

## Neural correlates of shape-from-shading

Chuan Hou\*, Mark W. Pettet, Vladimir Y. Vildavski, Anthony M. Norcia

*The Smith-Kettlewell Eye Research Institute, 2318 Fillmore Street, San Francisco, CA 94115, USA*

Received 14 June 2005; received in revised form 14 October 2005

### Abstract

Visual evoked potentials were recorded during presentation of a single stimulus that generated bi-stable perceptual alternation between two different three-dimensional percepts. One interpretation (asymmetric) changed depth structure from flat to corrugated in depth and the other (symmetric) had the appearance of a flat surface translating laterally behind a set of apertures. Responses during perception of the asymmetric three-dimensional structure contained larger negative components than did responses during perception of the symmetric three-dimensional structure. Control experiments suggest that the interpretation of depth structure is selected after junction information caused by the interplay between shading and object shape is extracted.

© 2005 Elsevier Ltd. All rights reserved.

*Keywords:* Shading; 3D-shape; Junctions; Bi-stable perception; Visual evoked potentials

### 1. Introduction

Shading, combined with boundary contour information provides a powerful monocular cue for three-dimensional (3D) shape. The ability to perceive shape-from-shading has been studied extensively using psychophysical methods (Aks & Enns, 1992; Braun, 1993; Enns & Rensink, 1990; Kleffner & Ramachandran, 1992; Ramachandran, 1988a, 1988b; Sun & Perona, 1997, 1996), but the underlying neural mechanisms are poorly understood. An early functional imaging study in humans used shading gradients to portray a series of convexities and concavities that were either lit from above or from the side (Humphrey et al., 1997). The observers reported stronger and more stable percepts of 3D when the lighting was from above, as previously described (Ramachandran, 1988b). A comparison of the activation in the two lighting conditions indicated differential responses in areas V1, V2, and V3. These results are consistent with a physiological study of V1 and V2 neurons of macaque monkeys (Lee, Yang, Romero, & Mumford, 2002) which

demonstrated that similar stimuli evoked a contextual pop-out response (120–320 ms after stimulus onset), although this response was influenced by higher-order stimulus attributes and task experience. Human evoked response data (Mamassian, Jentzsch, Bacon, & Schweinberger, 2003) suggest that the direction of lighting is disambiguated as early as 100 ms, which is consistent with activation of early visual areas. A more recent fMRI study, using attention to either shading or color as the task has, on the other hand, emphasized the intra-parietal sulcus as being important in the perception of shape-from-shading (Taira, Nose, Inoue, & Tsutsui, 2001).

Although these results provide important clues about the neural mechanisms that process shading cues and extract shape information, there are still some open questions. The physiological and imaging studies have used lighting-direction or the shape of the shading gradient to vary the 3D percept. Here we varied the consistency of shading and boundary information to study the time course of shading/border interactions. Stimuli whose border and shading cues were consistent with a 3D interpretation elicited enhanced visual evoked potentials (VEP) responses and associated spectral components that were not present in control conditions that had inconsistent border and

\* Corresponding author. Tel.: +1 415 345 2073; fax: +1 415 345 8455.  
E-mail address: [chuanhou@ski.org](mailto:chuanhou@ski.org) (C. Hou).

shading cues, but were matched for motion transients. Of greater interest, in the course of the experiments, we found a new bi-stable stimulus for which the same images yielded two different 3D interpretations, one alternating between flat and 3D and the other alternating between two translationally symmetric states with the same depth ordering. These stimuli have allowed us to separate response activity associated with the different depth interpretations from that due to low-level junction cues present in the stimuli.

## 2. Methods and materials

### 2.1. Observers

A total of 16 visually normal observers, aged 19 to 49 years participated. Eleven subjects participated in each experiment. Five of the subjects took part in both experiments. Each observer had or was corrected to 20/20 or better acuity in each eye, had normal stereopsis on the Titmus stereo test and was fully refracted for the viewing distance. The research followed the tenets of the World Medical Association Declaration of Helsinki and informed consent was obtained from the subjects after explanation of the nature and possible consequences of the study. The research was approved by the Institutional Human Experimentation Committee.

### 2.2. Stimuli

Stimuli were generated by a MacIntosh G4 computer equipped with an nVidia GeForce 2 graphics card and were

presented on a 19-in., multi-synch monochrome CRT monitor (MRHB2000, Richardson Electronics) set to a graphics mode of  $800 \times 600$  pixels, 72 Hz vertical refresh, at 256 gray levels (video bandwidth, 150 MHz). We varied the spatial relationship between shading and object border cues, either making the two consistent with a 3D interpretation or a predominantly two-dimensional (2D) interpretation. We used several variants of Ramachandran's windowed grating (Ramachandran, 1988a), as illustrated in Fig. 1 (see Supplementary Materials for animations). The shading information comprised a square-wave grating (denoted the *carrier*). The carrier pattern was a 0.32 cpd square-wave grating with 95% contrast and mean luminance of  $123 \text{ Cd/m}^2$ . The carrier was partially hidden by an occluding gray frame of  $123 \text{ Cd/m}^2$  with three horizontal apertures, each  $3^\circ$  tall, with  $3^\circ$  between each aperture. The carrier was windowed by this mid-gray occluder whose aperture (denoted the *envelope*) running along the direction of the carrier followed a serrated path, with the period of the serrations matching the period of the grating. When the peaks of the envelope serrations aligned with the black-white transitions of the carrier, the stimulus looked like a folded surface (denoted the 3D image). This surface could appear as a set of triangular corrugations resembling a series of adjacent rooftops or a fan-fold card standing upright. When the carrier and envelope were misaligned, the stimulus appeared as a relatively flat, jagged, black-and-white-striped ribbon.

Note that none of the stimuli were actually 3D—they contained monocular cues that were either consistent with or largely inconsistent with a 3D percept. Prior to beginning the recordings, we reviewed each of the stimuli with

