Computation and measurement of slant specified by linear perspective

Casper J. Erkelens

One of the striking features of vision is that we can experience depth in two-dimensional images. Since the Renaissance, artists have used linear perspective to create sensations of depth and slant. What is not known is how the brain measures linear perspective information from the retinal image. Here, an experimental technique and geometric computations were used to isolate slant related to linear perspective from slant induced by other cues. Grid stimuli, designed to induce strong impressions of slant, were sufficiently simple to allow accurate predictions on the basis of numeric computations. Measurement of slant about the vertical axis as functions of slant depicted on the screen and slant of the screen relative to the observer showed that linear perspective explained 95% of the slant judgments. Precision and accuracy of the judgments suggest a neural substrate that is able to make highly accurate comparisons between orientations of lines imaged at different retinal locations. The neural basis of slant from the linear perspective has not yet been clarified. Long-range connections in V1, however, and cells in V2, V4, lateral occipital cortex, and caudal intraparietal sulcus have features that suggest an involvement in slant perception.

Introduction

Despite the fact that pictures are flat, depicted surfaces can give a strong impression of receding from the picture plane. For more than half a century, researchers have tried to disentangle the many sources of information that cause slant and depth perception in pictures (Gibson, 1950; Freeman, 1966; Braunstein & Payne, 1969; Blake, Bülthoff & Sheinberg, 1993; Knill, 1998; Saunders & Backus, 2006). Linear perspective, one of the sources, has been the subject of lively debates in science (Flock, 1965; Freeman, 1965; Gibson, 1978) as well as art (Gombrich, 1975; Panofsky, 1991; Damisch, 1994). A hot-debated issue has been, for instance, the natural or cultural origin of linear perspective. Irrespective of its evolutionary explanation, linear perspective has been identified as a strong cue to slant. It is so powerful that it rivals and even dominates other slant cues if these are signaling conflicting slants (Cook, Hayashi, Amemiya, Suzuki, & Leumann, 2002; Papathomas, 2002; van Ee, Adams & Mamassian, 2003; Knill, 2007; Rogers & Gyani, 2010). The present study explores retinal sources of information that form the basis of slant induced by linear perspective and discusses neural circuitry capable of detecting and coding the required information. To study slant from the linear perspective in pictures in great detail, computational and experimental techniques have been employed that enable the separation of linear perspective–related slant from slant induced by other cues. Pictures are inherently ambiguous because, from any viewing point, each shape in a picture produces a retinal shape that could have been produced by an infinite number of shapes in three-dimensional (3D) space, the shape in the picture being just one of them. The human visual system disambiguates retinal shape by applying a priori simplicity constraints such as regularity, symmetry, and compactness (Pizlo, 2008). It is this assumption-based modus operandi of vision that enables seeing depth and slant in pictures (Erkelens, 2012).

If a trapezoid is drawn on a fronto-parallel screen (Figure 1a), observers perceive either the fronto-parallel trapezoid or a slanted shape that is generally understood to be rectangular. Slant of the rectangle signaled by linear perspective (ν) depends on slant φ depicted on the screen as well as the observer’s position relative to the screen (Figure 1b). Both screen distance and screen orientation (σ) affect ν. In this study, computations and experiments of slant are limited to variations of σ because this manipulation enables the separation of contributions to slant specified by the different cues (Erkelens, 2013).

