Determinants of neural responses to disparity in natural scenes

Yiran Duan

Department of Psychology, Stanford University, Stanford, CA, USA

Alexandra Yakovleva

Department of Psychology, Stanford University, Stanford, CA, USA

Anthony M. Norcia

Department of Psychology, Stanford University, Stanford, CA, USA

We studied disparity-evoked responses in natural scenes using high-density electroencephalography (EEG) in an event-related design. Thirty natural scenes that mainly included outdoor settings with trees and buildings were used. Twenty-four subjects viewed a series of trials composed of sequential two-alternative temporal forced-choice presentation of two different versions (two-dimensional [2D] vs. three-dimensional [3D]) of the same scene interleaved by a scrambled image with the same power spectrum. Scenes were viewed orthostereoscopically at 3 m through a pair of shutter glasses. After each trial, participants indicated with a key press which version of the scene was 3D. Performance on the discrimination was >90%. Participants who were more accurate also tended to respond faster; scenes that were reported more accurately as 3D also led to faster reaction times. We compared visual evoked potentials elicited by scrambled, 2D, and 3D scenes using reliable component analysis to reduce dimensionality. The disparity-evoked response to natural scene stimuli, measured from the difference potential between 2D and 3D scenes, comprised a sustained relative negativity in the dominant response component. The magnitude of the disparity-specific response was correlated with the observer's stereoacuity. Scenes with more homogeneous depth maps also tended to elicit large disparity-specific responses. Finally, the magnitude of the disparity-specific response was correlated with the magnitude of the differential response between scrambled and 2D scenes, suggesting that monocular higher-order scene statistics modulate disparity-specific responses.

Introduction

We live in a three-dimensional (3D) world and constantly and effortlessly construct 3D percepts from two-dimensional (2D) retinal images. This efficiency has been shaped over evolutionary and developmental time scales through exposure to sensory stimuli encountered in the natural environment. Natural scenes give rise to many cues for depth, some of which are monocular, but one of the most important cues is the binocular disparity cue arising from the slightly different viewpoints of the two eyes (Wheatstone, 1838). Depth induced by binocular disparity can be more compelling, robust, and immersive than depth perception induced by monocular cues (Patterson & Martin, 1992; Wheatstone, 1838). By using carefully designed stimuli such as bars, gratings, and random dot stereograms in highly controlled experimental settings, much progress has been made toward identifying perceptual and neural representations of disparity (Backus, Fleet, Parker, & Heeger, 2001; Parker, 2007; Welchman, 2016).

It has long been hoped that the insights gained from reduced stimuli in controlled experimental settings will generalize to an understanding of responses measured in complex natural viewing situations (Carandini et al., 2005; Felsen & Dan, 2005). However, both neural and perceptual results suggest that natural image responses are not readily predictable from responses to simple stimuli. For example, computational models based on cell properties derived from simple stimuli typically explain only 30% to 40% of the variance of responses to natural scenes (David, Vinje, & Gallant, 2004). The sensitivity of complex cells to the presence of their preferred features is higher in natural images

Citation: Duan, Y., Yakovleva, A., & Norcia, A. M. (2018). Determinants of neural responses to disparity in natural scenes. *Journal of Vision*, *18*(3):21, 1–19, https://doi.org/10.1167/18.3.21.

https://doi.org/10.1167/18.3.21 Received May 30, 2017; published March 29, 2018 ISSN 1534-7362

than in random stimuli, and this is not predicted by a standard model of complex cells (Felsen, Touryan, Han, & Dan, 2005). The response properties of higher visual areas are likely to be even more closely associated with the characteristics of natural stimuli (Tanaka, 1996).

Turning to the case of binocular vision, stereoacuity measurements using real depth instruments (e.g., the Howard-Dolman apparatus and the Frisby Stereo Test) often yield better thresholds than those using simulated depth (Howard, 1919; McKee & Taylor, 2010; Zaroff, Knutelska, & Frumkes, 2003). Natural image stimuli dominate artificial stimuli in perceptual rivalry even when the images are matched for contrast, luminance, and energy (Baker & Graf, 2009). The classic repetition-related change in hemodynamic response for 2D planar images is surprisingly weaker when observers are viewing real-world 3D objects (Snow et al., 2011). These discrepancies may arise because natural vision, and especially stereoscopic vision, differs from the situation of reduced cue experiments in at least three aspects: (a) "Pure disparity" does not exist in the natural environment where depth cues other than disparity are also available. (b) Visual stimuli are rarely presented in isolation, and the brain activation from a complex scene may not be a linear summation of the activation of individual simple components (such as spots, lines, edges, surfaces) in the scene. (c) Viewing devices such as stereoscopes can create cue conflicts (Hoffman, Girshick, Akeley, & Banks, 2008; Howard, 1919; McKee & Taylor, 2010).

