

Texture and haptic cues in slant discrimination: reliability-based cue weighting without statistically optimal cue combination

Pedro Rosas

Department of Psychology, University of Leuven, Tiensestraat 102, B-3000 Leuven, Belgium, and Max Planck Institute for Biological Cybernetics, Spemannstrasse 38, D-72076 Tübingen, Germany

Johan Wagemans

Department of Psychology, University of Leuven, Tiensestraat 102, B-3000 Leuven, Belgium

Marc O. Ernst and Felix A. Wichmann

Max Planck Institute for Biological Cybernetics, Spemannstrasse 38, D-72076 Tübingen, Germany

Received June 14, 2004; revised manuscript received November 25, 2004; accepted November 29, 2004

A number of models of depth-cue combination suggest that the final depth percept results from a weighted average of independent depth estimates based on the different cues available. The weight of each cue in such an average is thought to depend on the reliability of each cue. In principle, such a depth estimation could be statistically optimal in the sense of producing the minimum-variance unbiased estimator that can be constructed from the available information. Here we test such models by using visual and haptic depth information. Different texture types produce differences in slant-discrimination performance, thus providing a means for testing a reliability-sensitive cue-combination model with texture as one of the cues to slant. Our results show that the weights for the cues were generally sensitive to their reliability but fell short of statistically optimal combination—we find reliability-based reweighting but not statistically optimal cue combination.

© 2005 Optical Society of America

OCIS codes: 330.4060, 330.5510, 330.6100, 330.7310.

1. INTRODUCTION

In sensor fusion it is typically assumed that the different sources of information relative to the variable of interest are combined systematically to obtain an estimate of the variable. Clark and Yuille¹ classified the combination rules as belonging to the class of either weak-fusion or strong-fusion models. In weak fusion the data coming from the sensors are assumed to be independent, and they are linearly combined to obtain a single estimate. In strong fusion the “modules” processing the input data from the sensors are not independent; that is, the estimates from each module are not independent, and the fusion mechanism might be nonlinear.

In vision science the currently most popular models for depth perception are weak-fusion models, partly because strong fusion is difficult to characterize.² In weak fusion for depth perception the final depth estimate results from a weighted average of the independent depth estimates obtained from each source of information or cue, such as disparity, texture, or motion-parallax.^{3–5} A more complex model than weak fusion is the modified weak-fusion model.⁴ In this model a specific type of interaction between cues promote them “to be on equal footing [so that] the values obtained from them are commensurate” (Ref. 4, p. 392). Except for this interaction, called promotion, cues are assumed to be independent. In this paper we will thus not make a distinction between weak fusion and

modified weak fusion. In these models the weights are determined by the reliability of the corresponding cue: A more reliable cue should have a larger weight in the combined estimate.^{4,5} In effect they are reliability-based weighted-average models. Young *et al.*⁶ reported evidence for a reliability-based weighting of cues in shape from texture and motion, the kinetic depth effect. The cues were made less reliable by either randomly changing the shape of the texel or jittering the otherwise smooth motion of a cylinder shape. Goodale *et al.*⁷ trained gerbils to jump from one platform to another. They reported that a decrease in the size of the landing platform induced the gerbils to increase the number of their vertical head movements. The authors interpret this behavior as a compensation for the loss of information from the looming of the landing surface. To estimate the weights of cues empirically Young *et al.*⁶ proposed a technique they named perturbation analysis in which a small discrepancy is introduced in the depth depicted by one of the cues. The larger the change in the depth percept induced by the perturbation, the larger the influence of the cue (its weight). It is important not to introduce too-large perturbations, as the visual system is assumed to use a robust weighting rule, that is, to ignore or at least nonlinearly reweight cues if they are too discrepant from the other cues.

Assuming independence of cues and independent nor-

mally distributed internal noise for each cue, the minimum-variance unbiased estimator of depth is a weighted average. The weights in this average are proportional to the inverse of the variances associated with the cues.^{4,5} Ernst and Banks⁵ formulated depth-cue combination as a maximum-likelihood estimator (MLE) relating discrimination thresholds to the variance of the underlying estimator and assuming cumulative Gaussian psychometric functions. From these assumptions they predicted that the resulting variance of the combined cues estimator should be smaller than the variance corresponding to each cue alone, if observers combined haptic and visual information in a “statistically optimal” fashion in the MLE sense (in the present article we will use the term “optimal” in this sense). By asking subjects to discriminate the height of a bar with haptic and visual information (binocular disparity) they collected data supporting statistically optimal cue combination. To change the reliability of the visual cue, the authors changed the binocular disparity of the random dots used in the stimulus.

In a recent paper Rosas *et al.*⁸ showed that different texture types elicit different performance in a slant-discrimination task. This effect was observed by measuring the discrimination performance of the slant of a textured plane at different slant levels. An example of this effect is illustrated in Fig. 1. The obtained systematic rank-order based on performance provides a way of testing the reliability-based weighting: In cue combination the weight assigned to the texture cue should be larger

when a texture type that allows good slant-from-texture performance is mapped on the plane. Furthermore, if the more strict “statistically optimal” combination rule holds, the measured psychometric functions for the texture-only condition should be related to the weight for the texture cue [see Eq. (3) below]. The aim of this paper is to test both of these propositions using a slant-discrimination task.

In Section 2 we describe a slant-discrimination experiment in which the texture and haptic cues were depicting the same slant. This experiment allowed us to test the prediction of an unbiased-minimum-variance-estimate (optimal) model for the change in slant-discrimination thresholds when both cues are combined. In addition, the results were used to determine a range of discrepancies between cues when conducting the perturbation analysis experiment described in Section 3. The perturbation analysis experiment allowed us to determine empirically the weights assigned to each cue.

2. EXPERIMENT 1: SLANT DISCRIMINATION WITH CONSISTENT CUES

A. Stimuli

The stimuli used in this experiment consisted of slanted planes onto which different texture types were mapped. The (virtual) planes could be touched by the subjects with a PHANToM force-feedback device (SensAble Technologies) attached to their index finger. This device simulates the forces corresponding to the given plane creating in the observer a compelling sensation of a real slanted plane. The graphic rendering (perspective projection) and texture mapping were done with the OpenGL library. The setup was carefully calibrated such that haptic and graphic planes were aligned.

