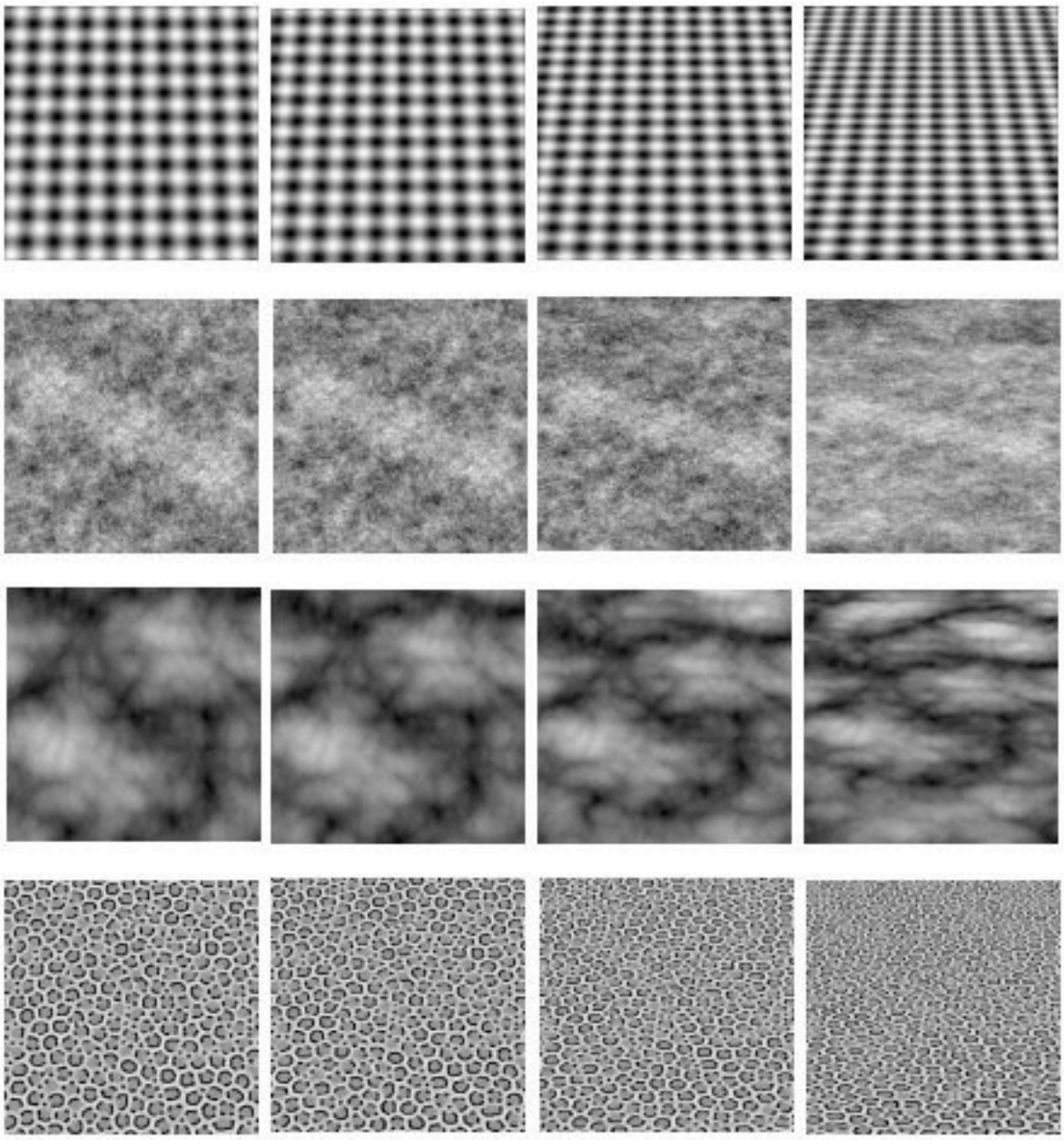

Figure 1: (cont.)

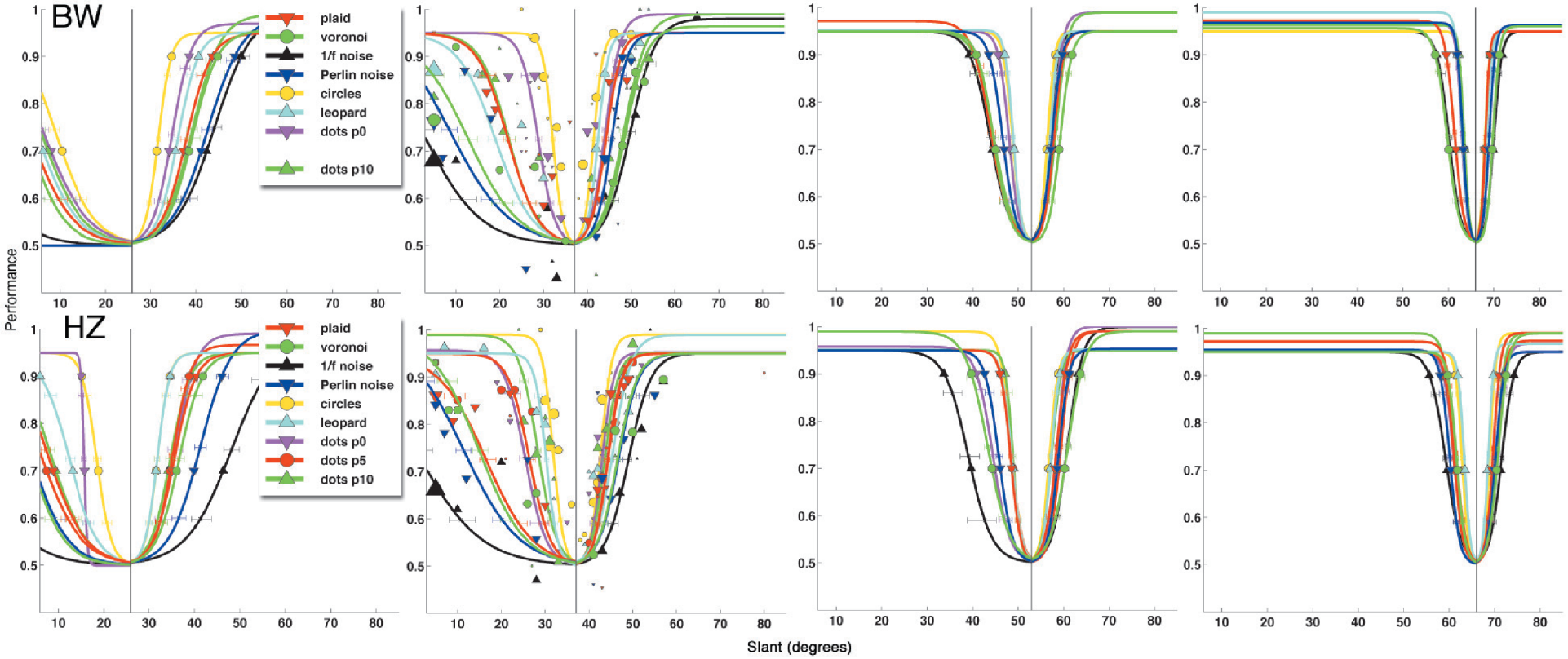
BW
HZ
plaid
voronoi
1/f noise
Perlin noise
circles
leopard
dots p0
dots p5
dots p10
Performance
Slant (degrees)