Therefore, to understand disparity processing in the context of everyday life, we need to present our visual system with ecologically relevant natural images. Although this might be considered challenging because of the relative lack of stimulus control over the multiple depth cues, real-world scenes are actually highly regular and thus exploitable in laboratory studies (Geisler & Diehl, 2002; Simoncelli & Olshausen, 2001). Surprisingly, only a handful of studies have studied neural correlates of stereopsis using naturalistic images. The first such study used electroencephalography (EEG) and source localization to identify brain areas relevant to depth perception in natural images (Fischmeister & Bauer, 2006). In agreement with functional imaging studies in humans, they observed higher activations in the parietal cortex extending into occipital regions while processing binocular disparity cues. They also demonstrated the feasibility of adopting more realistic alternatives to stimuli based solely on one type of depth cues. More recently, a functional magnetic resonance imaging study has shown that binocular disparity increases intersubject correlations of brain networks and enhances the experience of immersion when viewing complex 3D movies (Gaebler et al., 2014). Furthermore, visual search task response times in a complex natural space are significantly shorter when binocular depth information is available, with area V3A showing greater activation during search tasks containing binocular cues (Ogawa & Macaluso, 2015).

In the present study, we used high-density EEG recordings and an event-related design to compare responses evoked by 2D natural and scrambled images to determine how natural scene structure affects the disparity-specific spatiotemporal response distribution. We then compared these responses to those generated by matching intact natural images presented in 3D to determine how the simple addition of disparity modulates the response to natural scenes. Finally, based on a growing body of psychophysical evidence describing a large range of individual difference in stereoacuity in the normal population (Bosten et al., 2015; Howard, 1919), we compared differential 3D versus 2D response across participants. We found robust differences in spatiotemporal response patterns between each of the three levels of image structure (random 2D, natural 2D, natural 3D). By analyzing responses to individual scenes, we found a correlation between the structure of the scene depth map and the magnitude of disparity response elicited. By analyzing 2D versus 3D differential responses in individual observers, we found a correlation between brain responses and perceptual sensitivity to disparity. In addition, the higher-order image structure in natural scenes modulated disparity responses both for scenes and observers.

Methods

Participants

Twenty-four healthy adults (13 men) aged between 18 and 33 years (mean = 24.3 years) participated in this study. All participants had normal or corrected-to-normal visual acuity, and the average logMar visual acuities of their left and right eyes were each 0 (corresponding to a Snellen acuity of 20/20). They reported no difficulty perceiving stereoscopic depth when viewing 3D pictures, and their average stereoacuity as measured by Randot® stereotest (Stereo Optical, Inc., Chicago, IL) was 27.06 arcsec. The study was approved by the Stanford University Institutional Review Board, and all participants gave written informed consent prior to the experiment. The procedure is in accordance with the Declaration of Helsinki.

Figure 1. Thumbnails of experimental stimuli.

Stimulus presentation and trial structure

Thirty high-quality stereo image pairs of outdoor scenes from McCann (2015) were used. Briefly, each image pair was collected using two camera station points spaced 65 mm apart, mimicking the average distance between the two eyes of adult males (Dodgson, 2004). The corresponding depth map for each image was obtained using a scanning laser range finder. The outdoor scenes included trees, lawns, buildings, signs, and fences. All images were resampled to a resolution of 1,920 pixels width × 1,080 pixels height. The images were presented at a 3-m viewing distance, and this resulted in images of natural size (orthostereoscopic presentation) with minimal conflict between vergence and accommodation. A scrambled version of each scene was generated by applying the Portilla-Simoncelli algorithm (http://www.cns.nyu.edu/~lcv/texture/) to the monocular half images. This procedure produces an image without recognizable content but with an identical power spectrum and second-order correlations over locations, scales, and orientations (Portilla & Simoncelli, 2000). A comprehensive description of the natural-scene capture pipeline can be found in Burge, McCann, and Geisler (2016). Thumbnails of the 30 images used can be seen in Figure 1.

Image pairs were presented using in-house software on a Sony Bravia (model XBR-65HX929) 3D TV (143.4 × 80.7 cm) at a resolution of 1,920 × 1,080 pixels. Active shutter glasses were used to present separate images to each eye: They were either a stereo-pair in the 3D condition or copies of the right eye image in the two 2D conditions. A single trial was defined as a sequence of four epochs, during which two versions of the same scene (2D vs. 3D) were presented sequentially, with each natural scene image preceded by a scrambled version of the that image (see Figure 2). The scrambled images are always pairs of 2D scrambled versions of the right eye images. During each trial, a cross was placed in the center of the scrambled image at the plane of the screen when viewed stereoscopically. Stereo disparities of the intact images were rendered behind the fixation cross. The participant was instructed to maintain fixation on the cross and to reduce blinks and

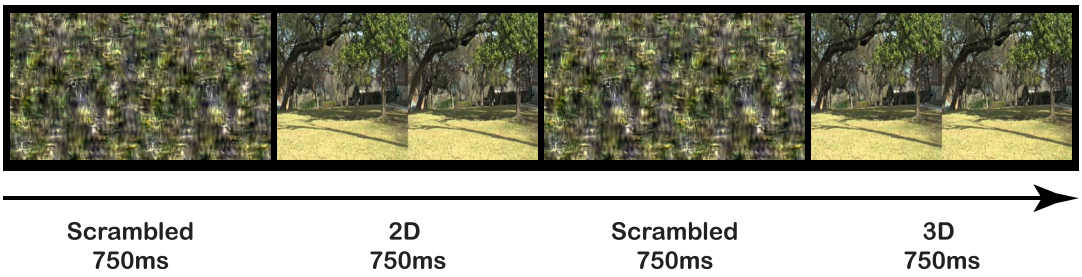

Figure 2. Experiment trial structure. The 2D images display identical image pairs to left and right eyes. The 3D images display 3D stereo pairs to left and right eyes. A block consisted of 60 trials, with each scene showing up twice. A uniform gray background was displayed between each trial.