The texture types used in all our experiments were a subset of the ones described and tested by Rosas *et al.*⁸ We used three different textures: a texture composed of circles, which tends to allow the best slant-discrimination performance; a leopard skinlike pattern;⁹ and “Perlin” (coherent) noise¹⁰ (see Fig. 1).

B. Methods

The experimental setup was similar to the one used by Ernst and Banks⁵: The image from a SONY GDM 500 21-in. monitor was reflected on a mirror. The image subtended approximately 10 deg of visual angle and was monocularly viewed through an aperture of 2.6-cm diameter. The subject’s position was controlled by means of a head and chin rest located approximately 50 cm from the image. The subjects used an eye patch to obtain monocular viewing. The aperture was aligned such that the subject’s open eye was looking through the center of the viewing aperture.

Five subjects participated in standard temporal two-alternative forced-choice experiments. On every trial, one of the three conditions to be tested (texture only, haptic only, and texture and haptic cues) was selected randomly and independently. For the texture-only condition, the two images were shown sequentially for 1250 ms with a ramp-in, maximum-contrast, ramp-out timing of

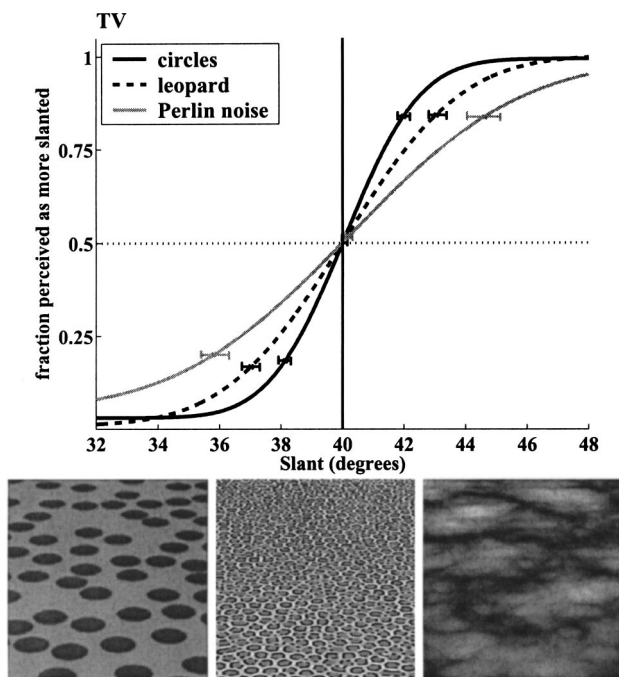


Fig. 1. Effect of texture type on slant discrimination. Top, the psychometric functions are significantly different for three texture types (error bars represent 68% confidence intervals). Bottom, texture patterns used to obtain the data. Left to right: circles, leopard skinlike, and Perlin noise textures. For this subject (TV) the task was easier when the circle texture was mapped onto the slanted planes while discriminating slant near 40 deg (the standard, depicted as a solid vertical line), reflected in the steepest psychometric function. Her worst performance was obtained with Perlin noise.

125 ms–1000 ms–125 ms to avoid afterimages. The interstimulus interval was 400 ms. No haptic object was available to the subjects in this condition.

For the trials in which the haptic cue was present (either a haptic-only or a haptic-and-texture trial), a small green sphere was shown on the center of the image (matching the center of the plane such that this position was the same throughout the experiment), indicating to the subject that she or he had to touch the surface in order to start the trial. Once the subject touched the surface, the green sphere disappeared from the screen and the graphic object was ramped in and out as described before. If no texture was to be shown in such trial, a gray field was projected onto the plane. Subjects were instructed to explore the plane freely during the trial. After the 1250 ms corresponding to the stimulus presentation time, both the graphic and the haptic objects disappeared. Subjects had to respond which interval contained the more slanted plane by pressing one of two (virtual) buttons, that were displayed on the screen after the stimuli had disappeared; the subjects' finger was not visible to the subject during the experiment. No feedback was provided.

Slant discrimination was tested for each of the three texture types described above against two comparison standards: 27 and 40 deg away from a horizontal plane, with the top part tilted in depth away from the viewer.

A combination of adaptive and constant-stimuli procedures was used to collect data: The adaptive procedure¹¹ 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. The final estimation of the psychometric functions was made with the combined data collected with both procedures, typically between 500 and 700 trials for each psychometric function. All the fits were forced to cross chance performance (50% correct in two-alternative forced-choice) at the slant level of the standard. The fits were obtained with the Psignifit Toolbox, which implements the constrained-MLE method described by Wichmann and Hill,^{12,13} with a cumulative Gaussian function as underlying shape. The psychometric functions were obtained by fitting the general expression

$$\psi(x, \alpha, \beta, \lambda) = \lambda + (1 - \lambda)F(x, \alpha, \beta), \quad (1)$$

where F represents the underlying function with range between 0 and 1, α and β the parameters of location and scale of the psychometric function, and λ the lapse rate. Wichmann and Hill¹² showed that level-independent lapses (e.g., eye blinks, wrong-button responses, etc.) can result in serious estimation errors in fitting psychometric functions (with MLE methods). Introducing λ as a (highly constrained) free parameter during the fit cures this bias problem. Thus not all psychometric functions asymptote at 1.0 on the data sets fitted (see Fig. 2). When obtaining discrimination thresholds from our fits [see Eq. (2)], we used the underlying function F , that is, the underlying performance of subjects when lapse rate is zero.

