Pooling speed information in complex tasks: Estimation of average speed and detection of nonplanarity

Maarten A. Hogervorst[1]
Department of Experimental Psychology, University of Oxford, Oxford, UK

Andrew Glennerster
Department of Physiology, University of Oxford, Oxford, UK

Richard A. Eagle
Department of Experimental Psychology, University of Oxford, Oxford, UK

To gain insight into how speeds are combined in structure-from-motion, we compared performance for estimating the mean speed and performance for detecting deviations from planarity. The stimuli showed a center dot surrounded by an annulus of dots. In one (plane) condition, the stimuli simulated a rotating plane. In a two alternative forced choice (2AFC) task, the subject had to choose in which of two stimuli the center dot moved in the plane. In another (cloud) condition, the same dot locations and speeds were used but now assigned to different dots. Such a stimulus resembles a translating and rotating cloud of dots. In this case, the subject had to choose the stimulus in which the center dot moved with the mean speed of the surrounding dots. Performance was measured as a function of deformation/slant. Although location and speeds were the same in both conditions, performance was much poorer in the cloud condition. Subsequent experiments and an ideal observer model point to a plausible explanation: in detecting deviations from planarity, the visual system can focus on the most reliable pieces of information (the slower dots, closest to the test dot). Although performance could benefit by taking more dots into account, performance barely improved with an increase in the number of dots. This may reflect a limited processing capacity of the visual system.

Keywords: structure from motion, depth, speed, spatial integration

Introduction

The importance of motion for the visual system stems largely from the fact that it contains valuable information about 3D layout and ego-motion (e.g., see Nakayama, 1985). Most psychophysical research into human processing of motion has been focused at the analysis of uniformly translating textures. However, the motion patterns associated with 3D structure-from-motion (SFM) and ego-motion are much more complex, and it is still largely unclear how human processing of uniform moving patterns is related to that for processing of more complex flow fields. Here we investigate how speed information is pooled over space in a SFM task.

In principle, a better estimate of the speed can be obtained by integrating over a larger area or more dots. However, human performance for discriminating speed (de Bruyn & Orban, 1988) or detecting changes in speed (Snowden & Braddick, 1991; Werkhoven, Snippe, & Toet, 1992) of a uniformly moving texture is found to be the same for stimuli containing a large number of dots as for stimuli containing only a single dot. These studies show that the visual system does not make effective use of the additional information supplied by the additional objects in the stimuli.

In more complex tasks, such as estimation of ego-motion or 3D SFM, it is essential to combine motion information from various locations and sometimes from different times. Also, the combination rule can be quite complex, as in SFM (see, e.g., Koenderink & van Doorn, 1991). However, this does not mean that the visual system uses the mathematically correct algorithm to estimate a property. Such a case is presented by Werkhoven and Koenderink (1991), who investigated human processing of angular 2D rotation. Their results suggest that subjects based their judgments of the rotation magnitude on the average of the speeds and did not take the eccentricities into account. Their study shows some improvement with increasing numbers of dots, although beyond 8 dots no further improvement is found. The latter finding suggests that some amount of spatial integration occurred, but that this is limited to a few dots. Verghese and Stone (1995, 1996, 1997) investigated human performance for speed discrimination using Gabor patches. Performance was found to increase with the number of patches (up to 6 patches were used). Intriguingly, no improvement was found when the area of a single Gabor patch was increased by the same amount. Their results suggest that performance improves with the number of independently treated entities rather than with the stimulus area.

DOI 10:1167/3.7.2 Received February 7, 2002; published August 19, 2003 ISSN 1534-7362 © 2003 ARVO

Other studies suggest that the visual system can integrate speed and direction information over a large number of dots. When humans are shown a stimulus containing many different local motion vectors, a unified global percept in the direction of the mean may arise if the range of component directions is 180 deg or less (Williams & Sekuler, 1984), and subjects can estimate the average direction within 1-2 deg (Watamaniuk, Sekuler, & Williams, 1989). Watamaniuk and Duchon (1992) performed an experiment in which subjects discriminated the average speed of two distributions of speeds with the same width. They found that thresholds were unaffected by the width of the speed distribution in the tested range. These results suggest that velocity can be averaged over many dots.

Performance in 3D SFM tasks is constrained by the accuracy with which 2D motions are represented within the visual system (e.g., Nakayama, 1985; Hogervorst, Kappers, & Koenderink, 1996). To make specific predictions, explicit assumptions have to be made about the accuracy with which speeds are represented. A simple assumption that has been used is that all speed measurements are independent. With assumptions about the magnitude of the noise in these measurements, the maximum accuracy with which structural properties and ego-motion can be deduced can be estimated (e.g., Koenderink & van Doorn, 1987). Similarly, one can determine for which amount of noise such a model reaches the same level of performance as the human subjects (e.g., Werkhoven & van Veen, 1995). Eagle and Blake (1995) have shown that the relative inability of subjects to estimate the depth of objects can be explained from the low accuracy of the visual system in processing accelerations. Hogervorst and Eagle (1998, 2000) have shown that misjudgments of the depth of objects can be explained from noise on the 2D motions (velocities and accelerations) when one takes into account the fact that certain 2D motions are more likely than others, when they arise from a rotating 3D object. The same model also explains thresholds for discriminating the depth of a rotating pair of rigidly connected hinged planes (Eagle & Hogervorst, 1999). In these analyses, estimates for the noise on the velocities and accelerations were derived directly from thresholds for discriminating speed and direction, and from thresholds for detecting changes in speed and direction of uniformly moving patterns. In their model, simple assumptions are used about the way in which these elementary motions are combined. To advance this approach further, it is necessary to determine how motion information from different spatial locations is combined.

This study is a first step in establishing how speeds are integrated across space in the recovery of surface structure. We determined human performance for two tasks that require the combination of different speed vectors. We compared human sensitivity to average speed with human sensitivity to detecting nonplanarity using stimuli that contained the same set of speeds and dot locations. This approach gives direct insight into the way in which speeds are pooled in both tasks. Finally, we compared the results with a model in which (independent) speed measurements are combined in an optimal way.

Methods

Tasks

Two stimuli were shown containing an annulus of moving dots and a central test dot. The subject indicated in which of the two stimuli (1) the test dot moved with the average speed of the surrounding dots or (2) the test dot lay in the plane defined by the movement of the surrounding dots.

Subjects

Three subjects participated in the experiments: MH (the author), who was fully aware of the objectives of the study, and JR and ES, who were naive to the objectives of the study. All subjects had normal or corrected-to-normal vision.

Apparatus

All stimuli were generated on a Silicon Graphics O2 workstation. They were presented on a 19-inch Silicon Graphics monitor whose screen resolution was 1,280 x 1,024 pixels at a frame rate of 75 Hz.

Stimuli

Measurements were obtained in two conditions in which the stimuli resembled a rotating plane and a rotating cloud of dots and will be referred to as plane and cloud conditions. For small rotation angles, both plane and cloud stimuli are compatible with rigid interpretations, in which depth is proportional to speed. For larger assumed rotations, they are not. (How small a rotation angle is compatible with a rigid rotation is a matter of tolerances [noise] in the visual system). Note, however, that the task does not require that the stimulus is interpreted as a (rigid) 3D object. Figure 1 shows schematically an example of both types of stimuli (see Figure 2 for demos).

In this section, the parameters in the standard settings (used in Experiment 1) are given. In a range of experiments, the effect of different parameters was investigated and changes from the standard settings are given in the appropriate sections. Each stimulus consisted of a central dot surrounded by an annulus of dots. All dots moved horizontally with different speeds. The dots were depicted as white dots, size 2 x 2 pixels, against a black background at high contrast, using standard subpixel interpolation.

To generate the stimuli in both the plane and the cloud conditions, a set of speeds and positions were chosen as if the dots were part of a rotating plane. In the cloud condition, the speeds were later assigned to different dots. In this way, we ensured that the stimuli in both conditions contained similar speeds and locations (see Figure 1).

Figure 2 shows example stimuli for the plane (a and b) and cloud (c and d) conditions. The process of generating a sequence began with the middle frame of the sequence, in which the dots were randomly positioned within an annulus. The rest of the sequence was created by applying an affine (shearing, stretching, and translation) transformation to the texture. In the *reference* stimulus, the center dot moved also in accordance with the affine transformation (Figure 2a and 2c). The speed of the center dot was also equal to the average speed. In the *signal* stimulus, the center dot moved either with a larger or a smaller speed (Figure 2b and 2d). The task of the subject was to indicate which of the two stimuli was the reference stimulus.

The speeds and locations were chosen as follows. In the first stage, the locations of the total of N number of dots (N = 49) in the middle frame of the stimulus sequence were chosen uniformly and randomly distributed within an annulus with inner radius of r_{min} and outer radius r_{max} (r_{min} = 100 pixels = 1.9 deg, r_{max} = 200 pixels = 3.8 deg). The angles of the positions in polar coordinates (r_i, α_i) were equally distributed from 0 to 360 deg with +/- 30% scatter: the angle α_i of a point with index i is given by

$$\alpha_i = \alpha_o + (i + 0.3 * (2 * random - 1)) / N * 360°,$$

in which *random* represents a random number between 0 and 1 and α_0 is fixed for all dots and randomly chosen. In the second stage, the center of mass was calculated and subtracted from the positions, such that the test dot, located at the center of the screen, coincided with the center of mass. A set of horizontal displacements S was assigned to the dots. The dots moved linearly from $x+S/2$ to $x–S/2$. The displacement S assigned to the dots was a linear function of the horizontal and vertical positions x and y:

$$S = Def * (x \cos\varphi + y \sin\varphi) + T, \quad (1)$$

in which Def is the deformation, φ the direction of deformation and T the overall displacement.

In the reference stimulus, the displacement of the center dot was T, and in the signal stimulus, the displacement of the center dot was $T + \delta S$. Because the test dot was in the center of mass, the test dot moved with the average speed in the reference condition, on which basis the signal could be discriminated from the reference stimulus.

In the plane conditions, the displacements were unaltered. In the cloud condition, the positions and displacements were initially chosen in the same way. However, the displacements were assigned to different dots (i.e., the displacements were interchanged, such that

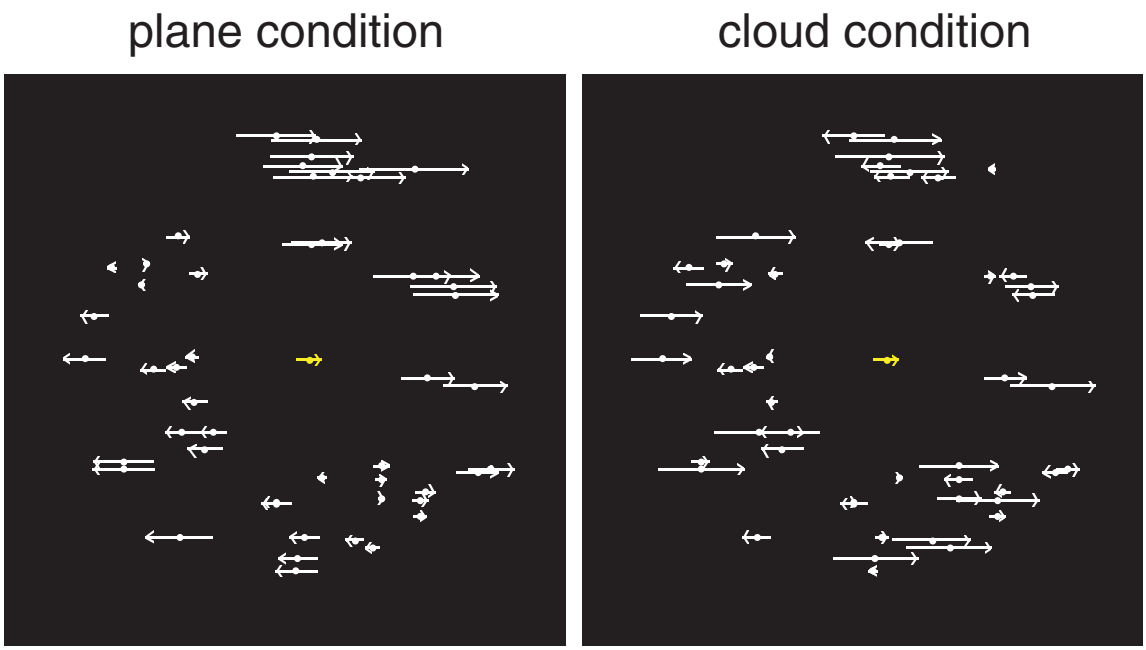

Figure 1. Flow fields of the stimuli used in the plane condition and in the cloud condition. The arrows indicate the displacements (velocities); the dots indicate the average position. Between the first and last frame of the stimulus sequence, the dots were displaced with constant velocity. The set of speeds and positions is the same in both conditions. In the plane condition, the velocities are consistent with a rotating plane; in the cloud condition, the speeds are assigned to different dots, and the stimulus is perceived as a rotating cloud of dots.

