

What is second-order vision for? Discriminating illumination versus material changes

Andrew J. Schofield

School of Psychology, University of Birmingham,
Birmingham, UK



Paul B. Rock

School of Psychology, University of Birmingham,
Birmingham, UK



Peng Sun

School of Psychology, University of Birmingham,
Birmingham, UK



Xiaoyue Jiang

School of Psychology, University of Birmingham,
Birmingham, UK



Mark A. Georgeson

School of Life and Health Sciences, Aston University,
Birmingham, UK



The human visual system is sensitive to second-order modulations of the local contrast (CM) or amplitude (AM) of a carrier signal. Second-order cues are detected independently of first-order luminance signals; however, it is not clear why vision should benefit from second-order sensitivity. Analysis of the first- and second-order contents of natural images suggests that these cues tend to occur together, but their phase relationship varies. We have shown that in-phase combinations of LM and AM are perceived as a shaded corrugated surface whereas the anti-phase combination can be seen as corrugated when presented alone or as a flat material change when presented in a plaid containing the in-phase cue. We now extend these findings using new stimulus types and a novel haptic matching task. We also introduce a computational model based on initially separate first- and second-order channels that are combined within orientation and subsequently across orientation to produce a shading signal. Contrast gain control allows the LM + AM cue to suppress responses to the LM – AM when presented in a plaid. Thus, the model sees LM – AM as flat in these circumstances. We conclude that second-order vision plays a key role in disambiguating the origin of luminance changes within an image.

Keywords: 3D surface and shape perception, shading, spatial vision, computational modeling

Citation: Schofield, A. J., Rock, P. B., Sun, P., Jiang, X., & Georgeson, M. A. (2010). What is second-order vision for? Discriminating illumination versus material changes. *Journal of Vision*, 10(9):2, 1–18, <http://www.journalofvision.org/content/10/9/2>, doi:10.1167/10.9.2.

Introduction

The human visual system is sensitive to variations of second-order cues such as modulations of the local contrast (CM) of textured stimuli. This is true for both moving (see Baker, 1999 for an early review) and static (Badcock, Clifford, & Khuu, 2005; Dakin & Mareschal, 2000; Georgeson & Schofield, 2002; Graham & Sutter, 2000; Henning, Hertz, & Broadbent, 1975; Larsson, Landy, & Heeger, 2006; Nachmias, 1989; Nachmias & Rogowitz, 1983; Schofield & Georgeson, 1999, 2003; Sutter, Sperling, & Chubb, 1995) stimuli, although here we concentrate on static cues. There is strong psychophysical evidence to suggest that static CM is detected separately from first-order luminance modulations (LMs). For example, there is no sub-threshold facilitation between the cues (Schofield & Georgeson, 1999), they can be

distinguished at detection threshold (Georgeson & Schofield, 2002), lateral interactions are different for the two cues (Ellemberg, Allen, & Hess, 2004), their channel structure is different (Ellemberg, Allen, & Hess, 2006), noise masking is doubly dissociated (Allard & Faubert, 2007), they make separate contributions to global form detection (Badcock et al., 2005) and different contributions to contour linking processes (Hess, Ledgeway, & Dakin, 2000). Finally, although most retinotopic visual areas respond to both LM and CM, there is preferential fMRI adaptation for CM in the higher areas (specifically VO1, LO1, and V3a; Larsson et al., 2006).

It is also clear, however, that CM and LM are integrated or partially integrated in some cases. For example, contrast modulations of high-contrast grating carriers mask LM signals (Henning et al., 1975; Nachmias & Rogowitz, 1983), but modulations of low contrast noise carriers do not (Schofield & Georgeson, 1999). LM masks

the detection of CM in noise carriers but not vice versa (Ellemberg et al., 2006; Schofield & Georgeson, 1999), and similar asymmetric interference has been found for global form detection (Badcock et al., 2005). The orientation of first-order stimuli affects the perceived orientation of second-order stimuli (Morgan, Mason, & Baldassi, 2000). The signal types combine at low contrasts to improve perceptual accuracy (Smith & Scott-Samuel, 2001). Further, tilt and contrast reduction after-effects transfer between LM and CM (Georgeson & Schofield, 2002), as does the tilt illusion (Smith, Clifford, & Wenderoth, 2001). Finally, we have previously shown that LM and CM interact in the perception of shape from shading (Schofield, Hesse, Rock, & Georgeson, 2006).

The physiological evidence for independent first- and second-order mechanisms is less clear-cut and comes mainly from studies using moving stimuli. Mareschal and Baker (1998) found cells in cat area 18 that are responsive to second-order stimuli, but these also responded to first-order stimuli: suggesting early integration. However, typically, preferred frequencies for the two cues were slightly different. They concluded that such cells were likely to take their input from independent first- and second-order sub-mechanisms (see also Song & Baker, 2006; Zhou & Baker, 1996). Further, in physiology, it is common to search for cells using first-order stimuli. Any cell that is then found to be sensitive to second-order cues will, by definition, also be sensitive to first-order stimuli. Finally, second-order signals may be extracted in another visual area; V3a has been implicated in second-order processing for both static (Larsson et al., 2006) and moving stimuli (Ashida, Lingnau, Wall, & Smith, 2007). Perhaps second-order signals are extracted in V3a and fed back to V1/V2.

Despite the above evidence for separate but interacting first- and second-order mechanisms, psychophysically human vision is an order of magnitude less sensitive to CM than LM (Schofield & Georgeson, 1999) and similar, if less extreme, results have been found for motion in cat area 17/18 (Hutchinson, Baker, & Ledgeway, 2007; Ledgeway, Zhan, Johnson, Song, & Baker, 2005; Mareschal & Baker, 1998; Zhou & Baker, 1996) and monkey MT (Albright, 1992). This suggests that CM is something of a secondary cue, and it is not yet clear why the independent detection of static second-order cues is beneficial to human vision. We now address this question.

Human vision presumably obtains some advantage from processing first- and second-order cues independently and indeed from detecting second-order cues at all. Johnson and Baker (2004) measured the relationship between patterns of LM and CM in natural scenes and found the two cues to be highly correlated on an unsigned magnitude metric. This implies that CM variations tend to occur alongside LM. However, Schofield (2000) performed a similar analysis using a signed metric and found that whereas the two cues may be strongly correlated within a single image the sign of the correlation varies between

images, such that they are uncorrelated over an ensemble of images. Taken together, these results suggest that CM is an informative cue in natural images, but that information may be conveyed by its relationship with LM rather than its mere presence.

In this paper (as previously, Schofield et al., 2006), we prefer to use the term amplitude modulation (AM) over CM because although they are mathematically equivalent when presented alone, when combined with LM they can be interpreted as distinct image properties with AM being the better description for our purposes. Schofield et al. (2006) showed that LM and AM are yoked whenever an albedo-textured surface is shaded or in shadow (see Figure 1 for a natural example of such shading and Schofield et al., 2006 for a full account of the yoking between these cues). Albedo textures represent locally smooth surfaces whose local reflectance changes creating a visual texture. So LM + AM represents a strong cue to shading/shadows when certain textured surfaces are present.

People see sinusoidal shading patterns as sinusoidally undulating surfaces (Kingdom, 2003; Pentland, 1988; Schofield et al., 2006; Schofield, Rock, & Georgeson, submitted for publication) even though such surfaces only give rise to sinusoidal shading in restricted circumstances. We presume that the luminance component of the LM + AM signal is coded as a shading pattern and then interpreted as a corrugated surface via shape from shading (Christou & Koenderink, 1997; Erens, Kappers, & Koenderink, 1993; Horn & Brooks, 1989; Kleffner & Ramachandran, 1992; Langer & Bülthoff, 2000; Ramachandran, 1988; Todd & Mingolla, 1983; Tyler, 1998) whereby luminance level is equated with surface gradient such that the parts of the surface that are most luminous are seen as being oriented toward the illuminant. When the direction of the illuminant is unknown, humans assume a lighting-from-above prior (Adams, Graf, & Ernst, 2004; Brewster, 1826; Mamassian & Goutcher, 2001; Ramachandran, 1988; Rittenhouse, 1786; Sun & Perona, 1998). Our earlier results (Schofield et al., 2006) with LM + AM sinusoids are consistent with this interpretation, except that we now propose an illumination prior that is a mixture of diffuse and point source lighting (Schofield et al., submitted for publication; Schofield, Rock, Georgeson, & Yates, 2007).

The filter–rectify–filter model used by Schofield (2000) to extract second-order cues from natural images was sensitive to AM, and it seems likely that natural images containing positively correlated first- and second-order cues are dominated by shadows and shading. However, what is the composition of those images that contain negatively correlated cues?

Transparent overlays also give rise to second-order cues in natural stimuli (Fleet & Langley, 1994). The specific case of a semi-opaque light (or milky) transparency is pertinent here. Those parts of a textured surface that are obscured by such a transparency suffer an increase in mean luminance (e.g., if the base color of the overlay is white, its luminance will be higher than the mean