movements to a minimum. Each image presentation epoch lasted 750 ms, and the overall duration of a trial was 3 s. A block consisted of 60 trials, in which each version of a scene was shown twice in the two different sequences (2D first or 3D first). A total of four blocks of trials was administered to each participant, and the scene presentation order within each block was randomized. A two-alternative temporal forced-choice procedure was used in which the participants were instructed to respond with a button press to indicate whether the second or fourth stimulus epoch contained a 3D image.

EEG acquisition and preprocessing

The EEG data were collected using 128-channel HydroCell Geodesic Sensor Nets and a NetAmps 400 system (Electrical Geodesics Inc., Eugene, OR). The EEG was bandpass filtered from 0.1 to 50 Hz and digitized at a rate of 432 Hz (Net Amps 400 TM, Electrical Geodesics). Individual electrodes were adjusted until impedances were below 50 kΩ before starting the recording. Artifact rejection was performed offline according to a sample-by-sample threshold procedure to remove noisy electrodes and replace them with the average of the six nearest neighboring electrodes. On average, less than 5% of the electrodes were substituted; these electrodes were mainly located near the forehead or the ears, and substituting them is unlikely to affect our results. The EEG was then re-referenced to the common average of all the remaining electrodes. Epochs with more than 15% of the data samples exceeding 30 μV were excluded on a sensor-by-sensor basis. Typically, these epochs included movements or blinks. A 131-ms delay between the onset of EEG recording and the stimulus onset caused by the EEG recording system (66 ms) and the BRAVIA monitor (65 ms) has been corrected in analysis.

Statistical analysis

Reliable-component analysis

Reliable-component analysis (RCA) is a newly developed technique that aims to combine electrode potentials linearly to reduce the dimensionality of high-density EEG data and to identify distributed sources of neural activity (Dmochowski, Greaves, & Norcia, 2015; Dmochowski, Sajda, Dias, & Parra, 2012). RCA is based on the fundamental assumption underlying evoked responses, namely, that the signal of interest is spatiotemporally reproducible across trials (Dmochowski & Norcia, 2015). RCA works by obtaining a linear spatial filter $\mathbf{W}$ by explicitly maximizing the ratio of across- to within-trial covariance, for example,

$$\text{argmax } \rho(W) = \frac{W^T R_{across} W}{W^T R_{within} W}$$

where R_{within} denotes the within-trial covariance matrix and R_{across} denotes the across-trial covariance matrix. The solution is known to be a conventional eigenvalue problem:

$$(R_{across}^{-1} R_{within}) W = \frac{1}{\rho} W$$

When performing the eigenvalue decomposition, we regularized the within-trial pooled covariance by keeping only the first K dimensions. In the present data, $K = 6$ corresponded to the "knee" of the eigenvalue spectrum representation of R_{within}. The bulk of the across-trial reliability is captured in the first C dimensions., where C is much less than the number of electrodes. The proportion of reliability explained by the first C reliable components (RCs) can be quantified by the following measure:

$$\eta(C) = \frac{\sum_{i=1}^{c} \lambda_i}{\sum_{i=1}^{D} \lambda_i}$$

The first C column of the weight matrix $\mathbf{W}$ is then used to project the original EEG data into the component space.

Data from all 24 subjects were considered for learning the RCs. Because the overall accuracy was greater than 90%, both correct and incorrect trials were used. For each trial, we extracted the first 1,500 ms of the recording corresponding to the presentation of a scrambled scene followed by the first presentation of an intact natural scene, with the 2D or 3D scene versions being equiprobable. For the purposes of the current study, the second half of each trial was not analyzed because that neural response may be subject to a priming effect that does not occur for the first image presentation (i.e., when the subject saw a 2D image during the first interval, he or she could expect to see a 3D image during the second interval). RCs were learned for scrambled images during the initial 750-ms interval (5,760 trials) and separately for the 2D and 3D images presented in the second 750-ms interval (2,880 trials each).