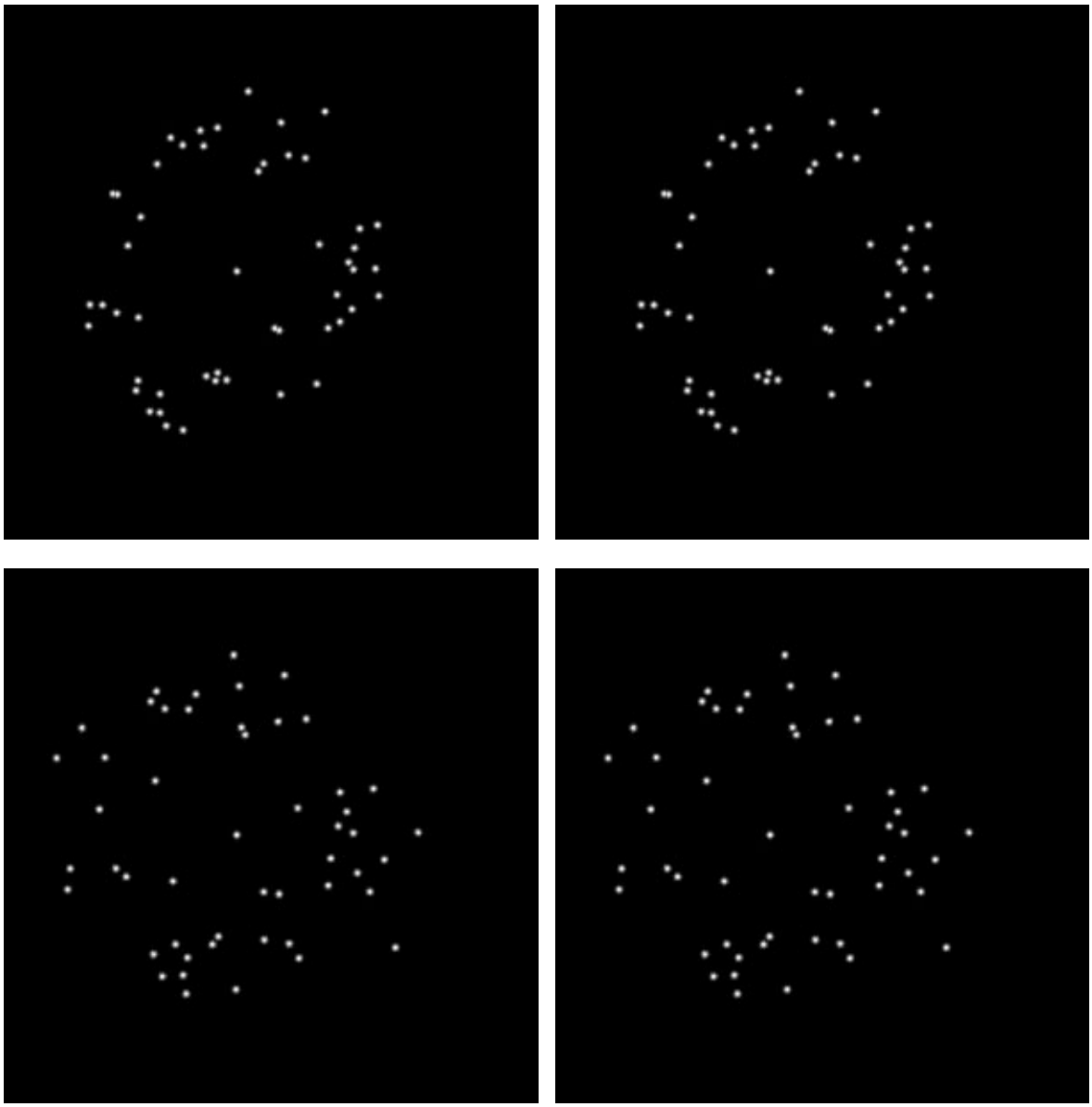

Figure 2. Example stimulus movies for a deformation *Def* of 0.5, a translation *T* of 0.75 and a speed difference δT of 0.4. Figure a) shows the reference plane stimulus, b) the signal plane stimulus, c) the reference cloud stimulus, and d) the signal cloud stimulus.

the displacement of dot 1 was assigned to dot 6 and the displacement of dot 6 was assigned to dot 1). This ensured that the positions and the displacements were the same in both conditions.

In the standard setting, the direction φ was chosen randomly. A φ of 0 leads to a horizontal compression (horizontally tilted plane), while a φ of 90 deg leads to a horizontal shearing motion (a vertically tilted plane). In the standard setting, the overall displacement *T* was randomly chosen between 0.47 and 1.42 deg (corresponding to mean speeds between 1.42 and 4.27 deg/s). Each stimulus sequence consisted of 25 frames and lasted 0.33 s.

Figure 2 shows example stimuli of a reference plane stimulus (a), a signal plane stimulus (b), a reference cloud stimulus (c), and a signal cloud stimulus (d), for which $T = 0.75$, $\varphi = 45$, $Def = 0.5$, $\delta T = 0.4$. Note that in each trial each sequence was shown only once.

Procedure

A two interval forced choice (2IFC) design with an adaptive staircase method with a maximum likelihood procedure (Snoeren & Puts, 1997; Watson & Pelli, 1983) was used to determine the threshold displacement δT at which the subject could discriminate the signal stimuli from the reference stimuli with 81% probability. At any

given trial, the absolute value of the offset displacement δT was set to the maximum likelihood estimation of the threshold and its sign was chosen randomly. At the start of each staircase, δT was set to twice the estimated threshold level. The subject was seated in a dimly lit room at 70 cm from the screen with the left eye covered and the right eye aligned with the center of the screen, with his/her head on a chin rest. At each trial, a reference stimulus and a signal stimulus were shown in random order separated by a blank interval (showing a black screen) that lasted 0.4 s. After this, the subject indicated which of the two stimuli represented the reference stimulus by pressing the left or right mouse button. In the plane condition, the reference stimulus was defined by the fact that the test dot moved with the local speed of the plane, as well as with the average speed of the surrounding dots. In the cloud condition, the reference stimulus was defined by the fact that the test dot moved with the average speed of the surrounding dots. Feedback was provided in the form of a tone that sounded after a wrong answer was given. This ensured that subjects were using a (near) optimal strategy in each condition. In each session, thresholds were determined for several conditions simultaneously in which the conditions were randomly interleaved. In each condition, a threshold was calculated after 80 trials. The thresholds presented are the geometric averages (average on a logarithmic scale) of the thresholds obtained in five or more sessions. Errors correspond to SEM of these values. We also present the geometric average of the thresholds of the three subjects (labelled as "Average"). In the latter case, the error estimates presented in the figures correspond to the square root of the sum of squared (individual) errors.

The thresholds and radii are reported in dimensionless units, in which 1 unit equals 100 pixels (the inner radius of the stimulus in the standard setting), 1.9 deg, or a speed of 5.7 deg/s (in case of the thresholds). This would especially make sense if performance would be scale independent (i.e., independent of the viewing distance). This assumption seems reasonable, considering that scale independence holds for many tasks, including visual acuity, contrast sensitivity, speed discrimination thresholds, magnitude of the motion after effect, receptive field size, etc. (see, e.g., Johnston & Wright, 1985). We discuss this issue further in the "Results" section of Experiment 2a. Regardless of whether performance is scale independent, the thresholds can be transformed into whatever units are preferred.

The deformation broadens the width of the speed distribution (in the standard setting the width [SD] equals about 5 times the deformation in deg/s; i.e., for deformation of 1, the width equals 5 deg/s). The speeds are not normally distributed. The shapes of the distributions used in the various experiments are shown in Figure 6.

Experiment 1: Standard Conditions

Experiment 1a: The Test Dot in the Center

In the first experiment, thresholds were compared for the cloud and the plane conditions. Thresholds were measured for deformations of 0, 0.1, 0.2, 0.4, and 0.8 (corresponding to a width in the speed distribution of 0, 0.5, 1.0, 2.0, and 4.0 deg/s). Thresholds for the plane and cloud conditions were measured in separate sessions. Within each session, thresholds were obtained for all deformations by interleaving the staircase procedures.

Results

Figure 3 shows the thresholds as a function of the deformation for both conditions for all subjects. Thresholds increase with increasing magnitudes of deformation. For larger deformations the thresholds increase approximately in proportion to the amount of deformation (i.e., a slope of one in Figure 3).

Thresholds for the cloud condition are much higher than for the plane condition. For a deformation of 0.8, the thresholds are on average 2.1 times higher for the cloud condition than for the plane condition. These factors are 1.5 (MH), 3.3 (JR), and 1.8 (ES) for the individual subjects. All subjects show the same trends, although the absolute levels differ considerably (thresholds of JR and ES are about twice as high as the thresholds of MH).

The magnitude of thresholds obtained in the cloud condition is surprisingly high. In fact, these thresholds approach the largest relative speeds in the distribution (dotted line in Figure 3). This means that subjects can only reliably indicate in which of the two stimuli the test dot moves with the average speed when the speed in the signal stimulus approaches the fastest speed or slowest speed in the distribution (i.e., when the speed of the test dot is at the edge of the distribution of speeds).

That such large differences exist between the cloud conditions and the plane conditions may come as a surprise, because both stimuli contain a similar set of speeds and locations. Indeed, as will be shown later, these results are difficult to reconcile with a model in which the speed information of *all* dots is optimally combined. Instead, the results of Experiment 2 and the modelling exercise will show that these results are consistent with the idea that the visual system focuses on the most relevant pieces of information in the stimulus.

For a deformation of zero, the thresholds obtained in the plane and the cloud conditions are about the same. This is to be expected because the stimuli are the same: all dots move with the same speed. The fact that thresholds are somewhat higher (for MH and ES, but not for JR) in the cloud condition suggests that subjects use a somewhat

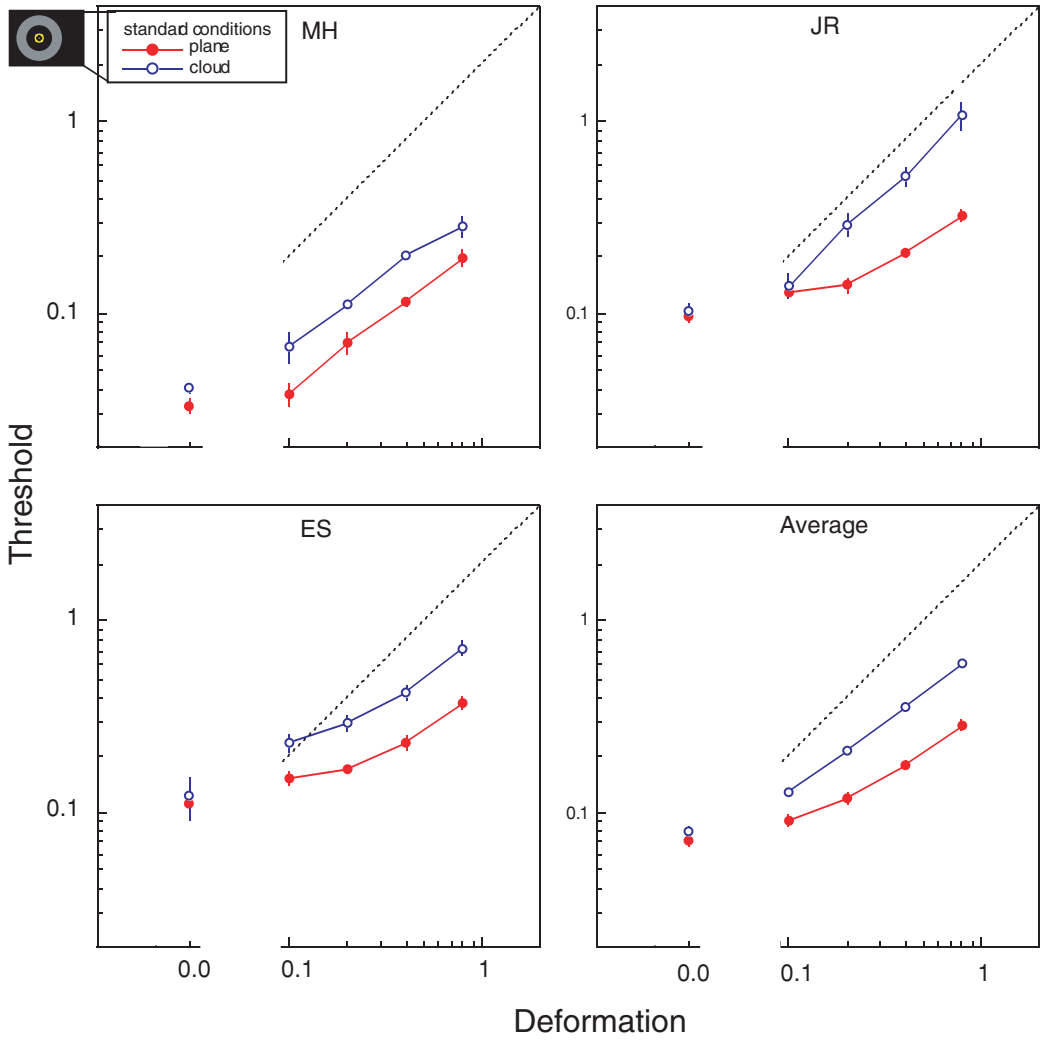

Figure 3. The threshold displacement as a function of the deformation for the plane and the cloud stimuli under standard conditions for the three subjects and the (geometric) average. The straight dotted line corresponds to a level for which the speed of the center dot equals the largest or smallest speeds in the distribution. Its slope of one corresponds to Weber behavior. Note also that the data for a deformation of zero are plotted on the left (in the otherwise logarithmic plots).

different strategy in the two cases. Still, this is a rather small effect.