C. Results

In Fig. 2 we show typical examples of the psychometric functions obtained. The horizontal axis shows the slant in degrees and the vertical axis shows the fraction of trials

in which the comparison stimulus was perceived as more slanted than the standard (the value of the standard is indicated with a vertical solid line). Error bars correspond to 68% confidence intervals obtained from a parametric bootstrap.¹³ Different types of curves and symbols represent the performance for the different conditions: texture only, haptic only, and texture and haptic cues. The unbiased-minimum-variance-estimator model predicts that the psychometric functions for both cues together should be steeper than the psychometric functions obtained from either cue by itself, because the underlying estimator should have less variance than the estimator for single cues. In the examples shown in Fig. 2 only subject PR's data seem to be consistent with the optimal model.

According to the unbiased-minimum-variance model the following relation holds for single-cue discrimination thresholds and combined-cues thresholds (from Eq. (7) in Ref. 5):

$$\tau_{th}^2 = \frac{\tau_t^2 \tau_h^2}{\tau_t^2 + \tau_h^2}, \quad (2)$$

where τ_t stands for the discrimination threshold when texture was the only available cue, τ_h the discrimination threshold for haptic only, and τ_{th} when both cues were present in the stimulus. Here the thresholds are defined as the difference between the comparison stimulus judged more slanted 84% of the time and the point of subjective equality (PSE). It is important to note that when comparing the empirical results for combined cues with the predictions obtained with Eq. (2), we assume that the thresholds measured in the single-cue condition are representative of the reliability of the cue in the combined-cues condition. This critical assumption is not unique to our experimental design but is standard in all cue-combination research. As described in Subsection 2.B we did, however, attempt to discourage subjects from using different strategies in the single-cue and combined-cue conditions by randomly interleaving all conditions within a block of trials.

In Fig. 3 we depict the measured thresholds for each cue independently (open symbols) and for both cues combined (filled circles). In addition Fig. 3 shows the thresholds predicted by the statistically optimal model as filled squares. There are two plots per subject displaying the data for both standards tested. The vertical axes contain the thresholds on a log scale, while the texture type tested is on the horizontal axes. Note that for subject AO at standard 40 deg, the vertical scale is different from the other plots, because of her small thresholds. The error bars represent 68% confidence intervals estimated by a parametric bootstrap routine.¹³ By visual inspection of the overlap of the confidence intervals we see that in general our results are not consistent with the unbiased-minimum-variance model: Ignoring the six cases in which the prediction of the statistically optimal model cannot be distinguished from the performance based on a single cue, we observe 3 cases of optimal performance (subject PR, circles at 40 deg and Perlin noise at both standards) and 21 cases of suboptimal performance. The cases of suboptimal performance do not follow a single rule of

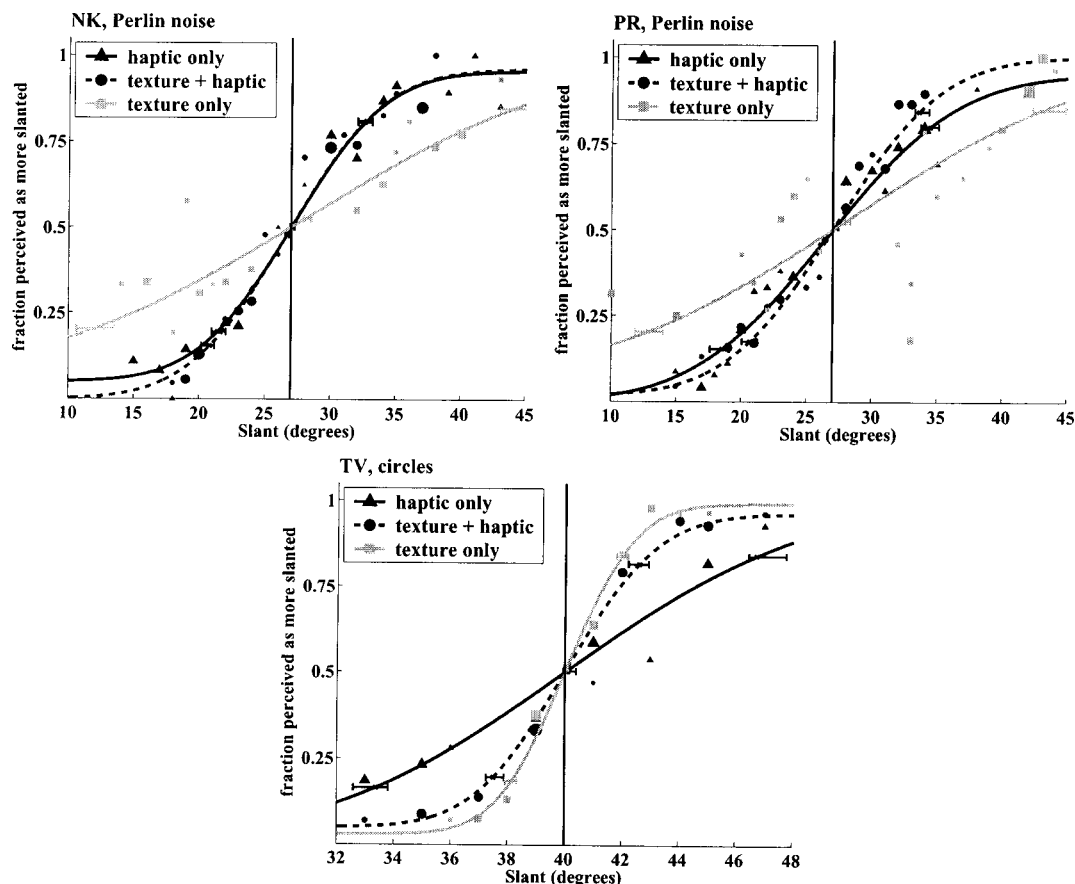


Fig. 2. Examples of the psychometric functions for slant discrimination obtained for texture only (solid gray curve for the fitted function, gray squares for data), haptic only (solid black curve for fit, black triangles for data) and texture and haptic cues (dashed black curve for fit, black circles for data) for subjects NK, PR, and TV. On each plot, the horizontal axis represents slant in degrees and the vertical axis shows performance as the fraction of the comparison stimuli perceived as more slanted than the standard. Standards are indicated with a vertical solid line. Error bars represent 68% confidence intervals. The size of the data points is proportional to the number of trials recorded.

combination: There are cases of performance between the thresholds for each cue. This is particularly puzzling since the observers could have done better by simply ignoring the weak cue. Also there are eight cases consistent with a veto rule in which subjects “followed” one of the two cues instead of combining them when both were present. In five of these cases they selected the best cue for the task.