Waveform permutation testing

To compare the waveforms between different experimental conditions, we projected the sensor data averaged across all the trials within each condition onto the first three spatial filters maximizing reliability over the separate 2D, 3D, or scrambled trials. Differences between the resulting waveforms were identified by a permutation test devised by Blair and Karniski (1993) and described in detail in Appelbaum, Wade, Vildavski, Pettet, and Norcia (2006). Specifically, to determine at which time points the RC amplitudes differ between 2D and 3D scenes, the differences between 2D and 3D were calculated for each subject at each time point, resulting in a $m \times n$ difference matrix $\mathbf{Y_\Delta}$, where m is the number of time points and n is the number of subjects. The mean and variance across subjects are denoted as $\boldsymbol{\mu_\Delta}$ and σ^2_Δ. Then, a vector of t scores was obtained through the following statistic:

$$t = \frac{\mu_\Delta}{\sqrt{\frac{\sigma^2_\Delta}{n-1}}}$$

From the above vector, we determined the longest consecutive sequence of t scores having a p value <0.05, and this longest sequence is denoted as t_L. If there are no differences between the experimental conditions, then the sign of the difference between 2D and 3D at each time point would be positive or negative in a random fashion. Therefore, we can simulate the distribution of the difference matrix under the null hypothesis by randomly permuting the signs of the columns of $\mathbf{Y_\Delta}$. Considering 10,000 permutations of signs for the columns $\mathbf{Y_\Delta}$, we accumulate a permutation sample space of $\mathbf{Y^*_\Delta}$ and a nonparametric reference distribution for $\mathbf{t^*_L}$. The critical value t_C is then determined by the top 5% cutoff in the reference distribution of $\mathbf{t^*_L}$. We reject the null hypothesis if the length of any consecutive sequence of significant t scores in the original, non-randomized data exceeded t_C (Appelbaum et al., 2006). Because each permutation sample contributes only its longest significant sequence to the reference distribution, this procedure implicitly compensates for the problem of multiple comparisons and is a valid test for the omnibus hypothesis of no difference between the waveforms at any time point. Furthermore, this test not only detects significant departures from the null hypothesis but also localizes the time periods when such departures occur. However, because the correction procedure is tied to the length of the data and the somewhat arbitrary choice of keeping familywise error at 5%, we therefore also present the uncorrected significance values visualized as red to yellow color maps in the figures. By evaluating the data using both statistical approaches, we are better able to identify time periods when the responses depart from the null hypothesis.

Spatial topographies

To depict the spatial topographies of the RCs, we examined the scalp projection of the activity recovered by the filters. Specifically, let $\mathbf{W}$ denote a matrix whose columns represent the weight vectors generated by RCA. The projections of the recovered sources onto the sensor data are given by $A = R_{within}W(W^TR_{within}W)^{-1}$. The columns of A represent the pattern of electric potentials that would be observed on the scalp if only the source signal recovered by $\mathbf{W}$ was active, informing us of the approximate location of the underlying neuronal sources (Haufe et al., 2014; Parra, Spence, Gerson, & Sajda, 2005).

Quantification of 2D versus 3D spatiotemporal differences

Each individual scene or subject can exhibit or elicit a different spatiotemporal response pattern under 2D versus 3D conditions. To quantify these differences in component space, for each scene (subject), we averaged trials within each condition (96 trials for each scene and 120 trials for each subject) and used the weights of the first RC component to project the 128-channel sensor data from the 2D and 3D responses onto the dimension-reduced component space. The response difference between the 2D and 3D stimuli—the disparity response—was quantified by the Euclidean distance between the waveforms. Waveform permutation testing was performed to localize the time period when significant differences occurred.

Multiple regression analysis

A multiple regression model was used to explore the underlying factors related to the individual scene and subject differences in disparity responses. For individual scenes, we determined whether response time, response amplitude to low-level image statistics, response amplitude to the high-level scene structure (the "sceneness" of the stimuli), and median and standard deviation of the corresponding depth map of each scene affect how different its 2D versus 3D responses are. The response amplitude driven by low-level image statistics was quantified by as the Euclidean distance between the corresponding waveform generated by the onset of the scrambled image and the zero vector. The response to the higher-order image structure that defines sceneness was quantified as the differential response between the 2D natural scene response and the scrambled image response that preceded it, again through the Euclidean distance between the waveforms. For individual subjects, the independent variables included age, stereoacuity, response time for each scene, response amplitude to the scrambled version of the scene, and the sceneness metric of the stimuli.

Results

The goal of this study is to understand how retinal disparity in natural scenes modulates neural responses. By employing the RCA methods described above, we reduced the high-dimensional sensor-space data to a small set of components that correspond to the most reliable cortical sources of stimulus-related activity.

Behavioral results

The accuracy for discriminating 2D from 3D images in the two-alternatives force-choice task was greater than 90%, with a mean response time of 2,981 ms from the onset of the initial scrambled image. Eleven participants responded during the presentation of the fourth interval (2,250 to ~3,000 ms), and 13 responded after presentation of the last image (>3,000 ms). Although not instructed to respond quickly, a significant negative correlation was found between the participants' response time and accuracy averaged over scenes ($r = -0.65$, $n = 24$, $p < 0.001$); participants were faster when they were more accurate. When looking at the reaction time and accuracy for each scene averaged across participants, again, a significant negative correlation was found ($r = -0.66$, $n = 30$, $p < 0.001$). Scenes that were reported more accurately as 3D also led to faster response times.

RC selection

Averaged across 2D and 3D responses, the descending eigenvalues corresponding to the first five RCs were 0.093, 0.068, 0.025, 0.021, and 0.015, and the reliability explained by the first three RCs was collectively 80.89% (Figure 3). Consequently, we chose to retain the first three RCs but focus primarily on the first RC component for individual differences and scene-level analysis as each of these subanalyses used a smaller fraction of the entire data.