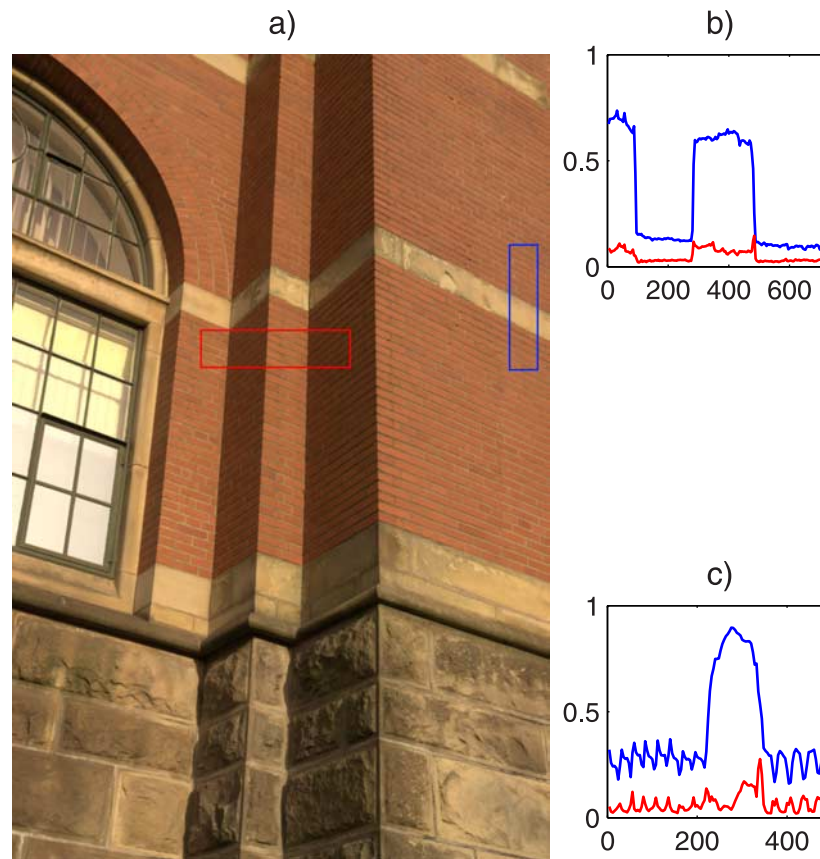


Figure 1. (a) A natural image showing part of a building on the University of Birmingham campus. The building “steps” out twice working left to right and the orientation of the faces produces shading but not cast shadows. The brick sections are, approximately, a reflectance texture of the type described in the text. The image also shows gross reflectance changes, most notably the strips of sandstone among the red brick sections. The red and blue boxes show approximate sampling regions for the traces of (b) and (c), respectively. The red section of (a) was extracted and rotated so that the shading edges were vertical. The blue section of (a) was extracted and rotated so that the sandstone edges were vertical. Sample sections were also converted to grayscale. (b) Mean (blue line) and standard deviation (red line) of the gray level values in each column of the rotated red section. Mean pixel values are a measure of luminance whereas their standard deviation measures luminance amplitude or range. Transitions of high to low luminance (LM) are clearly mimicked by changes in luminance amplitude (AM) and the two cues are positively correlated. (c) Mean and standard deviation for the columns in the rotated blue section of (a); here the transition to high luminance in the sandstone section is not mirrored by a change in standard deviation.

luminance of the texture) but a decrease in local amplitude (the difference between the light and dark parts of the texture will fall due to the blurring caused by the semi-transparent medium). This configuration exhibits negatively correlated LM and AM (LM – AM; note, however, that if the transparency is dark, LM and AM will again be positively correlated). The notion that LM – AM is a possible cue for transparency is supported by the qualitative description of such stimuli given by Georgeson and Schofield (2002; they used the term LM – CM). If LM – AM is seen as a cue to transparency, then the overall perception is likely to be of flat surfaces although the semi-transparent regions may be seen as being in front of the main surface. LM – AM might also be interpreted as a material change, as there is no restriction on the relationship between LM and AM when two surfaces comprising materials with different textures are abutted (see Figure 1).

The idea that LM – AM may be interpreted as either a material change or as an overlaid transparency was given empirical support by our previous finding that this cue is seen as flat when presented in a plaid with LM + AM (Schofield et al., 2006). LM + AM is, by contrast, seen as a shading cue and is therefore perceived as corrugated in depth via shape from shading. However, when presented alone LM – AM is also seen as corrugated albeit less strongly (less reliably) than LM + AM. Why might LM – AM be seen as flat in some cases and corrugated in others? There are cases where undulating surfaces *can* produce negatively correlated LM and AM. An example of such a surface would be a physically textured (rough) surface under certain illumination conditions (see Figure 2 of Schofield et al., 2006). Thus, we previously concluded that whereas LM + AM is a strong cue to shading, LM – AM is rather ambiguous when seen alone. However, when intimately associated with LM + AM as in the case of a

plaid stimulus where the two cues are necessarily presented with the same texture carrier the interpretation of LM + AM as being due to shading seems to force the interpretation of LM – AM as being due to some sort of material change (Schofield et al., 2006).

The notion that the relationship between LM and AM provides a key for separating shading and shadows from material changes has important implications for human vision and applications in machine vision. In principle, a given image can arise from an infinite number of scene and lighting combinations. Human vision may make considerable use of stored knowledge about the world in a top-down fashion to correctly interpret visual scenes. However, natural images may also contain cues that can be used to disambiguate the incoming luminance variations via bottom-up processes. Specifically, luminance variations are ambiguous; they may result from changes in illumination (shadows and shading) or changes in surface reflectance. If human vision were only sensitive to luminance, its ability to distinguish these possibilities on the basis of low-level cues would be greatly restricted. Barrow and Tenebaum (1978) showed how some progress can be made toward the separation of illumination and reflectance in a “luminance only” system, but they also highlighted the potential benefits of being sensitive to other cues and the importance of understanding how cues relate to one another in real-world stimuli. Others have shown that hue can be used to separate illumination from reflectance changes (see, for example, Kingdom, 2003; Olmos & Kingdom, 2004; Tappen, Freeman, & Adelson, 2005). Here we consider the use of AM as a cue to separate the luminance changes due to variations in surface reflectance from those due to variations in illumination or shading, and we provide a simple bottom-up model—based on both the filter–rectify–filter model of second-order vision (Wilson, Ferrera, & Yo, 1992) and the processing scheme for envelope neurons proposed by Zhou and Baker (1996)—that can account for our psychophysical results.

In our earlier study (Schofield et al., 2006), we asked observers to make relative depth judgements about pairs of probe points from which we derived normalized gradients before reconstructing perceived surface profiles: we did not measure perceived depth directly. Thus, we were unable to express perceived depth in absolute terms, unable to measure differences in depth between stimuli with very different signal strengths, and unable to distinguish between low-relief and unreliable depth percepts. Further, participants in our earlier experiments reported that the depth probe task felt artificial because the probe markers did not appear to be attached to the surface. We avoided these problems here by asking observers to match the properties of a haptic surface to the perceived corrugations in a co-located visual stimulus. This task felt natural to participants and gave direct and absolute estimates of perceived depth amplitude.

We report three experiments. In the first experiment, we fixed the position of the haptic cue based on the results of a pilot study and asked observers to set the amplitude of the haptic undulations to match the perceived surface undulations. Our previous study (Schofield et al., 2006) only measured depth profiles at two levels of LM (for fixed AM) and found little difference between these conditions. We now measure perceived depth amplitude (PDA) as a function of signal strength, varying LM and AM together (Experiment 1) yielding a better understanding of how LM and AM interact at different signal strengths. In Experiments 2 and 3, we fixed the contrast of the LM cue and measured PDA as a function of AM signal strength in both plaid (Experiment 2) and single component (Experiment 3) stimuli, exploring the role of AM in more detail. We also present a biologically plausible model providing a good fit to the data suggesting that human performance in this task can be explained by a bottom-up system that first detects and then integrates first- and second-order information.

General methods

We introduce a new method for assessing shape from shading. Observers viewed sinusoidal visual stimuli while stroking a sinusoidally corrugated haptic stimulus and were asked to set the depth amplitude of the haptic stimulus to match the visually perceived surface. Visual stimuli comprised various combinations of LM and AM as described below. After a short training session, this method felt very natural to the observers. However, the method relies on the assumption that observers would perceive sinusoidal luminance patterns as sinusoidal corrugations with the same spatial frequency. This assumption is supported by our previous depth mapping experiments (Schofield et al., 2006), the findings of Pentland (1988), and results from a gauge figure experiment reported elsewhere (Schofield et al., submitted for publication). There is also a danger that the haptic stimulus might alter the visual experience, perhaps acting as a training stimulus (Adams et al., 2004). We think that this is unlikely partly because results from the haptic match task are similar to those obtained with other methods (Schofield et al., 2006 and Schofield et al., submitted for publication). Further, while we do not doubt that haptic stimuli can be used to alter visual perception, we see no reason why such cross-modal influence should be mandatory. Here we made it clear that observers should treat the visual stimulus as the fixed reference and set the haptic stimulus to match it. Other than being a sinusoid of the same frequency as the visual cue, there was no systematic

manipulation of the haptic stimulus to entrain the visual percept.

Visual stimuli

We follow Pentland (1988) and Kingdom (2003) in using sinusoidal shading patterns with no occluding boundaries. Stimuli were not rendered surfaces. Studies of shape perception more typically use images of rendered (or real) objects, irregular shapes, or sections thereof. We used grating stimuli and random noise textures for the following reasons: (1) shading is known to be a relatively weak or secondary cue to shape and can be dominated by other cues including object outlines. Thus, the outlines of rendered objects or blobs can influence both the perceived surface shape (see Knill, 1992) and the strength of the depth percept. (2) We need to simulate textured surfaces in our stimuli, but if these had been rendered, then geometric distortions in the texture would have been an additional cue to shape. Our noise textures were isotropic, providing no cue to shape. (3) With gratings, it is very easy to control the phase relationship between LM and AM and the amount of AM. (4) The use of gratings made it easy for us to cue which component was to be matched to the haptic probe.

Visual stimuli were formed from isotropic binary visual noise with a Michelson (and r.m.s.) contrast of 0.1, onto which we imposed sinusoidal modulations of luminance and amplitude. Noise elements comprised 2×2 screen pixels and subtended 0.06 degrees of arc at the 57-cm viewing distance. We imposed five types of sinusoidal modulation onto these noise textures: (a) LM-only (Figure 2a) comprising luminance modulations added to the noise pattern with no variation in AM, (b) AM-only (Figure 2b) comprising amplitude modulated noise, (c) LM + AM alone (Figure 2c), (d) LM – AM alone (Figure 2d), and (e) plaid stimuli comprising LM + AM on one oblique and LM – AM on the other (Figure 2e). Except when AM modulation depth was zero, we did not test plaids composed of the same cues (i.e., both LM + AM) on both diagonals. In the case of plaids, either the LM + AM or LM – AM component could be designated as the test cue making a total of 6 test conditions in all (but not all conditions were tested in every experiment). Test cues were presented in one of two orientations; left oblique or right oblique ($\pm 45^\circ$). The wavelength of the modulations was 25 mm (spatial frequency = 0.4 c/deg). The contrast of the LM signals and the modulation depth of the AM signals varied between experiments and conditions. Stimuli were presented in a modified ReachIN haptic workstation (Reachin AB, Sweden) depicted in Figure 3. Visual stimuli were presented on a 17" Sony Trinitron CPD G200 CRT monitor (Sony, Japan) mounted at an angle of 45° above a horizontal half-silvered mirror. Observers looked into the mirror at a downward angle and thus perceived the visual stimulus to be beneath the mirror and approximately

perpendicular to their line of sight. A hood prevented the observer from viewing the monitor directly. Observers were asked not to tilt their heads to one side but, except for the need to sit close to the workstation and the limitations imposed by the hood, viewing position was not physically constrained. Stimuli were viewed in the dark such that observers could not see their own hand beneath the mirror. Viewing was binocular and so the visual stimulus provided stereoscopic cues to flatness. However, a robust percept of shape from shading can be derived from such stimuli (Schofield et al., 2006).