Convergence between lines of the trapezoid in association with an assumption of parallelism and horizontal compression of the trapezoid in association with an assumption of the width of the rectangle are sources of slant information available to vision. Several
studies have shown that humans indeed exploit perspective convergence (also called parallelism) and compression (also called foreshortening) as cues to slant perception (Freeman, 1966; Cutting & Millard, 1984; Cumming, Johnston & Parker, 1993; Andersen, Braunstein, & Saidpour, 1998; Knill, 1998; Todd, Thaler, & Dijkstra, 2005; Saunders & Backus, 2006, 2007). Obliqueness of angles in association with an assumption of orthogonality has also been named as a cue to slant (Saunders & Backus, 2007). Because vertical lines border the slanted strips of the grids that are used as stimuli (Figure 1c), the obliqueness of angles is based on the same information as perspective convergence and is therefore not an independent cue. Another potential cue to slant is provided by the relative sizes of the vertical line elements of the grids (Figure 1c). This cue is not independent either. Perspective convergence, obliqueness, and vertical size ratio covary as functions of slant of the grid depicted on the screen (\(\phi\)) and slant of the screen (\(\sigma\)). These cues are extracted from the perspective shape of the proximal stimulus and as a group are called “perspective cue” in this study. Horizontal compression is derived from the width of the proximal stimulus and is called “compression cue” for short.

Besides the perspective and compression cues, many other cues contribute to the perception of slant. Established cues are blur and brightness of lines, binocular disparity (in the case of binocular viewing), ocular vergence, and accommodation (Landy, Maloney, Johnston, & Young, 1995). These cues have in common that they indicate the slant of the physical stimulus and thus the slant of the screen but not the depicted slant of the stimulus on the screen. This class of cues is dubbed screen cue. Measuring perceived slant as functions of \(\phi\) and \(\sigma\) provides the possibility to separate contributions of perspective, compression, and screen cues from each other. Until now, the strength of perspective-related cues has been estimated by pitting them against a single other cue, mostly binocular disparity, neglecting contributions of other screen-related cues (Landy et al., 1995; van Ee et al., 2003; Hillis, Watt, Landy, & Banks, 2004; Welchman, Deubelius, Conrad, Büllhoff, & Kourtzi, 2005; Knill, 2007).

Methods

Subjects

The study included 20 subjects (19 first-year physics students and the author). All but the author were naive with respect to the purpose of the study. All subjects had normal or corrected-to-normal vision and gave informed consent in accordance with the Declaration of Helsinki.
Stimuli

The physical stimuli were grids of white (93 cd/m$^2$) lines against a black (approximately 2 cd/m$^2$) background displayed on a thin-film transistor liquid-crystal display monitor (21-in. LaCie 321, 1600 × 1200 pixels, 75 Hz). The display area of the monitor was approximately $43\frac{1}{8} \times 32\frac{1}{8}$. The grids were perspective projections of five differently slanted, planar vertical strips, each composed of four square elements. In the frontal view, each element measured $3\frac{1}{8} \times 3\frac{1}{8}$, and the depicted slants of the strips on the screen were $(-30^\circ, -15^\circ, 0^\circ, 15^\circ, 30^\circ), (-45^\circ, -30^\circ, 0^\circ, 30^\circ, 45^\circ), (-60^\circ, -45^\circ, 0^\circ, 45^\circ, 60^\circ), (-75^\circ, -60^\circ, 0^\circ, 60^\circ, 75^\circ)$, respectively, of the four grids that were used as stimuli (Figure 1c). Multislanted grids with different combinations of slant were chosen as stimuli to investigate the interaction between different slants. Slants induced by planar grids using similar methods were investigated in a previous study (Erkelens, 2013).