Figure 2

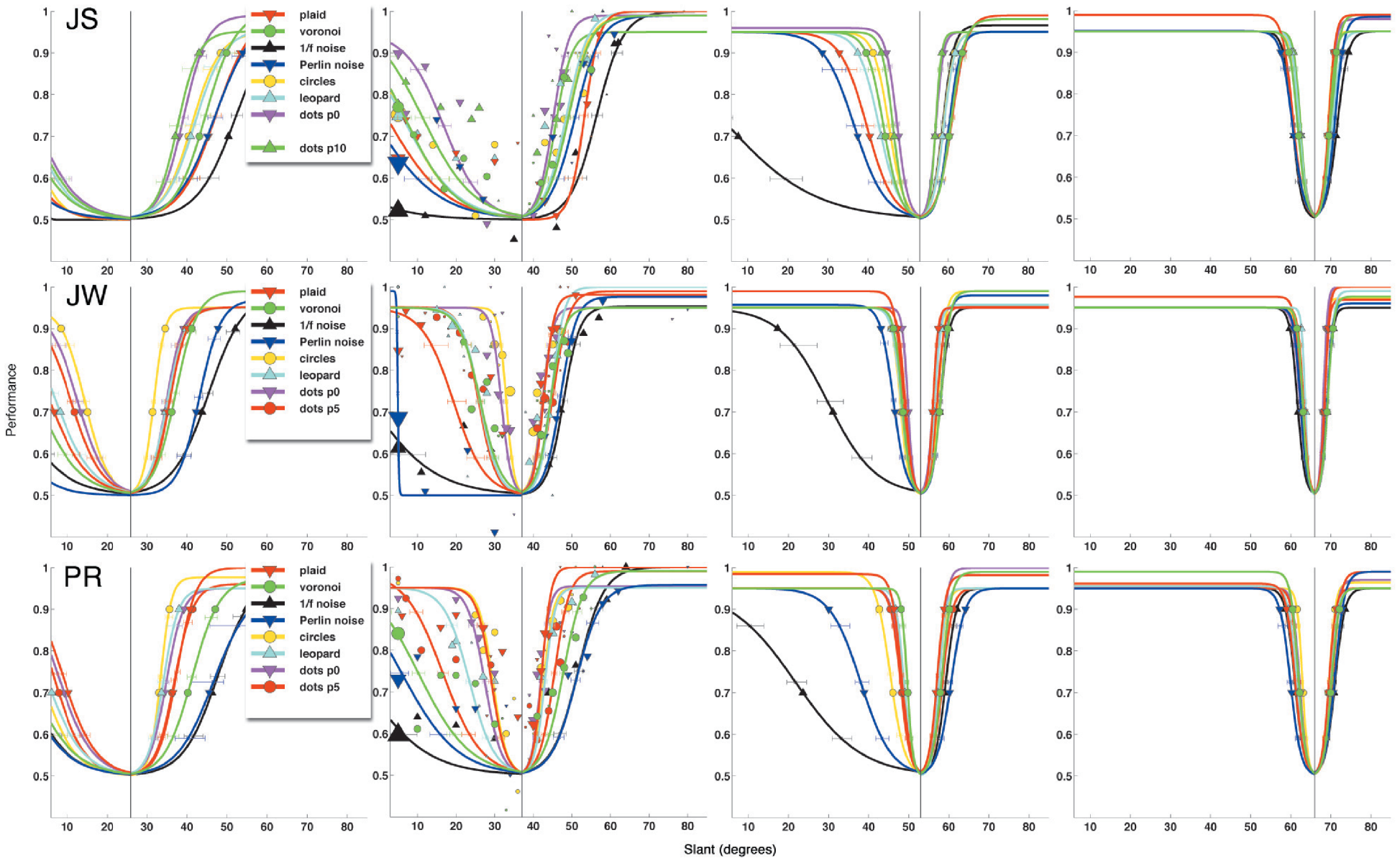

Figure 2 (cont.)

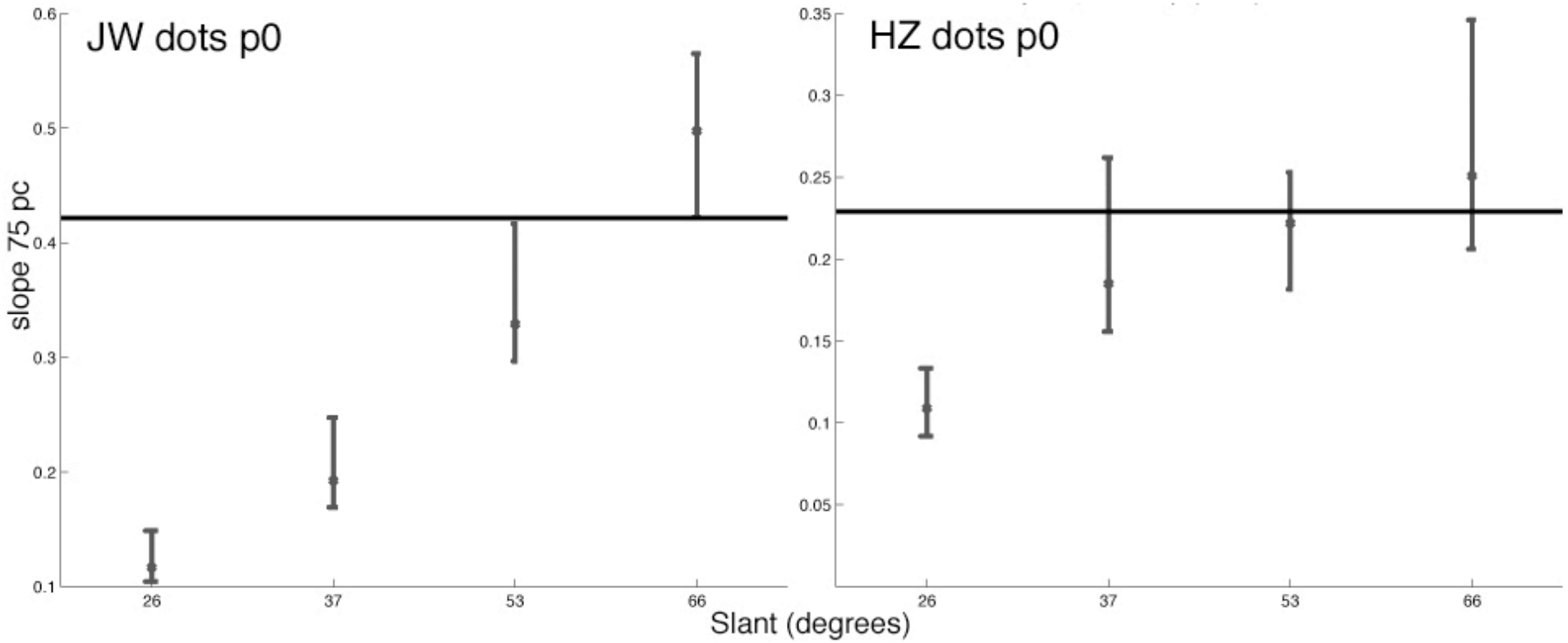
JW dots p0
HZ dots p0
slope 75 pc
0.6
0.5
0.4
0.3
0.2
0.1
0.35
0.3
0.25
0.2
0.15
0.1
0.05
26
37
53
66
Slant (degrees)