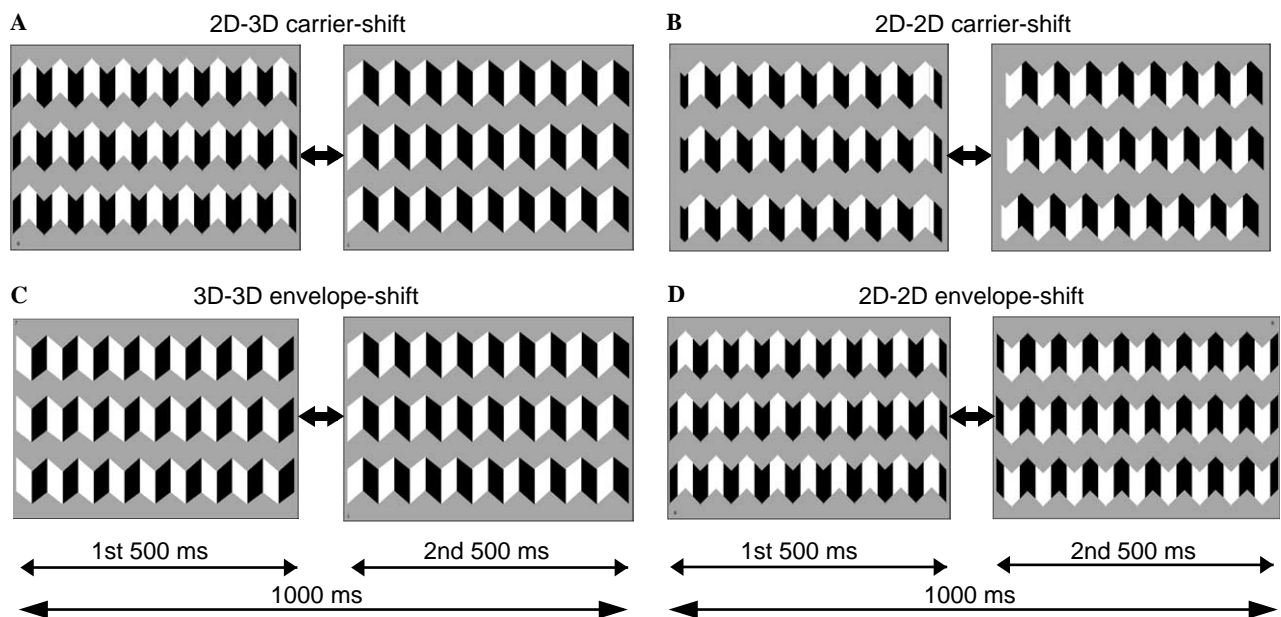


Fig. 1. Stimuli and experimental paradigm. (A) Test condition: 2D–3D carrier-shift, the alternation was between misaligned ( $\sim$ 2D image) and aligned states (3D image); only the carrier modulated by  $90^\circ$ . (B) Control condition 1: 2D–2D carrier-shift, the same grating motion (carrier shift of  $90^\circ$ ) was made symmetrically between misaligned states ( $\sim$ 2D images) again without an envelope shift. (C) Control condition 2: 3D–3D envelope-shift, the carrier remained constant and the envelope shifted by  $180^\circ$ , yielding an alternation between two 3D images. (D) Control condition 3: 2D–2D envelope-shift, the carrier remained static and the envelope was shifted through  $180^\circ$  symmetrically about the zero crossing of the carrier grating. The images were updated every 500 ms and repeated every second. See Supplementary Materials for animations of the stimuli (set QuickTime to “loop” for continuous viewing).

the observers to ensure that they were aware of the differences in apparent 3D between the different stimulus conditions. Perceived 3D was highest for images that had aligned carrier and envelopes (e.g., the second image of Fig. 1A, both images of Fig. 1C). These images will be referred to as “3D images”. Some observers reported partial depth for the images of Fig. 1B. None of the observers reported 3D percepts for the first image of Fig. 1A or the images of Fig. 1D. Because some observers may perceive residual depth in these images, we refer to them throughout as “approximately 2D images” (~2D images).

Time-locked evoked responses were elicited by periodic phase shifts of either the carrier or the envelope. The image pairs for each test and control condition are illustrated in Fig. 1. Shifting the carrier by  $90^\circ$  resulted in an alternation between an ~2D image and a 3D image if the envelope and carrier were misaligned in the first phase and aligned in the second phase (Fig. 1A), or an alternation between two mirror symmetric ~2D images if the carrier and envelope were misaligned in both phases (Fig. 1B). Shifting the envelope by  $180^\circ$  resulted in an alternation between two 3D images when the carrier and envelope

were aligned on both phases (Fig. 1C) or an alternation between two ~2D images when they were not (Fig. 1D). For simplicity, these conditions will be referred to as the 2D–3D carrier-shift condition, the 2D–2D carrier-shift condition, the 3D–3D envelope-shift condition and the 2D–2D envelope-shift condition, respectively. Observers viewed the stimulus from a distance of 70 cm and were instructed to hold their gaze on the center of the display (no fixation pattern was shown) but attend to the entire display. The VEP responses with and without explicit fixation are similar in both the 2D–3D and 2D–2D carrier-shift conditions and the magnitude of eye movements is much smaller than the extent of carrier motion (see Supplementary Materials). The image pairs were updated every 500 ms, repeating every 1 s. Each trial lasted 12 s with the first and last seconds discarded. Data from each experimental condition were collected in blocks of five trials. The blocks were chosen at random with respect to stimulus condition. A total of 10 trials per condition were collected in the first experiment (Figs. 2–5, 100 one-second stimulus cycles), and 25 trials (250 one-second stimulus cycles) were collected in the second experiment.

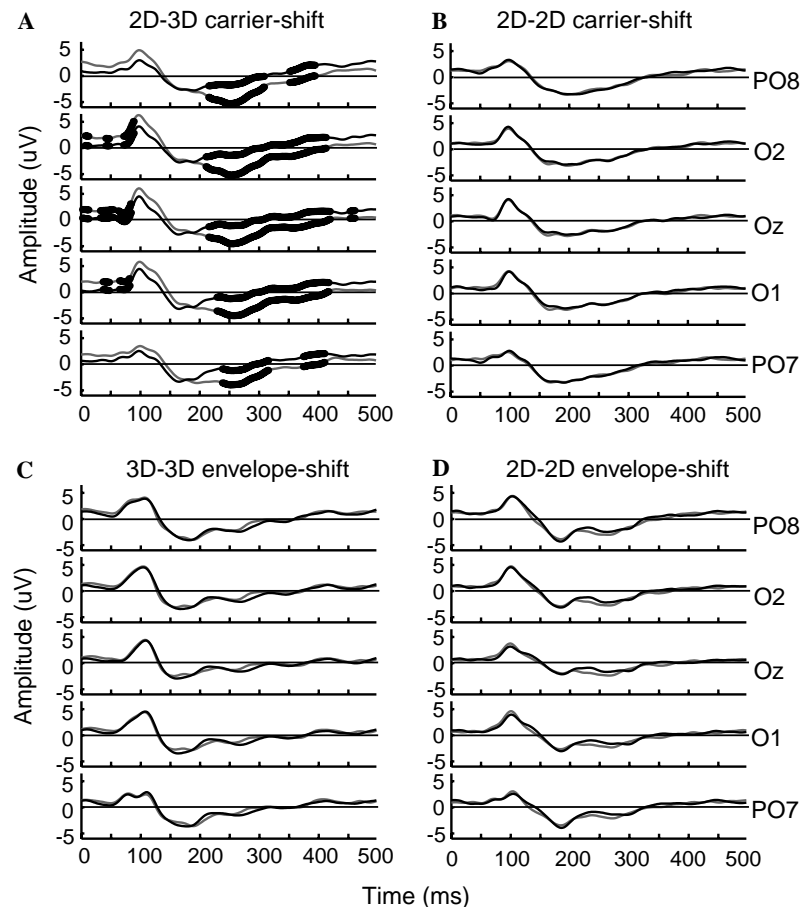


Fig. 2. Responses to the first and second 500 ms transitions of each stimulus type averaged over 11 observers. The dark lines indicate the first 500 ms transition and the gray lines indicate the second 500 ms transition based on the stimulus set of Fig. 1. The black dots indicate significant differences between the waveforms identified by permutation testing (see Section 2). Each row plots data from a different recording channel (from bottom to top: PO7, O1, Oz, O2, and PO8). (A) Plots data from the 2D–3D carrier-shift condition. The transition from the ~2D to the 3D image (gray lines) evoked an increased negativity between approximately 200–400 ms. (B–D) Plot data from the 2D–2D carrier-shift, 3D–3D envelope-shift, and 2D–2D envelope-shift conditions. There were no significant differences between the responses to the first and second 500 ms transitions of each of these controls.