Stimuli were calibrated against the monitor's gamma characteristic using lookup tables in a BITS++ attenuation device (CRS, UK), which also served to enhance the available gray level resolution to the equivalent of 14 bits. Values in the lookup tables were determined by fitting a four-parameter monitor model to luminance readings recorded with a CRS ColourCal photometer. Problems in presenting AM stimuli associated with the adjacent pixel non-linearity (Klein, Hu, & Carney, 1996) were avoided by using a high bandwidth monitor and noise samples with relatively low contrast but relatively large element size. However, the noise elements were unlikely to be large enough to produce a noticeable clumping artifact (Smith & Ledgeway, 1997; see Schofield & Georgeson, 1999 for a full discussion of these issues).

Haptic stimuli

Haptic stimuli were presented via a Phantom-Desktop (SensAble Technologies, MA, USA) force feedback device located beneath the mirror and consisted of a virtual surface collocated with the visual stimulus. Haptic surfaces had sinusoidal undulations in the direction of the visual test cue. The spatial frequency of the undulations matched that of the visual stimuli. Observers held the Phantom's stylus like a pen with their dominant hand and stroked the surface. The Phantom provided physical resistance whenever the observer tried to move the stylus tip through the virtual surface. Three markers were added to the visual stimulus: one at the center and two at opposite corners of the stimulus, so that the alignment of the three markers indicated the direction in which observers should stroke the haptic surface in order to feel the undulations. We verified that distances specified in the haptic stimuli were faithfully reproduced by the Phantom. Visual and haptic stimuli were generated on the same PC.

Visual cursor

We ensured that the location, orientation, and spatial frequency of the haptic stimuli matched the visual stimuli well. However, we also conducted a pilot experiment to verify that observers could reliably match the position of

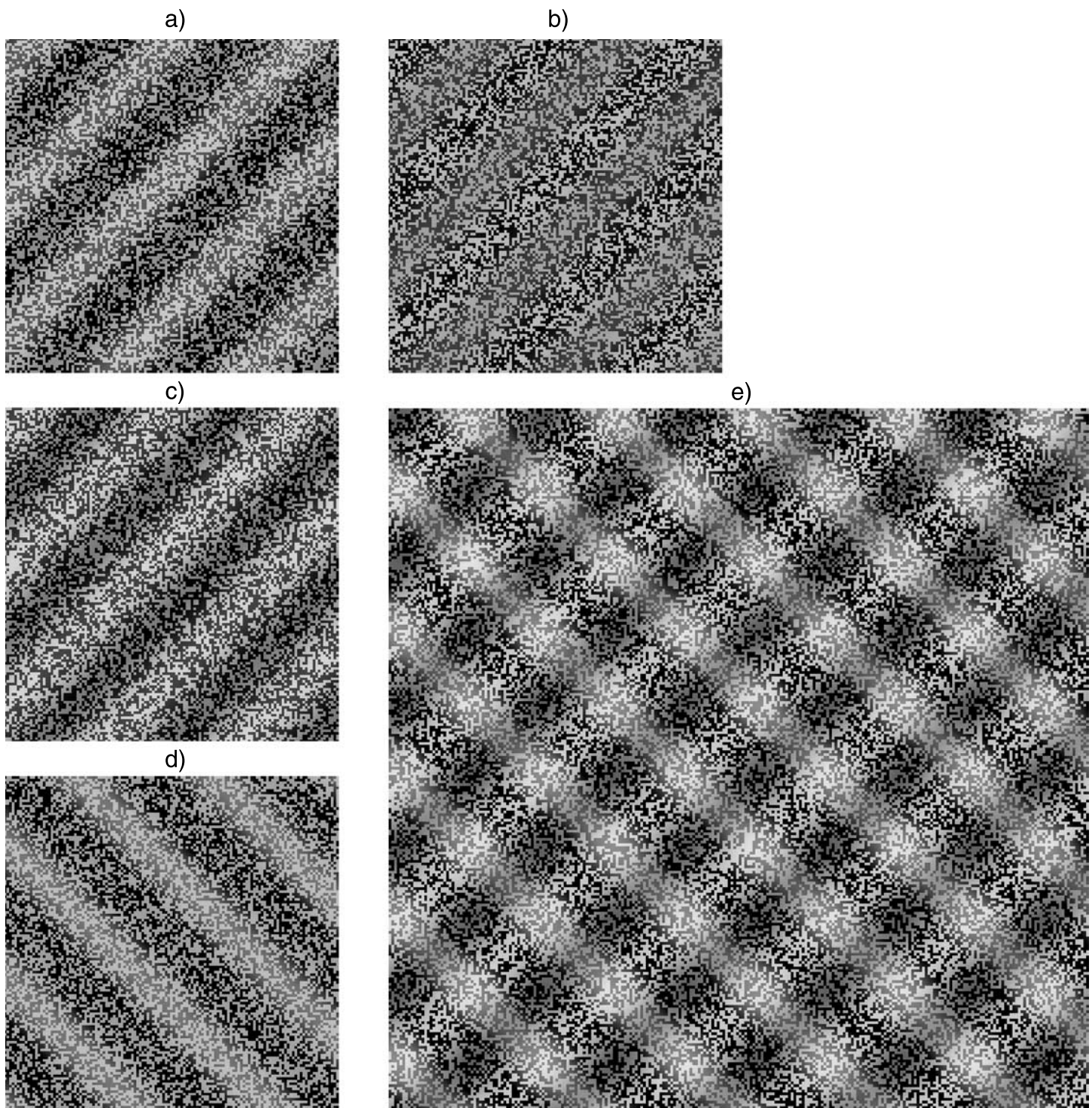


Figure 2. Extracts from sample stimuli. (a) LM-only, formed by arithmetically adding a luminance grating to spatial binary noise. (b) AM-only, formed by modulating the amplitude (standard deviation) of the noise. (c) LM + AM only, formed by combining the cues of (a) and (b) in-phase, equivalent to multiplicative shading. (d) LM – AM only, formed by combining the cues of (a) and (b) in anti-phase. (e) LM + AM and LM – AM in a plaid configuration; here LM + AM is on the right oblique. Note that noise contrast has been increased from 0.1 to 0.3 to aid presentation.

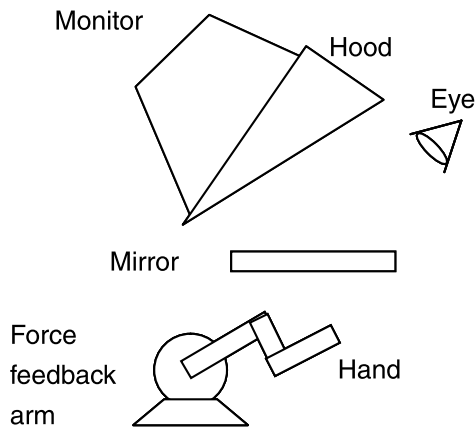


Figure 3. Sketch of the ReachIN workstation with additional hood; support structure not shown.

the haptic undulations to visual features. In this experiment, the visual stimuli consisted of a horizontal luminance grating and observers were asked to adjust the position of the peaks in the haptic stimuli to match the position of the luminance peaks. In the absence of any visual feedback as to the location of the stylus tip, observers were unable to match the positions on the visual and haptic stimuli with any reliability (standard deviation of match positions = 0.288 wavelengths). However, reliable position matches were possible on the introduction of a visual cursor that tracked the tip of the stylus (standard deviation of match positions = 0.041 wavelengths). A cursor was therefore included in all the experiments. We conclude that co-registration of the haptic and visual stimuli is not sufficient to allow reliable position matching in the absence of visual feedback. Further, although we have not tested this directly, we suspect that precise co-registration is not necessary if feedback is provided. We note, for example, that computer users can reliably place a pointer at a specified screen location despite a gross mismatch between the physical positions of the pointer and “mouse”.

Position of haptic stimulus

Prior to the main experiments, we asked observers to adjust the position of a haptic stimulus to match that of the perceived corrugations in the visual stimuli. These settings were then used to determine the precise relative position of the visual and haptic stimuli in the main experiments such that haptic peaks were always aligned with perceived surface peaks. Typically, perceived surface peaks (and hence haptic peaks) are offset from the luminance peaks (see Schofield et al., 2006). Details of how these measurements were performed can be found in Experiment 1 of Schofield et al. (submitted for publication). We measured

offsets (the difference between the position of the luminance peaks and the haptic peaks) for LM + AM, LM – AM, LM-only, and AM-only in the single oblique condition and LM + AM when presented as part of a plaid stimulus. AM-only offsets were measured relative to peaks in the amplitude signal. We then applied the appropriate offsets between our visual and haptic stimuli on a per condition and observer basis. However, we could not measure offsets for LM – AM stimuli in the plaid configuration as observers saw this cue as flat and therefore could not identify any surface peaks against which to make a match. Instead, we used the LM + AM offsets when testing LM – AM in a plaid.

Main adjustment task

In the experiments reported below observers adjusted the amplitude of the haptic surface up or down by pressing one of two keys on a numeric keypad. A third key toggled the step size for adjustments between 2 and 0.5 mm (half-height amplitude). Observers heard a long tone for each 2-mm adjustment and a short tone for each 0.5-mm adjustment. Observers could not drive the amplitude of the haptic surface below zero and received an auditory warning of any attempt to do so. Estimates of PDA were calculated as the median of at least 5 measurements.

Observers

Five observers took part in the experiments. With the exception of author PR, observers were naive to the purposes of the experiment and were paid for their time. Author PS was a naive observer at the time of the study. Author AJS contributed some additional data to Experiment 2. All observers had normal or corrected-to-normal vision and no physical disability or injury. Observers held the stylus in their dominant hand: JG is left-handed; the remaining observers are right-handed.

Experiment 1: Perceived depth amplitude versus overall signal strength

In this experiment, we considered the effect of overall signal strength on the PDA of visual stimuli. We also varied the relative phase of the LM and AM cues at the test orientation, and we compared two component (plaids) with single component stimuli (gratings). The LM contrast and AM modulation depth were equal in any given stimulus, consistent with multiplicative shading for in-phase pairings.

Methods

Signal strength, governing both LM component contrast and AM component modulation depth, was varied in multiples (0.1, 0.4, 0.8, 1.6, 3.2, and 4.0) of each observer's AM detection threshold as measured in separate sessions using a staircase method (Levitt, 1971) and a two-interval forced-choice design. In this pilot experiment, stimuli consisted of AM gratings presented alone. Note that our AM gratings are identical to the CM gratings often used to study second-order vision. The mean AM threshold across observers was 0.086, and this is consistent with the literature on second-order vision (Schofield & Georgeson, 1999). Stimuli consisted of plaids comprising LM + AM on one diagonal and LM – AM on the other (Figure 2e), LM + AM presented alone (Figure 2c), or LM – AM presented alone (Figure 2d). Because they contain two orientation components, plaids had greater overall contrast and modulation depth than single component stimuli. Many of the stimuli in this experiment contained sub-threshold levels of AM, but their LM components were likely to be supra-threshold because thresholds for LM in visual noise are about an order of magnitude lower than AM (CM) thresholds (Schofield & Georgeson, 1999).