Figure 3

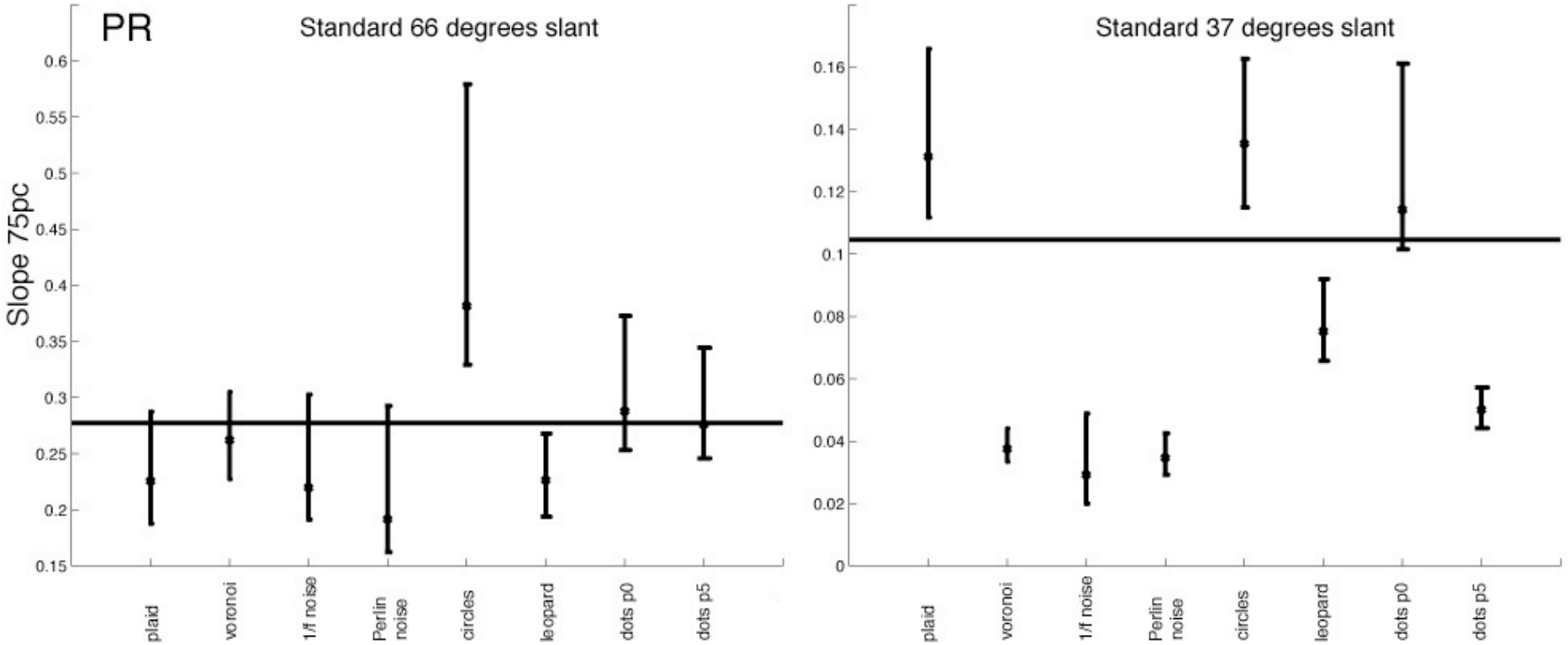
PR
Standard 66 degrees slant
Slope 75pc
0.6
0.55
0.5
0.45
0.4
0.35
0.3
0.25
0.2
0.15
plaid
voronoi
1/f noise
Perlin noise
circles
leopard
dots p0
dots p5
Standard 37 degrees slant
0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
plaid
voronoi
1/f noise
Perlin noise
circles
leopard
dots p0
dots p5