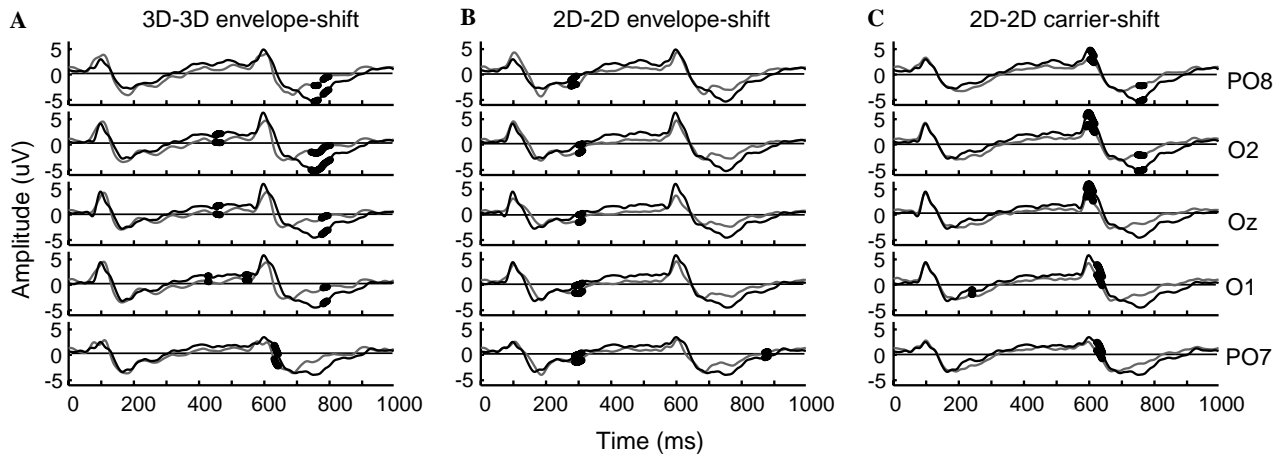


Fig. 3. Comparison of the responses to the test condition (2D–3D carrier-shift, dark lines) and each of three control conditions (gray lines). (A) The response to the same 3D image, presented at 500 ms in both 2D–3D carrier-shift (dark lines) and 3D–3D envelope-shift (gray lines) was significantly larger in the 2D–3D carrier-shift condition. (B) The response to the same ~2D image, presented at 0 ms in both 2D–3D carrier-shift (dark lines) and 2D–2D envelope-shift conditions (gray lines) was significantly different at 300 ms. (C) The test condition (dark lines) differed from the 2D–2D control (gray lines) as early as 100 ms after the transition to the 3D image at 500 ms in the test condition. The test and control stimuli differed by a 45° difference in relative phase of the carrier and envelope at both 0 and 500 transitions, but only the second (3D) transition resulted in a response difference.

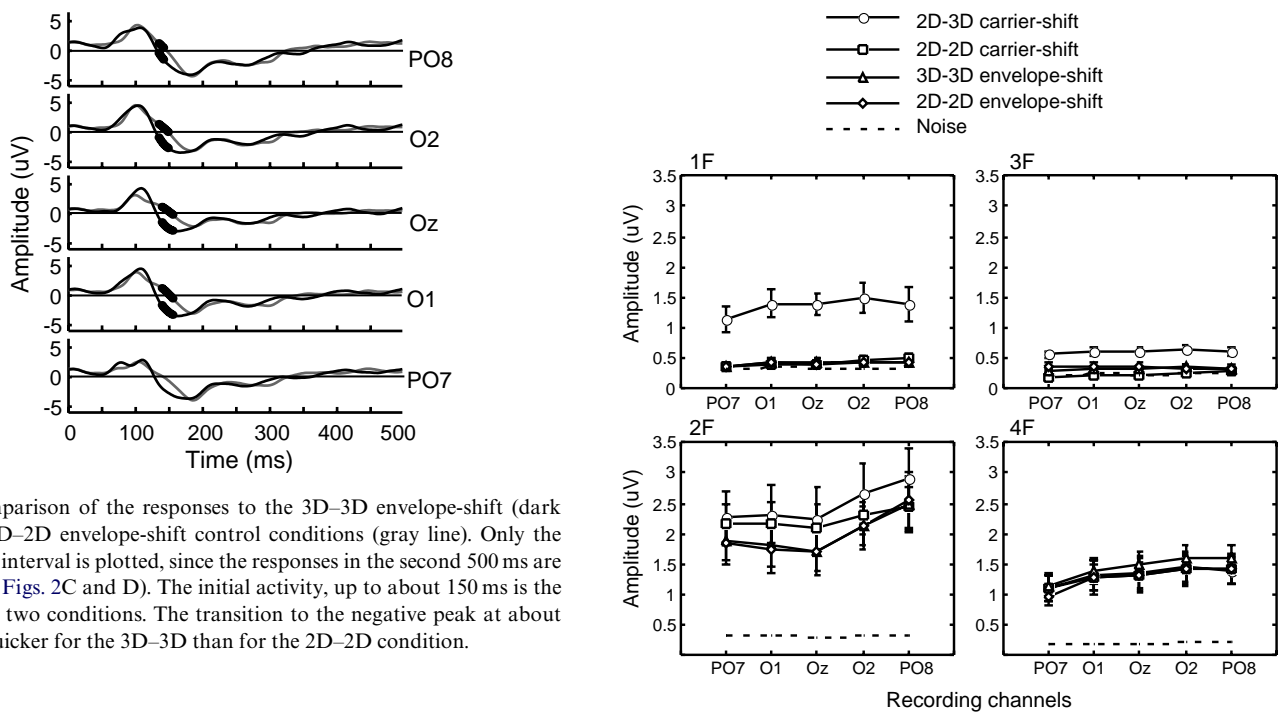


Fig. 4. Comparison of the responses to the 3D–3D envelope-shift (dark line) and 2D–2D envelope-shift control conditions (gray line). Only the first 500 ms interval is plotted, since the responses in the second 500 ms are similar (see Figs. 2C and D). The initial activity, up to about 150 ms is the same in the two conditions. The transition to the negative peak at about 150 ms is quicker for the 3D–3D than for the 2D–2D condition.

### 2.3. VEP recording

The EEG (electroencephalogram) was recorded with Grass gold-cup electrodes. The skin was prepared with Omni-Prep, and 10–20 conductive cream (D.O. Weaver) was applied. Electrode impedances were between 2 and 10 kΩ. VEPs were recorded during binocular viewing over five derivations PO7, O1, Oz, O2, and PO8 each referenced to Cz, according to international 10–10 electrode placement system (Nuwer et al., 1999). The electroencephalogram was amplified 50,000× with Grass Model 12 amplifiers and digitized to 16 bits accuracy at a sampling rate of 432 Hz. Analog filter settings were 0.3–100 Hz, measured at –6 dB points.