Although we can conclude from these data that the weights were not statistically optimal, we cannot dismiss the reliability-based weighting in general. In Section 3 we describe a perturbation analysis experiment that we conducted in order to obtain the weights for both cues. This allows us to test whether the weights change following the reliability of the texture cue, albeit not statistically optimal.

3. EXPERIMENT 2: PERTURBATION ANALYSIS

A. Stimuli and Methods

The stimuli and methods used for this experiment were similar to the ones described for the previous experiment. The only difference was that a small discrepancy was introduced in the slant depicted by each cue in the stimuli

to allow the estimation of weights for each cue according to perturbation analysis. Perturbation analysis was introduced by Young *et al.*⁶ as a methodology to estimate the weight of a depth cue within the framework of the weak-fusion model. In this model, cue combination is proposed to be a statistically robust mechanism for parameter estimation. The robustness implies that the depth estimates derived from every cue are integrated linearly only when they do not differ substantially. If the discrepancy increases, then the influence of the most discrepant cue on the fused percept decreases, and “it should be no surprise that [the cues] interact in complex ways.” (Ref. 4, p. 395.) If the discrepancy is small, the weight of the perturbed cue is derived from the amount of change in the percept given the amount of perturbation (see Fig. 4). Thus if we are interested in studying the visual system under normal or “ecological” conditions, we should avoid large conflicts between cues.

Four subjects who participated in the previous experiment took part in the perturbation analysis experiment. Four levels of discrepancies were introduced. To determine the amount of discrepancy to be introduced we considered the performance of the particular subject in the consistent-cues experiment such that the maximum perturbation corresponded to approximately 80%–83% cor-

rect for the best condition observed for that subject in the consistent-cues experiment. For example, for subject TV and a standard of 40 deg, a maximum discrepancy of 2 deg was taken, considering her performance for texture only (texture circles; see Figs. 1 and 2). One slant level was tested per subject; we selected the one in which the differences in performance elicited by the texture types were larger. Typically 500 trials were collected per psychometric function.

B. Results

In Fig. 4 we show an example of the effect of the perturbation analysis and how the slope of the change of the PSEs defines the weight for a particular cue.

To estimate the weights of the cues we first fitted a family of parallel psychometric functions (“common model”) for all the (“individual”) data sets obtained for each observer for a perturbed cue, including the corresponding data set from the consistent-cues experiment (see Wichmann¹⁴ for further details). Given the linearity of the weak-fusion model—and the small perturbations introduced that should make the cue-combination mechanism operate in its linear range—the psychometric functions *should* be parallel, and our analysis did indeed show that: With the exception of the data of subject AO, the vast majority of the data sets (67 of 72) could be fitted sat-

isfactorily to their respective common model. [The common models obtained for AO’s data failed in our tests of goodness of fit even when only two data sets were considered for each perturbed cue. The failure of the parallelism could be explained by the robust-estimator effect. However, considering that two individual data sets of the perturbation analysis (0.75 deg difference) represent a perturbation level of less than 70% correct in her best condition in the consistent-cues experiment and 55% correct for her worst performance in that experiment, we think it unlikely that in such conditions the cues were so highly discrepant that the robust estimator effect was present. An alternative explanation is the effect of the slant level in the discrimination performance. This effect is such that the more slanted the plane, the easier the discrimination and the steeper the psychometric function.^{8,15} We have implicitly assumed that for the amount of perturbation we can dismiss this effect and assume parallelism between the psychometric functions. Such assumption seems justified by observing the results for the other subjects. However, AO’s data show that she is highly sensitive to change of slant (note her thresholds in the consistent-cues experiment and the small confidence intervals obtained for her data in this experiment). Then the “accelerating” performance for increasing slants might be observed for her even for small changes of slant,