Experimental procedure

A chin rest restricted head movements so that the center of the forehead (the “cyclopean eye”) was positioned at the center of projection of the stimuli, when the screen was fronto-parallel ($\sigma = 0^\circ$). Viewing was binocular. The setup was placed in a normally lit room because pictures are usually viewed against visible frames and backgrounds. The monitor was centered on a turntable at a distance (center of the forehead to the center of the turntable) of 57 cm. A second turntable, centered within a protractor and carrying a vertical rectangular pad (23 × 16 cm), was placed between the center of the monitor and the chin rest at a distance of 21 cm from the forehead. Subjects oriented the pad with both hands to indicate the perceived slant of a vertical strip of the grid. The vertical angle between the bottom of the stimuli and the upper edge of the pad was $38^\circ$ so that the subjects alternated their fixation between up and far and near and down during the adjustment of the pad. The slant of the screen $\sigma$ was varied between $-75^\circ$ and $75^\circ$ in steps of $15^\circ$. Slants of the four stimuli were measured in separate blocks. Combinations of strip slant and screen orientation were selected in a fully randomized order. To familiarize the subjects with the task, 3 pretrials were ran without a stimulus on the screen. Subjects were asked to match the slant of the screen while they viewed the screen binocularly. The subjects received feedback on their slant settings. Results of pretrials were evaluated and the trials repeated if the settings were too inaccurate. The purpose of practicing and feedback was to ensure that the slant settings reflected perceived slant as much as possible. After some practice, all subjects were able to judge the slants of the screen within error margins of $5^\circ$ of slant and without a noticeable bias. Viewing condition and method of slant judgment were extensively tested in a previous study (Erkelens, 2013). The subjects were instructed to judge the slant of individual strips of the grids relative to their heads in the real experiment and to ignore the slant of the screen.

Results

Computation of slant specified by cues

Perspective, compression, and screen cues predict relationships between $v$, $\phi$, and $\sigma$ (Figure 2) that are described by the following analytical equations (derivations are presented in the Appendix):

\begin{align}
\text{Perspective: } v_p & = \arctan(\tan \sigma + r \sec \sigma \tan \phi) \quad (1) \\
\text{Screen: } v_s & = \sigma \quad (2) \\
\text{Compression: } v_c & = \pm \arccos(\cos \phi \cos \sigma) \quad (3) \\
\text{Disambiguated compression: } v_d & = \arccos(\cos \phi \cos \sigma) \text{sign}(v_p) \quad (4)
\end{align}
strip, although slopes were smaller than 1, indicating that slant was underestimated. Matched slant was just weakly dependent of $\sigma$ (showing shallow slopes) for the highly slanted strips ($\phi = \pm 60^\circ, \pm 75^\circ$). Both observations indicate that perceived slant followed slant specified by perspective, at least up to a certain extent. To assess how accurate perceived slant was described by slant specified by a linear combination of cues, all data obtained from slant judgments in response to the different stimuli were accumulated in a single set. Six stimulus slants were presented twice in two of the stimuli ($\phi = \pm 30^\circ, \pm 45^\circ, \pm 60^\circ$). To decide whether all sets of data could be included in the analyses, paired sets were compared for statistical differences.

None of the paired sets showed significant differences (repeated-measures analysis of variance: all $F_{1,20} < 0.09, p > 0.77$). Stimulus slant $\phi = 0^\circ$ was presented in all stimuli and thus measured four times. Again, the data did not significantly differ, $F_{3,40} = 0.75, p = 0.53$. The absence of statistical differences between perceived slants in response to equal depicted slants in different stimuli showed the absence of any between-strip interaction. In turn, the absence of between-strip interaction showed that the perceived slant of each strip was fully based on the geometry of the strip itself. Because of this result, all further analyses were conducted with inclusion of all data sets.

To compare perceived slant with predictions of the assumption-based slant models, mean slant judgments were arranged and displayed as functions of $\phi$ and $\sigma$ (Figure 4a). Visual inspection showed a remarkable resemblance between matched slant and slant specified by the perspective cue (Figure 2). Between-subjects variability indicated by standard deviations (SDs) was modest for many combinations of $\phi$ and $\sigma$ (Figure 4b). Standard deviations were smaller than 5\(^\circ\) when both $\phi$ and $\sigma$ were smaller than or equal to $\pm 30^\circ$. Between-subjects variability was largest when $\phi$ and $\sigma$ were large and opposite of sign, suggesting that variability was related to the difference in slant indicated by the perspective and screen cues.