The thresholds increase with increasing deformation. This is one of several reasons why, in our model, we assume that judgments are based on the relative speeds of the dots (i.e., relative to the speed of the test dot) and that the uncertainty increases with an increase in speed. Note that we allowed subjects to track the dots with their eyes, which means that the retinal speed of the dots is not known. However, the relative speeds of the dots are known, and there are indications in the data that these were the key factor determining performance. This can be seen by looking more closely at the percentage correct score as a function of mean speed. The mean speed was randomly chosen between 1.4 and 4.3 deg/s (0.25 and 0.75 in dimensionless units). To determine whether the mean speed influences performance, we calculated the fraction of correct answers over those trials in which the dots moved with a given mean speed. We calculated the fraction correct for low (1.4 to 2.4 deg/s), medium (2.4 to 3.3 deg/s), and high mean speeds (3.3 to 4.3 deg/s). The fractions correct are plotted in Figure 4. For subjects JR and ES, performance on trials in which the mean speed was low was somewhat better than on trials in which the mean speed was high. Subject MH, on the other hand, shows no effect of mean speed. The corresponding d' values decrease by 15% (JR), 0% (MH), 17% (ES), and 10% (on average). The stimulus duration was relatively short (333 ms) and the speed direction was randomized. The fact that performance was somewhat worse for faster mean speeds is predictable given that higher retinal velocities are associated with higher speed discrimination thresholds (e.g., de Bruyn & Orban, 1988). It is likely that the mean velocity is not fully nulled by eye movements, given that tracking accuracy deteriorates with increasing speed (Collewijn & Tamminga, 1984). Still, the effect is

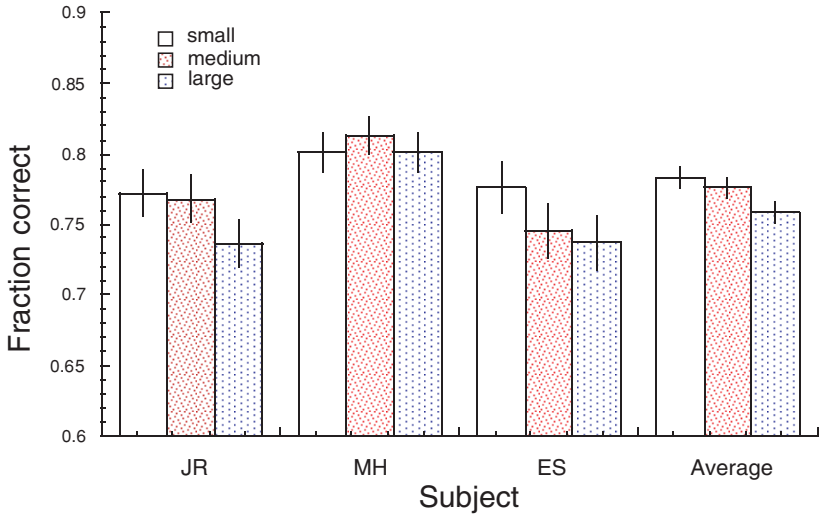

Figure 4. The fraction correct of all answers given in the standard conditions with the mean speed falling within a certain range for the three subjects and the average over all subjects. The fractions correct were calculated separately for slow mean speeds (1.4 to 2.4 deg/s), medium speeds (2.4 to 3.3 deg/s), and high mean speeds (3.3 to 4.3 deg/s).

relatively small given that the mean speed in the category with the highest speeds is twice as high as the mean speed of the category with the lowest speeds. The thresholds are certainly not proportional to the mean speed. The effect of mean speed is small relative to the influence of the deformation. In summary, performance appears to be largely determined by *relative* rather than *absolute* speed.

Experiment 1b: The Test Dot Outside the Center

To make the test dot in the reference stimulus move with the average speed and with the local speed of the plane (in the plane condition), the dot was put in the center of the stimulus in the standard settings. This may represent a special case. In this experiment, we tested the importance of having the test dot in the center of mass. For this purpose, we obtained thresholds for stimuli in which the test dot was outside the annulus, at 300 pixels to the right of the center of the screen (5.7 deg, x=3 in dimensionless units). As before, in the cloud condition, the test dot moved with the average speed of the other dots in the reference stimulus. In this setting, two different types of plane conditions were run. In the first condition, the signal stimulus was defined by the fact that the test dot moved with the average speed of the other dots. Of course, this speed differed considerably from the local speed of the plane. Data for this condition are shown as closed squares in Figure 5 ("planeOutAv"). In another condition, the subjects had to choose the stimulus in which the test dot moved with the local speed of the plane (i.e., as if the dot was on the plane containing the other dots). Data for this condition are shown as closed red triangles in Figure 5 ("planeOutJP," in which JP stands for "judge plane"). To facilitate comparison, the data of Figure 3 (Experiment 1a) is replotted alongside the data of Experiment 1b in Figure 5.

Results

First, notice that the thresholds obtained in the conditions with the test dot outside the annulus (open blue squares in Figure 5, "cloudOut") are higher than those with the test-dot in the center (open blue circles in Figure 5, "cloud"). The thresholds are on average 1.6 times higher. This shows that not only the distribution of speeds, but also the spatial distribution is important, even though, in principle, the locations can be discarded when estimating average speed.

Second, notice that the thresholds for estimating average speed are even higher when the stimulus depicts a rotating plane (closed squares in Figure 5, "planeOutAv"). The thresholds are on average 2.1 times higher than the thresholds obtained in the standard cloud condition (open blue circles in Figure 5, "cloud"). Subjects might in this case confuse the average speed and the local speed of the plane. Alternatively, subjects may base their estimate of the average speed on part of the stimulus. This would not impair performance much in the cloud condition, because each part contains a representative sample of the speed distribution, whereas this is not the case for the plane condition.

Thirdly, thresholds for judging the local speed of the plane are higher with the test dot outside the center (closed red triangles in Figure 5, "planeOutJP") than with the test dot in the center (closed red circles in Figure 5, "plane"). They differ on average by a factor 2.7. This is to be expected because it involves extrapolation that is generally less robust than interpolation.

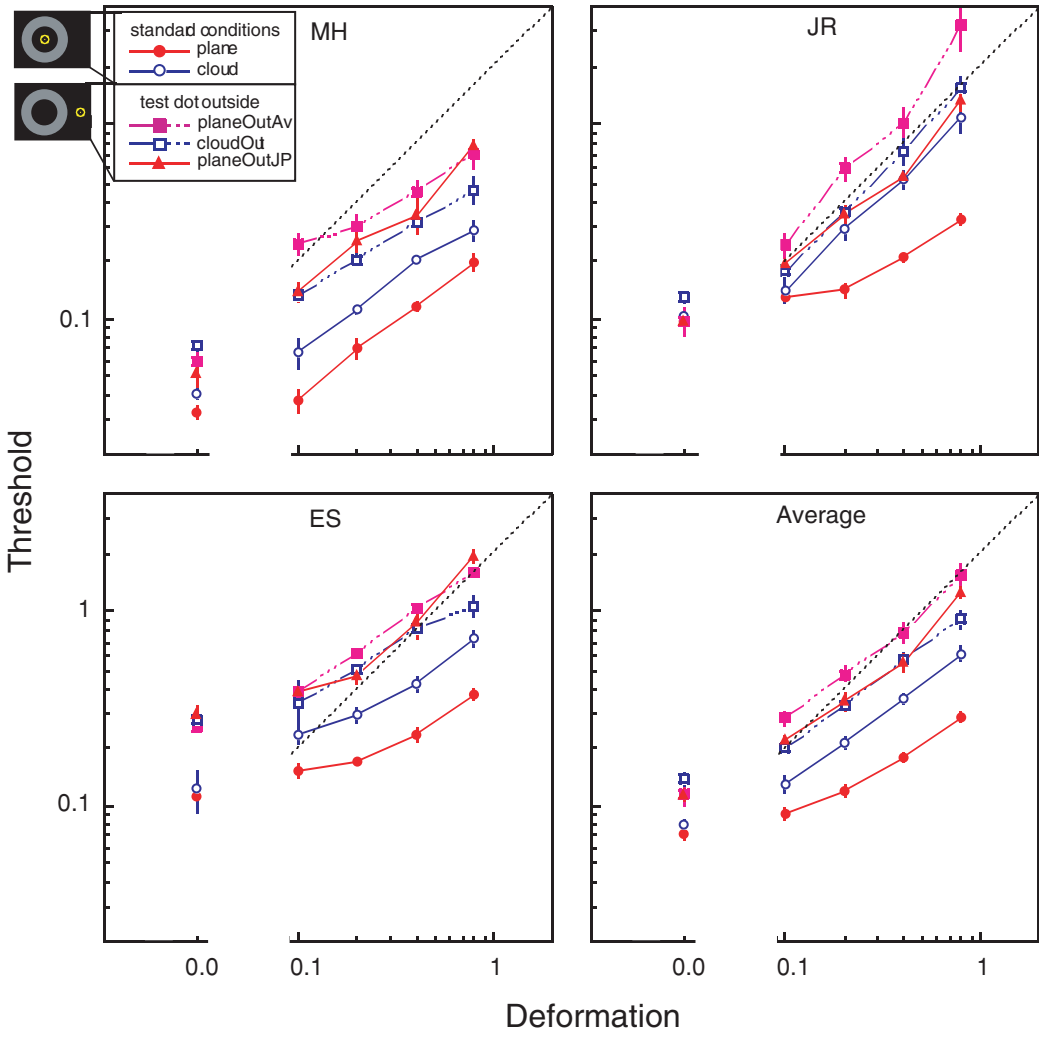

Figure 5. The threshold displacement as a function of deformation for the conditions in which the dot was placed outside the annulus, along with standard conditions for the three subjects and the (geometric) average. Condition "planeOutAv" refers to a condition in which the stimulus had to be chosen in which the dot moved with the average speed (of the dots in the annulus), while the stimulus represented a rotating plane. In condition "planeOutJP," the stimulus had to be chosen in which the test dot moved with the local speed of the plane at the position of the test dot. The straight dotted line corresponds to a level for which the speed of the center dot equals the largest or smallest speeds in the distribution. Its slope of one corresponds to Weber behaviour. Note also that the data for a deformation of zero are plotted on the left (in the otherwise logarithmic plots).

Experiment 2: Which Dots are Used?

We wished to investigate further how the planar configuration of dots in Experiment 1a conferred an advantage. One possibility is that certain dots provide more useful information and that, in the plane condition, these dots are easier to find. Here, we determined whether the judgments were based on part of the stimulus and, if so, which parts. This was done by measuring thresholds for stimuli in which the dots were restricted to particular parts of the visual field.

Experiment 2a: Which Eccentricities Contribute Most?

In this experiment, we varied the inner and outer radii of the annulus. The deformation was chosen around 0.4, and it took a random value between 0.3 and 0.5. The deformation value was jittered to obtain a similar level of uncertainty about the deformation as in Experiment 1, in which the conditions with different deformations were mixed. Experiment 1 showed that the threshold varies approximately linearly with deformation (especially for small deviations from the average). Therefore it can be assumed that the threshold obtained in this mixed condition is close to the threshold for a deformation of

0.4. In each session, thresholds were measured in 5 conditions, for which the inner and outer radii (r_{min}, r_{max}) differed. These were respectively (1, 1.5), (1, 2), (1, 3), (2, 3) and (2.5, 3), in which 1 unit corresponds to 1.9 deg, and in which (1, 2) resembled the standard setting. Note that for the first 3 conditions, the inner radius is the same, whereas for the last 3 conditions, the outer radius is the same. In all conditions, the dot density was kept constant and was the same as in the standard setting.

The speed distributions and spatial distributions are schematically shown in Figure 6. The speed and spatial distributions are similar. This is because the fraction of dots with a certain relative speed is proportional to the cross-section at a certain x-value. To keep the dot density constant, the number of dots N in the annulus was 20, 49, 132, 82, and 45 for conditions 1 to 5.