Figure 4

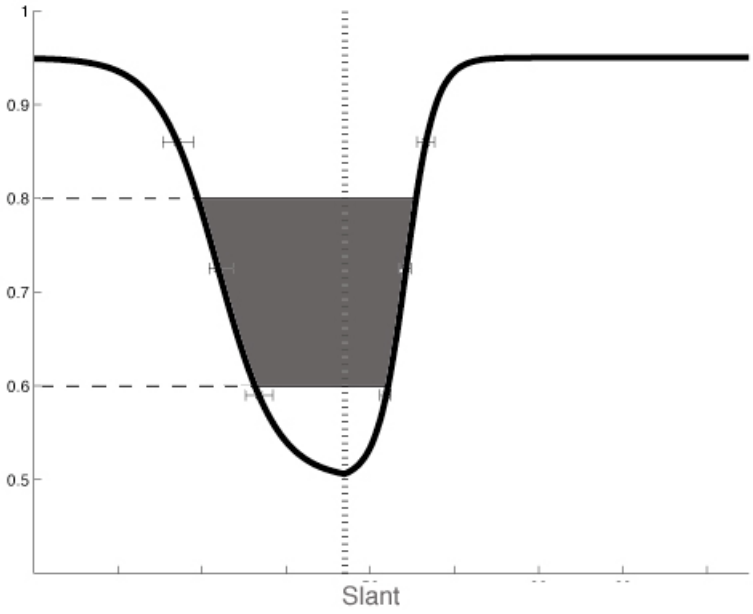

Figure 5

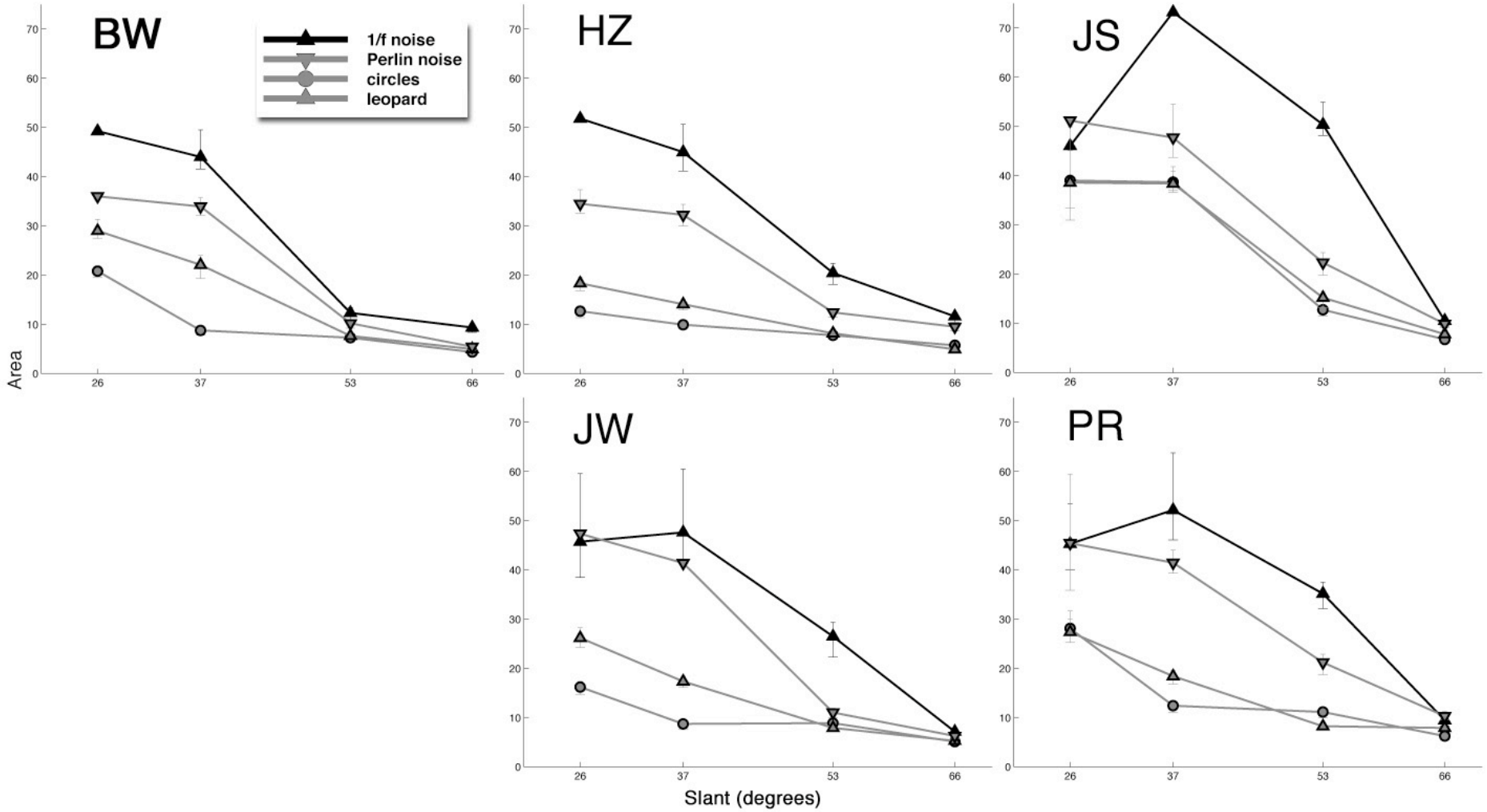
BW
HZ
JS
JW
PR
1/f noise
Perlin noise
circles
leopard
Area
Slant (degrees)
0
10
20
30
40
50
60
70
26
37
53
66

Figure 6

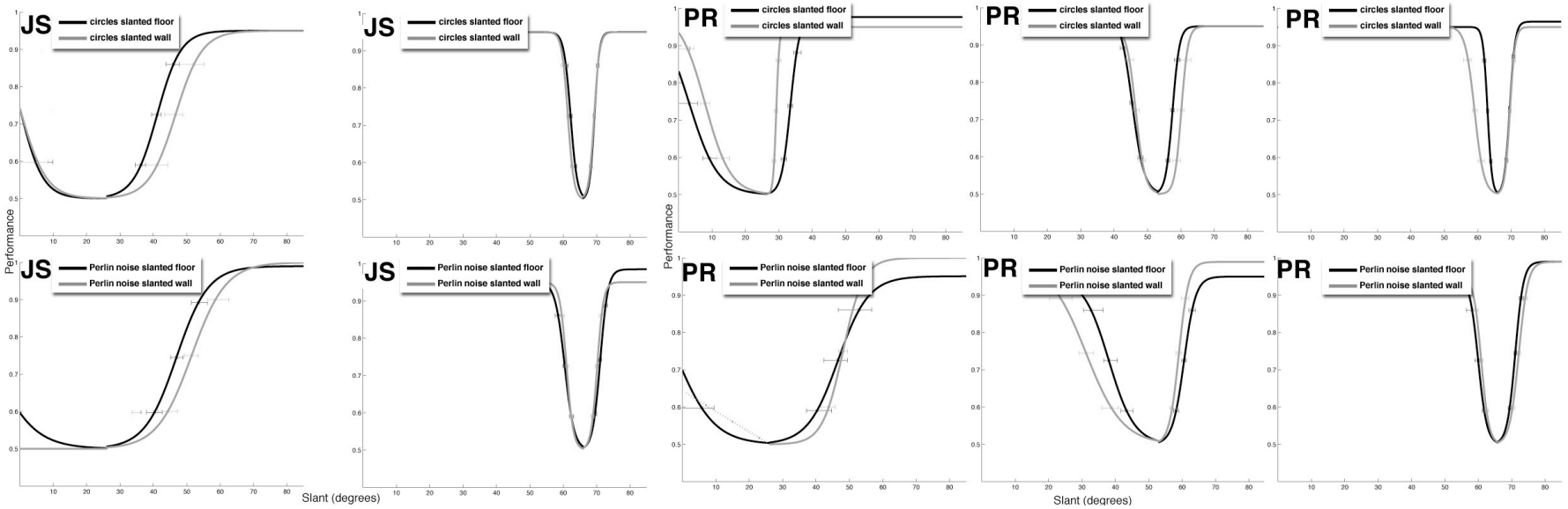
JS
PR
circles slanted floor
circles slanted wall
Perlin noise slanted floor
Perlin noise slanted wall
Performance
Slant (degrees)