Results and discussion

Figure 4 shows the results of Experiment 1 averaged over the five observers. Mean PDA was low for weak stimuli regardless of their composition and remained low for LM – AM at all signal levels when this cue was part of a plaid (squares in Figure 4b). However, when LM – AM was presented alone (squares in Figure 4a), PDA increased with signal strength. PDA also increased with signal strength for LM + AM whether presented alone (circles in Figure 4a) or in a plaid (circles in Figure 4b). Although the variances were high, we note that PDA rises to a level significantly above zero for all cues except LM – AM presented in a plaid (error bars in Figure 4 represent 95% confidence intervals). PDAs for strong LM + AM gratings tend to be greater than those for LM + AM presented as part of a plaid despite the fact that overall luminance contrast was higher for the latter stimulus. This trend can also be seen in weaker stimuli where components of a plaid produced lower PDAs than single grating stimuli. For single obliques, strong LM + AM gratings produced somewhat greater PDAs than LM – AM gratings, but only when AM was above threshold. Perceived depth for LM + AM was also greater than for LM – AM in plaid stimuli and this seemed to hold down to signal levels where AM was below threshold (between 0.4 and $1 \times$ AM threshold). Lines in Figure 4 show predictions of the model described later.

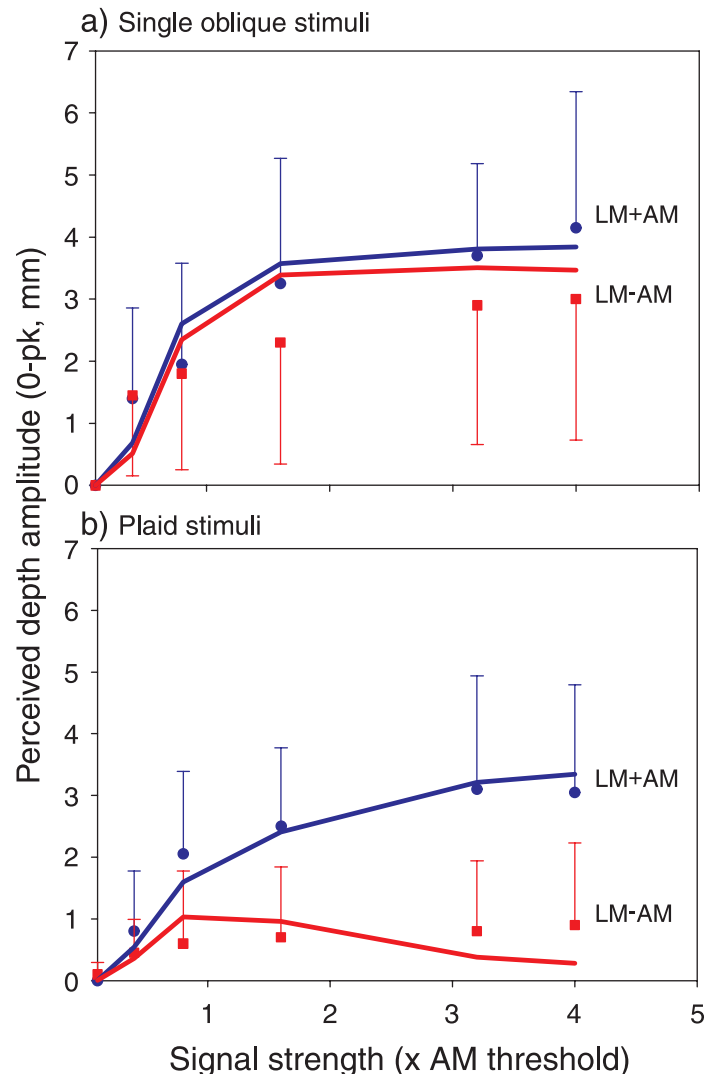


Figure 4. Experiment 1. Perceived depth amplitude as a function of overall signal strength: (a) single oblique stimuli, (b) plaid stimuli. Blue circles show the perceived amplitude of LM + AM mixes; red squares show LM – AM mixes. X-axis shows signal strength as a multiple of AM threshold. Error bars represent 95% confidence intervals and are drawn single-sided to aid interpretation. Lines show predictions of the “shading-channel” model; blue and red for LM + AM and LM – AM, respectively (see description of model for details).

Noting that the plots of Figure 4 are approximately linear against log signal strength, we estimated (with linear regression) the slope of the relationship between log signal strength and PDA separately for each participant and each stimulus type. Figure 5 plots the mean slope for each stimulus type and their associated 95% confidence intervals. Slopes for LM – AM were not significantly different from zero regardless of the configuration used (one-sample, one-way t -test: LM – AM only, $t = 2.55$, $df = 4$, $p > 0.05$; LM – AM in plaid, $t = 1.16$, $df = 4$,

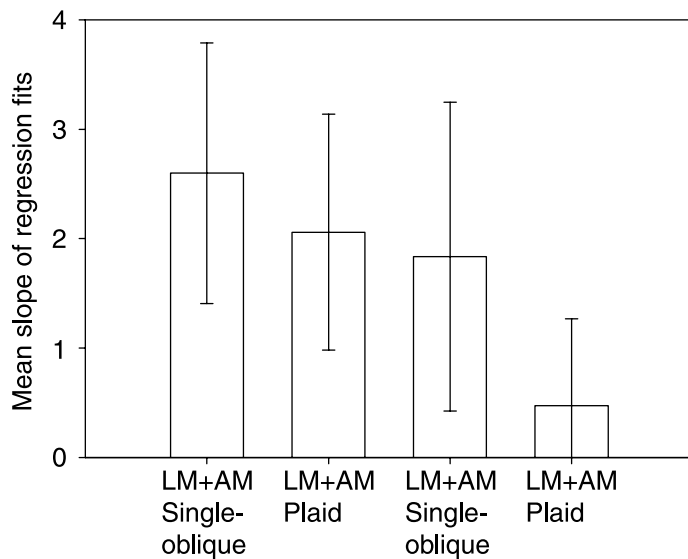


Figure 5. Mean slopes for regression fits to individual data from Experiment 1 for each of the four test conditions. Error bars represent 95% confidence intervals.

$p > 0.05$). LM + AM stimuli produced significant slopes (LM + AM only, $t = 4.26$, $df = 4$, $p < 0.05$; LM + AM in plaid, $t = 3.74$, $df = 4$, $p < 0.05$). A repeated measures ANOVA (with Greenhouse–Geisser correction) showed that there were significant differences between the mean slopes across the four conditions ($F = 8.57$, $df = 1.6, 6.38$, $p < 0.05$). Bonferroni corrected post-hoc paired comparisons showed that slopes for LM – AM in a plaid were significantly lower than those for the LM + AM conditions (LM – AM in a plaid vs. LM + AM in plaid, $t = 6.2$, $df = 4$, $p < 0.05$; LM – AM in plaid vs. LM + AM only, $t = 5.32$, $df = 4$, $p < 0.05$). The difference in slopes between LM – AM in a plaid and this cue presented alone was significant prior to Bonferroni correction but not after ($t = 3.1$). None of the other pairings were significantly different suggesting that LM – AM presented alone produces behavior similar to that of LM + AM.

Taken together, these results show that LM – AM is seen as a shape-from-shading cue when presented on its own. PDAs for this cue are about the same as those for LM + AM in a plaid but below those for LM + AM presented alone. When LM – AM is presented as part of a plaid, however, it is seen as quite flat. Inspecting individual data revealed that most observers saw this condition as almost completely flat even at high signal strength and that the slope observed in Figure 4 is largely due to one observer who saw this stimulus as conveying some depth. By contrast LM – AM alone was seen as quite corrugated by all but one observer and the two LM + AM conditions were seen as corrugated by all observers. PDAs naturally converge toward zero as signal strength is reduced. PDAs for single components converge at about the point where the AM signal falls below threshold. PDAs for the two

members of a plaid converge at a point below the measured AM detection threshold; this could be due to probability summation that may serve to increase the visibility of AM in plaid stimuli above that of single orientation components. It is clear that LM is the dominant cue for depth perception in shaded textures but that its relationship with AM and the overall configuration of the stimulus is also important. We now investigate the specific role of AM in more detail.

Experiments 2 and 3: Effect of AM modulation depth on perceived depth amplitude

In these experiments, we varied AM strength while keeping LM contrast constant. We thus assessed the ability of AM to influence perceived depth.

Methods

Visual stimuli were diagonally oriented gratings and plaids with a fixed LM contrast of 0.2 and several AM modulation depths (0, 0.1, 0.2, 0.4). Again, we varied the phase relationship between LM and AM. In Experiment 2, we tested plaid stimuli only. Experiment 3 tested single component stimuli including AM-only gratings (see Figure 2b). When we devised Experiment 2, we considered the LM + AM and LM – AM components to be distinctly different stimulus types. We therefore did not test the case where the AM signal was zero (i.e., an LM-only vs. LM-only plaid). We later realized that these cues form a continuum running from strong negative AM to strong positive AM, with LM-only (AM modulation depth = 0) representing the midpoint on this continuum. We thus added the AM = 0 case to the test battery for Experiment 3 and tested an additional observer in Experiment 2 including the AM = 0 case.

Results

Figure 6 shows PDA as a function of AM modulation depth. Blue squares show the results for plaid stimuli (Experiment 2); red circles and green triangles show the single component results (Experiment 3). There was no effect of test orientation (left or right oblique) so we averaged across this condition.