In the plane condition, the speed distribution is linked to the spatial distribution. In the cloud condition, the speed distribution and spatial distribution can be decoupled. We tested the effect of each in turn. Thus, in one set of sessions, the speed distribution was the same as in the standard setting, while the spatial distribution was varied, in the same way as in the plane conditions used in this experiment (i.e., the annuli shown in Figure 6). Data for this condition are shown as open diamonds in Figure 7 ("cloudpos"). In another set of sessions, the spatial distribution was the same as in the standard setting (i.e., with r_{min} = 1, r_{max} = 2), while the speed distribution was varied, that is the same as in the plane conditions used in this experiment (i.e., the lines shown in Figure 6). Data for this condition are shown as open blue squares in Figure 7 ("cloudspeed"). In this condition, the number of dots was kept constant at N = 49.

Results

Figure 7 shows the thresholds for the 5 annuli shown in Figure 6, for the plane condition, and the two types of cloud condition described above. In the plane condition (closed red circles in Figure 7, "plane"), performance does not improve when the outer radius is increased while the inner radius is held constant (first 3 annuli). When the inner radius is increased, thresholds rise (last 3 annuli). This shows that subjects rely largely on the dots closest to the test dot. Performance does not improve when the plane is extended outwards. This confirms that subjects are probably using the dots closest to the test dot to carry out the task.

In the cloud condition, when the speed distribution was held constant and only the spatial distribution was varied (open diamonds in Figure 7, "cloudpos") thresholds are (on average) constant, for all the annuli we tested. This is in agreement with the idea that, in principle, only the speed distribution is important to solve the task. Moving the surrounding dots to a somewhat larger eccentricity does not influence the thresholds. In the cloud condition, when the spatial distribution was held constant and the speed distribution was varied (open blue squares in Figure 7, "cloudspeed"), thresholds gradually increase going from left to right in Figure 7. The threshold increases when the outer radius is increased (first 3 annuli) as well as when the inner radius is increased (last 3 annuli). This suggests that judgments are not based on a particular (fixed) subset of dots.

In the plane conditions, the average thresholds are 47% higher in the 2→3 conditions than in the 1→1.5 conditions. On the basis of scale independent performance, one might have expected the thresholds to be twice as large. That this was not found is probably because the stimuli in the two conditions are not fully scaled versions of each other (varying in dot size, speeds, and the number of dots). In particular, the fact that the number of dots in the 2→3 annulus conditions was much higher than in the 1→1.5 annulus conditions may have played a role (82 vs. 20 dots, respectively). This

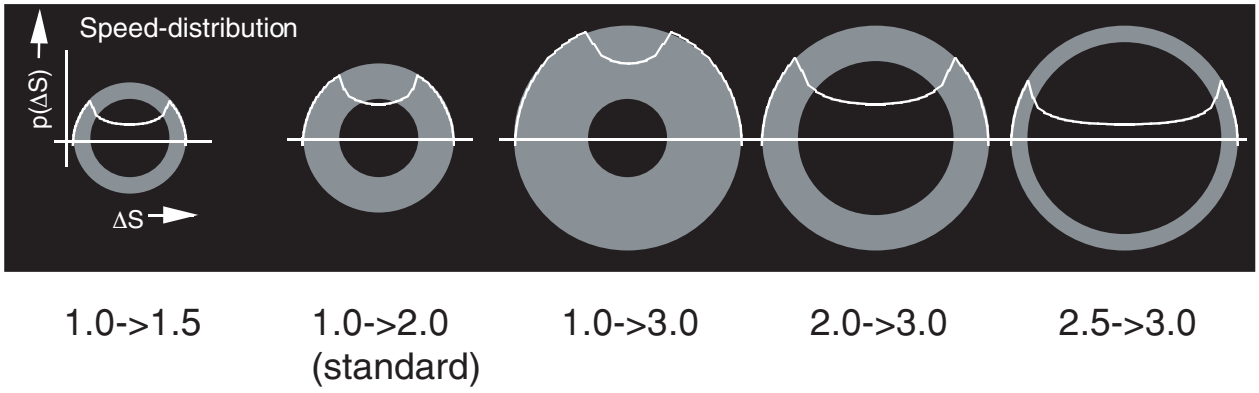

Figure 6. A schematic drawing of the distributions of locations and speeds used in Experiment 2a. The gray annuli represent the distributions of locations. The white line drawings show what the relative speed distributions look like. The location and speed distributions are similar. In the plane conditions, these distributions are linked. In the cloud conditions, these can be varied independently. In one set of conditions, "cloudpos," the speed distribution was held constant while the position distribution was varied, while in "cloudspeed," the position distribution was held constant and the distribution of speeds was varied.

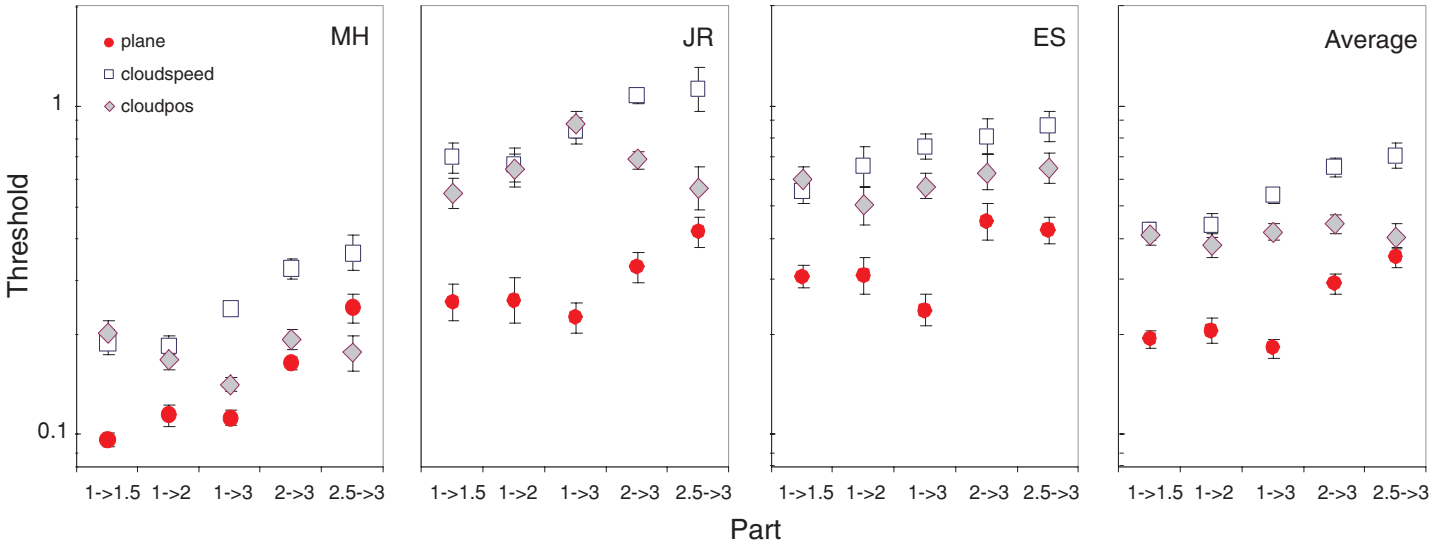

Figure 7. The thresholds for the various distributions of speeds and locations sketched in Figure 6 for the 3 conditions. Plotted are the data of the three subjects and their (geometric) average.

hypothesis fits with the findings of Experiment 3 that show that performance improves (by a small amount) as the number of dots is increased. This (small) improvement may account for the fact that the thresholds in the 2→3 conditions are not twice the thresholds in the 1→1.5 conditions.

Experiment 2b: Which Segments Contribute Most (Only Planar Conditions)?

Experiment 2a showed that in the plane condition, subjects base their judgments primarily on the dots that are closest to the test dot. In this experiment, we investigated whether this could be narrowed down further. We obtained thresholds in the plane condition only for stimuli containing dots in certain segments. As in Experiment 2a, the deformation was chosen randomly between 0.3 and 0.5. The direction of the deformation φ (see Equation 1) was either approximately 0 or 90 deg (i.e., it was either at a random angle between −10 and +10 deg, or between 80 and 100 deg). (A direction of 0 deg leads to a horizontal compression, while a direction of 90 deg leads to a horizontal shear transformation). To produce the stimuli, we generated a set of locations and speeds as in Experiment 1a (standard settings). Then a subset of the dots was shown whose locations fell within a certain segment of the annulus. The segments in which the dots were shown were either around an angle α of 0 and 180 deg, or around 90 and 270 deg, and were 45 deg wide. A total of 4 conditions were used varying in φ and α: $(\varphi, \alpha) = (0, 90), (90, 0), (0, 0)$, and (90, 90), schematically depicted in the top of Figure 8.

Results

Figure 8 shows the thresholds for various combinations of deformation directions and segment types, along with the thresholds obtained in Experiment 1 for a deformation of 0.4 (i.e., with the full annulus visible and the deformation direction chosen at random). The thresholds for the conditions with $(\varphi, \alpha) = (0, 90)$ and (90,0) are about the same as those obtained in the standard settings, suggesting that these segments contain sufficient information to carry out the task as well as in the standard condition. The thresholds for the conditions with $(\varphi, \alpha) = (0, 0)$ and (90, 90), on the other hand, are significantly higher. The average data on the right show that performance in these conditions is equally poor in these two conditions (green columns). The stimuli in conditions (0, 90) and (90, 0) (i.e., data shown by the red columns) contain dots that move with the lowest relative speeds, whereas the stimuli in conditions (0, 0) and (90, 90), shown by the green columns, contain dots moving with high speeds. This suggests that in the plane condition, performance is determined mainly by those dots that have the smallest relative speeds. Performance does not improve when other dots are shown as well.

The combined results from Experiment 2a and 2b indicate that in the plane condition, judgments are mainly based on the dots closest to the test dot moving with the slowest (relative) speeds. Although, in principle, it would be possible to improve performance by using more dots, the visual system appears to be incapable of this. Instead, it focuses on the best pieces of information. Assuming that the uncertainty in the speed measurements increases with increasing relative speed, it is best to use the dots with the smallest relative speeds. Also, because in general it cannot be assumed that the stimulus underlies a

perfect plane, it makes sense to use the dots closest to the test dot to infer the local speed of the plane.

In the cloud condition, judgments are largely based on the speed distribution and are largely (but not fully) independent of the spatial distribution. Judgments appear not to be based on any *particular* subset of dots (see "cloudpos" results in Figure 7). This might be because for estimation of the average speed all dots are equally informative.

Experiment 3: Effect of the Number of Dots

Models in which the noise on the different speed samples is independent from one another predict a decrease in the threshold with an increase in the number of dots. As will be shown later, a model that combines such measurements optimally predicts that thresholds decrease with an increase in the number of dots, proportional to $1/\sqrt{N}$ (sometimes referred to as probability summation).

In Experiment 3, we varied the number of dots. As in the original Experiment (1a), the test dot was in the center of mass in the plane condition. In the cloud condition, we did not apply this constraint to allow for situations in which, apart from the test dot, the stimulus contained only one other dot (such a constraint would make the task fairly trivial when $N = 2$, because then the task would amount to determining whether all 3 dots are part of the same line). A control experiment showed that the use of this constraint did not change the thresholds for a number of dots larger than 2 (note that the test dot is close to the center even when this constraint is not applied).

Thresholds were obtained in 3 conditions. In a plane condition and a cloud condition, the deformation was chosen around 0.4, randomly between 0.2 and 0.6. In a third condition, there was no deformation (the stimulus showed a translating set of dots).

In the plane condition, thresholds were measured for $N = 3$, 6, 12, and 49 dots. In the cloud condition, deformation thresholds were obtained for $N = 2$, 3, 4, 8, 16, and 49 dots. In the condition *without* deformation, thresholds were obtained for $N = 1$, 2, 3, 4, 8, 16, and 49 dots.

Results

Figure 9 shows the thresholds as a function of the number of dots for the 3 conditions. The solid line has a slope of -0.5 and indicates the decrease with an increase in the number of dots (N) predicted by probability summation ($\sim 1/\sqrt{N}$). The highest thresholds were obtained in the cloud condition. Remarkably, performance did not improve when the number of dots increased from 2 to 49 dots. One might suppose that such poor performance arises from using a limited number of randomly chosen dots. However, the fact that thresholds remain constant is incompatible with such a

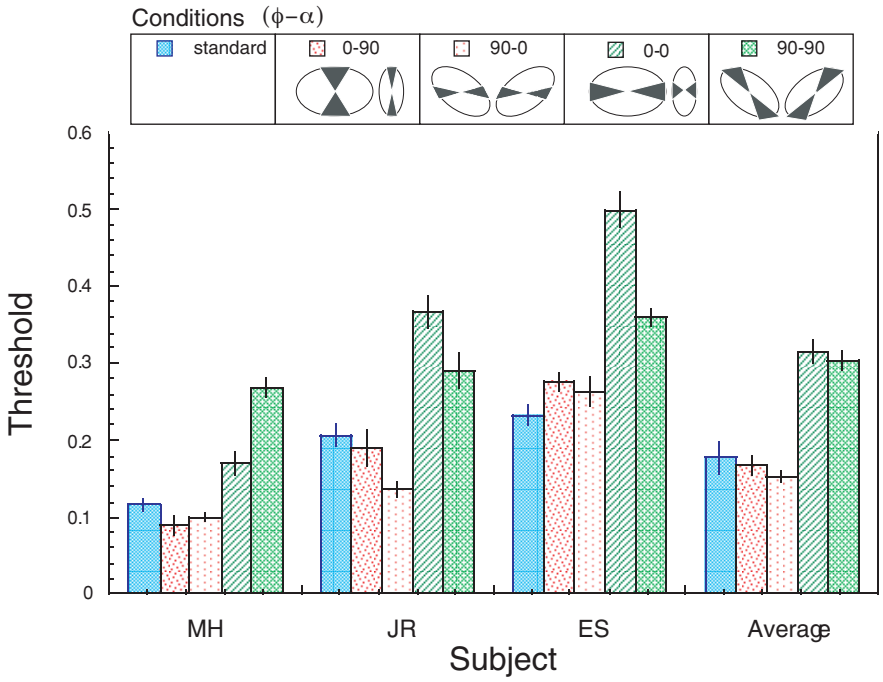

Figure 8. The thresholds obtained in Experiment 2b in which only those dots of the plane stimuli were shown that fell within certain radial segments. Thresholds are shown for the various combinations of segments and deformation directions, schematically outlined at the top of the figure (for details, see text), along with the thresholds obtained in Experiment 1 for a deformation of 0.4.