Fig. 5. Grand scalar average amplitudes of the VEP signal ( $n = 11$ ) for the first (1F), second (2F), third (3F), and fourth (4F) harmonics as a function of recording channels. The test stimulus (2D–3D carrier-shift; circles) had significant responses at both odd harmonics (1F and 3F) and even harmonics (2F and 4F), while the three control stimuli, 2D–2D carrier-shift (squares), 3D–3D envelope-shift (triangles) and 2D–2D envelope-shift (diamonds) showed significant responses only at even harmonics (2F and 4F). The dash lines indicate the noise level determined at two frequencies adjacent to the response frequency of interest. Error bars are  $\pm 1$  standard error of mean.

The recordings in the second experiment (Figs. 7 and 8) were sorted according to the observers’ perception of the 2D–3D carrier-shift stimulus. At the beginning of each

stimulus presentation, observers waited until they perceived “3D lateral motion” (see details below; VEPs during bi-stable perception) and then initiated recording by pressing a button designated for the 3D lateral motion percept. When their percept changed to a “3D on/off” percept, observers switched to a second button designated for the 3D on/off percept. Each time their percept switched between the 3D on/off and 3D lateral motion states, they pressed the appropriate button again until the end of the 10 s trial. Records of button presses were used to sort the data for off-line analysis into separate data arrays corresponding to the two percepts. The redirection of the raw data into the appropriate array happened 500 ms before each button press to adjust for decision and motor response latencies. The results did not vary significantly for latency adjustments in the range 400–800 ms (data not shown). The 3D lateral motion and 3D on/off data arrays were analyzed and averaged separately as described below.

#### 2.4. Time domain analysis

The data in Fig. 2–4 were obtained by a combination of digital filtering and averaging in the frequency domain, followed by a synthesis of the filtered frequency components into a time-averaged waveform. For each subject, stimulus condition, and channel, raw EEG recordings for each 10 s trial were partitioned into five sequential epochs of 2 s duration. These epochs were averaged together (over the 2 s epoch) within each trial, and then across trials to obtain a single grand average waveform for each observer, channel, and stimulus condition. Each grand average waveform was decomposed by discrete Fourier transformation and digitally filtered by zeroing the data at 53 Hz and above to remove 60 Hz and other high-frequency noise. A new waveform corresponding to a single stimulus cycle (1 s duration) was then reconstructed by an inverse discrete Fourier transformation that only included harmonics of the 1 Hz stimulus frequency up to 53 Hz. The filter thus did not pass frequencies of 0.5, 1.5, 2.5, 3.5 Hz, etc., since these frequencies were not harmonics of the stimulus frequency and were therefore EEG noise. The filtering procedure thus allowed us to reject more EEG noise from the average (at the non-harmonic frequencies) than would conventional low-pass filtering. These reconstructed waveforms were then averaged together across observers for each channel and stimulus condition.

Data in Fig. 7 were obtained by sorting and time-averaging without spectral decomposition and filtering (note 60 Hz noise). For each subject and channel, records of button presses were used to sort data samples from raw recordings into two separate time series data arrays corresponding to the 3D lateral motion or 3D on/off percepts. Each of these time series was then partitioned into successive epochs of 1 s duration (i.e., the stimulus cycle duration). These epochs were averaged together within each trial, then across trials, and then across observers to

obtain a single grand average waveform for each channel and perceptual category.

#### 2.5. Spectral analysis

The data in Figs. 5 and 8 were obtained by digital spectral analysis. For each subject, stimulus condition, and channel, raw scalp potentials for each trial were adaptively filtered (Tang & Norcia, 1995) to extract a running record of the amplitude and phase of the evoked response at a particular harmonic of the 1 Hz stimulus frequency. In this analysis, digital filters were fit to each 1 s epoch in succession to obtain a series of complex-valued coefficients representing the amplitude and phase of response components tuned to various harmonics of the stimulus frequency. These complex coefficients were averaged together across successive epochs in each trial, then across trials to obtain a single grand average complex coefficient for each subject, channel, harmonic, and stimulus condition. Scalar amplitudes derived from these coefficients were averaged together across subjects.

#### 2.6. Statistical analysis

Multivariate analysis of variance for repeated measures (Statistica GLM module, StatSoft) was used to assess significance of evoked potential amplitudes for within-subject factors and interactions (data in Figs. 5 and 8). We used Wilks’ ratio to compute  $F$  ratios for multivariate and univariate tests of significance (Timm, 2002).

The permutation test of significance for waveform differences was based on previously described methods (Blair & Karniski, 1993). Under the null hypothesis of no difference between two conditions, the two responses waveforms are exchangeable for a given subject. By randomly permuting which subjects had their waveforms exchanged, we created a large set of permuted data sets. For each permutation, we calculated the  $t$  value (10 degrees of freedom) for each time point in the waveform. Then, the maximum  $t$  values (over time) from each permutation were accumulated into a reference distribution. Any time point from the original, un-exchanged data whose  $t$  value fell within the top 5th percentile of the reference distribution was deemed significant and implied rejection of the null hypothesis with a 5% probability of type 1 (false rejection) error. Such time points in the waveform plots in Figs. 2–4, and 7 are indicated with black dots. Since we used only the maximum  $t$  value from each permutation, no correction for multiple comparisons was necessary (Nichols & Holmes, 2001).

### 3. Results

#### 3.1. Consistent border and shading information leads to a larger evoked response

We first asked whether the VEP is sensitive to the relationship between border and shading information. To do

this we compared responses to stimuli where the phase of either the carrier or the envelope was periodically shifted while the other remained constant. If the visual system was responding only to contrast transients associated with the phase shift of the carrier, as in Figs. 1A and B, the response to the image updates of both carrier-shift conditions should be the same, since the carrier translations are the same. If on the other hand, the relationship between the shading and border has been encoded, the responses will differ. A similar argument applies in the envelope-shift conditions of Figs. 1C and D.

Our first analysis found that the response waveform depends on the specific relationships between border and shading information. Transitions from either 3D to ~2D images or ~2D to 3D images evoked an initial positive response peaking at 100 ms (P100) followed by a longer latency negative response commencing after about 150 ms (Fig. 2A). However, both of these components were larger after the onset of the 3D image (Fig. 2A, gray lines). Because of the periodic alternation of the ~2D and 3D images and the sustained nature of the response, there may be wrap-around of the trailing negativity from the 3D onset-response onto the leading part of the ~2D onset response. This could magnify the apparent difference between the P100 responses. This point will be addressed again below when we consider the responses to the control stimuli.

In the remaining three panels of Fig. 2, which represent various control responses, it can be seen that the waveforms from the first and second 500 ms epochs of each condition do not differ from each other if the carrier is shifted symmetrically with respect to the envelope (Fig. 2B) or if the envelope is shifted symmetrically with respect to the carrier (Figs. 2C and D). As before, the waveforms were biphasic with a P100 component and a subsequent sustained negativity. The two phases of each of these three control stimuli exhibit both mirror and repetition symmetry (either carrier changing or envelope changing), so transitions between symmetric stimulus exemplars should evoke comparable population responses. On the other hand, the images comprising the test condition (2D–3D carrier-shift) do not have these kinds of symmetry. The population response is measurably different to these two images, indicating that the relative position of border and shading information has been encoded by mechanisms accessible by the VEP.

The possible effects of eye movements was assessed in a sub-set of observers by comparing responses generated with and without an explicit fixation point. VEP responses were similar with and without an explicit fixation point. Eye movements, if present at all, did not alter the difference in waveform between test and control conditions (see Supplementary Material for details of these control experiments).