2D versus scrambled natural scene responses

We first analyzed the difference in evoked response between 2D natural scenes and their scrambled versions. Any differences in these responses can be attributed to the high-order statistical regularities present in natural scenes. The evoked response waveforms and their corresponding spatial topographies for the first three RC components (RC1, RC2, RC3) are shown in Figure 4. Each RC component has a characteristic time course and topography, and significant differences between scrambled and intact natural images were present for each RC. These differences emerged at different time points, as reflected by the asterisks in the red/yellow bars on the horizontal axis.

RC1 was maximal at midline posterior electrodes over early visual cortex for both scrambled (Figure 4b) and intact 2D scenes (Figure 4c), with the response to 2D scenes being located slightly more anteriorly. Response amplitude increased sharply around 50 ms after the stimulus onset for both scrambled and intact 2D images. The 2D response reached an initial peak 20 ms later than the response for the scrambled images (105 ms vs. 85 ms). After this initial response peak, the response for the 2D image comprised a sustained deviation from zero, whereas the response to the scrambled image was more transient, as reflected by the decrease in response starting around 220 ms after stimulus onset. The response to scrambled images was significantly larger during 125 to 240 ms but dropped quickly after the initial peak, resulting a switch over of the two curves at 290 ms, where the differences were significant between 325 and 750 ms.

Both of the RC2 topographies showed maxima over the medial frontocentral cortex, with the topography of the scrambled image extending more toward the anterior-posterior direction (Figure 4e) and the topography of the 2D image extending more laterally (Figure 4f). Both time courses for RC2 showed a negativity at 100 ms followed by a positive peak at 200 ms. The response for scrambled scene went back to baseline after the second peak, whereas the response for the 2D scene sustained to the end of the stimulus presentation.

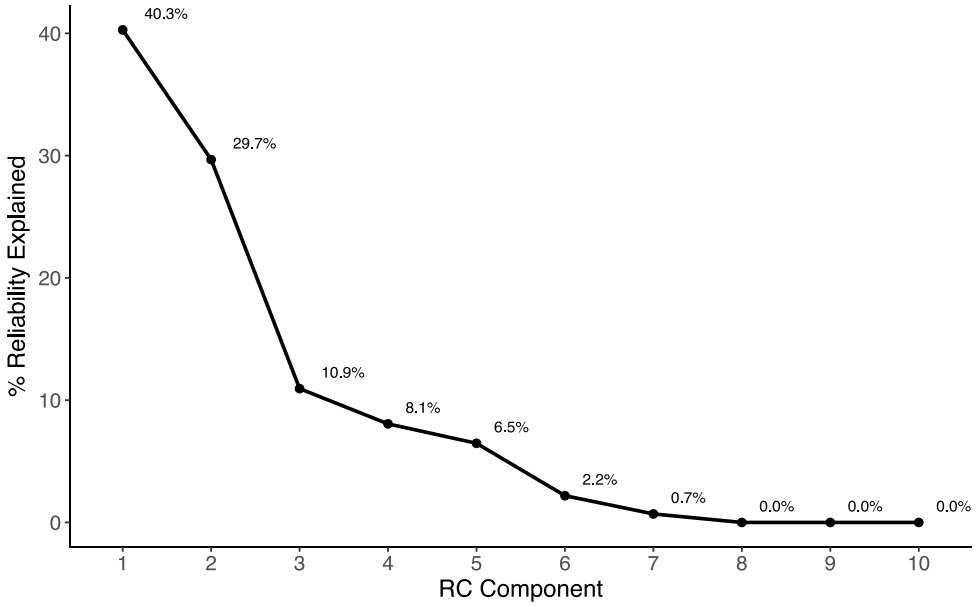

Figure 3. A scree plot of percentage of reliability explained by the first 10 RC components. The first three RC components explained 40.28%, 29.67%, and 10.94% trial-to-trial reliability, respectively, and the subsequent components explained decreasingly smaller proportions of the reliability.

The 2D response was significantly larger than that of scrambled scene between 315 and 750 ms (Figure 4d).

The RC3 topography of the scrambled image was maximal over the medial-parietal electrodes (Figure 4h), whereas that for the 2D images was right lateralized (Figure 4i). The RC3 waveform for scrambled images comprised a multiphasic pattern at about 70 to 170 ms that was not present for the 2D image. The response to the 2D scenes was significantly more negative between 320 and 480 ms (Figure 4g).

2D versus 3D natural scene responses

The main interest of the current study is to measure differences in the neural response to 2D versus 3D natural scenes. The first three RC waveforms and their corresponding spatial topographies are shown in Figure 5. The spatial distribution of 2D responses associated with the first component, replotted from Figure 4, peaked over the occipital-parietal cortex (Figure 5b). The corresponding maximum of the topography for the 3D scene response was displaced posteriorly (Figure 5c). Both RC1 and RC2 waveforms differed significantly between 2D and 3D conditions, starting at different time points, whereas RC3 did not. Waveform comparisons for RC1 showed that the onset of differential 2D versus 3D responses started at 95 ms, about 45 ms after an initially identical pattern of amplitude increase starting at 50 ms (Figure 5a). Compared with the response to 2D scenes, the response to 3D scenes initially peaked at 95 ms versus 105 ms. The differential response comprises a relative negativity for 3D versus 2D scenes that continues throughout the 750-ms image presentations. Activity of RC1 significantly discriminated 2D and 3D responses throughout the extended time period after the onset of the differential response.