To estimate the predictive quality of the models, the residual sum of squares as a fraction of the total sum of squares (RSS/TSS) was computed for fits of the models to the mean matched slant results (Figure 5a). The minimum values of RSS/TSS were 0.99 ($w_c = 0.05$) for compression, 0.69 ($w_c = 0.45$) for screen, 0.17 ($w_d = 0.55$) for disambiguated compression, and 0.05 ($w_d = 0.65$) for perspective. These values imply that the individual cues explain very different percentages of the matched slants: compression 1\%, screen 31\%, disambiguated compression 83\%, and perspective 95\%. An even slightly better fit to the data than provided by the perspective model was found by combining perspective and screen cues (Figure 5b). RSS/TSS reached a minimum value of 0.03 at the cue-weight combination

Figure 3. Judgments of perceived slant. Mean slant judgments ($n = 20$) for each of the five strips of the four grids. Error bars, ±SD. Slants of the strip: $\phi = 0^\circ$ (blue), $\pm 15^\circ$ (magenta), $\pm 30^\circ$ (cyan), $\pm 45^\circ$ (red), $\pm 60^\circ$ (green), $\pm 75^\circ$ (yellow).

The parameter $r$ in Equation 1 is the ratio between the distance of viewing and the distance of projection. The ratio was equal to 1 in the experiment. According to Equation 1, the perspective cue predicts that $v_p = \sigma$ if $\phi = 0^\circ$, implying that pictures of frontal surfaces retain their slant relative to the screen and not the viewer in oblique viewing directions. The perspective cue further predicts that $v_p$ approaches $\pm 90^\circ$ for all $\sigma$ if $\phi$ approaches $\pm 90^\circ$. This relationship implies that $v$ becomes independent of $\sigma$ for highly slanted surfaces, meaning that such surfaces retain their slant relative to the viewer irrespective of the direction at which the picture is viewed. The screen cue (Equation 2), including binocular disparity, indicates the slant of the screen and therefore is independent of $\phi$. In Equation 3, $v_c$ is a symmetric function of $\phi$ and $\sigma$, rendering the compression cue by itself ambiguous as a cue for perceived slant because it does not distinguish between leftward and rightward slant. In combination with sign information provided by the perspective cue, however, disambiguated compression (Equation 4) provides useful slant information that may serve as a cue.

**Experiment: slant judgments**

Observers judged the slants of depicted grids (Figure 1c) by matching the slant of a rectangular pad to that of one of the vertical strips of the grid. Matched slant (Figure 3) varied linearly with $\sigma$ for the central $\phi = 0^\circ$
$w_p = 0.60$ and $w_s = 0.10$. The minimum of 0.03 means that the linear combination of perspective and screen cues explains 97% of matched slant. The combination of disambiguated compression and screen cues was a slightly less predictive combination of cues (Figure 5c). RSS/TSS reached a minimum of 0.10 at $w_d = 0.50$ and $w_s = 0.25$, explaining 90% of the data. The perspective cue by itself (Figure 5a) was almost as predictive as the combination of perspective and disambiguated compression (Figure 5d). For the latter combination, RSS/TSS reached a shallow minimum of 0.04 at $w_p = 0.80$ and $w_d = -0.15$, explaining 96% of the matched slants. The three-cue combination of perspective, disambiguated compression, and screen was as powerful as the two-cue combination of perspective and screen. RSS/TSS showed a minimum of 0.03 at $w_p = 0.70$, $w_s = 0.10$, and $w_d = -0.10$, explaining 97% of the matched slants.