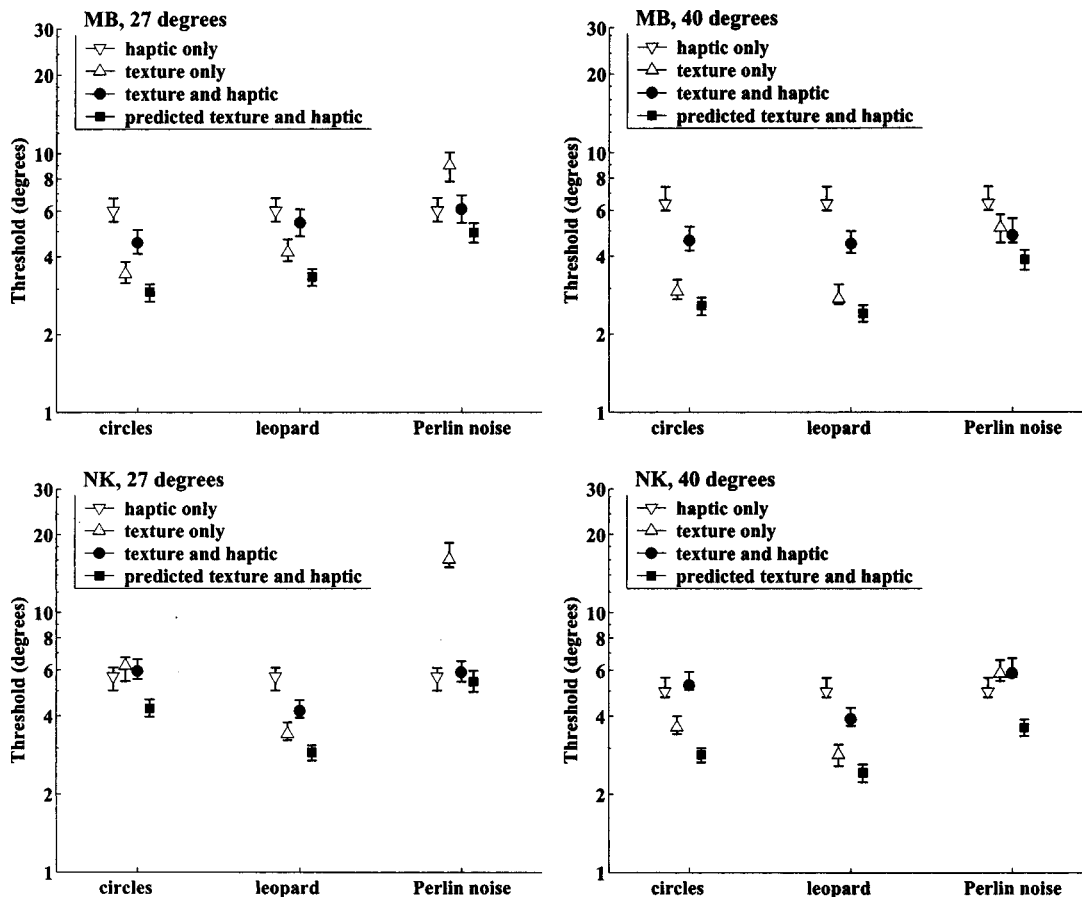


Fig. 3. Measured and predicted thresholds for slant discrimination with haptic and texture information provided by three different texture types for all five subjects. The vertical axes contain the thresholds in degrees on a log scale, and the texture types are on the horizontal axes. Error bars represent 68% confidence intervals estimated by bootstrap. Thresholds are defined as the difference between the stimulus judged 84% of the trials as more slanted and the PSE. (Fig. 3 continued next page.)

producing the failure of the parallelism. We opt to report the weights estimated for her data by using all the individual data sets obtained with the perturbation analysis, though they might represent a different mechanism than do the data observed for the other subjects.]

After determining the common model and the individual data sets consistent with the parallel hypothesis, we refitted the data sets, fixing their slope to that of the common model. From the new fits we obtained the PSEs to estimate the change of weight (see Fig. 5). We estimate a straight line connecting the PSEs by minimizing the

weighted least-squared error of the line fit, weighting each PSE point by its bootstrap standard deviation. To obtain normalized weights, we constrain the minimization process such that the weights for texture and haptic should be positive and add up to unity. As an example, in Fig. 5 we show the considered PSEs and the estimation of the weights for one subject. In Table 1 we summarize the results of weights for the texture cue for all subjects. We also include the weights predicted by the statistically optimal cue-combination model using the thresholds measured in experiment 1:

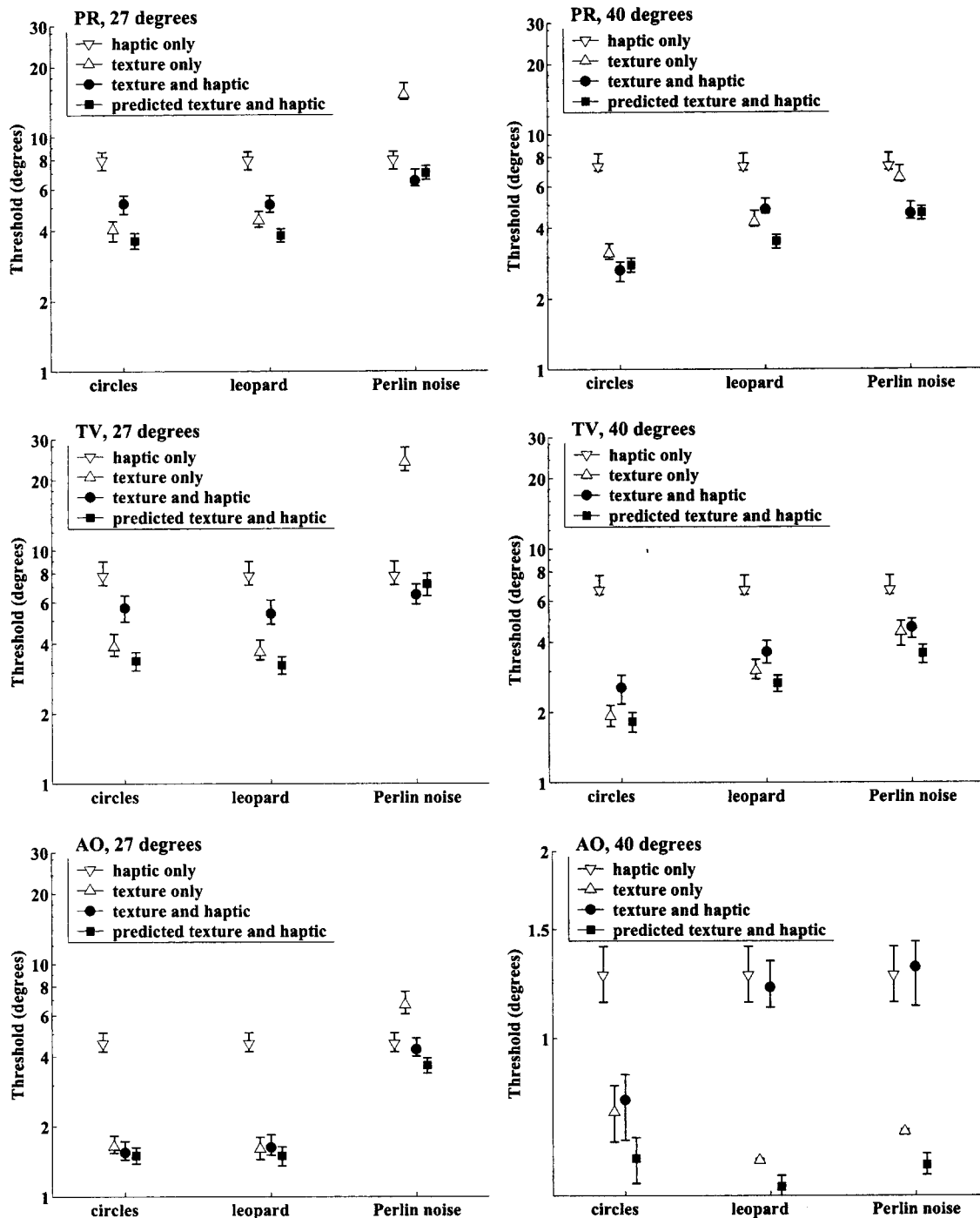


Fig. 3. (Continued).