Figure 7

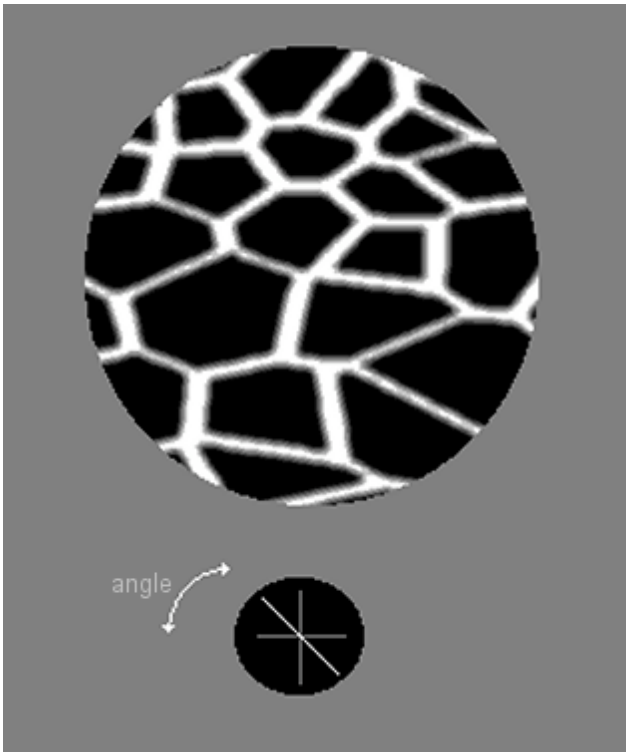

Figure 8

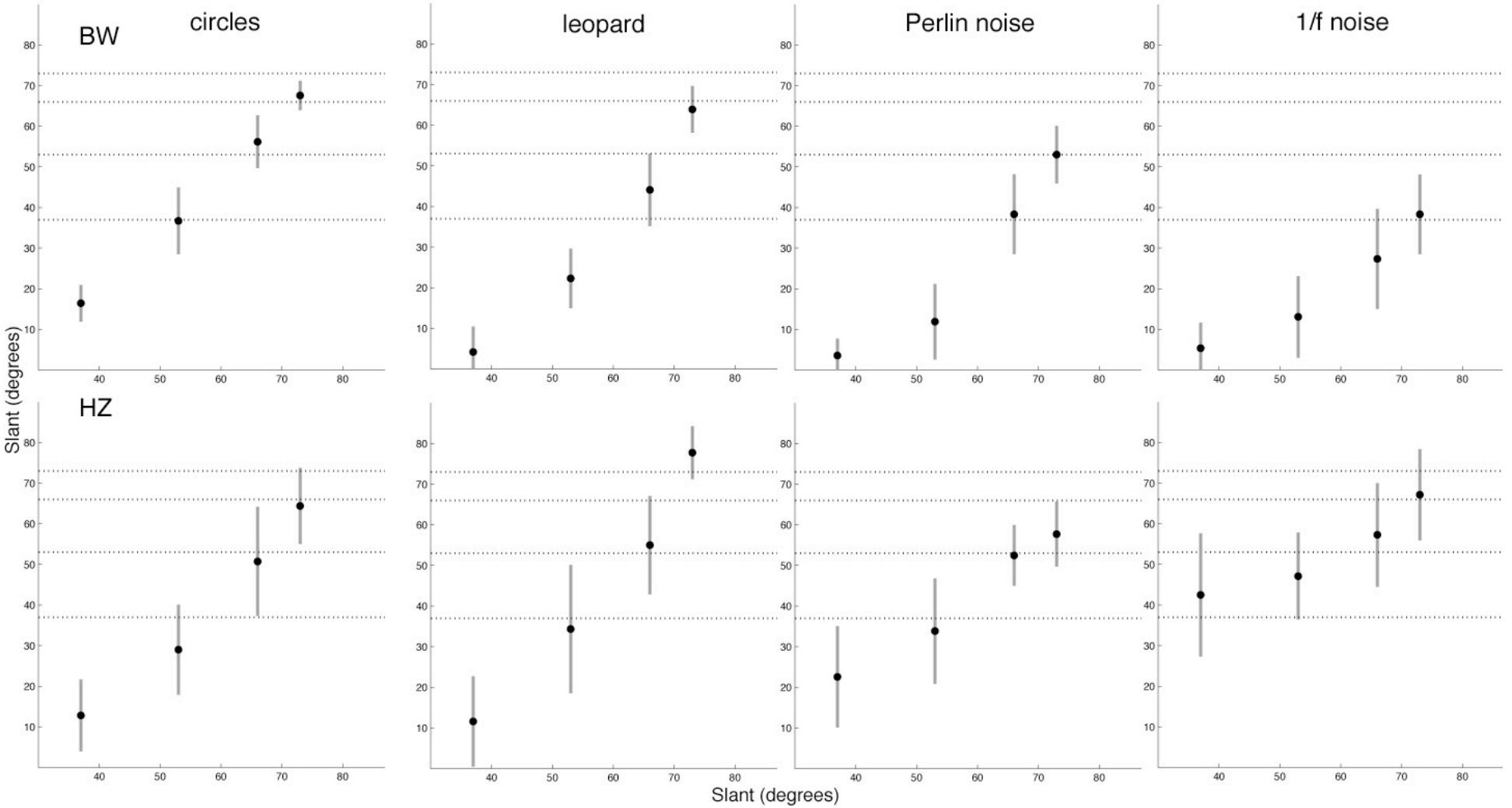
BW
HZ
circles
leopard
Perlin noise
1/f noise
Slant (degrees)
Slant (degrees)