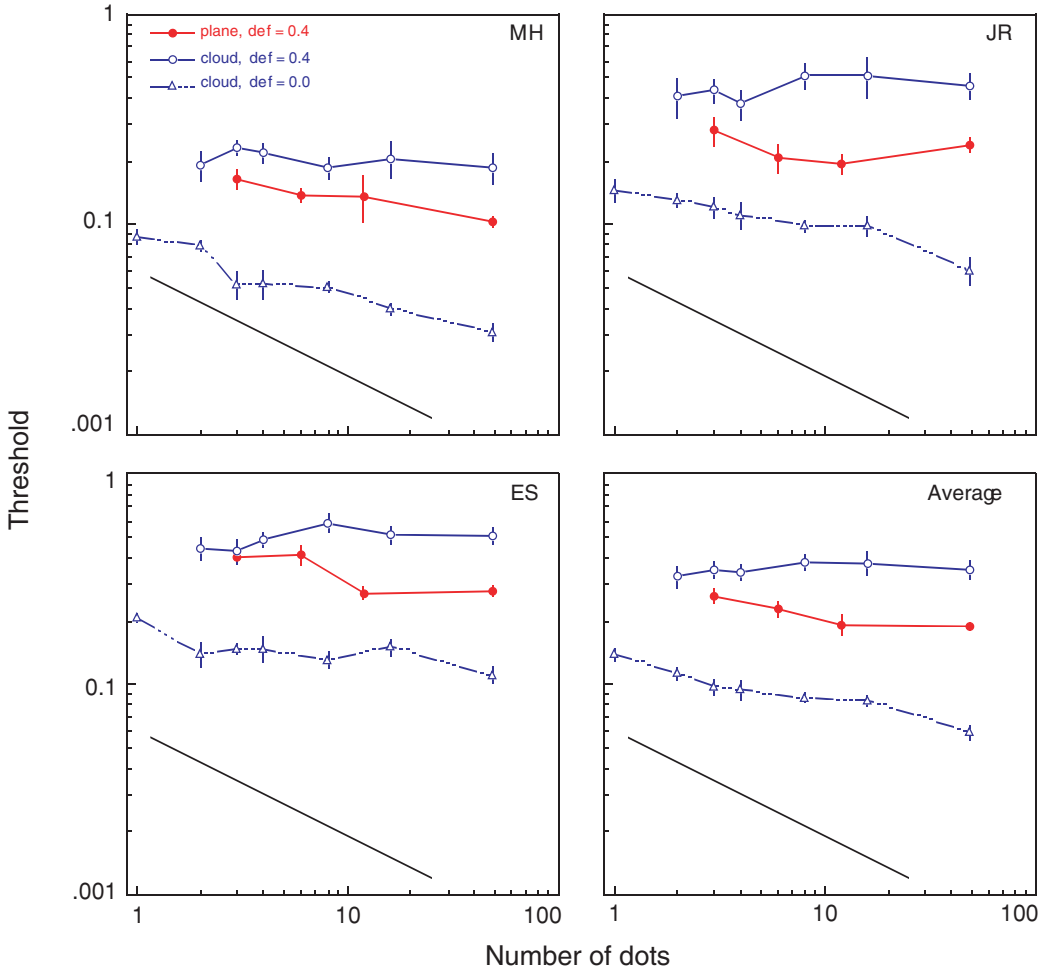

Figure 9. The thresholds as a function of the number of dots for the cloud condition without deformation, and for a plane and a cloud condition with a deformation around 0.4 (ranging from 0.2 to 0.6). The solid line has a slope of -0.5 and indicates the decrease with *N* predicted by probability summation.

strategy. With an increase in the number of dots, the average of the subset of dots would be more and more dissimilar from the real average, which would lead to an *increase* in the threshold.

In the plane condition, thresholds were somewhat lower than in the cloud condition (in agreement with previous results). The threshold shows a small decrease with an increase in the number of dots, although not as fast as predicted by probability summation (slopes of –0.25 (JR), –0.19 (MH), –0.48 (ES), and –0.28 [Average]). On average the threshold decreases only by some 40% when the number of dots is increased from 3 to 49. This slight improvement is likely because with an increase in the number of dots, the chance increases that dots fall within the "more useful" regions, as indicated by the results from Experiment 2 (which showed a threshold difference of a factor of 2 between the "better" and the "less informative" segments: see Figure 8). Thresholds were lowest for the uniformly translating dots (consistent with the results of Experiment 1a). Here, there was a small but consistent improvement as the number of dots was increased, although the slope is much shallower than the value of –0.5 predicted by probability summation (slopes of –0.15 [JR], –0.08 [MH], –0.17 [ES], and –0.13 [Average]). Werkhoven and Koenderink (1991) found that thresholds for discriminating 2D rotation initially decrease with $1/\sqrt{N}$ when the number of dots increased from 1 to 8, but level off for higher numbers of dots. Here, there is no evidence that the slope is initially –0.5 or that the slope changes with an increase in the number of points.

Control Experiment: Speeds or Change in Spatial Configuration?

It could be argued that instead of using relative speeds to solve the task, the subjects based their judgments on changes in spatial configuration over time. For example, with the planar displays it would have been possible to perform the task by determining whether the

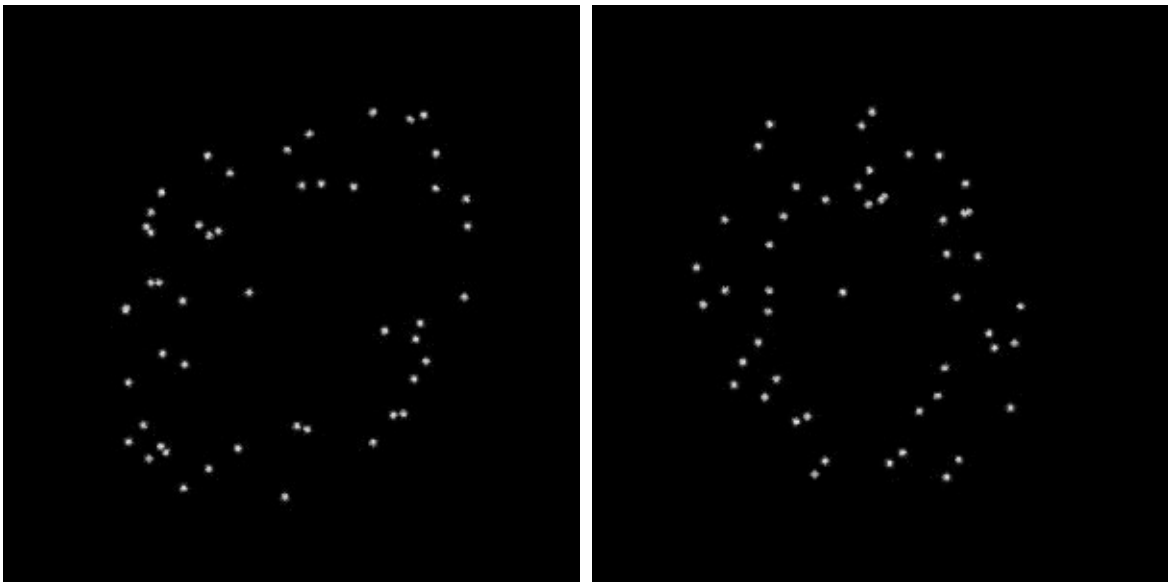

Figure 10. Movies of (plane) stimuli used in the control experiment showing a (signal) stimulus (a) containing dots with a limited lifetime (of 6 frames), and a stimulus containing dots (b) with unlimited lifetime (similar to the ones used in the main experiments).

affine shape has changed over time without necessarily measuring the speeds. Recent work by Lappin and his colleagues (e.g., Lappin, Donnelly & Kojima, 2001; Lappin & Craft, 2000) has shown that human observers are quite sensitive to such changes. The human visual system appears especially sensitive to certain differences in shape, such as convexity/concavity, parallelism, and co-linearity (Wagemans, Van Gool, Lamote, & Foster, 2000). It may well be the case that the subjects used such properties to perform the task in the plane conditions. In the cloud conditions, these alternative strategies would not work. This might be the reason why performance was so much worse in the cloud conditions than in the plane conditions. To test whether changes in shape rather than speeds were used by the subjects, we performed an experiment in which dots with a limited lifetime were used. Only stimuli of the plane condition type were used.

Method

The experiments were run on a PC using images containing Gaussian blobs (width of 1.3 pixels) to achieve subpixel accuracy, shown with a refresh-rate of 60 Hz. The stimuli were the same as those used in the standard condition: the inner and outer radius were 100 and 200 pixels, respectively, and the viewing distance was such (70 cm) that these amounted to the same visual angle as in the main experiments (100 pixels = 1.9 deg). As in the main experiments, each stimulus lasted 333 ms (i.e., each stimulus consisted of 20 frames). The deformation was randomly chosen between 0.4 and 0.6. A method of constant stimuli was used to obtain the thresholds. Thresholds were obtained in two conditions: (1) "unlimited lifetime" stimuli: similar to the stimuli that were used in the main experiments, and (2) "limited lifetime" stimuli in each dot was visible for a maximum of 6 frames (100 ms). In the latter condition, the dots had a lifetime of 6 frames. The initial lifetime was randomly chosen between 1 and 6 frames to make the dots disappear at random phases. When a dot disappeared, a new average location (the location in the middle frame) was chosen and a displacement was derived (using Equation 5). The average locations and displacements were used to calculate the position of the dot in each frame. Figure 10 shows examples of both types of stimuli. In each session one of the two conditions were probed and five stimulus levels were shown with 10 stimuli per level. Measurements were obtained from subjects ES and MH in three sessions for each condition (the two conditions were alternated).

Results

Figure 11 shows the results of this control experiment. The thresholds are expressed as a proportion of the deformation. Also shown are the thresholds obtained in Experiment 1 in the (standard) plane and the cloud conditions, expressed in the same units. The thresholds in the "limited lifetime" condition are the same as the thresholds obtained in the "unlimited lifetime" condition. These in turn are the same as the thresholds obtained in the standard plane condition (Experiment 1)

using a somewhat different experimental set up. The thresholds obtained in the standard cloud condition (Experiment 1, for a deformation of 0.4) are about twice as large as the thresholds obtained in the plane conditions, reconfirming our main result that performance is much worse in the cloud condition than in the plane condition.

These results show that subjects' judgments are mainly based on speed (and position) measurements rather than on changes in the (affine) spatial structure.

Modelling: Optimal Combination of Speed Measurements

In this section, we set out the theoretical limits on performance for the cloud and plane conditions. We show how uncertainty in the individual speed measurements would limit performance in both conditions. The analysis shows that this limitation on its own cannot explain the difference in performance found in the two conditions. However, together with the idea that in the plane conditions the visual system focuses on the more informative dots, the difference in performance can be explained.

Our model is an "ideal observer" model that makes no assumptions about the underlying physiology. It calculates the predictions of an ideal signal combination rule. Its failures are evidence of neural constraints that prevent the human observer from making optimal use of the available information. This is a rather different exercise from generating a physiologically inspired model of motion detection as, for example, Yuille, Grzywacz,

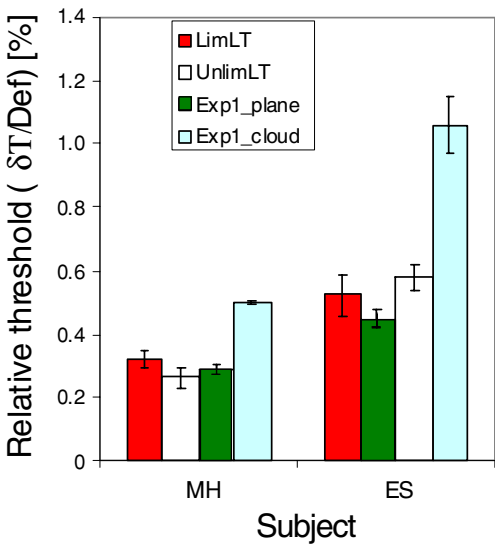

Figure 11. Thresholds obtained in the control experiment using limited lifetime stimuli (LimLT: lifetime of 100 msec) as well as unlimited lifetime stimuli (UnLimLT) along with thresholds for a deformation of 0.4 obtained in Experiment 1 for the plane (Exp1_plane) and cloud (Exp1_cloud) conditions. The thresholds are expressed as a proportion of the deformation.