Experiment 2: Plaids. For plaid stimuli, PDA increased with signed modulation depth such that stimuli were seen as increasingly flat for negative modulation depths (LM – AM) and increasingly corrugated for positive modulation depths (LM + AM). There was a pronounced increase in

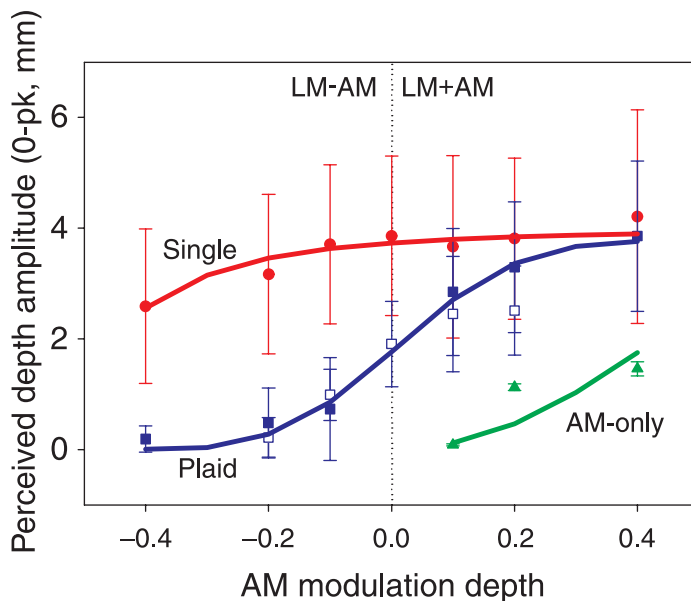


Figure 6. Experiments 2 and 3. Perceived depth amplitude as a function of AM modulation depth and sign. X-axis shows AM modulation depth; negative values mean that the AM cue was in anti-phase with the LM cue (LM – AM). Green triangles, AM-alone. Red circles, single oblique LM and AM signals. Blue squares, LM and AM presented as a plaid. Note that when AM was in anti-phase with LM on the test oblique (negative values) the non-test oblique had an in-phase mix with an equally strong AM cue (and vice versa). Open squares, results for observer AJS for plaid stimuli including the case where AM modulation depth was zero—i.e., LM-only on both obliques. Lines represent model fits for the “shading-channel” model. Except for the open squares, data points are the means of 5 observers and error bars represent 95% confidence intervals. For AJS (open blue squares), error bars represent the standard deviation of individual depth estimates.

PDA around AM = 0. A repeated measures ANOVA (with Greenhouse–Geisser correction) showed that the overall change in PDA was significant ($F = 42.468$, $df = 1.493$, 7.464 , $p < 0.01$) and Bonferroni corrected post-hoc paired-samples t -tests showed that anti-phase stimuli (LM – AM) produced significantly lower PDAs than in-phase stimuli (LM + AM). Results from the one observer (AJS) tested with AM = 0 (open square symbols in Figure 6) suggest that PDAs for LM-only plaids fall nicely on the continuum from LM – AM to LM + AM.

Experiment 3: Single components. There was much less variation in PDA with AM modulation depth in the single component stimuli. Here we found only a gradual increase in PDA with AM modulation depth and hardly any increase at all among LM + AM stimuli. The overall trend was not significant (Greenhouse–Geisser corrected ANOVA, $F = 4.013$, $df = 1.583$, 7.916 , $p = 0.069$). There were no significant differences between any of the levels tested for the single component stimuli (based on Bonferroni corrected paired t -tests). Paired-sample t -tests between

AM-only stimuli (triangles) and single component mixed stimuli (filled circles) with equivalent levels of AM suggest that the AM-only stimuli were seen as significantly flatter than LM/AM mixes regardless of the phase relationship in the mix (based on paired samples t -tests corrected using Horn’s multistage Bonferroni method). Similarly, PDAs for LM + AM in a plaid were significantly greater than their AM-only counterparts. In contrast, PDAs for LM – AM stimuli in a plaid were not significantly greater than those for AM-only. Finally, we note that LM – AM stimuli in a plaid are seen as significantly less corrugated than the equivalent single component stimuli but that the differences between LM + AM in plaid and single component configurations are not significant.

Discussion

Taken together, the results of Experiments 2 and 3 show that LM – AM was seen as flat when shown in a plaid with LM + AM but was seen as corrugated otherwise. PDAs for LM – AM and LM + AM stimuli tend to be similar at low AM modulation depths. This result is to be expected because these cues become identical as AM modulation depth approaches zero. However, while PDAs for the LM + AM and LM – AM gratings (at a single orientation) were almost identical for AM modulation depth in the range -0.1 to $+0.1$, those for the plaid stimuli varied significantly over this range.

We note that LM + AM stimuli also appear a little less corrugated in a plaid than they do as single components, and although these differences are not significant, some discussion is merited. We note particularly that plaid stimuli with little or no AM signal have a doubly corrugated or “egg box” appearance. The PDA of such stimuli in a given direction is likely to vary with position along the orthogonal axis and this may reduce the average PDA. Single component stimuli appear as single corrugations whose PDA does not vary with position in the direction orthogonal to the modulations. It is possible that the “egg box” effect accounts for the observed difference between plaid and single component stimuli in the LM + AM case. However, there is an alternative explanation based on mutual suppression between obliques and we discuss this next.

Model

We constructed a model to explain our data. The purpose of the model is to demonstrate that the observed effects can be predicted by bottom-up mechanisms involving biologically plausible second-order processes. The model (shown in Figure 7) is intended to represent one spatial frequency tuned “shading channel” within a

multichannel scheme. It is based on the processing scheme for envelope sensitive neurons proposed by Zhou and Baker (1996) and the filter–rectify–filter (FRF) model of second-order vision (Wilson et al., 1992) and has similarities with the three-stage model proposed by Henning et al. (1975). The first-stage comprises a bank of linear filters tuned to multiple spatial frequencies and orientations. These filters share a gain control mechanism. The second-stage consists of a bank of rectifiers followed by linear filtering (the RF of the FRF scheme) taking their

input from high-frequency first-stage filters. This stage extracts the AM cue and is not directly subject to gain control. At the third stage, we take a weighted sum of the outputs of like-oriented linear and FRF channels, producing behavior like that of Zhou and Baker’s (1996) envelope neurons. This final stage is subject to gain control. We envisage that separate signals for first- and second-order cues are available at the points marked LM and AM, respectively, and that these signals support the detection of these cues.

a) Shading channel model

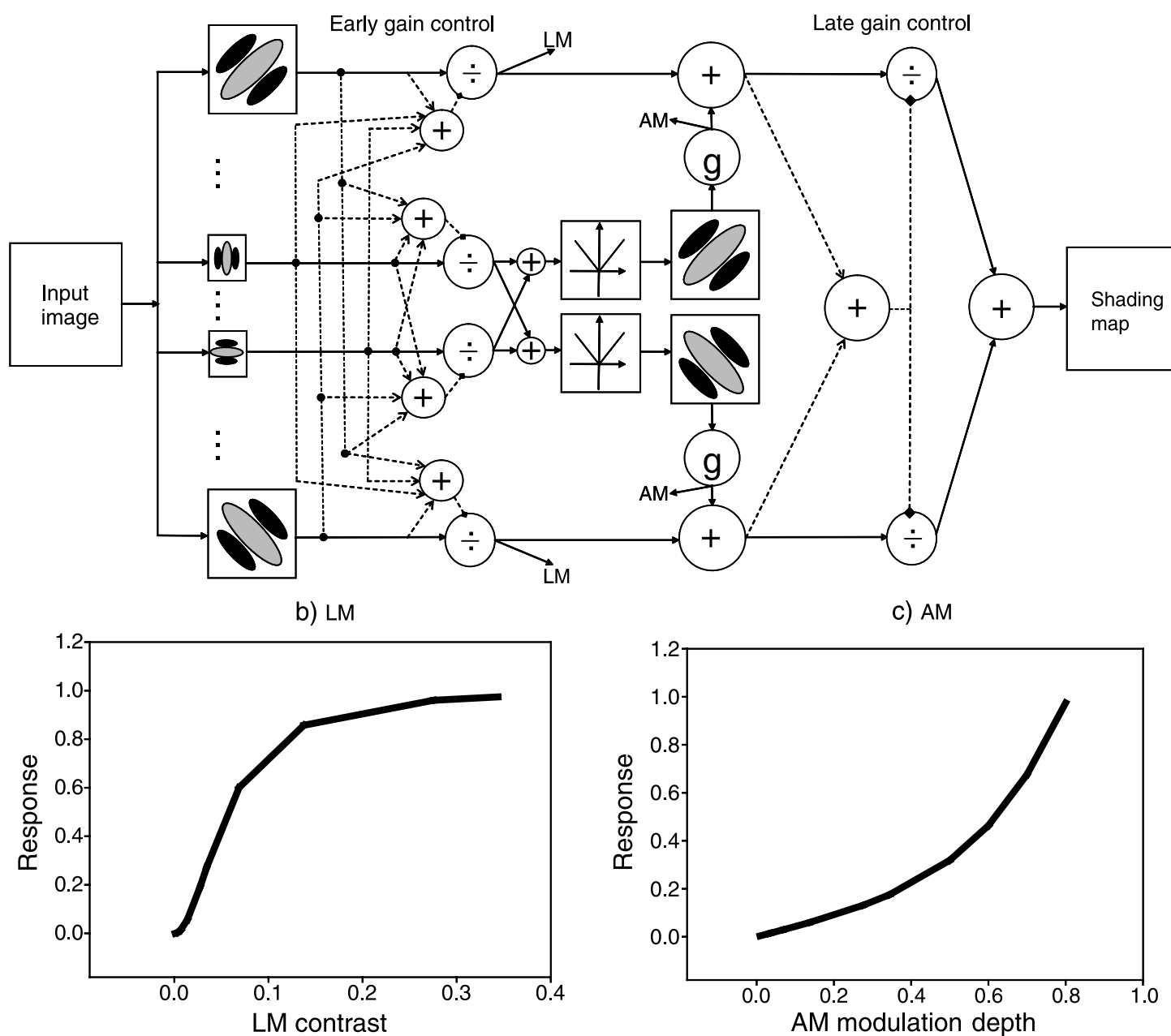


Figure 7. (a) Schematic diagram of the “shading-channel” model; see text for description. (b) Input–output response for first-order (LM) sub-channel. (c) Input–output response for second-order (AM) sub-channel.

We now address the biological plausibility of the proposed scheme, considering the following components: Linear first-stage filtering with gain control, rectification, independent outputs, weighted summation between sub-mechanisms, final gain control.