Figure 9

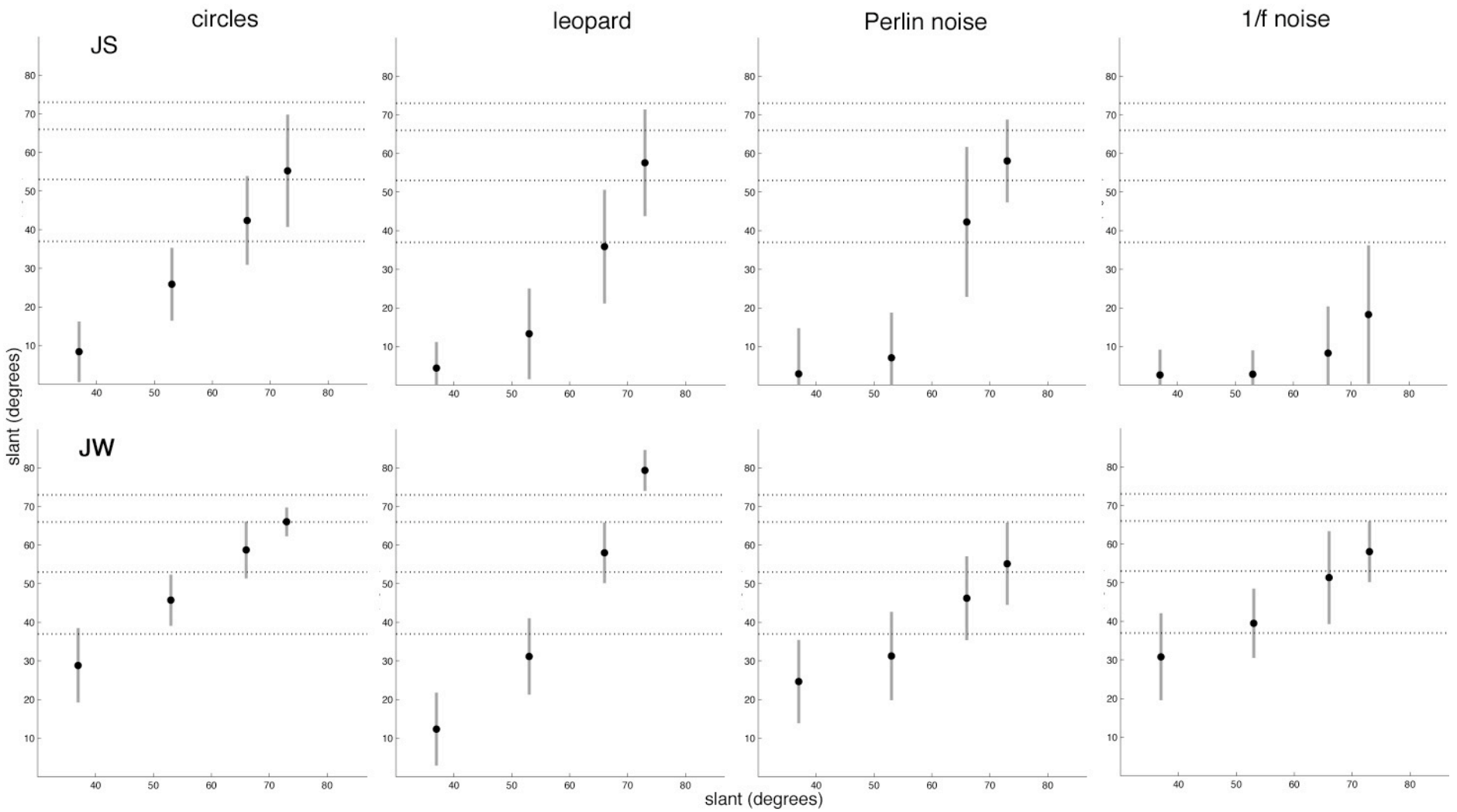

Figure 9 (cont.)