The RC2 topographies differed in terms of the location of their positive maximum, with the maximum response to 2D images being located more anteriorly than that for 3D images (Figure 5e, f). RC2 also showed significant differences between the two conditions. In terms of response dynamics, there is a continuous, ramplike increase in amplitude for 3D response after stimulus onset, peaking at about 450 ms (Figure 5d). By contrast, the 2D response had a sharp negative peak at about 100 ms. Although the waveform differences from 50 to 150 ms did not pass the multiple comparison correction, we believe the effect is still of interest. The permutation test run correction for significant differences is good for detecting sustained differences but may not be powerful enough to detect transient differences. Following this negative peak, the amplitude of 2D response increased in a ramplike fashion, but the response was significantly smaller than for the 3D response from 240 to 490 ms.

The topography of RC3 showed right-hemisphere lateralization (Figure 5h, i), but waveforms did not differ between the two conditions (Figure 5g).

From the overall picture of the three RC components, it was evident that the largest differences between experimental conditions were captured by the first few RCs, as the waveform differences gradually

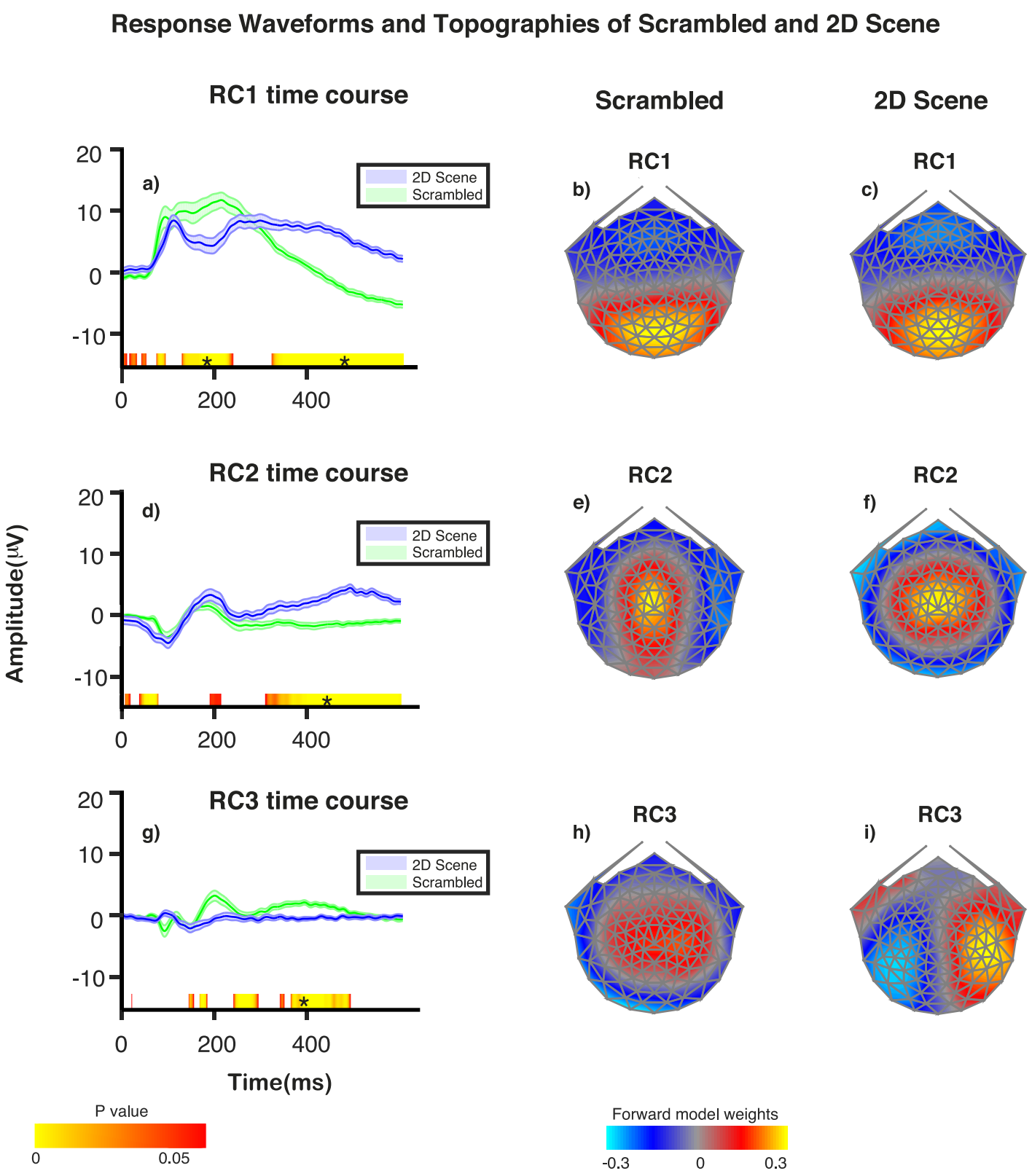

Figure 4. Comparison of intact 2D (blue) and scrambled (green) image onset responses. The pale blue and green lines plot the standard error of the mean. The RC waveforms were compared between the two conditions by a waveform permutation test, with the red/yellow bars on the horizontal axis indicating the time points at which a significant difference occurred, uncorrected for multiple comparisons. Differences surviving multiple comparison correction are reflected by the asterisks ($N = 24$, $p < 0.05$, corrected). RC1: The temporal dynamics differed between scrambled and 2D scenes during 125 to 240 ms and 325 to 750 ms. Both spatial topographies exhibited poles over the occipital cortex. RC2: The temporal dynamics differed from 315 to 750 ms. The RC2 topographies both showed increased activity over the medial frontocentral cortex, with the topography of the scrambled image extending more toward the anterior-posterior direction and the topography of the 2D image extending more laterally. RC3: The temporal dynamics differed from 320 to 480 ms. The topography of the 2D scene showed a right hemisphere lateralization.