Linear first-stage filtering with gain control. Linear spatial frequency channels were first proposed by Campbell and Robson (1968) and are now accepted as the basis for early visual processing. More recent evidence suggests that while such mechanisms are approximately linear they have a non-linear transfer function, which is expansive for low-input values and compressive for larger inputs (Legge & Foley, 1980). This compression is now thought to be due a contrast gain control mechanism that pools input from many channels and across space (Foley, 1994) and has been proposed as an explanation for the compressive behavior of simple cells in primary visual cortex (Albrecht & Geisler, 1991; Heeger, 1992, 1993). However, the pooling process is far from uniform: masking (and indeed facilitation) depends on the relative, frequency, orientation, and spatial locations of the test and mask stimuli giving rise to complex patterns of behavior (Foley, 1994; Meese, 2004; Meese, Challinor, Summers, & Baker, 2009). Specifically, a given channel receives most masking from channels tuned to similar frequencies and orientations although the orientation tuning of masking is very broad (Foley, 1994). Thus we apply cross-channel gain control to our first-stage filters. Each filter has its own gain control pool with equal weight being given to all orientations in the pool but less weight given to frequencies distant from the preferred frequency of the filter in question. Because of the simple nature of our stimuli, we only modeled first-stage filters tuned to the image equivalent of 0.4 and 16 c/deg and $\pm 45^\circ$. First-stage responses are given by

$$R_i = \frac{C_i^p}{s_1^q + (\sum C_a^q + w \sum C_b^q)}, \quad (1)$$

where C_i is the pre-gain control response of the i th filter, C_a is the response of all filters with the same preferred frequency as the i th filter, C_b is the response of filters with preferred frequency different to that of the i th filter, w is the weight applied to off-frequency filters in the gain pool, p and q represent exponents on the forward and gain control terms, respectively, and s_1 is the semi-saturation constant. In line with other similar models, we set p and q to 2.0 (e.g., Meese et al., 2009); s_1 and w were free parameters. Application of this gain control mechanism results in a first-stage transfer function that initially accelerates and then saturates (Figure 7b) broadly consistent with both psychophysical “dipper” experiments (Legge & Foley, 1980) and physiology (Albrecht & Geisler, 1991; Ledgeway et al., 2005).

Rectification. Nonlinear FRF channels similar to our rectification stage (where the first filters are found in the

first-stage of our model) have been proposed to explain the detection of contrast modulations (our AM; Wilson et al., 1992) and various texture segmentation phenomena (Graham & Sutter, 2000; Landy & Bergen, 1991). Although the FRF mechanism is now widely accepted as the means by which second-order cues are detected, debates continue about the wiring between first- and second-stage filters and the shape of the rectifying non-linearity. Within the context of our limited model and following Sutter et al. (1995) and Dakin and Mareschal (2000), we connect our second-stage filters to only the high-frequency first-stage filters according to

$$S_i = f_i \left(\left| \sum f_{hf}(I) \right|^\gamma \right), \quad (2)$$

where f_i is a second-stage filter with the same spatial frequency and orientation as the i th first-stage filter (but only low-frequency second-stage filters are implemented), f_{hf} are the high-frequency first-stage filters, $|\cdot|$ represents rectification, and γ governs the shape of the rectifier. We sum first-stage filter responses across orientation and after application of the gain control (Equation 1). Graham and Sutter (2000) suggest that γ should be about 3.5; however, this is based on psychophysical results that depend on the operation of the whole mechanism. Ledgeway et al. (2005) note that cells responsive to second-order cues demonstrate an accelerating transfer function and do not saturate. We used a linear rectifier ($\gamma = 1$) but tested the transfer function of our model in respect of AM signals and found it to accelerate as the cube of input strength with no saturation (see Figure 7c). This lack of saturation can explain why CM stimuli do not mask themselves (Schofield & Georgeson, 1999). We believe that the early gain control mechanism and linear rectifier serve to produce the cubic transfer function in the FRF network. It should be noted that cell responses to second-order stimuli are likely to saturate at some point if both the carrier and modulation signals are high enough. Due to the simplicity of our stimuli, we only implemented second-stage filters at 0.4 c/deg and $\pm 45^\circ$.

Independent outputs. It should be noted at this point that second-order detection could, in principle, be achieved by a single stage of non-linear filtering but that this would prevent the independent processing of first- and second-order cues. In the Introduction section, we describe a considerable body of evidence to suggest that the cues are detected independently. We will not rehearse that argument here, but it is our basis for proposing a separate second-order mechanism. However, the finding that cells responsive to first- and second-order cues have different preferred frequencies for the two cues strongly suggests the existence of separate sub-mechanisms (Mareschal & Baker, 1998). Given that we will shortly propose the integration of first- and second-order cues, the evidence for independent detection also leads us to propose that the

outputs of the mechanisms are separately available. If the first-order signals were extracted prior to the summation stage, this would explain why CM does not mask LM as second-order signals have no direct access to the first-stage gain control mechanism. This “separate signals” hypothesis is somewhat at odds with the physiological evidence. Although cells responsive to only first-order and both first- and second-order cues have been found, there is little or no physiological evidence for the existence of cells responsive to second-order signals only, but (as discussed in the [Introduction](#) section) this may be due to sampling biases.

Weighted summation between sub-mechanisms. For motion at least, there is compelling physiological evidence for cells that linearly sum first- and second-order information (Hutchinson et al., 2007; Ledgeway et al., 2005; Mareschal & Baker, 1998; Zhou & Baker, 1996). Hutchinson et al. (2007) explicitly tested for interactions between the two cues and found that cell responses were dependent on the phase relationship between the two cues, strongest for in-phase stimuli and considerably weaker for anti-phase stimuli. They used stimuli that produced equally strong responses when presented alone. Our AM cues were weaker (compared to threshold) than our LM cues so we should expect a weaker interaction. We note that our second-order mechanism is inherently insensitive. That is, by the time our relatively weak carrier has been filtered and the envelope extracted the response to the AM cue is very low—about 1/30th of the equivalent LM response. In order to provide some differentiation between LM + AM and LM – AM and to give the model more flexibility, we introduced a gain term (or weight) on the output of the second-stage filters. However, it is the overall sensitivity to AM relative to that for LM that matters. The output of each “shading channel” after the sum is given simply by

$$D_i = R_i + gS_i, \quad (3)$$

where g is the gain term for the second-order mechanisms. Only low-frequency first-stage filters and their corresponding second-stage filters are included at this stage.

Final gain control. The final gain control process is the most speculative part of the model, but its existence and position are fundamental to the successful operation of the model. It is these mechanisms that turn the relatively poor differentiation between LM + AM and LM – AM for single gratings into the relatively strong differences found for plaids. Its position, after summation, is key to this. If it acted before LM and AM were summed, then there would be no difference in signals to drive the “winner take all” behavior that the model needs to describe the plaid data. External justification for late gain control is provided by late interactions between the cues as noted in the [Introduction](#) section, most notably the transfer of the contrast-reduction after-effect and the tilt after-effect

(Georgeson & Schofield, 2002). Several authors have linked simultaneous masking with sequential adaptation (Foley & Chen, 1997; Meese & Holmes, 2002). So evidence for a crossover of adaptation could be taken as evidence of gain control. However, based on the evidence for independent detection, this would have to take place after an initial detection stage. The final response of the model is given by

$$U_i = K \cdot \frac{D_i^p}{s_2^q + \sum D_j^q}, \quad (4)$$

where D_i is the output from the i th “shading channel”, D_j is j th channel’s input to the gain control pool, s_2 is the semi-saturation constant, and exponents p and q were again set to 2.0. K is a final scaling factor used to equate the range of model outputs to the human data but with no influence on the shape of the model output curves.

Implementation

For the purpose of fitting the data, the model was implemented analytically. That is, we calculated ideal filter responses based on the stimulus parameters: we did not actually filter images. We subsequently implemented a “filter-based” version of the model that was capable of processing natural images (see later text). A final consideration is how to relate model output to measured PDAs. If we assume that the final output of the model described above is fed into a shape-from-shading module, then the model output up to that point can be thought of as a conditioned shading signal. That is, LM is assumed to be a shading signal, but its efficacy is modulated both by the presence of AM with the same orientation and the context provided from other orientations. For the purposes of model fitting, we assume a linear relationship between the input and output of the hypothesized shape-from-shading module (Pentland, 1988) such that the contrast of the input signal at any orientation gives the perceived depth of surface undulations in that direction up to a scale factor; K in [Equation 4](#).

Operation of the model

When an LM/AM mix is presented on only one oblique, the action of the normalization stage is largely irrelevant as there are only two channels, one of which has no output. In this case, AM will have a slight modulatory effect on the shading signal determined by the overall sensitivity of the AM channel. LM – AM will hence be seen as less corrugated than LM + AM, but the difference

Parameter	Value
w	0.23
s_1	0.029
s_2	0.25
g	3.0

Table 1. Model parameters.

will be small. When an LM/AM plaid is presented to the model, the stronger LM + AM signal will dominate the weaker LM – AM signal at the final gain control stage, driving its output down but the mutual inhibition will also limit the LM + AM signal to a value below that which would be obtained for LM + AM alone.

Model fits

The model described above has four free parameters: w , the weight applied to off-frequency maskers in the gain control of Equation 1, the semi-saturation constants s_1 and s_2 , and the second-stage gain term g . Noting that, due to arbitrary scaling, the maximum theoretical output of the model prior to the multiplier K is 1 we simply set $K = 4$ to match the maximum mean PDA. The remaining parameters were fit to the data for Experiments 2 and 3 using the *fminsearch* function in *Matlab* (The Mathworks, MA). Fitted parameter values are shown in Table 1 and the fits are shown as lines in Figure 6. The model fits the data well. A key characteristic of the model is that it allows LM – AM to be seen as relatively strongly modulated in depth when presented alone but flat when presented in a plaid. The model highlights the continuous nature of the relationship between LM and AM. Even in the plaid case adding weak AM does not produce an abrupt change in perceived depth amplitude.

We also used the model to predict the results of Experiment 1. Here PDA was measured as a function of AM threshold. The model has no concept of threshold so we added an extra parameter T , which represents the base AM modulation depth from which model “threshold” multiples were calculated. This parameter was used to fit the model to the data of Experiment 1 but with no further adjustment of the other parameters. Model predictions are shown as lines in Figure 4. The model provides a good fit to the data.

The gain term g is of interest only because it relates to the overall sensitivity of the second-order mechanism. Of more interest is the relative sensitivity of the two mechanisms. We recorded output strengths for LM-only and AM-only gratings at contrast/modulation depth = 0.2. These were 0.93 and 0.09, respectively, making second-order sensitivity 1/10th that of first order, and correctly predicting the ratio found by Schofield and Georgeson (1999) on noise carriers with contrast = 0.1 (as used here).