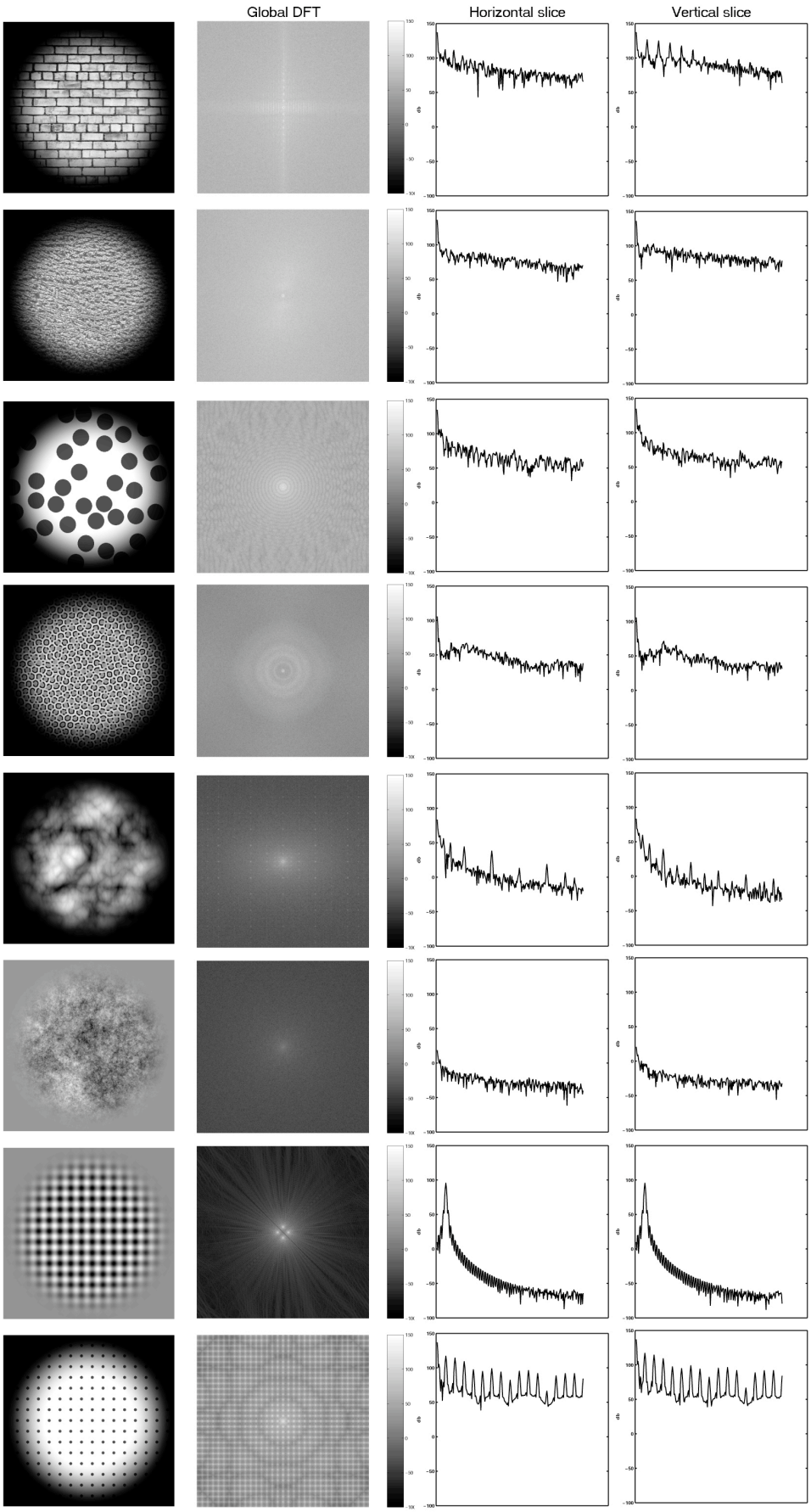

Figure 10

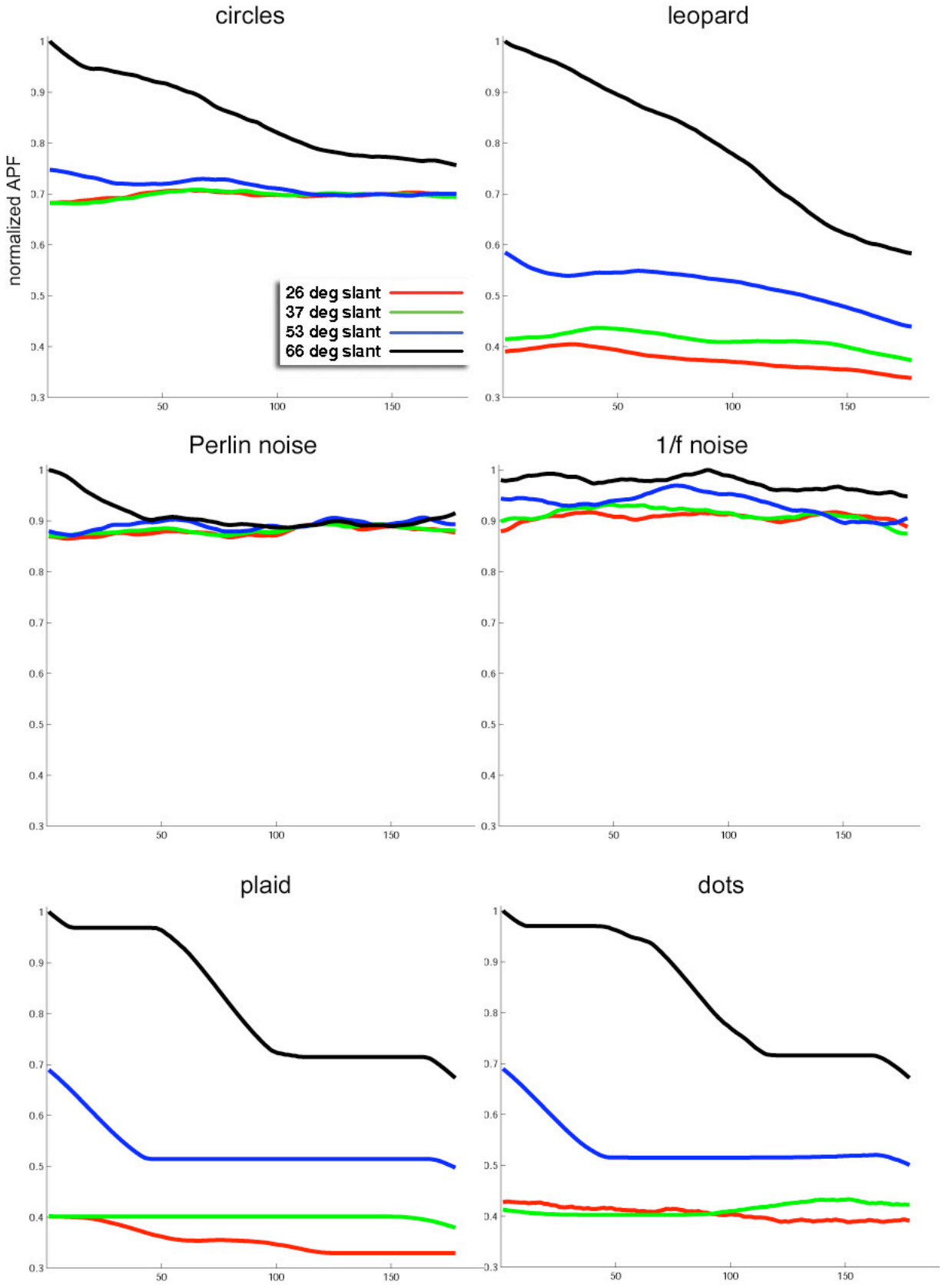
circles
leopard
Perlin noise
1/f noise
plaid
dots
normalized APF
26 deg slant
37 deg slant
53 deg slant
66 deg slant