To estimate confidence intervals for the cue weights, the regression analysis was repeated for matched slant data of individual subjects (Table 1). For the group of 20 subjects, 99% confidence intervals are about equally large as one SD. Residual errors computed from the individual data (Table 1) were larger than those computed from the mean data (Figure 5). The difference is explained by the reversed order of fitting and averaging operations. The one-cue fits show that perspective was twice or more as powerful as each of the other cues in explaining matched data. Addition of the screen cue to the model resulted in a modest increase in its explanatory power, namely, from 88% (err = 0.12) to 92% (err = 0.08). Extension of the model with disambiguated compression caused minor changes in the weighting factors but no improvement in power. Subjects showed considerable differences in the contributions of the various cues to perceived slant. The relatively large SDs of the cue weights express the intersubject variability. The combination of perspective and screen cues with or without disambiguated compression explained perceived slant up to a very high degree without showing many intersubject differences, as is expressed by the small SDs of the residual errors. From the one-cue computations, disambiguated compression seemed a strong cue. In combination with the perspective cue, it did not make a significant contribution to explaining the experimental results. The discrepancy between the single- and multiple-cue
results is caused by the fact that disambiguated compression shows relationships to $\phi$ and $\sigma$ that are partially similar to those of perspective. To judge the quality of the best model predictions given by the combination of perspective and screen cues in detail, residual errors were computed as functions of $\phi$ and $\sigma$ for the optimal cue-weight combination (Figure 6). Absolute residual errors and SDs (Figure 4b) were of the same order of magnitude.

Table 1. Mean ($\pm$SD) cue weights ($w$) and minimum residual errors ($res = \text{minimum of RSS/TSS}$) computed from fits of slant models to the matched slant data of individual subjects. The slant models were based on linear combinations of perspective ($P$), screen ($S$), compression ($C$), and disambiguated compression ($D$) cues.

<table>
<thead>
<tr>
<th>Number of cues</th>
<th>$w$</th>
<th>$res$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$P$</td>
<td>0.64 ($\pm$0.08)</td>
<td>0.12 ($\pm$0.05)</td>
</tr>
<tr>
<td>$D$</td>
<td>0.56 ($\pm$0.08)</td>
<td>0.24 ($\pm$0.07)</td>
</tr>
<tr>
<td>$S$</td>
<td>0.44 ($\pm$0.16)</td>
<td>0.70 ($\pm$0.15)</td>
</tr>
<tr>
<td>$C$</td>
<td>0.05 ($\pm$0.02)</td>
<td>0.99 ($\pm$0.01)</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$P$</td>
<td>0.59 ($\pm$0.09)</td>
<td>0.08 ($\pm$0.04)</td>
</tr>
<tr>
<td>$S$</td>
<td>0.13 ($\pm$0.12)</td>
<td></td>
</tr>
<tr>
<td>$P$</td>
<td>0.71 ($\pm$0.12)</td>
<td>0.12 ($\pm$0.05)</td>
</tr>
<tr>
<td>$D$</td>
<td>$-0.08$ ($\pm$0.10)</td>
<td></td>
</tr>
<tr>
<td>$D$</td>
<td>0.50 ($\pm$0.08)</td>
<td>0.15 ($\pm$0.06)</td>
</tr>
<tr>
<td>$S$</td>
<td>0.23 ($\pm$0.13)</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$P$</td>
<td>0.59 ($\pm$0.17)</td>
<td>0.08 ($\pm$0.04)</td>
</tr>
<tr>
<td>$S$</td>
<td>0.16 ($\pm$0.13)</td>
<td></td>
</tr>
<tr>
<td>$D$</td>
<td>$-0.02$ ($\pm$0.14)</td>
<td></td>
</tr>
</tbody>
</table>

Table 1. Mean ($\pm$SD) cue weights ($w$) and minimum residual errors ($res = \text{minimum of RSS/TSS}$) computed from fits of slant models to the matched slant data of individual subjects. The slant models were based on linear combinations of perspective ($P$), screen ($S$), compression ($C$), and disambiguated compression ($D$) cues.

The excellent agreement between matched and predicted slant indicates that the combination of perspective, screen, and disambiguated compression cues explains the perceived slant of pictured grids. The ratios between the weights of the cues show that perspective cues prevail for grid stimuli depicted on a slanted screen. It may be that the ratio was driven to an upper limit by the instruction given to the subject to ignore the slant of the screen. Such an effect of instruction would imply that the cue weights are, up to a certain extent, under voluntary control. The choice to look into a picture versus to look at a picture illustrates a certain degree of freedom. Manipulation of the weights may explain the fact that subjects underestimated slant specified by perspective by about 35%. The strong underestimation was measured in all subjects.