### 3.2. Effects of prior stimulus history

We next directly compared the responses in the 2D–3D carrier-shift condition with each of the control conditions in order to further examine stimulus configuration effects. If

image structure per se is the only determinant of the evoked response waveform, the response to the second half of the 2D–3D carrier-shift condition (the 3D image of Fig. 1A) should be the same as that to the second half of the 3D–3D control of Fig. 1C, since the images are the same. The only difference is the prior history of stimulation: in the 2D–3D condition, the 3D image was preceded by a ~2D image, but in the 3D–3D condition, it was preceded by a translationally symmetric 3D image. As can be seen in Fig. 3A, which compares the 2D–3D carrier-shift responses (Fig. 1A) to the 3D–3D envelope-shift responses (Fig. 1C), these responses differ as early as 120–150 ms at PO7 and at 230–320 ms on the other channels. Note that the latencies here are with respect to the 500 ms update. The response to the 3D image preceded by a 3D image reaches its maximum negativity about 200 ms earlier than the response to the same 3D image preceded by an ~2D image. The initial positivity starts slightly later, but the transition to the negativity occurs more quickly. Maintaining a translationally shifted representation of the same depth structure (as in the 3D–3D condition) may require less computational effort than generating a new depth structure every 500 ms, as is necessary in the 2D–3D condition. This may explain why the response returns to baseline more quickly in the 3D–3D condition compared to the 2D–3D condition. These results thus imply a form of visual memory for 3D shape.

A similar effect of prior stimulus history can be seen in the comparison of the responses to identical ~2D images—the first image of the 2D–3D condition in Fig. 1 and the first image in the 2D–2D envelope condition of Fig. 1D. The relevant data are shown in Fig. 3B for the 0–500 ms range. Here the responses differ briefly at around 300 ms. In the 2D–3D condition, the ~2D image presented at 0 ms had been preceded by a 3D image, but in the 2D–2D condition, the same image had been preceded by a ~2D image.

### 3.3. Clarification of the differential response at 100 ms

We now address the wrap-around ambiguity for the P100 response mentioned above for Fig. 2A by comparing the 2D–3D carrier-shift condition (Fig. 1A) to the 2D–2D carrier-shift control condition of Fig. 1B as shown in Fig. 3C. In making this comparison, we observe the same two configural effects observed in Fig. 2A: the second half of the 2D–3D condition has a larger P100, and a larger sustained negativity beginning about 150 ms after the image update at 500 ms compared to the 2D–2D control. Note that there is no response difference between conditions during the first 500 ms of the traces in Fig. 3C, where both conditions have the same ~2D stimulus. Second, note also that the first and second half of the 2D–2D carrier-shift control (Fig. 2B) do not differ. Hence, the P100 to the ~2D half of the 2D–3D stimulus is not diminished by any negative wrap-around from the second half of the stimulus. We can now see that the difference between the P100 in the first and second halves of the 2D–3D stimulus is entirely due to enlargement of the P100 for the 3D half of the stimulus.

The critical role of *alignment* of border and shading information is indicated by the fact that different responses are obtained from the 3D image compared to a control  $\sim$ 2D image that differs in a  $45^\circ$  shift of the carrier with respect to the envelope (data between 500 and 1000 ms in Fig. 3C), but that the responses to the two other  $\sim$ 2D images that also differ by only a  $45^\circ$  phase shift are the same (data from the 0 to 500 ms interval in Fig. 3C). The timing of the significant differences between the responses suggests that information about the relative alignment of the shading and border thus is available as early as 100 ms.

### 3.4. Comparison of 2D–2D and 3D–3D conditions

In Fig. 3, we compared the responses to the physically asymmetric 2D–3D carrier shift condition to three different controls, each of which had one or more axes of symmetry. A comparison of the two envelope shift conditions allows us to compare two symmetric stimulus conditions that differ in whether both phases support 3D (e.g., Fig. 1C) or not (Fig. 1D), but that are otherwise matched. This comparison is shown in Fig. 4 where it can be seen that the slope of the transition from the P100 to the sustained negativity was much steeper in the 3D–3D condition than in the 2D–2D condition. This difference in processing speed may be due to increased activity specific to 3D or to differences in responses to the particular local feature arrangements. This question is addressed in an experiment described below that used a display that was perceptually bi-stable.

### 3.5. Quantitative analysis of response symmetry in the frequency domain

Both carrier and envelope were animated by symmetric oscillations of forward and reverse motion. The scalp potential should include responses from topographically intermingled direction-selective mechanisms that respond first to the forward motion, then to the reverse motion, or local contrast mechanisms which receive comparable input modulations. Since either of these responses occur twice per forward–reverse cycle, the steady state waveform will be dominated by spectral components that are even harmonics of the stimulus frequency (denoted  $2F$ ,  $4F$ ,  $6F$ , etc.). If on the other hand, there are mechanisms sensitive to the relationship between the envelope and carrier, either locally or globally, then these responses will differ between the two halves of the stimulus cycle. These responses will repeat themselves at the period of the full stimulus cycle, and will therefore add odd harmonics ( $1F$ ,  $3F$ ,  $5F$ , etc.) to the response spectrum. Thus, while oscillation of the envelope or carrier in all three control stimuli should generate VEP responses at the even harmonics, only 2D–3D alternation should generate responses at the odd harmonics since this is the only asymmetric stimulus.

As expected, the responses in all conditions contained significant even harmonic components (Fig. 5). However, the 2D–3D carrier-shift condition was the only case where

significant odd-harmonic components were observed. MANOVA of the interaction between harmonic and stimulus condition indicated a significant effect,  $F(1,10)=7.91$ ,  $p=0.03$ . This interaction did not significantly depend on recording channel,  $F(4,7)=0.67$ ,  $p=0.64$ , despite noticeably larger amplitudes at O2 and PO8. The noise estimates in Fig. 5 were the mean amplitudes of the pair of adjacent frequency components located  $\pm 0.5$  Hz from the corresponding response harmonics of interest.

### 3.6. VEPs during bi-stable perception

We have thus far established that the VEP reflects the activity of mechanisms that are sensitive to the relationship between shading and border cues and that the response waveform to a given image depends strongly on the prior history of stimulation. In addition, when compared to controls, the 2D–3D stimulus pair generates odd harmonic responses consistent with asymmetric responses to the  $\sim$ 2D–3D images in the pair. At this point, we may be tempted to conclude that the asymmetric response comes from different depth percepts elicited by the two images in the 2D–3D condition. However, the image pair in this condition has another asymmetry that we must consider. The black and white polygons are shaped like arrows in the  $\sim$ 2D image, but are parallelograms in the 3D image. This change of shape could drive the observed response asymmetry, since the parallelograms in the control image pairs simply translate without changing shape. This shape-change confound was addressed in the following experiment.

After prolonged viewing, we noted that the perception of an alternation between a  $\sim$ 2D (relatively flat) and 3D (corrugated) images in the 2D–3D carrier-shift condition gives way to a second percept in which the black-and-white gratings always appear flat but are seen to oscillate laterally behind an occluding gray window. We will refer to the first percept as the “3D on/off” percept and the second percept as the “3D lateral motion” percept. Fig. 6 schematically illustrates these two perceptual states. The existence of a bi-stable percept provides the means to dissociate activity associated with the different 3D interpretations from the particular local border/shading relationships, since the stimulus never changes, only its interpretation does. Specifically, any increase in response asymmetry during the 3D on-off percept can be attributed to depth processing mechanisms, since the shape-change (arrows vs parallelograms) is constant both in the stimulus and perceptually.