Watamaniuk and McKee have done (e.g., Yuille and Grzywacz, 1988; Grzywacz and Yuille, 1991; Grzywacz, Watamaniuk, & McKee, 1995). Their model uses assumptions about the coherence of motion within a region to help solve the correspondence and aperture problems. It also incorporates physiologically plausible components such as Gabor filters in the motion detection stage. We have taken a different approach and simply considered the theoretical limits on performance in the two conditions we examined, imposed by (i) noise in measuring speeds of individual dots and (ii) the spatial layout of the dots (plane or cloud).

We assume that performance in our experiments depends on the estimation of two properties: (1) average speed and (2) the local speed of the plane. We assume that in the cloud condition, the task is based on estimation of the average speed. Although in principle it is also possible to use the average speed in the plane condition, it is more likely that subjects in that case use an estimate of the local speed of the plane. In the reference stimulus, the central dot moves with a speed at which it is perceived to lie in a plane formed by the surrounding dots.

We show here the accuracy with which these two properties can be derived when the speeds are available to the system with limited accuracy. We assume that

1) uncertainty in the measurements of the positions of the dots is negligible relative to the uncertainty in the speed measurements,
2) noise on every speed measurement is drawn from a Gaussian probability distribution,
3) and noise in the speed measurements are independent from each other.

The measurements consist of the positions and speeds $(x,y,S)_i$ of all points $i = 1 \ldots N$. The speeds S_i are measured with uncertainty σ_i (which may differ from dot to dot).

The best estimate of the average speed S_m is simply equal to the average of the speeds (i.e., the mean): $S_m = \Sigma S_i / N$. This leads to an uncertainty in the average speed estimation of σ_m given by: $\sigma_m^2 = \sum \sigma_i^2 / N^2$, or written differently:

$$\sigma_m = \sqrt{\langle \sigma^2 \rangle} / \sqrt{N}, \qquad (2)$$

that is, the uncertainty in the average speed estimate is equal to the square root of the average squared sigma divided by the square root of the number of measurements (the bracket < > indicates the average).

When estimating the local speed of the plane, we first have to estimate what the plane looks like. The best estimate of the plane follows from a least squares fit to the data points (the whole procedure is similar to fitting a line to a 2D data set). The best estimate of the local speed of the plane at the test location t is obtained by fitting a

plane $S = ax + by + t$ through the points $(x,y,S)_i$. Given that the noise is drawn from a Gaussian distribution, the optimal way to do this is by minimizing χ^2 given by

$$\chi^2 = \sum \left(\frac{S_i - t - a\,x_i - b\,y_i}{\sigma_i} \right)^2 . \quad (3)$$

Setting the partial derivatives to t, a and b to zero leads to a set of linear equations (e.g., see Press, Flannery, Teukolsky, & Vetterling, 1996) that can easily be solved (see "Appendix"). The equation describing the uncertainty in the estimate of t, σ_t is rather complex. In the model simulations, the exact equations are used. To give some intuitive idea, we also derived an approximation for the case that the test dot lies is the center of mass. In that case it is in close approximation equal to (see "Appendix"):

$$\sigma_t = \sqrt{1 / \left\langle 1/\sigma^2 \right\rangle} / \sqrt{N} . \quad (4)$$

This model — in which the (independent) speed measurements are optimally combined — gives the following (general) predictions:

a. If the noise on all measurements is equal, the uncertainty in the average speed and the local planar speed estimate is the same (Equations 4 and 2, respectively).
b. The predicted thresholds decrease with increasing numbers of dots (with one over the square root of N: $\sim 1/\sqrt{N}$). This follows from the assumption that the speed measurements are independent and that all measurements are taken into account. This is sometimes referred to as probability summation.
c. If the noise in the speeds differs from dot to dot, better performance is predicted for estimating the local planar speed than for estimating the average speed. For example, suppose that the sigma for individual dots (σ_i) spans the range $\sigma_o - \Delta/2$ to $\sigma_o + \Delta/2$, where σ_o is the average value and Δ the range. In that case, the uncertainty in the average speed estimate, σ_m, becomes larger than the average σ_o: $\sigma_m^2 = \sigma_0^2 + \Delta^2/12$ (using Equation 2) whereas the uncertainty in the local speed estimate, σ_t, becomes smaller than the average σ_o: $\sigma_t^2 = \sigma_0^2 - \Delta^2/4$ (using Equation 4).

Quantitative Predictions

In order to make quantitative predictions, extra assumptions will be used. We will assume that the system makes independent measurements of the relative speeds of the dots (relative to that of the test dot). The results show that performance is more dependent on relative speed than on absolute speed (see Experiment 1 and Figure 4). We will further assume that the width of the noise distribution, σ_i, increases with increasing (relative) speed and has the following form:

$$\sigma_i = k\,\Delta S_i + c , \quad (5)$$

in which k accounts for a proportional increase of the noise with speed, ΔS_i is the relative speed of the dot, and c is a plateau level of σ_i as ΔS_i becomes small. Up to high speeds (64 deg/s), thresholds for speed discrimination (de Bruyn & Orban, 1988) can be modelled by a similar expression (see Hogervorst & Eagle, 1998).

Model and Data Compared

The results of the experiments are compared here with the predictions of the model. The model predictions are based on (Monte Carlo) simulations: each is an average over 30 sample stimuli. Each time, a sample set of positions and speeds is calculated. For each of these sets, the model comes up with a prediction of the threshold. The final prediction is the geometric average over all 30 sample stimuli.

Experiment 1: Standard Conditions

That the thresholds in the cloud condition are much higher than in the plane condition cannot be explained with a model in which the magnitude of the noise on each speed measurement is the same [prediction (a) from the previous section]. However, because we assume that the noise increases with increasing speeds, the noise differs from dot to dot and this may explain the difference in thresholds between the two conditions [prediction (c)].

Figure 12 shows the average threshold data along with the predictions of several versions of the model. The assumption that the noise in the speed measurements increases with increasing speed (approaches Weber behaviour) predicts the observed increase in thresholds with increasing deformation in both conditions.

Standard experiment (1a): cloud condition

In practice, we can obtain an estimate of the Weber fraction for estimating the average speed by finding those values of k and c for which the property $\sigma_m \sqrt{N}$, the uncertainty in the average speed times $\sqrt{N}$, equals the thresholds obtained in the cloud condition (using Equation 5). This leads to very high values of k: 31% (MH), 94% (JR), 70% (ES), and 58% (Average), with values for the c of 0.04 (MH), 0.09 (JR), 0.13 (ES), and 0.08 (Average). The average threshold data is plotted in Figure 12 along with the fitted line ("cloudAll"). The parameter k can be compared directly with Weber fractions for speed discrimination, which are in the order of 5% to 8% (McKee, 1981; de Bruyn & Orban, 1988). Note that we use the factor $\sqrt{N}$ to compare the threshold obtained in this experiment with thresholds for speed discrimination with the same number of dots (because no effect of number of dots was found here or in the experiments of de Bruyn & Orban, discussed with

Experiment 3 "Results" below). Comparing the fitted *k* values with Weber fractions for speed discrimination (of uniformly translating textures) shows that performance for estimating the average speed is remarkably poor. This means that subjects are poor at judging the average speed of a cloud of dots when it is also rotating (a possible 3D interpretation of the stimulus): the rotation interferes with the estimation of the average speed. The deforming cloud data relate to the results from Watamaniuk and Duchon (1992), who performed experiments in which subjects had to discriminate the average speed of two sets of dots whose speed distributions were equal in width. They obtained thresholds for Gaussian speed distributions with moderate widths (up to 22% of the mean speed), and found that thresholds were *unaffected* by the width of the distribution.

The speed distributions of the stimuli from the standard cloud conditions with a deformation of 0.1 have a similar width (e.g., for medium overall speeds the width is 23%). For this magnitude of deformation, the thresholds are significantly higher than for zero deformation. To compare performance levels one might calculate thresholds as a fraction of the speed of the reference stimulus. In the study by Watamaniuk and Duchon, Weber fractions were around 8%. In a condition that is comparable to a stimulus used by Watamaniuk and Duchon (1992) (cloud condition with deformation of 0.1), Weber fractions expressed in this way are about 13% for the fastest stimulus speed. Note however, that subjects were allowed to track the stimulus. Therefore, this is not a very meaningful number. Figure 4 shows, for example, that thresholds change very little with overall speed, whereas they change radically with stimulus deformation (Figure 12).

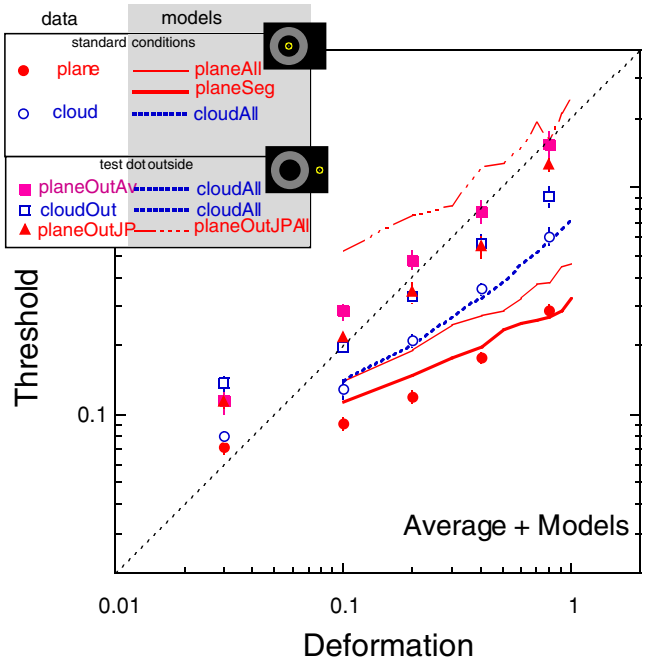

Figure 12. The average threshold data from Experiment 1 along with predictions from various models, in which all dots are taken into account, "All" or just the slow segments "Seg" (see top of Figure 8). The straight dotted line corresponds to a level for which the speed of the center dot equals the largest or smallest speeds in the distribution. Its slope of one corresponds to Weber behaviour.

However, their and our experiments differ in many ways. In their experiment, subjects had to compare the average speed of two successively shown speed distributions with the same width. In our experiments, the speed of one dot had to be compared with the average of a number of dots moving with different speeds. Also, subjects had to compare the speeds of two elements (the target dot and the cloud dots) that were visible at the same time. Another difference is that, in Watamaniuk and Duchon's experiments, the dots moved within a stationary aperture, with continuous replacement of dots, whereas in our experiment, the outline of the group of dots moved. Finally, in our experiments, subjects were free to track the test dot. This meant that in our experiments there were two (retinal) motion directions, whereas there was only one in the experiments of Watamaniuk and Duchon. The different motion directions might inhibit each other leading to different speed processing in the two cases.

Standard experiment (1a): plane condition

We used the same noise model to predict the thresholds in the plane condition ($k = 0.58$, $c = 0.08$). This prediction is also plotted in Figure 12 ("planeAll"). This model does not predict the thresholds obtained in the plane condition very well. Although the model predicts the thresholds to be somewhat lower than in the cloud condition [consistent with prediction (c) from the previous section], the observed difference is much larger. This model takes all speed measurements into account. However, Experiments 2 and 3 indicate that in the plane condition, performance is determined primarily by the dots in the slow segments close to the test dot. We therefore calculated the predictions of the model in which only the dots in the slow segments were taken into account. The same noise model was used, but only dots in segments within +/- 45 deg from the deformation direction were taken into account ("planeSeg", see top in Figure 8). Although the fit is not perfect, this model predicts the data fairly well (especially for large amounts of deformation). The main point is that the difference in thresholds between the plane and the cloud conditions can be accounted for by using only a subsection of the dots ("the best ones") in the plane condition, and using all dots in the cloud condition.

Test dot outside (1b)

The model does not take the positions of the dots into account when estimating the average speed. However, in the cloud condition, the threshold was

considerably higher when the test dot was outside the annulus (see Figure 12).

When the dots depict a plane and the task is to estimate the average speed, the thresholds are even higher. This indicates that it is not fully correct to discard the positions in the model. In the condition in which the local speed of the plane has to be judged, the observed thresholds are lower than predicted by the model. In this case, it is not obvious which subset of dots should be used for the judgment. We therefore used a model in which all dots were used. (If a subset of dots can be found that is relatively more informative than others, and only this subset is used by the model, the predictions will be lower).