Figure 11

	Areas (raw values)								
	plaid	vrnoi	1/f noise	Perlin noise	circles	leop	dotp0	dotp5	dotp10
BW									
26 deg	32.1 (2.0/ 3.2)	34.2 (2.7/ 3.9)	49.3 (19.3/70.2)	36 (NaN/ NaN)	20.8 (1.2/ 1.1)	29 (1.5/ 2.3)	26 (1.6/ 1.9)		30.5 (1.6/ 2.1)
37 deg	21.1 (2.2/ 1.4)	33.6 (2.5/ 2.1)	44 (2.4/ 5.5)	34 (1.8/ 1.8)	8.79 (1.0/ 0.8)	22.1 (2.8/ 1.9)	14.2 (1.1/ 1.1)		24.5 (2.0/ 1.0)
53 deg	12.9 (1.3/ 1.2)	13.9 (1.3/ 1.0)	12.3 (1.4/ 1.1)	10.2 (1.0/ 0.6)	7.29 (0.6/ 0.5)	7.67 (0.8/ 0.7)	8.93 (0.8/ 0.4)		12.7 (1.3/ 0.7)
66 deg	6.32 (0.6/ 0.6)	9.52 (0.8/ 0.5)	9.37 (1.0/ 0.7)	5.47 (0.5/ 0.4)	4.38 (0.4/ 0.3)	5.01 (0.4/ 0.3)	4.93 (0.5/ 0.4)		4.36 (0.4/ 0.4)
HZ									
26 deg	26 (1.4/ 1.3)	31.2 (2.1/ 2.7)	51.8 (15.3/43.7)	34.5 (1.9/ 2.9)	12.7 (1.3/ 0.9)	18.4 (1.5/ 1.3)	19.1 (1.8/ 1.7)	27.1 (1.4/ 1.8)	24.6 (1.6/ 1.2)
37 deg	26.1 (2.6/ 1.8)	28.9 (2.6/ 1.9)	45 (3.8/ 5.6)	32.3 (2.3/ 2.1)	9.94 (0.6/ 0.5)	14.1 (1.2/ 1.0)	16.5 (2.3/ 1.7)	16.8 (1.7/ 1.2)	14 (1.0/ 0.9)
53 deg	9.41 (0.6/ 0.4)	15.9 (1.2/ 1.2)	20.4 (2.3/ 2.0)	12.4 (1.0/ 0.8)	7.79 (0.7/ 0.5)	8.18 (0.7/ 0.5)	13.1 (1.3/ 1.2)	10.1 (0.8/ 0.3)	8.01 (0.7/ 0.9)
66 deg	7.52 (0.6/ 0.6)	8.56 (0.6/ 0.3)	11.6 (1.3/ 0.7)	9.54 (0.8/ 0.9)	5.75 (0.4/ 0.3)	4.95 (0.4/ 0.3)	8.92 (1.1/ 0.9)	8.87 (0.6/ 0.6)	8.48 (0.5/ 0.4)
JS									
26 deg	47 (8.9/15.1)	42.7 (5.8/13.6)	46.1 (1.2/15.0)	51.2 (15.3/38.9)	39 (8.1/10.9)	38.6 (5.1/ 6.6)	34.3 (2.5/ 4.6)		34.3 (4.0/ 6.0)
37 deg	47.3 (2.9/ 5.0)	38.7 (1.9/ 2.6)	73.2 (33.4/44.5)	47.7 (4.0/ 6.8)	38.7 (1.7/ 2.3)	38.4 (1.8/ 3.5)	26.4 (2.3/ 1.6)		31.3 (2.5/ 1.9)
53 deg	19.8 (2.1/ 1.3)	15.8 (1.1/ 0.7)	50.4 (2.2/ 4.6)	22.3 (2.5/ 2.0)	12.8 (1.3/ 0.7)	15.2 (2.0/ 1.3)	9.07 (0.9/ 0.8)		10.1 (0.9/ 0.6)
66 deg	8.15 (0.6/ 0.5)	7.53 (0.8/ 0.5)	10.5 (1.1/ 0.9)	9.88 (0.8/ 0.7)	6.74 (0.7/ 0.4)	7.8 (0.7/ 0.5)	6.85 (0.7/ 0.4)		7.25 (0.5/ 0.4)
JW									
26 deg	28.3 (1.6/ 1.9)	32.1 (2.6/ 4.5)	45.7 (7.3/13.8)	47.4 (11.7/46.3)	16.2 (1.5/ 1.1)	26.2 (1.9/ 2.0)	20.9 (2.2/ 1.6)	22.7 (1.7/ 1.4)	
37 deg	22 (3.3/ 1.6)	17.1 (1.6/ 1.0)	47.6 (6.2/12.9)	41.4 (NaN/ NaN)	8.73 (1.0/ 0.8)	17.4 (1.2/ 1.0)	10.1 (1.0/ 0.7)	18 (1.5/ 1.0)	
53 deg	6.57 (0.6/ 0.4)	9.08 (0.5/ 0.4)	26.5 (4.2/ 2.9)	11 (1.0/ 0.9)	8.9 (0.7/ 0.4)	7.93 (0.7/ 0.5)	5.98 (0.5/ 0.4)	7.87 (0.6/ 0.5)	
66 deg	5 (0.5/ 0.4)	5.9 (0.5/ 0.3)	7.15 (0.7/ 0.6)	6.29 (0.6/ 0.5)	5.13 (0.4/ 0.3)	5.3 (0.4/ 0.3)	5.48 (0.5/ 0.4)	5.54 (0.5/ 0.5)	
PR									
26 deg	26.1 (1.2/ 1.3)	38 (3.3/ 5.2)	45.3 (5.3/ 8.2)	45.5 (9.5/14.0)	28.1 (1.5/ 3.6)	27.4 (2.1/ 2.7)	25 (1.5/ 1.5)	28.3 (1.3/ 1.0)	
37 deg	12.1 (1.0/ 0.8)	33.9 (2.0/ 1.4)	52.2 (6.1/11.6)	41.5 (2.2/ 2.5)	12.4 (1.3/ 1.1)	18.4 (1.6/ 1.1)	13.7 (1.1/ 0.8)	26.2 (1.8/ 1.3)	
53 deg	8.1 (0.7/ 0.6)	8.02 (0.5/ 0.6)	35.2 (3.0/ 2.3)	21.2 (2.5/ 1.6)	11.2 (1.2/ 1.0)	8.26 (0.8/ 0.5)	8.81 (0.5/ 0.5)	9.88 (0.9/ 0.6)	
66 deg	7.94 (0.9/ 0.7)	7.69 (0.7/ 0.4)	9.47 (1.2/ 0.8)	10.3 (0.8/ 1.0)	6.26 (0.6/ 0.4)	7.91 (0.7/ 0.6)	7.27 (0.6/ 0.3)	7.89 (0.6/ 0.4)	

Table 1

	Areas (ratios)								
	plaid	voronoi	1/f noise	Perlin noise	circles	leop	dotp0	dotp5	dotp10
BW									
26 deg	0.65	0.69	1.00	0.73	0.42	0.59	0.53		0.62
37 deg	0.43	0.68	0.89	0.69	0.18	0.45	0.29		0.50
53 deg	0.26	0.28	0.25	0.21	0.15	0.16	0.18		0.26
66 deg	0.13	0.19	0.19	0.11	0.09	0.10	0.10		0.09
HZ									
26 deg	0.50	0.60	1.00	0.67	0.24	0.35	0.37	0.52	0.48
37 deg	0.50	0.56	0.87	0.62	0.19	0.27	0.32	0.32	0.27
53 deg	0.18	0.31	0.39	0.24	0.15	0.16	0.25	0.19	0.15
66 deg	0.15	0.17	0.22	0.18	0.11	0.10	0.17	0.17	0.16
JS									
26 deg	0.64	0.58	0.63	0.70	0.53	0.53	0.47		0.47
37 deg	0.65	0.53	1.00	0.65	0.53	0.52	0.36		0.43
53 deg	0.27	0.22	0.69	0.31	0.17	0.21	0.12		0.14
66 deg	0.11	0.10	0.14	0.13	0.09	0.11	0.09		0.10
JW									
26 deg	0.59	0.67	0.96	0.99	0.34	0.55	0.44	0.48	
37 deg	0.46	0.36	1.00	0.87	0.18	0.36	0.21	0.38	
53 deg	0.14	0.19	0.56	0.23	0.19	0.17	0.13	0.17	
66 deg	0.10	0.12	0.15	0.13	0.11	0.11	0.12	0.12	
PR									
26 deg	0.50	0.73	0.87	0.87	0.54	0.52	0.48	0.54	
37 deg	0.23	0.65	1.00	0.79	0.24	0.35	0.26	0.50	
53 deg	0.16	0.15	0.67	0.41	0.21	0.16	0.17	0.19	
66 deg	0.15	0.15	0.18	0.20	0.12	0.15	0.14	0.15	

Table 2

	circles	leopard	Perlin Noise	1/f noise
BW	0.95	0.94	0.89	0.77
HZ	0.92	0.89	0.86	0.71
JS	0.90	0.80	0.86	0.63
JW	0.93	0.93	0.82	0.79
Average	0.93	0.89	0.86	0.73

Table 3