We asked observers (11 observers, 5 of whom participated in the first experiment) to indicate their perceptual state by holding one button down during the 3D on/off percept, and a second one during the 3D lateral motion percept while the VEP was recorded. The data from the two perceptual states were then sorted contingent on the perceptual state and adjusted by 500 ms for the latency of reaction time following perceptual shifts. Time averaged waveforms and response spectra similar to those in the first

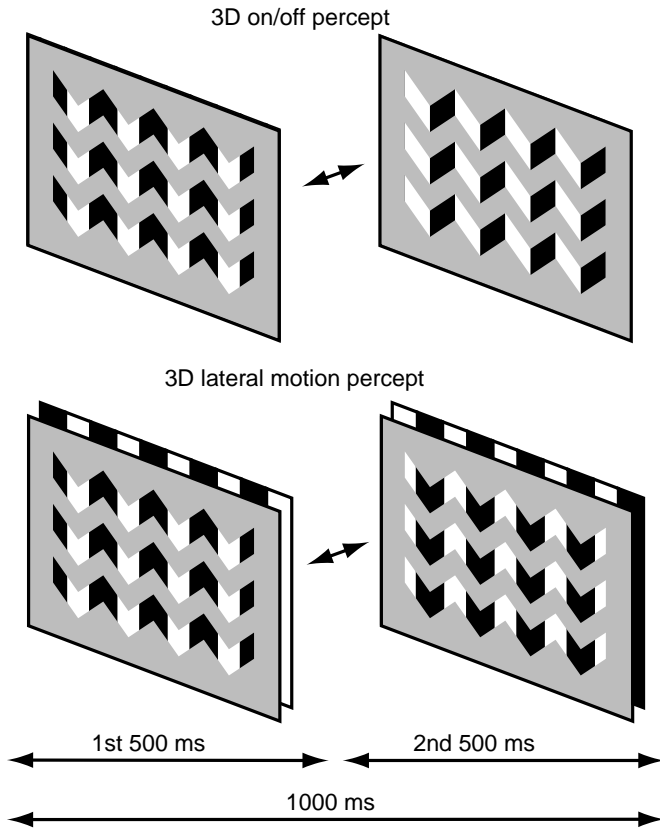


Fig. 6. Schematic illustration of the two perceptual interpretations of the 2D–3D carrier-shift stimulus. After prolonged viewing, the 2D–3D carrier-shift condition in the first experiment became bi-stable and alternated between a percept of the appearance and disappearance of 3D corrugation (top panel, as shown in the first experiment) and a second percept in which the black-and-white gratings appeared flat and were seen to oscillate laterally behind an occluding gray window (bottom panel). Note that the local junction relationships in the lower panel have been changed for illustration purposes. They were always present, as in the upper panel. Supplemental Material file “2D–3D carrier-shift” shows an animation of the stimulus (set QuickTime to “loop” for continuous viewing).

experiment were then calculated for each percept. Data from this experiment were accumulated from 25 ten-second trials (250 one-second stimulus cycles). Observers tended to spend about half of the time in either perceptual state after this initial period—the proportion of 3D on/off versus 3D lateral motion percepts was 51% versus 49%, respectively.

### 3.7. Waveforms during 3D on/off and 3D lateral motion percepts

There was no significant difference in response between the 3D lateral motion percept and the 3D on/off percept for the first 500ms interval of the stimulus cycle. This interval starts at the transition to the ~2D image phase that is always seen as a relatively flat surface. Differences between the responses during 3D on/off and 3D lateral motion percepts were seen for the second half of the stimulus cycle, which does support a corrugated 3D interpretation. The sustained negativity beginning about 150 ms after the onset of the 3D image was consistently larger in the 3D on/off

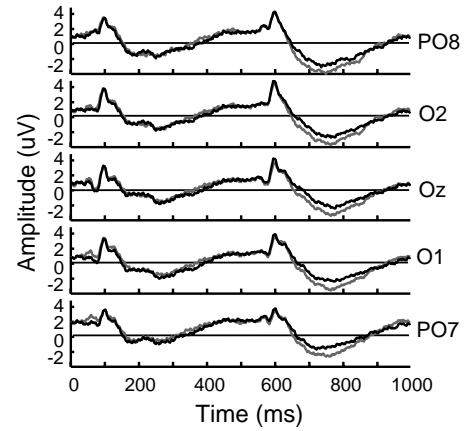


Fig. 7. Responses recorded during the 3D on/off and 3D lateral motion perceptual states averaged over 11 observers and plotted for each recording channel. (Note: these averages were accumulated without the benefit of digital filtering applied to data from the first experiment, and are therefore contaminated with 60 Hz noise; see details in Section 2). The dark lines indicate the response during the 3D lateral motion percept, and the gray lines indicate the response during the 3D on/off percept. The 3D on/off percept showed a more negative response than 3D lateral motion percept starting from about 150 to 350 ms during the second 500 ms transition.

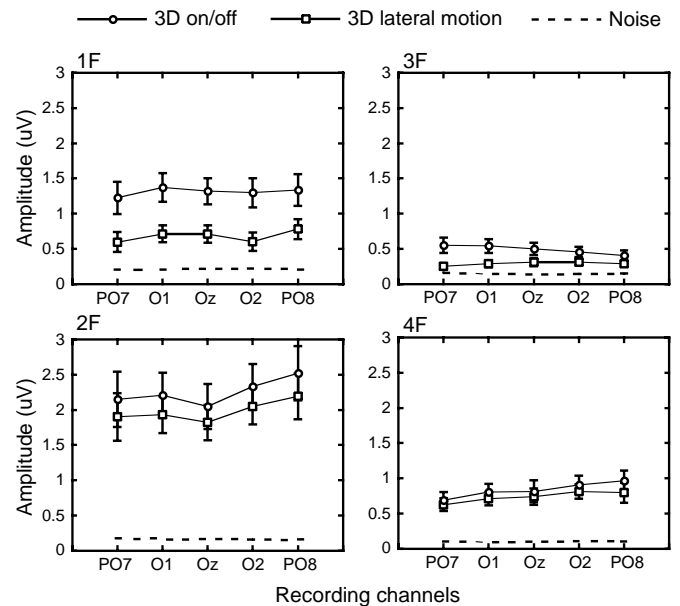


Fig. 8. Grand scalar average amplitudes of the VEP signal ( $n=11$ ) recorded during the 3D on/off percept and 3D lateral motion percept for the first (1F), second (2F), third (3F), and fourth (4F) harmonics as a function of recording channel. The 3D on/off percept (circles) was associated with more odd harmonic power (1F and 3F) than 3D lateral motion percept. Both percepts were associated with equivalent even harmonics (2F and 4F) in all recording channels. The dash lines indicate the noise level determined at two frequencies adjacent to the response frequency of interest. Error bars are  $\pm 1$  standard error of mean.

percept than in the 3D lateral motion percept. However, the point-by-point differences were not statistically significant at the  $p < 0.05$  level. The permutation test is a conservative test of significance, since it is designed to be valid for a very large number time points. There was no difference in the

P100 component between the two percepts in the second interval of the stimulus cycle.