Experiment 2: Which Dots Are Used?

Figure 13 replots the average threshold data from Experiment 2a (different annuli) along with the model predictions. In the cloud condition, all dots were taken into account and in the plane condition only the dots within certain (slow) segments were taken into account.

In the cloud conditions with varying speed distribution ("cloudspeed"), the predictions are somewhat too low (by 33% on average). This occurs because of the way the model curve is fixed at one point. In this case, the model threshold for the 1→2 condition was taken as equal to the model threshold for the 1→2 condition in Experiment 1a, with a deformation of 0.4. Although this value is a good fit to the empirical data for that Experiment 1a, it is not a good predictor of the threshold in this Experiment 2a: thresholds for this condition are 34% higher in Experiment 2a (accounting almost exactly for the discrepancy of data and model in Figure 13). It is not clear why the threshold levels turn out to be different. One difference is that here the deformation ranges from 0.3 to 0.5, whereas in Experiment 1a, it was fixed at 0.4. Although this is taken into account by the model, it does not predict a difference in thresholds. The difference may be due to increased uncertainty in the subject's expectation. The important thing is that the model gives a good qualitative prediction of the pattern of results: the measured and predicted thresholds rise with a similar slope going from left to right in Figure 13.

In the plane conditions, the model "planeSeg" takes only dots in certain (slow) segments into consideration. These are the segments shown to be most useful to subjects in Experiment 2b (Figure 8). The predictions are shown in Figure 13. This model predicts a gradual increase in thresholds going from left to right in Figure 13. The empirical data, however, show that the thresholds are independent of the outer radius. This indicates that the judgments are not based only on the dots in certain segments, but also (primarily) on those dots that are close to the center dot. When the inner radius is increased, the thresholds rise. This is in accordance with a model that is based on the slowest dots, closest to the center dot (indicated by model "plane" in Figure 13).

We have assumed in our model that the shape of the surface is known (a plane). An alternative explanation for the importance of the dots closest to the test dot is that, in general, surfaces tend to vary spatially. Therefore, in principle, it makes sense to restrict the interpolation to a region around the test location. However, our modelling

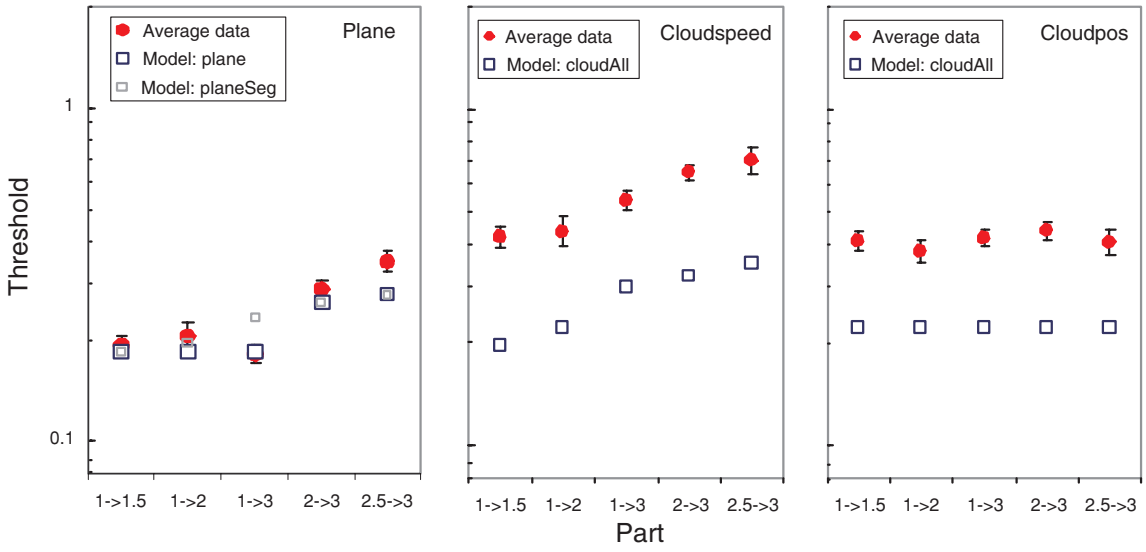

Figure 13. The average thresholds from Experiment 2a along with the model predictions. The left panel shows results for the plane condtion. The center panel shows results for the "cloudspeed" condition in which the speed distribution was the same as in the plane conditions. The right panel shows results for the "cloudpos" condition, in which the distribution of locations was the same as in the plane condition. In the cloud conditions, the model takes all dots into account ("cloudAll"), whereas in the plane conditions, only certain segments (containing the slowest dots) are taken into account ("planeSeg"). Model "plane" takes into account only dots that lie within certain segments *and* that are close to the test dot (see text).

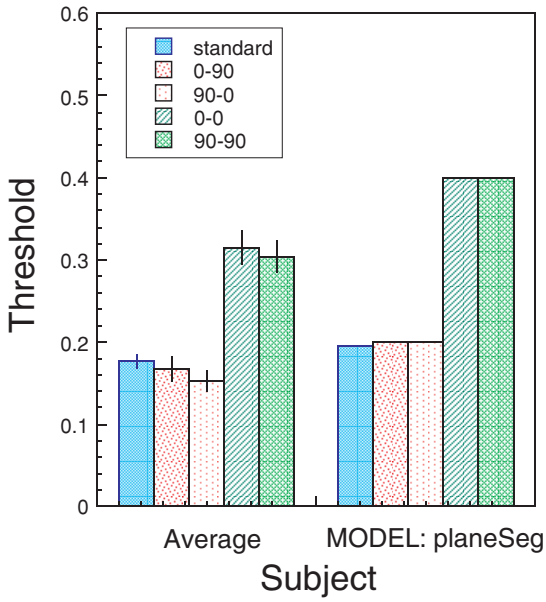

Figure 14. The average threshold data from Experiment 2b along with the model predictions (taking only slow segments into account). The various conditions refer to different combinations (φ, α) of the deformation direction φ and segment direction α (see top of Figure 8).

takes into account only the uncertainty on speed estimates of different dots.

In the cloud conditions in which the spatial distribution is changed, the threshold level remains constant. Again, the predicted threshold level is a bit (19%) too low. Because the model does not take the dot location into account, it predicts the same thresholds for all annuli, in agreement with the results.

Figure 14 replots the average threshold data from Experiment 2b (in which only certain segments of the plane were shown) along with predictions from the model (using all the available dots). In agreement with the results, the model predicts the thresholds to be higher in the conditions in which only fast dots are shown (in the direction of the deformation). This is because the uncertainty in the speed estimate increases with increasing speed. Even the threshold levels are predicted fairly well.

Experiment 3: Effect of the Number of Dots

The one aspect of the model that completely fails to account for the data is the lack of variation in threshold as the number of dots in the stimulus is varied. The model predicts that the thresholds should decrease with the square root of the number of dots, N. The results of Experiment 3 show that this is not the case. In the plane condition with nonzero deformation, there is a small decrease in threshold. However, this decrease can be explained because with an increase in the number of dots, there is an increase in the chance of finding a slow dot close to the test dot (see Experiment 2). In the cloud condition, only a small (but statistically significant) decrease was found in the zero deformation condition. In the model described here, we have assumed that the thresholds are independent of the number of dots. This assumption is in accordance with the results of de Bruyn and Orban (1988), who found that thresholds for speed discrimination were not higher when one rather than many dots were used. Such a lack of improvement with an increase in the number of dots was also incorporated into the model used by Hogervorst and Eagle (1998, 2000) and Eagle and Hogervorst (1999) that was successful in explaining performance in structure-from-motion experiments.

Discussion

Our results show that subjects are relatively good at estimating the local speed of a plane and relatively poor at estimating the average speed of a set of dots. This was shown in Experiment 1, where two types of stimuli were compared: dots had the same distributions of speeds and locations in each case but the speeds were assigned to different dots. The results show that the difference in performance in the two conditions probably stems from the fact that for estimation of the average speed all dots have to be taken into account, whereas for estimating the local speed of a plane, it is possible to restrict consideration to a limited number of dots. In the latter case, given certain simple assumptions, some dots supply better speed information than others (those dots with the slowest relative speed, closest to the test dot). Our results suggest that the visual system is able to focus on this information.

The model we propose succeeds in accounting for a number of important aspects of the data. This model takes into account only slow dots close to the test dot when estimating planar speed ("planeSeg" in Figure 12) and takes all the dots into account when estimating average speed ("cloudAll"). Both model and data show:

- better performance for judgments of the local speed of a plane than for the average speed of a set of unstructured dots (see Figure 12);
- a rise in thresholds with increasing deformation (see Figure 12);
- a similar pattern of thresholds when parts of the stimulus are removed (see Figure 13 and Figure 14).

Not only is the pattern of thresholds captured well, the model even gives very reasonable quantitative predictions. There are only two free parameters in the model, factor k that accounts for a proportional increase of the noise with speed and the plateau level c for speed discrimination thresholds (see "Model" section), which

were derived from the data in Experiment 1 and used for all the modelling.

There are also notable aspects of the data for which we have no explanation. Most strikingly, our results show that performance varies very little with the number of dots in the stimulus, despite the extra information these dots carry. We therefore used a model in which this lack of improvement is (somewhat artificially) taken into account (the predicted threshold is taken to be the threshold of the optimal combination model multiplied by the square root of the number of dots in the stimulus). Strictly speaking, this means that the noise in the speed measurements is no longer independent (the noise is correlated), or that the speeds are no longer combined optimally.

Another perplexing aspect of the data is that the noise level required to explain performance here is so high. Factor k (Equation 5), equal to 58% for the average subject, can be compared directly with Weber fractions for speed discrimination, which are around 5% (for medium speeds). This means that much higher noise levels have to be assumed to account for our results than for thresholds for speed discrimination. A similar model used by Hogervorst and Eagle (1998, 2000) was successful in modelling structure-from-motion thresholds and biases in perceived depth. They used estimates of the noise that were directly derived from human velocity and acceleration thresholds for uniform moving patterns (for small viewing angles, the noise was equal to these estimates, and for large viewing angles, the noise was twice as much). In our model, the noise is more than 10 times as high (for the average subject). One difference between our model and their model is that we use relative speeds as input, whereas Hogervorst and Eagle use speeds expressed in screen coordinates. The results show that this is more appropriate in our case (see Figure 4). Still, in the studies by Hogervorst and Eagle, the hinged planes rotated around their hinge, and the hinge did not show any additional translation. Whether absolute speeds, retinal speeds, or relative speeds (and relative to which reference) are more appropriate as input in structure-from-motion algorithms needs to be determined in future studies. The fact remains that the noise levels required to explain the results here are much higher than the noise levels required to explain the results in the studies by Hogervorst and Eagle and basic motion discrimination thresholds (using uniformly moving patterns). The reason for this is unclear.

Our model uses the assumption that (independent) noise in the speed measurements limits performance. However, there is a large amount of evidence (e.g., Legge & Campbell, 1981; McKee, Welch, Taylor, & Bowne, 1990) showing that motion thresholds are much lower with than without a reference frame, indicating that the assumption of independent speed measurements does not hold. For instance, the fact that thresholds for two dots moving in anti-phase are about twice as small as thresholds for two dots moving in phase (e.g., Hogervorst, Kappers & Koenderink, 1995; Hogervorst, 1996; Lappin et al., 2000) shows that human judgments of velocity fields (including structure-from-motion fields) are not derived from independent velocity measurements of the image features. Also our finding that performance improves little with an increase in the number of dots indicates that the assumption of independent speed measurements is too simple. It has been suggested that human estimates of structure-from-motion and ego motion are based more directly on the structure of the velocity field (e.g., optic flow), using higher order derivatives of the velocity field rather than the zero-order component. Such models appear to give a more realistic description of human velocity processing.

It is possible that other models of pooling motion signals, such as the motion coherence model of Grzywacz and colleagues (e.g., Yuille and Grzywacz, 1988; Grzywacz & Yuille, 1991; Grzywacz, Watamaniuk, & McKee, 1995), would also predict a difference in performance between cloud and plane conditions. For example, the model of Grzywacz et al. (1995) seeks (in general) to assign a single motion vector to each image location by using neighbouring regions. The model is likely to fare better with a spatially coherent pattern such as a plane than it would with a stimulus in which very different motion speeds and directions are present within the neighbourhood of each point, such as in the cloud condition. Our model does not rule out the possibility that other factors, such as those considered by Grzywacz et al. (1995) , are important. Rather, we have provided a quantitative account setting out the limits of performance that would be expected given noise in speed estimates and use of different subsets of dots. Our model is more like an "ideal observer" model that analyzes what information is available to perform the task at hand, and indicates what limits visual processing. The advantage of this ideal observer approach is that failures of the model provide evidence of neural constraints that prevent the human observer from making optimal use of the available information.