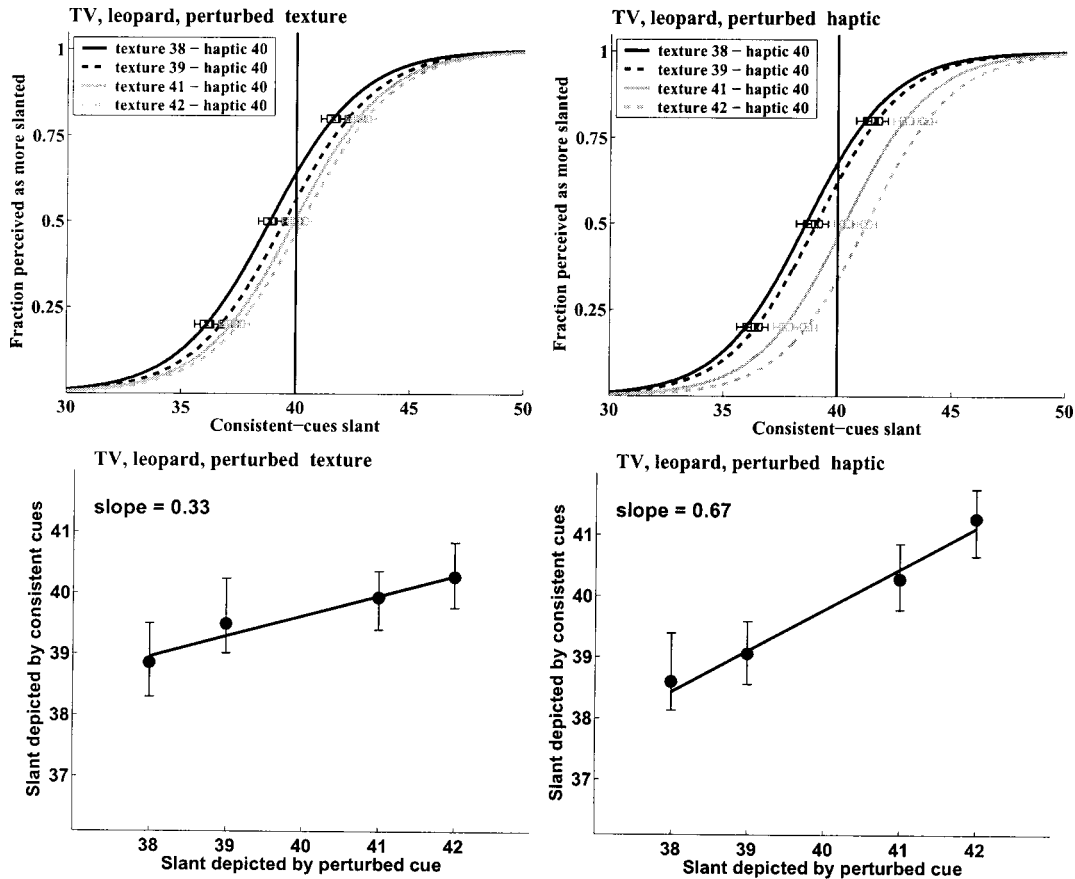


Fig. 4. Example of the effect of the perturbation analysis for subject TV. Top left, effect of introducing a discrepancy in the slant depicted by the texture cue while the haptic cue was depicting 40 deg. Top right, reverse, that is, perturbing the haptic cue while the texture cue was fixed at 40 deg slant. Bottom, PSEs obtained from the psychometric functions of the top row. The slope (obtained by weighted-square-root linear fit) represents the weight given to the perturbed cue. Error bars represent 95% confidence intervals. This figure is analogous to, for example, Fig. 7 of Landy *et al.*⁴

$$\omega_t^p = \frac{\tau_h^2}{\tau_t^2 + \tau_h^2}. \quad (3)$$

To assess the variability of the weight estimation, we computed the linear fit 1000 times using samples of the full distributions of bootstrap values of the PSEs obtained with Psignifit. To obtain a measurement of goodness of fit of the constrained straight line, we computed the mean of the normalized residuals of the linear fit over the full distribution of 1999 values for the PSEs obtained from Psignifit. In Table 1 we show the standard deviation of the 1000 estimated values for the weights and the mean value of the normalized residuals.

We observe that for three of the subjects the weight for texture cue (ω_t) diminishes as the texture type is less reliable for the task, following the prediction of a reliability-sensitive cue combination (note that in the rank order for NK, leopard skin was a better texture for the task than circles, while the reverse is in general the case for the other subjects as well as the rank order reported by Rosas *et al.*⁸). However, the weights do not change as predicted by the unbiased-minimum-variance estimator model (ω_t^p). In general, the weight given to the texture cue by the subjects is smaller than that predicted (ω_t^p) with the exception of subject AO, whose data seem consistent with visual capture.

4. DISCUSSION AND CONCLUSIONS

We have studied the problem of cue combination for depth perception using a slant-discrimination task. We were interested in testing predictions of the class of models that propose a weighted average as the cue-combination rule, with weights assigned to the cues according to the reliability of the sources of information. We introduced a manipulation of the reliability of the texture cue to slant that consists of using different texture types known to elicit different performance in slant discrimination.⁸ A particular formulation of the weighted-average model predicts a performance consistent with a minimum-variance estimator.⁵ We tested this model with a consistent-cues experiment and obtained results that do not support such a cue-combination rule in more than 80% of our data. A less strict prediction is that the weight of the texture cue diminishes for less reliable texture types. By performing a perturbation analysis experiment we estimated the weights for texture and haptic cues, observing a behavior generally consistent with such prediction. In short, our data suggest that the visual system is sensitive to the reliability of cues to depth, although the weights (representing the influence of a cue in the percept) are not statistically optimal in the sense of constructing an unbiased-minimum-variance estimator of depth. Alternatively, of course, there remains the possibility that subjects at-

tempted to combine cues optimally but did not have an accurate estimate of the variance of the individual cues.