### 3.8. Enhanced response asymmetry during the 3D on/off percept

We can extend the logic of our frequency domain interpretation of responses to symmetric and asymmetric *stimuli* (Fig. 6) to an analysis of responses recorded during symmetric and asymmetric *percepts*. The linking hypothesis here is that symmetric percepts involve translations within the same population or representation, but that asymmetric percepts involve translations between different populations or representations with different response characteristics.

The amplitudes of the first four harmonics of the stimulus frequency were computed separately from VEPs recorded during the two perceptual states. Fig. 8 shows these response amplitudes for the 3D on/off and 3D lateral motion percepts plotted for the different recording channels averaged over eleven observers. Similar to the previous data from the 2D–3D carrier-shift condition (Fig. 5), the 3D on/off percept (circles) elicited more odd harmonic power ( $1F$  and  $3F$ ) than the 3D lateral motion percept, while both percepts evoked equivalent even harmonics ( $2F$  and  $4F$ ). While this interaction between odd/even harmonics and percept was significant,  $F(1,10) = 5.72$ ,  $p = 0.04$ , it did not depend on recording channel,  $F(4,7) = 1.45$ ,  $p = 0.31$ . However, unlike the control responses from the first experiment, the 3D lateral motion percept responses in the second experiment had clearly evident odd harmonic power.

In both the time domain and frequency domain data, the 3D-specific responses that occurred in the first experiment were similar to those that differentiated the changing 3D percept supported by the bi-stable display in the second experiment. The enhanced longer latency negative response appears to be elicited by the percept of changing 3D structure, and not by changes in shape or local configuration of carrier and envelope, since the stimulus never changed. These results point to the existence of neural responses correlated with shape-from-shading that can be dissociated from low-level junction cues.

## 4. Discussion

The utility of shading information as a cue for 3D shape could only arise during the evolution of the visual system due to systematic correlations between shading information and object shape. In complex visual environments, these cues are likely to be relatively unambiguous by themselves and are supported by other shape cues such as those afforded by binocular disparity and motion parallax. In reduced cue situations, such as ours, shading still provides a powerful cue for 3D structure, but this structure can be ambiguous. We were able to exploit this ambiguity to show that the percept of changing 3D structure leads to an enhanced, sustained evoked response over the posterior scalp.

The importance of transitions between different perceived depth structures is clearest in the case of the ambiguous percepts: one of them involved transitions between a relatively flat surface and a corrugated 3D surface (Fig. 6, top) and the other involved perceptually symmetric transitions between two versions of the same depth structure, a flat grating moving laterally behind an aperture (Fig. 6, bottom). A similar effect was seen when comparing the 2D–3D stimulus to its 3D–3D control. In this case the response to the same image structure in the second half of the display was larger if it had been preceded by a flat percept than if it had been preceded by an image that gave rise to a mirror symmetric 3D lateral motion percept. These two effects are clearly non-linear, in that the response to the same image update depends on the prior *stimulus* history in Fig. 3A and the prior *perceptual* history in Fig. 7. It thus appears that the VEP is a sensitive indicator not only of the similarity of the population response to different stimuli, but to different perceptual states, which in our case are internal representations of different depth structures.

Our recordings with the ambiguous stimulus indicate that the specifically 3D activity is an enhancement of an underlying configuration-specific response (the odd-harmonics and the negativity between approximately 150 and 300 ms) occurring when boundary and shading information are aligned. This conjunction of border and shading information may first be detected locally as the conjunction of two low-level cues for image discontinuity—a luminance border in the carrier that is coincident with a sudden change of direction of the bounding contour. This information becomes available no later than 100 ms (Fig. 3C). Sharp transitions in object borders are always associated with shadow discontinuities, but the converse is not always true: sharp luminance transitions can come about from surface markings that are independent of object border shape. Breaking up the relationship between luminance transitions on the surface of objects and their borders is an effective means of natural and artificial camouflage.

The representation of the coincidence of border and shading information is independent of perceptual state up to about 150 ms, although it is available no later than 100 ms (cf. Fig. 2A). The increased negative activity after this point during the changing depth percept may represent the additional processing overhead needed to elaborate a changing depth representation as opposed to one that is simply a translation of a rigid, three-dimensional entity.

### 4.1. Comparison to other bi-stable percepts

The dynamics of the switching between perceptual states in our experiment on bi-stability is reminiscent of that observed during prolonged inspection of moving plaids comprised of superimposed gratings drifting in different directions (Hupe & Rubin, 2003). They found that plaids appear to cohere upon initial presentation, but after a considerable delay, perception alternates symmetrically between coherent and transparent interpretations. In their

work, symmetric alternation between states only occurs after the initial, long-duration percept of the coherent state. Our effect is similar—the initial changing-3D state persists for a long duration before giving way to symmetric alternation with the lateral motion percept. Rubin and Hupe (2005) have suggested that perceptual bi-stability across very different stimuli, e.g., their motion plaids, the Necker cube, the face-vase illusion, binocular rivalry, etc., is consistent with pervasive, competitive neural mechanisms that are designed to enforce unitary interpretations of ambiguous stimuli. The use of reduced cue stimuli such as ours increases the level of ambiguity of the shading cue and thus promotes multiple perceptual interpretations. These interpretations are mutually exclusive in terms of perceived depth structure and it is thus not surprising that we observe bi-stable alternation.

#### 4.2. Comparison to other EEG/MEG-based studies of bi-stability

There have been many EEG and MEG (magnetoencephalogram) based studies of visual processing during the spontaneous reversals of perceptually ambiguous stimuli undergoing binocular rivalry (see Brown and Norcia, 1997 for review). The available studies with non-rivalrous ambiguous stimuli have concentrated on the mechanisms underlying the spontaneous switch between states during continuous viewing (Basar-Eroglu, Struber, Kruse, Basar, & Stadler, 1996; Basar-Eroglu, Struber, Stadler, Kruse, & Basar, 1993; Isoglu-Alkac et al., 1998, 2000; Muller et al., 2005; Struber & Herrmann, 2002). Of most relevance to the present study, Kornmeier and Bach (2004, 2005) have developed a method that does not involve spontaneous reversals. In their method, successive short presentations of flashed ambiguous stimuli are either perceived as consistent or inconsistent with the previously seen state. This presentation mode allows the experiment to time-lock response averaging to stimulus onset, rather than to the observer's button press. We were able to time-lock our responses because we used periodically modulated stimuli. They have found a negative component in the latency range of our enhanced negativity that was related to perceptual switching. We did not exclude data after switches, so some switch-related activity may be present in our recordings. However, our data includes switches between both states, not just the switches to one state or the other, so it is unlikely that all of the activity we attribute to perceived depth is switch-related. Moreover, the negativity we observe is more sustained than the switch-related negativity of Kornmeier and Bach.