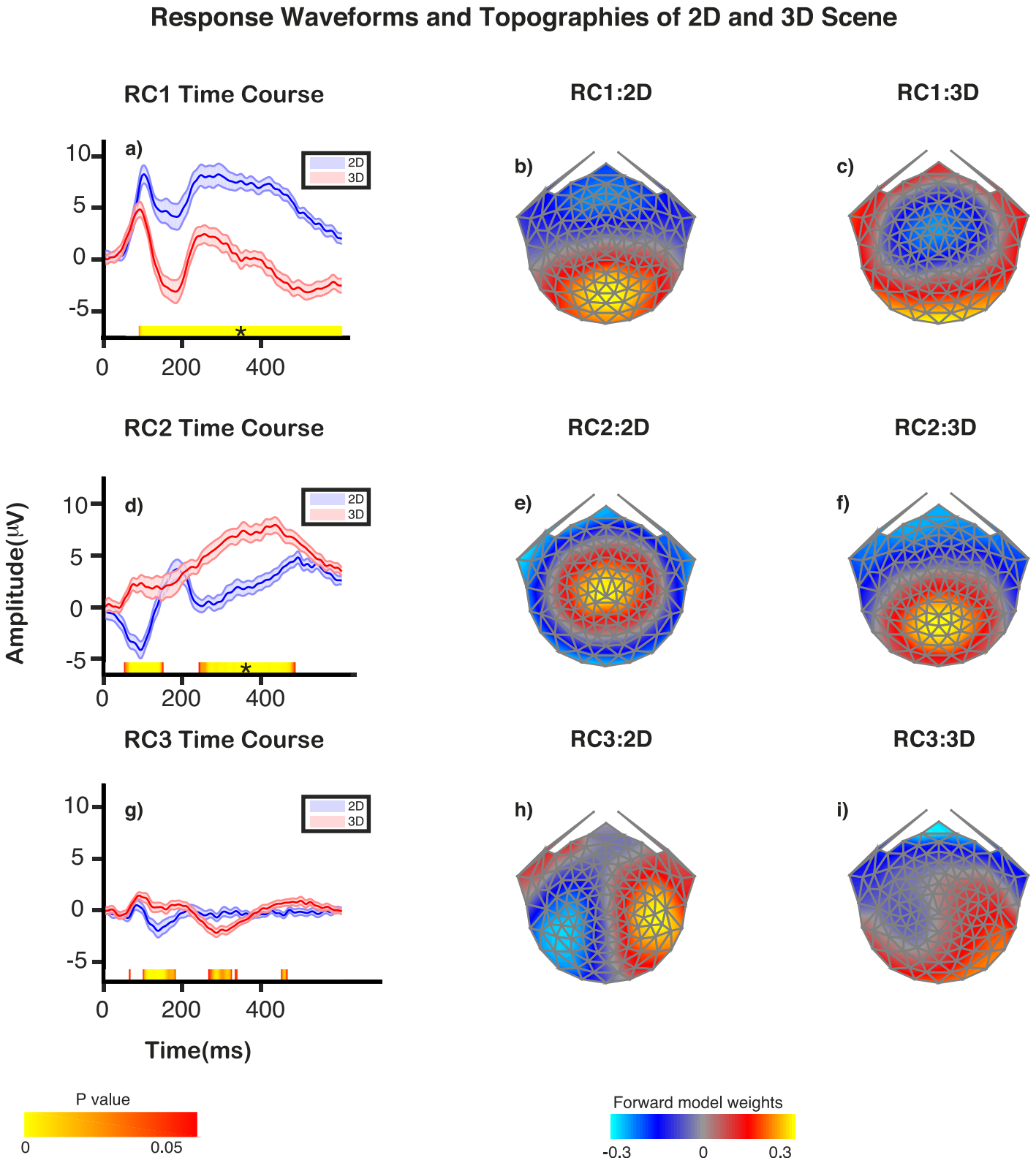

Figure 5. Comparison of intact 2D (blue) and intact 3D (red) image onset responses. The pale blue and red lines plot the standard error of the mean. The RC waveforms were compared between the two conditions by waveform permutation tests. The yellow/red bars on the horizontal axis indicate the time points at which a significant difference occurred, uncorrected for multiple comparisons. Runs of significant values surviving multiple comparison correction are indicated by the asterisks ($N = 24$, $p < 0.05$, corrected). RC1: The temporal dynamics differed between 2D and 3D scenes as early as 95 ms. Spatial topographies exhibited poles over the occipital-parietal cortex, with that of 3D scenes being more posteriorly displaced. RC2: The temporal dynamics differed from 240 to 490 ms. Both RC2 topographies showed increased activity over the medial-parietal cortex, with the topography of the 3D image more toward the occipital-parietal junction. RC3: The temporal dynamics did not differ between conditions. The spatial topographies for both conditions showed a right-hemisphere lateralization.

decreased in magnitude as we proceed further down the RC spectrum (see Figure 8 for a direct comparison of the waveforms of the scrambled, 2D, and 3D image responses).

Scene-level differences

Each scene differs not only in terms of its 2D layout and content but also in its 3D depth structure. It is thus natural to ask whether the magnitude of the disparity response elicited by different scenes depends on the content of the scene. To answer this question, we projected the sensor data into the space of the first RC separately for each scene. The magnitude of the disparity response was quantified by the Euclidean distance between each 2D and 3D waveform. Figure 6 ranks each scene according to the distance calculated. A substantial number of the 30 scenes showed significant run-corrected differences between the 2D and 3D responses. There was a spectrum of disparity-specific response magnitudes, with the strongest response being five times as large as the weakest response in terms of the Euclidean distance metric.

A multiple linear regression was then performed to explain the variations of the differential disparity response (Euclidean distance between 2D and 3D responses per scene) based on scene response time, scene response amplitude to low-level image statistics (the scrambled scene response magnitude), the sceneness of the scene (the difference between scrambled and 2D image responses), and median and the standard deviation of the depth maps of each scene. A significant regression equation was found, $F(4, 25) = 2.822$, $p = 0.038$, with an adjusted R^2 of 0.24, where the standard deviations of the depth maps and sceneness were each significant predictors of disparity response magnitude. Less variability in the depth map and stronger response to sceneness predicted larger disparity responses ($p = 0.044$ and 0.004, respectively, see Table 1).

Individual differences