A number of hypotheses, not mutually exclusive, might explain the failure of the statistically optimal cue combination in our case. First, the task used by Ernst and Banks⁵ was to discriminate the height of an object by grasping it, while our subjects used one finger to touch and freely explore a slanted plane. This difference in task induces potentially different dynamics of the combination of cues. In the study by Ernst and Banks⁵ the subjects acquired the full haptic percept faster than in our study, because our subjects had to explore the plane (in 1250 ms) to obtain the haptic-based information of slant. However, what to predict from this observation is not straightforward, because it would, at least, depend on the moment at which the decision is made by the observers: An “early” decision would weaken the haptic cue given that for our

subjects the haptic information was acquired over time. That is, however, inconsistent with our data, as most of the measured haptic weights are larger than predicted from the optimal model. A “later” decision, in which the haptic cue captures the combination would predict low texture weights, something that we find for three of four subjects, but it is not consistent with the data of subject AO (but see remark about subject AO’s data above) and is partially inconsistent with the data of subject PR. Clearly, studying the temporal dynamics of the cue-combination process is an important future challenge for cue-combination research.

Another difference between the studies is the type of visual cues used: While disparity can provide metric information about depth given knowledge of the viewing distance,⁴ this is not the case for the texture cue. It is thus not inconceivable that (precise) metric information may

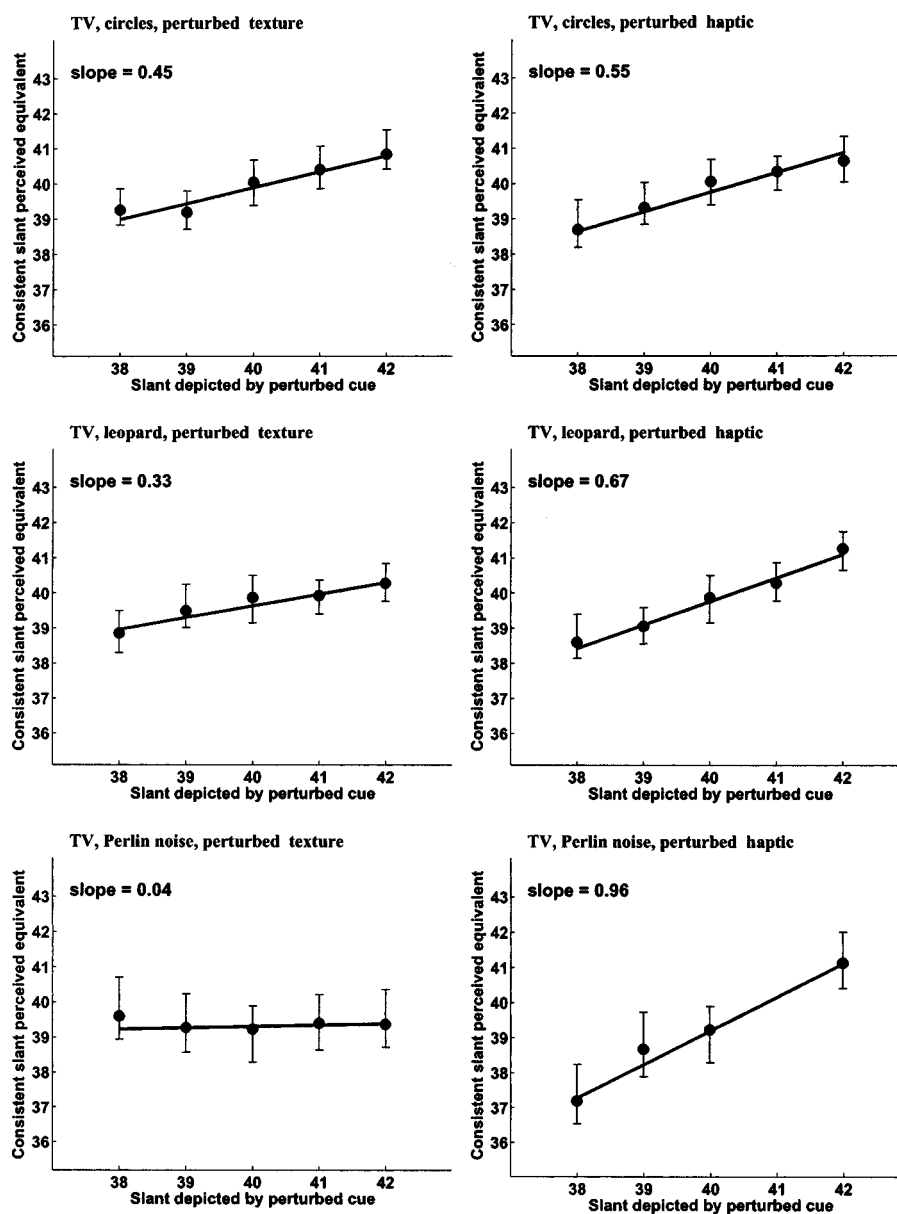


Fig. 5. Estimated weights as the slope of change in the PSE given a change in the slant depicted by the perturbed cue for subject TV: left column, perturbed texture; right column, perturbed haptic; one texture type per row. Error bars represent 95% confidence intervals obtained by bootstrap.