#### 4.3. Comparison with single-unit studies

Previous single-unit studies have found neural correlates of boundary detection derived from feature differences or discontinuities in V1 (Chaudhuri & Albright, 1997; Grosf, Shapley, & Hawken, 1993; Lamme, Rodriguez-Rodriguez,

& Spekreijse, 1999; Marcus & Van Essen, 2002; Sillito, Grieve, Jones, Cudeiro, & Davis, 1995). The extra negativity evoked by the 3D image is the most consistent aspect of our responses that is attributable to depth processing per se. It is reminiscent of the pop-out response (120–320 ms) of macaque V1/V2 neurons to shape-from-shading stimuli studied by Lee and co-workers (Lee et al., 2002), although the latency and duration of our negativity (200–400 ms) are somewhat longer.

#### Acknowledgments

This work was supported by NIH EY06579 and a Rachel C. Atkinson Fellowship (CH).

#### Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at [doi:10.1016/j.visres.2005.10.017](https://doi.org/10.1016/j.visres.2005.10.017).

#### References

- Aks, D. J., & Enns, J. T. (1992). Visual search for direction of shading is influenced by apparent depth. *Perception & Psychophysics*, *52*(1), 63–74.
- Basar-Eroglu, C., Struber, D., Kruse, P., Basar, E., & Stadler, M. (1996). Frontal gamma-band enhancement during multistable visual perception. *International Journal of Psychophysiology*, *24*(1–2), 113–125.
- Basar-Eroglu, C., Struber, D., Stadler, M., Kruse, P., & Basar, E. (1993). Multistable visual perception induces a slow positive EEG wave. *International Journal of Neuroscience*, *73*(1–2), 139–151.
- Blair, R. C., & Karniski, W. (1993). An alternative method for significance testing of waveform difference potentials. *Psychophysiology*, *30*(5), 518–524.
- Braun, J. (1993). Shape-from-shading is independent of visual attention and may be a 'texton'. *Spatial Vision*, *7*(4), 311–322.
- Brown, R. J., & Norcia, A. M. (1997). A method for investigating binocular rivalry in real-time with the steady-state VEP. *Vision Research*, *37*(17), 2401–2408.
- Chaudhuri, A., & Albright, T. D. (1997). Neuronal responses to edges defined by luminance vs temporal texture in macaque area V1. *Visual Neuroscience*, *14*(5), 949–962.
- Enns, J. T., & Rensink, R. A. (1990). Influence of scene-based properties on visual search. *Science*, *247*(4943), 721–723.
- Grosf, D. H., Shapley, R. M., & Hawken, M. J. (1993). Macaque V1 neurons can signal 'illusory' contours. *Nature*, *365*(6446), 550–552.
- Humphrey, G. K., Goodale, M. A., Bowen, C. V., Gati, J. S., Vilis, T., Rutt, B. K., et al. (1997). Differences in perceived shape from shading correlate with activity in early visual areas. *Current Biology*, *7*(2), 144–147.
- Hupe, J. M., & Rubin, N. (2003). The dynamics of bi-stable alternation in ambiguous motion displays: a fresh look at plaids. *Vision Research*, *43*(5), 531–548.
- Isoglu-Alkac, U., Basar-Eroglu, C., Ademoglu, A., Demiralp, T., Miener, M., & Stadler, M. (1998). Analysis of the electroencephalographic activity during the Necker cube reversals by means of the wavelet transform. *Biological Cybernetics*, *79*(5), 437–442.
- Isoglu-Alkac, U., Basar-Eroglu, C., Ademoglu, A., Demiralp, T., Miener, M., & Stadler, M. (2000). Alpha activity decreases during the perception of Necker cube reversals: an application of wavelet transform. *Biological Cybernetics*, *82*(4), 313–320.
- Kleffner, D. A., & Ramachandran, V. S. (1992). On the perception of shape from shading. *Perception & Psychophysics*, *52*(1), 18–36.

- Kornmeier, J., & Bach, M. (2004). Early neural activity in Necker-cube reversal: Evidence for low-level processing of a gestalt phenomenon. *Psychophysiology*, *41*(1), 1–8.
- Kornmeier, J., & Bach, M. (2005). The Necker cube—an ambiguous figure disambiguated in early visual processing. *Vision Research*, *45*(8), 955–960.
- Lamme, V. A., Rodriguez-Rodriguez, V., & Spekreijse, H. (1999). Separate processing dynamics for texture elements boundaries and surfaces in primary visual cortex of the macaque monkey. *Cerebral Cortex*, *9*(4), 406–413.
- Lee, T. S., Yang, C. F., Romero, R. D., & Mumford, D. (2002). Neural activity in early visual cortex reflects behavioral experience and higher-order perceptual saliency. *Nature Neuroscience*, *5*(6), 589–597.
- Mamassian, P., Jentzsch, I., Bacon, B. A., & Schweinberger, S. R. (2003). Neural correlates of shape from shading. *Neuroreport*, *14*(7), 971–975.
- Marcus, D. S., & Van Essen, D. C. (2002). Scene segmentation and attention in primate cortical areas V1 and V2. *Journal of Neurophysiology*, *88*(5), 2648–2658.
- Muller, T. J., Federspiel, A., Horn, H., Lovblad, K., Lehmann, C., Dierks, T., et al. (2005). The neurophysiological time pattern of illusionary visual perceptual transitions: a simultaneous EEG and fMRI study. *International Journal of Psychophysiology*, *55*(3), 299–312.
- Nichols, T. E., & Holmes, A. P. (2001). Nonparametric permutation tests for functional neuroimaging: A primer with examples. *Human Brain Mapping*, *15*, 1–25.
- Nuwer, M. R., Comi, G., Emerson, R., Fuglsang-Frederiksen, A., Guerit, J. M., Hinrichs, H., et al. (1999). IFCN standards for digital recording of clinical EEG. The International Federation of Clinical Neurophysiology. *Electroencephalography and Clinical Neurophysiology Supplement*, *52*, 11–14.
- Ramachandran, V. S. (1988a). Perceiving shape from shading. *Scientific American*, *259*(2), 76–83.
- Ramachandran, V. S. (1988b). Perception of shape from shading. *Nature*, *331*(6152), 163–166.
- Rubin, N., & Hupe, J. M. (2005). Dynamics of perceptual bistability: Plaids and binocular rivalry compared. In D. Alais & R. Blake (Eds.), *Binocular rivalry* (pp. 137–154). London: Bradford.
- Sillito, A. M., Grieve, K. L., Jones, H. E., Cudeiro, J., & Davis, J. (1995). Visual cortical mechanisms detecting focal orientation discontinuities. *Nature*, *378*(6556), 492–496.
- Struber, D., & Herrmann, C. S. (2002). MEG alpha activity decrease reflects destabilization of multistable percepts. *Brain Research Cognitive Brain Research*, *14*(3), 370–382.
- Sun, J., & Perona, P. (1997). Shading and stereo in early perception of shape and reflectance. *Perception*, *26*(4), 519–529.
- Sun, J. Y., & Perona, P. (1996). Preattentive perception of elementary three-dimensional shapes. *Vision Research*, *36*(16), 2515–2529.
- Taira, M., Nose, I., Inoue, K., & Tsutsui, K. (2001). Cortical areas related to attention to 3D surface structures based on shading: An fMRI study. *Neuroimage*, *14*(5), 959–966.
- Tang, Y., & Norcia, A. M. (1995). An adaptive filter for steady-state evoked responses. *Electroencephalography and Clinical Neurophysiology*, *96*(3), 268–277.
- Timm, N. H. (2002). *Applied multivariate statistics*. New York: Springer-Verlag.