The conclusion of the residual error analysis is that, either by itself or in combination with other cues, the perspective cue best explains the data. The analysis shows that perspective is the most powerful source of information for slant perceived in the grid stimuli. Computations of slant specified by compression showed that it is an ambiguous cue to slant. Multiplying $v_c$ (Equation 3) by the sign of $v_p$, disambiguates $v_d$ (Equation 4) such that it correctly signals leftward and rightward slant (Figure 2). Thus, compression can act as a cue to slant only if it includes sign information provided by the perspective cue or another unambiguous cue. Disambiguated compression by itself explains a high percentage of the data. At first sight, the high percentage suggests that disambiguated compression is a powerful cue to slant. However, its ability to serve as a cue depends on the presence of the perspective cue (Figure 7). Furthermore, it hardly contributes to explaining perceived slant if it is present in combination with perspective (Figure 5d; Table 1). As a result, disambiguated compression is an insignificant cue to slant in perspective drawings.
Retinal stimulus for slant from linear perspective

Equation 1 expresses slant from perspective convergence in terms of depicted and screen slants; however, it does not relate slant to a retinal measure. To find the effective retinal stimulus for perceived slant, relationships were examined between slant specified by perspective and geometric quantities of the proximal stimulus. The proximal stimulus contains several quantities that are related to slant on assumptions of parallelism, orthogonality, and equal vertical line lengths. As an example, the relationship will be worked out for perspective convergence in detail to demonstrate how perspective slant may be implemented in the brain by a fairly simple neural mechanism. The choice for perspective convergence is mainly motivated by known properties of the early visual system and simplicity of the required computations. As will be shown later, slant from perspective convergence follows from comparing the orientations of two vertically separated oblique or horizontal lines in the stimulus. Any combination of lines contains the appropriate slant information, implying that $4 + 3 + 2 + 1 = 10$ pairs of lines signaled the slant of each strip of the grid stimuli. Orientation specificity is the dominant feature of neurons in V1. In monkeys, about 80% of the cells are orientation selective, in cat almost 100% (Hubel, 1988). These arguments make perspective convergence a potentially basic cue to slant. Slant from the assumption of orthogonality requires comparison of at least two angles of separated vertices. The fact that a single vertex angle is not sufficient is easily demonstrated in Figure 7a. In each slanted strip, the vertices of the extreme and interior elements have different angles, although the elements specify equal slants. Starting from orientation selectivity as the basis for neural processing, computations of angles require a second stage of neural processing and comparisons of angles a third stage. Together, these arguments show that oblique angles are not a simple cue to slant. It is even possible that an assumption of orthogonality is not used for slant perception in its own right but is a particular case of an assumption of symmetry. Sawada and Pizlo (2008) showed that subjects are good at detecting slanted symmetric shapes in skewed retinal images. Saunders and Knill (2001) provided convincing evidence that skew in association with an assumption of symmetry is a strong cue to surface slant. The computational model of Sawada and Pizlo (2008) suggests that defining symmetry axes, performing shear transformations, and computing compactness are necessary operations for the detection of symmetry. It is not very likely that this type of processing occurs early in the visual cortex.

Slant specified by the size ratios of vertical lines, another potential cue to slant, involves comparison of lengths of two adjacent vertical line pieces in the stimulus. Any pair of lines contains the appropriate slant, implying that 10 combinations of lines may signal the slant of each strip. Measurement of line lengths requires end-stopped cells that are modestly present in V1. Just 20% are found in Layers 2 and 3 and almost none elsewhere (Hubel, 1988). These arguments make the relative size of vertical lines a less basic cue to slant in comparison with perspective convergence.