Table 1. Measured and Predicted Weights for the Texture Cue

Subject/ Types of Cue	ω_t			
	Mean	Standard Deviation ^a	Normalized Residual ^a	ω_t^p
NK				
Circles	0.03	0.04	1.12	0.45
Leopard	0.08	0.06	1.32	0.73
Perlin noise	0.00	0.00	1.05	0.11
PR				
Circles	0.80	0.04	1.46	0.85
Leopard	0.56	0.06	0.99	0.75
Perlin noise	0.53	0.08	0.80	0.56
TV				
Circles	0.45	0.06	0.96	0.92
Leopard	0.33	0.06	0.91	0.83
Perlin noise	0.04	0.06	0.84	0.70
AO				
Circles	1.00	0.00		0.89
Leopard	1.00	0.00		0.89
Perlin noise	1.00	0.00		0.32

^aThe standard deviations of the measured weights were computed by bootstrapping a constrained linear function, and the residuals (in standard deviation units) were computed by using the distribution of bootstrap values of the PSEs. (We show only the weights for the texture cue because the texture and haptic weights add up to unity.)

permit optimal cue combination, whereas we do not observe optimal cue combination for the nonmetric slant-from-texture and haptic cues. Knill and Saunders¹⁶ studied slant discrimination based on texture and horizontal disparity information. Though they state that “on average, subjects appear to weight the two cues optimally” (p. 2551), they could not test the optimal cue combination for their subjects individually because of the large uncertainty in the threshold estimates derived from their data (p. 2549).

We can thus state a number of differences between the studies that point to factors that may be necessary to produce the conditions in which an optimal combination of cues is observed. Oruç *et al.*¹⁷ studied linear perspective and texture gradients in a slant-judgement task and found evidence of optimal cue combination, optimal combination of correlated cues, and suboptimal cue combination. Those results and the results reported here suggest that optimal combination is only one possibility, but certainly not the combination rule under all conditions. Furthermore, an optimal cue combination requires that the observers be able to estimate or represent accurately (even on a trial-by-trial basis) the underlying variance of the available cues. An open question is how the visual system would implement a mechanism capable of obtaining such information for a wide range of tasks and stimuli.

Our results support a weaker mechanism that does not represent the exact variance of the cues for every possible case (though it might match the optimal combination in some cases). Such mechanism appears able to rank order approximately the reliability of a set of visual cues and construct a depth percept in a manner consistent with the weak-fusion model.

ACKNOWLEDGMENTS

Part of this research was presented as a talk at the Fall Vision Meeting, Tucson, Arizona, October 2003. Financial support was provided by the Research Council at the Katholieke Universiteit Leuven (IDO/98/002) and the Fund for Scientific Research-Flanders (FWO G.0189.02 and G.0095.03). We are indebted to two anonymous reviewers and our topical editor, Michael Landy, for their helpful comments.

Corresponding author Pedro Rosas's e-mail address is pedro.rosas@tuebingen.mpg.de.

REFERENCES

1. J. J. Clark and A. L. Yuille, *Data Fusion for Sensory Information Processing Systems* (Kluwer, Boston, Mass., 1990).
2. A. J. Parker, B. G. Cumming, E. B. Johnston, and A. C. Hurlbert, “Multiple cues for three-dimensional shape,” in *The Cognitive Neurosciences*, M. S. Gazzaniga, ed. (MIT, Cambridge, Mass., 1995), pp. 351–364.
3. L. T. Maloney and M. S. Landy, “A statistical framework for robust fusion of depth information,” in *Visual Communications and Image Processing IV*, W. A. Pearlman, ed., Proc. SPIE **1199**, 1154–1163 (1989).
4. M. S. Landy, L. T. Maloney, E. B. Johnston, and M. Young, “Measurement and modeling of depth cue combination: in defense of weak fusion,” *Vision Res.* **35**, 389–412 (1995).
5. M. O. Ernst and M. S. Banks, “Humans integrate visual and haptic information in a statistically optimal fashion,” *Nature (London)* **415**, 429–433 (2002).
6. M. J. Young, M. S. Landy, and L. T. Maloney, “A perturbation analysis of depth perception from combinations of texture and motion cues,” *Vision Res.* **33**, 2685–2696 (1993).
7. M. A. Goodale, C. G. Ellard, and L. Booth, “The role of image size and retinal motion in the computation of absolute distance by the Mongolian gerbil (*Meriones unguiculatus*),” *Vision Res.* **30**, 399–413 (1990).
8. P. Rosas, F. A. Wichmann, and J. Wagemans, “Some observations on the effects of slant and texture type on slant-from-texture,” *Vision Res.* **44**, 1511–1535 (2004).
9. G. Turk, “Generating textures on arbitrary surfaces using reaction-diffusion,” in *Proceedings of the Computer Graphics Conference (SIGGRAPH '91)*, T. W. Sederberg, ed. (1991), Vol. 25, No. 4, pp. 289–298 (www.SIGGRAPH.org).
10. K. Perlin, “An image synthesizer,” *Comput. Graph.* **19**, 287–296 (1985).
11. A. B. Watson and D. G. Pelli, “QUEST: a Bayesian adaptive psychometric method,” *Percept. Psychophys.* **33**, 113–120 (1983).
12. F. A. Wichmann and N. J. Hill, “The psychometric function: I. Fitting, sampling, and goodness of fit,” *Percept. Psychophys.* **63**, 1293–1313 (2001).
13. F. A. Wichmann and N. J. Hill, “The psychometric function: II. Bootstrap-based confidence intervals and sampling,” *Percept. Psychophys.* **63**, 1314–1329 (2001).
14. F. A. Wichmann, “Some aspects of modelling human spatial vision: Contrast discrimination,” Ph.D. thesis (University of Oxford, Oxford UK, 1999).
15. D. C. Knill, “Discrimination of planar surface slant from texture: human and ideal observers compared,” *Vision Res.* **38**, 1683–711 (1998).
16. D. C. Knill and J. A. Saunders, “Do humans optimally integrate stereo and texture information for judgments of surface slant?” *Vision Res.* **43**, 2539–2558 (2003).
17. I. Oruç, L. T. Maloney, and M. S. Landy, “Weighted linear cue combination with possibly correlated error,” *Vision Res.* **43**, 2451–2468 (2003).