Processing natural images

It is useful to fit an analytical model to data, as done here. In particular, restricting the complexity of the model reduces the number of free parameters and this is useful for fitting purposes. However, it does not follow that the model will produce meaningful results when applied to real-world images such as that in Figure 1. Even if implemented with filters, the model described above would be useless in such an application because it has only two oriented channels at one spatial frequency. At best, it would produce plaid-like outputs for every image. We therefore implemented a more complete model with multiple orientation and frequency channels (both first and

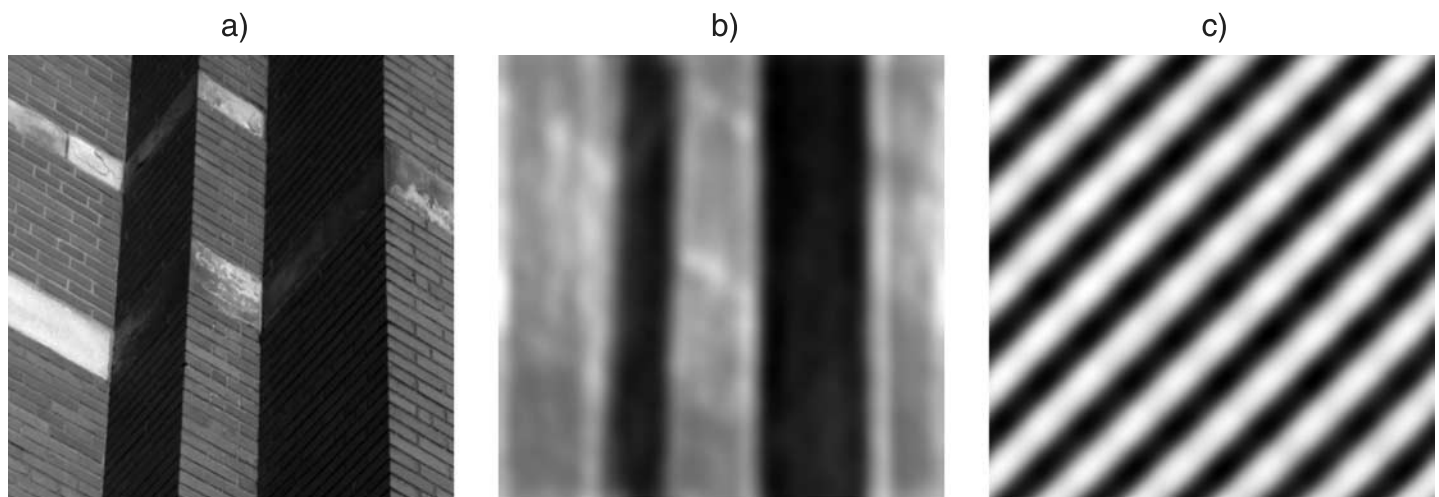


Figure 8. (b) Results of applying the multichannel shading model to an image of a section of wall (a) similar to that shown in Figure 1. (c) Results of applying the model to the plaid stimulus of Figure 2e.

second order) carried through to the final output. We used 3 frequency bands and 16 orientations; 48 channels in all. Apart from having multiple channels, the structure of the model was very similar to that of Figure 7, a key difference being that we dispensed with the early gain control stage and replaced it with a simple sigmoidal transfer function. We did this because we felt unable to model the subtle spatial interactions required of a full-blown gain control mechanism (Meese, 2004). This model captures the spirit of the “shading channels” described above. As might be expected, we find the model to be most effective in cases where $LM + AM$ and $LM - AM$ co-exist in the same scene. Figure 8a shows a sample input image and the resulting model output (Figure 8b). Figure 8c shows the result of processing the stimulus example shown in Figure 2e. In both cases, the model successfully separates shading (or perceived shading) from reflectance changes.

General discussion

The results presented here extend those of Schofield et al. (2006) by introducing a more natural depth matching task, new test conditions, and a computational model. Observers had to set the amplitude of haptic stimuli to match the properties of a visually perceived surface. Perceived depth amplitude increased with overall modulation strength (Experiment 1) for all stimuli containing LM except $LM - AM$ in a plaid. $LM - AM$ in a plaid was perceived as nearly flat across a range of signal strengths but, consistent with our previous findings, $LM - AM$ was seen as modulated in depth when presented alone. Note that, as we found previously, $LM - AM$ alone was seen corrugated, but less so than $LM + AM$ alone. This difference is smaller when measured with the haptic task. Keeping LM contrast constant while varying AM modulation depth (Experiments 2 and 3) allowed us to study the influence of AM on LM cues. Increased AM modulation depth did not greatly affect the PDA of LM when the two were presented in-phase and alone ($LM + AM$, circles to right of Figure 6). Anti-phase AM did reduce the PDA of the associated LM signal ($LM - AM$) but only slightly (circles to left of Figure 6). However, AM had a more marked influence on PDAs in the plaid configuration. Here increasing AM in-phase with LM produced a marked but saturating increase in PDA while anti-phase AM reduced PDA (squares in Figure 6). We stress that in these plaids $LM + AM$ and $LM - AM$ were seen together such that as AM was stronger in the $LM - AM$ component it also became stronger in the associated $LM + AM$ component and vice versa. The pattern of results observed would not necessarily hold if say the $LM - AM$ member of a plaid were fixed while the AM part of the $LM + AM$ cue was allowed to vary, although the model

would allow us to make predictions for this case. Amplitude modulations presented alone produced only a weak depth percept, but perceived depth amplitude did increase a little with AM modulation depth (triangles in Figure 6).

It is tempting to suggest that higher level cognitive processes must be at work in the interpretation of stimuli when, as here, the stimulus context is relevant to the interpretation of a particular cue: here $LM - AM$ was seen as flat only when present in a plaid with $LM + AM$. However, we have successfully modeled the data with an architecture that requires no top-down control and that could well be implemented in early visual areas such as V1 or V2 with the possible aid of V3a to process AM. The model combines LM and AM responses in an additive fashion within a given orientation/frequency band and then combines those responses across different orientations with gain control governing the balance between them. The resultant shading signal tends to be stronger when AM is presented in-phase with LM but is very weak when the anti-phase combination occurs in a plaid alongside an $LM + AM$ component. A multichannel version of the model was tested on natural images and worked well in conditions where $LM + AM$ and $LM - AM$ cues co-existed.

The model presents some challenges to our previous work on cue independence. We have previously argued quite strongly that LM and CM (in our current terminology AM) are detected independently (Georgeson & Schofield, 2002; Schofield & Georgeson, 1999), but our current model suggests relatively early summation and a lack of independence. We suggest that LM and AM are indeed detected independently and are thus (for example) discriminable at threshold but that they are summed for the purpose of disambiguating the role of the luminance cue at some stage beyond simple detection. Such a configuration would allow the two cues to interact in various ways both with each other and with other cues such as disparity and texture. Our proposal here is that the two cues are summed to aid the computation of shape from shading, and perhaps in other situations too, but we do not suppose that this summation is either ubiquitous or mandatory.

The model makes some clear predictions about interaction of LM and AM in shape from shading. If such processing is based on the early channel-like mechanisms with gain control, then we should expect interactions along the lines of those described above for a variety of interleaved stimuli. For example, we might expect it to be possible for $LM - AM$ to be seen as corrugated if presented alone in one part of a stimulus but flat in some other part of the same stimulus if it overlapped with $LM + AM$ in that region. We might expect some degree of spatial overlap to be necessary between $LM + AM$ and $LM - AM$ for the latter cue to be seen as flat but that the overlap need not be complete. We predict that plaids should behave as described above when their components are not orthogonal, but only if there is sufficient separation between the orientations that they fall into different orientation channels. We similarly expect $LM + AM$ and

LM – AM to dissociate if handled by different spatial frequency channels. Finally, adding an additional LM + AM component at another orientation should further suppress PDA for an LM – AM cue. We have yet to test these interesting predictions.

We presume that if AM is used to disambiguate LM in the way described above then this interaction should be driven by ecologically valid constraints. That is, LM – AM should be a reliable cue to a material change but only in the context of LM + AM cues. We have previously noted that visual texture can arise from a variety of sources and that the yoking of LM and AM (LM + AM) is only guaranteed for shaded albedo textures (Schofield et al., 2006). LM – AM can arise when a rough corrugated surface is shaded, although such an outcome is not guaranteed. However, it is highly unlikely that a doubly corrugated locally rough surface could give rise to LM – AM on one oblique and LM + AM on the other. We therefore conclude that the co-presentation of LM + AM and LM – AM confirms the former cue as shading of an albedo texture and the latter cue as due to reflectance changes within that texture.

Conclusion

In conclusion, second-order modulations (specifically modulations of local luminance amplitude/contrast) can affect the perception of shape from shading from luminance-modulated textures. In some cases, this influence is profound with the phase relationship between LM and AM determining the perceptual role of the luminance cue, flipping it from being used as a shading cue to a cue for material change. Given that luminance changes are ambiguous about their environmental causes, second-order vision may play an important role in the interpretation of luminance variations. Perhaps the need to compare these two cues is one reason why human vision is configured to detect AM (CM) cues separately from LM in the first place. In general, when AM varies in anti-phase with LM (LM – AM) surfaces are seen as flatter than when the two cues co-vary in phase (LM + AM). The flattening observed in LM – AM stimuli is most pronounced when it is presented in a plaid configuration with an LM + AM cue. However, this context effect does not require a top-down interpretation because it was possible to model key features of our data using bottom-up channel-like mechanisms.

Acknowledgments

This work was supported by EPSRC Grants GR/S07254/01 and EP/F026269/1 to AJS and GR/S07261/01

to MAG. We thank the two anonymous reviewers for their helpful comments.

Commercial relationships: none.

Corresponding author: Andrew J. Schofield.

Email: a.j.schofield@bham.ac.uk.

Address: School of Psychology, University of Birmingham, Birmingham, B15 2TT, UK.