Geometric computation reveals that slant specified by perspective convergence can be expressed in closed form if it is expressed in terms of the retinal convergence angle between separated lines ($\alpha$) and the angular distance between the lines ($\beta$) measured perpendicular to the mean orientation of the lines (Figure 8a). The relationship between $\alpha$, $\beta$, and $v_p$ is given by the following equations (derivations are given in the Appendix):
\[
t_p = \arctan\left(\frac{\tan a}{\tan b}\right),
\]

Equations 5 and 6 show that \(t_p\) is fully determined by a combination of the retinal angle \(a\) and the retinal separation \(b\) of lines. This implies that slant specified by perspective convergence can be extracted from retinal information alone, without requiring explicit knowledge of the viewing distance (Papathomas, 2002; Rogers & Gyani, 2010). The relationship between \(v_p\) and \(a\) and \(b\) explains why slant perceived in pictures increases with viewing distance (Erkelens, 2013). \(v_p\) increases because \(a\) remains constant and \(b\) decreases with viewing distance. Repploting of the matched slant data that were measured as functions of \(\phi\) and \(\sigma\) (Figure 4a, b) shows that all data obey the following relationship: matched slant = 0.65 \(v_p\) \(= 0.65\left(\arctan\left(\frac{a}{b}\right)\right)\). The single relationship describing all the data underlines that the combination of \(a\) and \(b\) is an effective stimulus for slant perception. The gain of 0.65 for all combinations of \(\phi\) and \(\sigma\) indicates a common source for the underestimation of all slant judgments (Erkelens, 2013).

**Thresholds of convergence-angle detection**

The relationship between \(a\) and \(v_p\) shows high slopes for small angles of retinal convergence and tapers off for larger angles (Figure 8b). Near the origin, the slope is about 19, implying that slants of 0°, 15°, and 30° are associated with retinal convergence angles of about 0°, 0.8°, and 1.7°, respectively. The subjects easily distinguished slant differences of 15°, implying that convergence angles (\(a\)) of just 0.8° at a separation (\(b\)) of 3° were detected and perceived as slant (Figure 8c), be it strongly underestimated. Slant SDs less than 5° measured for small angles of \(\phi\) and \(\sigma\) (Figure 4d) were associated with variability below 0.2° in retinal angles between lines projected 3° apart. The small variability suggests that the minimum convergence angle of 0.8° in the stimuli was well above all subjects’ thresholds for seeing slant from perspective convergence. The measured minimum slant differences imply that thresholds of perspective convergence involved in slant perception were probably of the same level as perceptual thresholds of orientation discrimination that have been measured for fronto-parallel lines and gratings (Heeley & Buchanan-Smith, 1998). From sensitivity experiments to sinusoidal curvature, Tyler (1973) concluded that the limit for orientation differences was about 0.33° within an area of integration of the order of 2.5°. In terms of the current slant data, his results imply that \(a = 0.33°\) in combination with \(b = 2.5°\) are limits for 2D orientation discrimination. Sullivan, Oatley, and Sutherland (1972) measured similar limits in a Vernier acuity task. The agreement between the limits measured for 2D orientation discrimination and 3D slant judgment suggests common circuitries for the neural processing of orientation and slant, in turn suggesting...
that processing of perspective convergence occurs at an early stage in the visual cortex.

**Neural coding of slant**

The power of perspective convergence for slant perception in pictures suggests a specific and not yet proposed mechanism for the detection and coding of orientation differences in the primary visual cortex. Coding of perspective convergence requires accurate measurement of small orientation differences between pairs of lines or contours that are imaged at different retinal locations. Differences in orientation can be as small as on the order of minutes of arc, whereas separations between retinal locations are on the order of degrees. The combination of both components, angle between two line orientations (a) and angular distance between two retinal locations (b), carries the signal for perceived slant under two conditions: (a) lines or contours must belong to the same shape and (b) orientation of the line connecting the two retinal locations is orthogonal to the mean orientation of the lines or contours measured at the two locations (Figure 8a).