Our model allows for quantitative predictions for the kinds of stimuli used in structure-from-motion and ego-motion tasks. Similar models have been successful in predicting human performance in a range of structure-from-motion and ego-motion tasks (e.g., Koenderink & van Doorn, 1987; Werkhoven & van Veen, 1995; Hogervorst & Eagle, 1998, 2000). This type of model consists of two stages: in the first stage, the noise is specified, and in the second stage, an optimal observer model is used to solve the task using the measurements. In the latter stage, assumptions or prior information may be used, although in our model no priors are used (equivalent to using flat priors).

The results show that deformations have a very deleterious effect on thresholds. Suggestions that the visual system's sensitivity to spatial structure is unaffected by affine distortions (e.g., Lappin & Craft, 2000) are not

compatible with this strong effect of deformation. By comparison, changes in translation speed had little effect on thresholds.

In summary, the processing capacity of the visual system appears to be limited. In some situations, such as when estimating the local speed of the plane, we suggest that the task is solved by focusing on the best pieces of information. In other situations, such as when estimating average speed, such a strategy is not possible and attention has to be paid to all pieces of information.

Appendix

Given the measured z values in a number of points $(x,y,z)_i$. for $i = 1..N$ and given that these measurements are taken from their real values with noise added from Gaussian distributions with widths σ_i, the object is to find a plane $z = t+ax+by$ that best represents the data. This is done by minimizing function χ^2 given by:

$$\chi^2 = \sum \left(\frac{z_i - t - a\,x_i - b\,y_i}{\sigma_i} \right)^2 \qquad \text{(A1)}$$

Ideally, the deviations from the plane are weighted by the inverse of the width σ_i. However, other weightings are also possible; for example, the weight could be made to vary with the distance from the origin to make it more local (as in splines). Minimization amounts to setting the partial derivatives to zero:

$$\partial\chi^2 / \partial a = \partial\chi^2 / \partial b = \partial\chi^2 / \partial t = 0 \,,$$

leading to the following set of linear equations:

$$\begin{pmatrix} \Sigma_{xx} & \Sigma_{xy} & \Sigma_x \\ \Sigma_{xy} & \Sigma_{yy} & \Sigma_y \\ \Sigma_x & \Sigma_y & \Sigma \end{pmatrix} \bullet \begin{pmatrix} a \\ b \\ t \end{pmatrix} = \begin{pmatrix} \Sigma_{zx} \\ \Sigma_{zy} \\ \Sigma_z \end{pmatrix}, \qquad \text{(A2)}$$

in which

$$\begin{array}{lll} \Sigma_{xx} = \sum x_i\,x_i / \sigma_i^2 & \Sigma_x = \sum x_i / \sigma_i^2 & \Sigma_{zx} = \sum z_i\,x_i / \sigma_i^2 \\ \Sigma_{xy} = \sum x_i\,y_i / \sigma_i^2 & \Sigma_y = \sum y_i / \sigma_i^2 & \Sigma_{zy} = \sum z_i\,y_i / \sigma_i^2 \\ \Sigma_{yy} = \sum y_i\,y_i / \sigma_i^2 & \Sigma = \sum 1 / \sigma_i^2 & \Sigma_z = \sum z_i / \sigma_i^2 \end{array}$$

The solution is written as

$$(a,\ b, t) = \mathbf{M} \bullet (\Sigma_{zx}, \Sigma_{zy}, \Sigma_z) \ ,$$

in which **M** is the inverse of the matrix displayed above. The task set in our experiment requires deduction of t. The best estimate of t follows from the solution of the matrix equation in which the measurements of z_i are used: $t = m_{31}\Sigma_{zx} + m_{32}\Sigma_{zy} + m_{33}\Sigma_z$. The noise in these measurements propagate into the noise on the estimate of t, σ_t, in the following way:

$$\begin{aligned} \sigma_t^2 &= \sum_i \left(\sigma_i * \frac{\partial t}{\partial z_i} \right)^2 \\ &= \sum_i \left(m_{31}\frac{x_i}{\sigma_i} + m_{32}\frac{y_i}{\sigma_i} + m_{33}\frac{1}{\sigma_i} \right)^2 \\ &= m_{31}^2\Sigma_{xx} + m_{32}^2\Sigma_{yy} + m_{33}^2\Sigma \\ &\quad + 2\left(m_{31}m_{32}\Sigma_{xy} + m_{31}m_{33}\Sigma_x + m_{32}m_{33}\Sigma_y \right) \end{aligned} \qquad \text{(A3)}$$

This is analogous to the derivation for fitting a line in 2D: $y = a\,x + t$, described in Numerical Recipes (Press et al., 1996). In the 2D case, the variance in the t is given by:

$$\sigma_t^2 = \frac{\Sigma_{xx}}{\Sigma_{xx}\Sigma - \Sigma_x^2} \,, \qquad \text{(A4)}$$

which reduces to

$$\sigma_t^2 = \frac{1}{\Sigma} = \sigma_t^2 \sqrt{N} = \frac{1}{\left\langle 1/\sigma^2 \right\rangle} \,, \qquad \text{(A5)}$$

when $\Sigma_x = 0$, in which the bracket $\langle\,\rangle$ stands for the average. Therefore, when the tilt is well defined and $\Sigma_x = 0$, the uncertainty in the local speed of the plane reduces to (A5). When one fits a plane $z = t$ to the data in a similar way as described above, one obtains:

$$t = \frac{\sum z_i / \sigma_i^2}{\sum 1 / \sigma_i^2} \,, \qquad \text{(A6)}$$

in which case the variance in t is described exactly (i.e., not an approximation) by (A5). This is a weighted average of all measurements, in which the weight is inversely related to the uncertainty in the measurement. When taking a normal average $m = \Sigma z_i/N$, the variance is simply

$$\sigma_m^2 = \sum \sigma_m^2 / N \,,$$

or

$$\sigma_m^2 \sqrt{N} = \left\langle \sigma^2 \right\rangle \ . \qquad \text{(A7)}$$

To appreciate what these equations mean, one could use the analogy of a number of resistances with magnitudes equal to σ_i^2. The total resistance corresponds to the variance in the speed estimate of the test dot. The variance in the local planar speed resembles a situation in which the resistances are in parallel, whereas the variance in the average speed estimation resembles a situation in which the N resistances are in series (actually, N of these series should be placed in parallel to account for the division by N). While the variance in the average speed estimate is determined equally by all variances, the variance in the local speed of the plane is determined largely by the smallest variance (the smallest resistance).

Acknowledgments

The work was funded by a project grant from the Biotechnology and Biological Sciences Research Council (No. 43/SO9621) and by a Royal Society University Research Fellowship awarded to Andrew Glennerster.

Commercial relationships: none.

Footnotes

1. Current address: TNO Human Factors, Kampweg 5, 3769 DE Soesterberg, The Netherlands.

References

Collewijn, H., Tamminga, E. P. (1984). Human smooth and saccadic eye movements during voluntary pursuit of different target motions on different backgrounds. *Journal of Physiology*, *351*, 217-250. [PubMed]

de Bruyn, B., & Orban, G. A. (1988). Human velocity and direction discrimination measured with random dot patterns. *Vision Research*, *28*, 1323-1335. [PubMed]

Eagle, R. A., & Blake, A. (1995). Two-dimensional constraints on three-dimensional structure from motion tasks. *Vision Research*, *35*, 2927-2941. [PubMed]

Eagle, R. A., & Hogervorst, M. A. (1999). The role of perspective information in the recovery of 3D structure-from-motion. *Vision Research*, *39*, 1713-1722. [PubMed]

Grzywacz, N. M., & Yuille, A. L. (1991). Theories for the visual perception of local velocity and coherent motion. In M. S. Landy & J. A. Movshon (Eds), *Computational models of visual processing* (pp. 231-252) Cambridge: MIT Press.

Grzywacz, N. M., Watamaniuk, S. N. J., & McKee, S. P. (1995). Temporal coherence theory for the detection and measurement of visual motion. *Vision Research*, *35*, 3183-3203. [PubMed]

Hogervorst, M. A., Kappers, A. M. L., & Koenderink, J. J. (1995). Detection of relative motion [Abstract]. *Perception*, *25*, 9a.

Hogervorst, M. A., Kappers, A. M. L., & Koenderink, J. J. (1996). Structure from motion: A tolerance analysis. *Perception & Psychophysics*, *58*, 449-459. [PubMed]

Hogervorst, M. A. (1996). *Limitation on the recovery of structure from motion.* Thesis, Utrecht University, Chapter 5, pp. 77-93, ISBN 90-393-1212-5.

Hogervorst, M. A., & Eagle, R. A. (1998). Biases in three-dimensional structure-from-motion arise from noise in the early visual system. *Proceedings of the Royal Society of London B, Biological Sciences*, *265*, 1587-1593. [PubMed]

Hogervorst, M. A., & Eagle, R. A. (2000). The role of perspective effects and accelerations in perceived three-dimensional structure-from-motion. *Journal of Experimental Psychology, Human Perception and Performance*, *26*, 934-955. [PubMed]

Johnston, A., & Wright, M. J. (1985). Lower thresholds of motion for gratings as a function of eccentricity and contrast. *Vision Research*, *25*, 179-185. [PubMed]

Koenderink, J. J., & van Doorn, A. J. (1987). Facts on optic flow. *Biological Cybernetics*, *56*, 247-249. [PubMed]

Koenderink, J. J., & van Doorn, A. J. (1991). Affine structure from motion. *Journal of Optical Society of America A*, *8*, 377-385. [PubMed]

Lappin, J. S., Donnelly, M. P., & Kojima, H. (2001). Coherence of early motion signals. *Vision Research*, *41*, 1631-1644. [PubMed]

Lappin, J. S., & Craft, W. D. (2000). Foundations of spatial vision: from retinal images to perceived shapes. *Psychological Review*, *107*, 6-38. [PubMed]

Legge, G. E., & Campbell, F. W. (1981). Displacement detection in human vision. *Vision Research*, *21*, 205-213. [PubMed]

McKee, S. P. (1981). A local mechanism for differential velocity detection. *Vision Research*, *21*, 491-500. [PubMed]

McKee, S. P., Welch, L., Taylor, D. G., & Bowne, S. F. (1990). Finding the common bond: Stereoacuity and the other hyperacuities. *Vision Research*, *30*, 879-891. [PubMed]

Nakayama, K. (1985). Biological image motion processing: A review. *Vision Research*, *25*, 625-660. [PubMed]

Press, W. H., Flannery, B. P., Teukolsky, S. A., & Vetterling, W. T. (1996). *Numerical Recipes in* C (2nd ed.) Cambridge, UK: Cambridge University Press.

Snoeren, P. R., & Puts, M. J. H. (1997). Multiple Parameter Estimation in an Adaptive Psychometric Method: MUEST, an Extension of the QUEST Method. *Journal of Mathematical Psychology*, *41*, 431-439. [PubMed]

Snowden, R. J., & Braddick, O. J. (1991). The temporal integration and resolution of velocity signals. *Vision Research*, *31*, 907-914. [PubMed]

Verghese, P., & Stone, L. S. (1995). Combining speed information across space. *Vision Research*, *35*, 2811-2823. [PubMed]

Verghese, P., & Stone, L. S. (1996). Perceived visual speed constrained by image segmentation. *Nature*, *381*, 161-163. [PubMed]

Verghese, P., & Stone, L. S. (1997). Spatial layout affects speed discrimination. *Vision Research*, *37*, 397-406. [PubMed]

Wagemans, J., Van Gool, L., Lamote, C., & Foster, D. H. (2000). Minimal information to determine affine shape equivalence. *Journal of Experimental Psychology: Human Perception and Performance*, *26*, 443-468. [PubMed]

Watamaniuk, S. N., & Duchon, A. (1992). The human visual system averages speed information. *Vision Research*, *32*, 931-941. [PubMed]

Watamaniuk, S. N., Sekuler, R., & Williams, D. W. (1989). Direction perception in complex dynamic displays: The integration of direction information. *Vision Research*, *29*, 47-59. [PubMed]

Watson, A. B., & Pelli, D. G. (1983). QUEST: A Bayesian adaptive psychometric method. *Perception & Psychophysics*, *33*, 113-120. [PubMed]

Werkhoven, P., & Koenderink, J. J. (1991). Visual processing of rotary motion. *Perception & Psychophysics*, *49*, 73-82. [PubMed]

Werkhoven, P., Snippe, H. P., & Toet, A. (1992). Visual processing of optic acceleration. *Vision Research*, *32*, 2313-2329. [PubMed]

Werkhoven, P., & van Veen, H. A. H. C. (1995). Extraction of relief from visual motion. *Perception & Psychophysics*, *57*, 645-656. [PubMed]

Williams D. W., & Sekuler, R. (1984). Coherent global motion percepts from stochastic local motions. *Vision Research*, *24*, 55-62. [PubMed]

Yuille, A. L., & Grzywacz, N. M. (1988). A computational theory for the perception of coherent visual motion, *Nature*, *333*, 71-74. [PubMed]