References

- Adams, W. J., Graf, E. W., & Ernst, M. O. (2004). Experience can change the “light-from-above” prior. *Nature Neuroscience*, 7, 1057–1058. [PubMed]
- Albrecht, D. G., & Geisler, W. S. (1991). Motion selectivity and the contrast-response function of simple cells in visual cortex. *Visual Neuroscience*, 7, 531–546. [PubMed]
- Albright, T. D. (1992). Form-cue invariant motion processing in primate visual cortex. *Science*, 255, 1141–1143. [PubMed]
- Allard, R., & Faubert, J. (2007). Double dissociation between first- and second-order processing. *Vision Research*, 47, 1129–1141. [PubMed]
- Ashida, H., Lingnau, A., Wall, M. B., & Smith, A. T. (2007). fMRI adaptation reveals separate mechanisms for first-order and second-order motion. *Journal of Neurophysiology*, 97, 1319–1325. [PubMed]
- Badcock, D. R., Clifford, C. W. G., & Khun, S. K. (2005). Interactions between luminance and contrast signals in global form detection. *Vision Research*, 45, 881–889. [PubMed]
- Baker, C. L., Jr. (1999). Central neural mechanisms for detecting second-order motion. *Current Opinion in Neurobiology*, 9, 461–466. [PubMed]
- Barrow, H. G., & Tenebaum, J. M. (1978). Recovering intrinsic scene characteristics from images. In A. Hanson & E. Riseman (Eds.), *Computer vision systems*. New York: Academic Press.
- Brewster, D. (1826). On the optical illusion of the conversion of cameos into intaglios, and intaglios into cameos, with an account of other analogous phenomena. *Edinburgh Journal of Science*, 4, 99–108.
- Campbell, F. W., & Robson, J. G. (1968). Application of Fourier analysis to the visibility of gratings. *The Journal of Physiology*, 181, 551–566. [PubMed] [Article]

- Christou, C. G., & Koenderink, J. J. (1997). Light source dependence in shape from shading. *Vision Research*, 37, 1441–1449. [PubMed]
- Dakin, S. C., & Mareschal, I. (2000). Sensitivity to contrast modulation depends on carrier spatial frequency and orientation. *Vision Research*, 40, 311–329. [PubMed]
- Ellemberg, D., Allen, H. A., & Hess, R. F., (2004). Investigating local network interactions underlying first- and second-order processing. *Vision Research*, 44, 1787–1797. [PubMed]
- Ellemberg, D., Allen, H. A., & Hess, R. F. (2006). Second-order spatial frequency and orientation channels in human vision. *Vision Research*, 46, 2798–2803. [PubMed]
- Erens, R. G. F., Kappers, A. M. L., & Koenderink, J. J. (1993). Perception of local shape from shading. *Perception & Psychophysics*, 54, 145–156. [PubMed]
- Fleet, D. J., & Langley, K. (1994). Computational analysis of non-Fourier motion. *Vision Research*, 34, 3057–3079. [PubMed]
- Foley, J. (1994). Human luminance pattern-vision mechanisms: Masking experiments require a new model. *Journal of the Optical Society of America A*, 11, 1710–1719. [PubMed]
- Foley, J. M., & Chen, C. C. (1997). Analysis of the effect of pattern adaptation on pattern pedestal effects: A two-process model. *Vision Research*, 37, 2781–2788. [PubMed]
- Georgeson, M. A., & Schofield, A. J. (2002). Shading and texture: Separate information channels with a common adaptation mechanism? *Spatial Vision*, 16, 59–76. [PubMed]
- Graham, N., & Sutter, A. (2000). Normalization: Contrast-gain control in simple (Fourier) and complex (non-Fourier) pathways of pattern vision. *Vision Research*, 40, 2737–2761. [PubMed]
- Heeger, D. (1992). Normalization of cell responses in cat striate cortex. *Visual Neuroscience*, 9, 181–197. [PubMed]
- Heeger, D. (1993). Modelling simple-cell direction selectivity with normalised, half-squared, linear operators. *Journal of Neurophysiology*, 70, 1885–1898. [PubMed]
- Henning, G. B., Hertz, B. G., & Broadbent, D. E. (1975). Some experiments bearing on the hypothesis that the visual system analyses spatial patterns in independent band of spatial frequency. *Vision Research*, 15, 887–897. [PubMed]
- Hess, R. F., Ledgeway, T., & Dakin, S. (2000). Impoverished second-order input to global linking in human vision. *Vision Research*, 40, 3309–3318. [PubMed]
- Horn, B. K. P., & Brooks, M. J. (1989). *Shape from shading*. Cambridge, MA: MIT Press.
- Hutchinson, C. V., Baker, C. L., Jr., & Ledgeway, T. (2007). Response to combined first-order and second-order motion in visual cortex neurons. *Perception*, 36, 305–306.
- Johnson, A. P., & Baker, C. L. (2004). First- and second-order information in natural images: A filter-based approach to image statistics. *Journal of the Optical Society Of America A, Optics, Image Science, and Vision*, 21, 913–925. [PubMed]
- Kingdom, F. A. A. (2003). Colour brings relief to human vision. *Nature Neuroscience*, 6, 641–644. [PubMed]
- Kleffner, D. A., & Ramachandran, V. S. (1992). On the perception of shape from shading. *Perception & Psychophysics*, 52, 18–36. [PubMed]
- Klein, S. A., Hu, Q. J., & Carney, T. (1996). The adjacent pixel nonlinearity: Problems and solutions. *Vision Research*, 36, 3167–3181. [PubMed]
- Knill, D. C. (1992). The perception of surface contours and surface shape: From computation to psychophysics. *Journal of the Optical Society of America A*, 9, 1449–1464. [PubMed]
- Landy, M. S., & Bergen, J. R. (1991). Texture segregation and orientation gradient. *Vision Research*, 31, 679–691. [PubMed]
- Langer, M. S., & Bulthoff, H. H. (2000). Depth discrimination from shading under diffuse lighting. *Perception*, 29, 649–660. [PubMed]
- Larsson, J., Landy, M., & Heeger, D. (2006). Orientation-selective adaptation to first- and second-order patterns in human visual cortex. *Journal of Neurophysiology*, 95, 862–881. [PubMed]
- Ledgeway, T., Zhan, C. A., Johnson, A., Song, Y., & Baker, C., Jr. (2005). The direction-selective contrast response of area 18 neurons in different for first- and second-order motion. *Visual Neuroscience*, 22, 87–99. [PubMed]
- Legge, G., & Foley, J. (1980). Contrast masking in human vision. *Journal of the Optical Society of America*, 70, 1458–1471. [PubMed]
- Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *Journal of the Acoustical Society of America*, 49, 467. [PubMed]
- Mamassian, P., & Goutcher, R. (2001). Prior knowledge on the illumination position. *Cognition*, 81, B1–B9. [PubMed]
- Mareschal, I., & Baker, C. L., Jr. (1998). Temporal and spatial response to second-order stimuli in cat area 18. *Journal of Neurophysiology*, 80, 2811–2823. [PubMed]

- Meese, T. S. (2004). Area summation and masking. *Journal of Vision*, 4(10):8, 930–943, <http://www.journalofvision.org/content/4/10/8>, doi:10.1167/4.10.8. [PubMed] [Article]
- Meese, T. S., Challinor, K. L., Summers, R. J., & Baker, D. H. (2009). Suppression pathways saturate with contrast for parallel surrounds but not for super-imposed cross-oriented masks. *Vision Research*, 49, 2927–2935. [PubMed]
- Meese, T. S., & Holmes, D. J. (2002). Adaptation and gain pool summation: Alternative models and masking data. *Vision Research*, 42, 1113–1125. [PubMed]
- Morgan, M. J., Mason, A. J. S., & Baldassi, S. (2000). Are there separate first-order and second-order mechanisms for orientation discrimination? *Vision Research*, 40, 1751–1763. [PubMed]
- Nachmias, J. (1989). Contrast modulated maskers: Test of a late nonlinearity hypothesis. *Vision Research*, 29, 137–142. [PubMed]
- Nachmias, J., & Rogowitz, B. E. (1983). Masking by spatially modulated gratings. *Vision Research*, 23, 1621–1629. [PubMed]
- Olmos, A., & Kingdom, F. A. A. (2004). A biologically inspired algorithm for the recovery of shading and reflectance images. *Perception*, 33, 1463–1473. [PubMed]
- Pentland, A. (1988). Shape information from shading: A theory about human perception. *Second International Conference on Computer Vision*, 4, 165–182.
- Ramachandran, V. S. (1988). Perception of shape from shading. *Nature*, 331, 163–166. [PubMed]
- Rittenhouse, D. (1786). Explanation of an optical deception. *Transactions of the American Philosophical Society*, 2, 37–42.
- Schofield, A. J. (2000). What does second-order vision see in an image? *Perception*, 29, 1071–1086. [PubMed]
- Schofield, A. J., & Georgeson, M. A. (1999). Sensitivity to modulations of luminance and contrast in visual white noise: Separate mechanisms with similar behaviour. *Vision Research*, 39, 2697–2716. [PubMed]
- Schofield, A. J., & Georgeson, M. A. (2003). Sensitivity to contrast modulation: The spatial frequency dependence of second-order vision. *Vision Research*, 43, 243–259. [PubMed]
- Schofield, A. J., Hesse, G., Rock, P. B., & Georgeson, M. A. (2006). Local luminance amplitude modulates the interpretation of shape-from-shading in textured surfaces. *Vision Research*, 46, 3462–3482. [PubMed]
- Schofield, A. J., Rock, P. B., & Georgeson, M. A. (submitted for publication). Is there a sun? Probing the default illuminant for shape from shading. *Journal of Vision*.
- Schofield, A. J., Rock, P. B., Georgeson, M. A., & Yates, T. A. (2007). Humans assume a mixture of diffuse and point-source lighting when viewing sinusoidal shading patterns. *Perception*, 36, 108–109.
- Smith, A. T., & Ledgeway, T. (1997). Separate detection of moving luminance and contrast modulations: Fact or artifact? *Vision Research*, 37, 45–62. [PubMed]
- Smith, A. T., & Scott-Samuel, N. E. (2001). First-order and second-order signals combine to improve perceptual accuracy. *Journal of the Optical Society of America A*, 18, 2267–2272. [PubMed]
- Smith, S., Clifford, C. W. G., & Wenderoth, P. (2001). Interaction between first- and second-order orientation channels revealed by the tilt illusion: Psychophysics and computational modelling. *Vision Research*, 41, 1057–1071. [PubMed]
- Song, Y., & Baker, C. L., Jr. (2006). Neural mechanisms mediating responses to abutting gratings: Luminance edges vs. illusory contours. *Visual Neuroscience*, 23, 181–199. [PubMed]
- Sun, J., & Perona, P. (1998). Where is the sun? *Nature Neuroscience*, 1, 183–184. [PubMed]
- Sutter, A., Sperling, G., & Chubb, C. (1995). Measuring the spatial frequency selectivity of second-order texture mechanisms. *Vision Research*, 35, 915–924. [PubMed]
- Tappen, M. F., Freeman, W. T., & Adelson, E. H. (2005). Recovering intrinsic images from a single image. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27, 1459–1472. [PubMed]
- Todd, J. T., & Mingolla, E. (1983). Perception of surface curvature and direction of illumination from patterns of shading. *Journal of Experimental Psychology: Human perception and performance*, 9, 583–595. [PubMed]
- Tyler, C. W. (1998). Diffuse illumination as a default assumption for shape-from-shading in graded images. *Journal of Image Science and Technology*, 42, 319–325.
- Wilson, H. R., Ferrera, V. P., & Yo, C. (1992). A psychophysically motivated model for two-dimensional motion perception. *Visual Neuroscience*, 9, 79–97. [PubMed]
- Zhou, Y.-X., & Baker, C. L., Jr. (1996). Spatial properties of envelope-responsive cells in area 17 and 18 neurons of the cat. *Journal of Neurophysiology*, 75, 1038–1050. [PubMed]