**Potential neural substrates**

To date, there are no reports of neural circuitries that qualify for both criteria. Yet there are reports of neural structures in V1 and V2, whose functions are still called a mystery (Allman, Miezin, & McGuinness, 1985; Anzai, Peng, & Van Essen, 2007) but whose properties suggest a role in the perception of slant. In the visual cortex of carnivores and primates, neurons selective for the orientation of visual edges are organized in orientation columns, which are vertical arrays of neurons that prefer the same orientation (Hubel & Wiesel, 1977; Blasdel & Salama, 1986; Bonhoeffer & Grinvald, 1991; Ohki, Chung, Ch’ng, Kara, & Reid, 2005; Ohki, Chung, Kara, Hübener, Bonhoeffer, & Reid, 2006). An important feature of V1 connections is the plexus of long-range horizontal connections most prominent in Layers 2 and 3, which enable neurons to integrate inputs from an area of cortex representing an area of visual field that is much larger than their classical receptive fields. Connections beyond their classical receptive fields are a prerequisite for neurons coding for slant. Although the extent and orientation dependence of long-range horizontal connections have been found to match the properties of salient contours and the geometry of natural scene contours (Gilbert & Wiesel, 1989; Li & Gilbert, 2002; Stetter, Das, Bennett, & Gilbert, 2002; Gilbert & Li, 2012), the function performed by this general feature of the cortical architecture remains obscure (Muir et al., 2011). Neurons selective for the orientation of visual edges at different retinal locations in combination with interneuron long-range connections form a structure that seems the natural candidate for the coding of slant. Cells in V2 responsive to combinations of orientations have been attributed the function of encoding contours and providing cues for surface segmentation (Anzai, Peng, & Van Essen, 2007). Showing optimal responses to different orientations (a) in different parts of their receptive fields (b) implies, however, that such cells may carry the required signal for the coding of 3D slant. Another component essential for the coding of slant is a structure signaling that detected combinations of orientations belong to a common shape. V4 and the lateral occipital cortex (LOC) are likely candidates because these areas play an important role in human object recognition (Kourtzi & Kanwisher, 2001; Silson et al., 2013). Recent high-density electroencephalogram recordings show that the LOC is critically involved in perceptual decisions about object shape (Ales, Appelbaum, Cottereau, & Norcia, 2013). V4 cells having 3D orientation tuning (i.e., tuning for specific slants) may also be involved in signaling a common shape to V2 or V1 orientation detectors (Hinkle & Connor, 2002). Cells responsive to linear perspective have been reported in the caudal intraparietal sulcus (CIP; Tsutsui, Jiang, Yara, Sakata, & Taira, 2001; Tsutsui, Taira, & Sakata, 2005). These neurons were also sensitive to binocular disparity and thus probably involved in the integration of both types of information. Nevertheless, the authors did not look for perspective-related activity in the occipital cortex as they did for binocular disparity. In conclusion, a number of neural structures in V1, V2, V4, LOC, and CIP show properties that are essential for the detection and coding of 3D slant. However, the neural activity of most structures has never been tested with specific slant stimuli. The present study provides stimuli and experimental techniques that enable the isolation of assumption-based slant from slant induced by screen-related cues. Computations and psychophysical results show that the combination of perspective-induced convergence (a) and angular separation (b) of lines or contours carries the information for 3D slant. Further studies on the neural processing of slant will provide important insight into networks mediating perspective cues for perception of the 3D world.

**Keywords:** human vision, depth perception, picture perception

**Acknowledgments**

The author thanks S. Dumoulin for inspiring discussions and helpful comments. Thanks also to P. Schiphorst for technical assistance.
Commercial relationships: none.
Corresponding author: Casper J. Erkelens.
Email: c.j.erkelens@uu.nl.
Address: Helmholtz Institute, Utrecht University, Utrecht, The Netherlands.

References