Psychophysical studies have reported a substantial range of individual differences in stereoacuity for simple stimuli, with the distribution of stereoacuity being positively skewed (Bosten et al., 2015; Coutant & Westheimer, 1993; Howard, 1919). Our participants had a range of "high-grade" stereopsis on the Rand-dot test that quantifies stereoacuity on the basis of a graded set of disparate circle targets. Among the 24 subjects, 18 showed relatively large differences between the 2D and 3D response amplitude that survived the run-corrected significance criterion (Figure 7). The rest showed very similar responses for both conditions, and their responses did not differ reliably between 2D and 3D natural scenes. Among those who showed differences between 2D and 3D conditions, there is also a spectrum of differences. For example, compared with subject 1147, who showed the biggest difference between 2D and 3D response, subject 1308 had a difference three times smaller in terms of the Euclidean distance metric.

A multiple linear regression was performed to explain the variations of the disparity response based on age, stereoacuity, response time, response amplitude to low-level image statistics, and the sceneness of the stimuli. A significant regression equation was found, $F(5, 18) = 4.922$, $p = 0.005$, with an adjusted R^2 of 0.46, where the stereoacuity and the sceneness were significant predictors of disparity response magnitude. Lower stereoacuity threshold and a stronger response to sceneness predicted a larger disparity response ($p = 0.038$ and 0.013, respectively, see Table 2). Sceneness was thus a significant predictor of both scene- and individual-level differences in disparity responses magnitude. Parameters related to disparity, per se (depth map variability), and stereoacuity were also predictive but to a lesser degree.

Discussion

The neural basis of disparity processing has been almost exclusively studied using artificial stimuli such as random-dot stereograms (RDS) or stimuli based on gratings, shading, or perspective alone. In the present study, we have extended these previous findings by using stereoscopic natural scenes. Our high-density EEG recordings allowed us to measure ensemble neuronal responses to natural scenes and to identify three sources that are sensitive to the presence of disparity to images of natural scenes. Moreover, we showed that the magnitude of the disparity-specific responses depends on the degree to which the high-level structure of the 2D scene itself elicits a differential response. In the following discussion, we will mainly focus on the interpretation of the first RC that summarizes more than 40% of the trial-to-trial reliability. The second component and its possible implications will be discussed at the end.

Natural scene content modulates the evoked response

Although images can be compactly described by the amplitude and the phase of their Fourier spectrum, the "content" of an image that results in its meaningful appearance is mainly determined by its phase structure (Field, 1987; Morgan, Ross, & Hayes, 1991; Oppen-

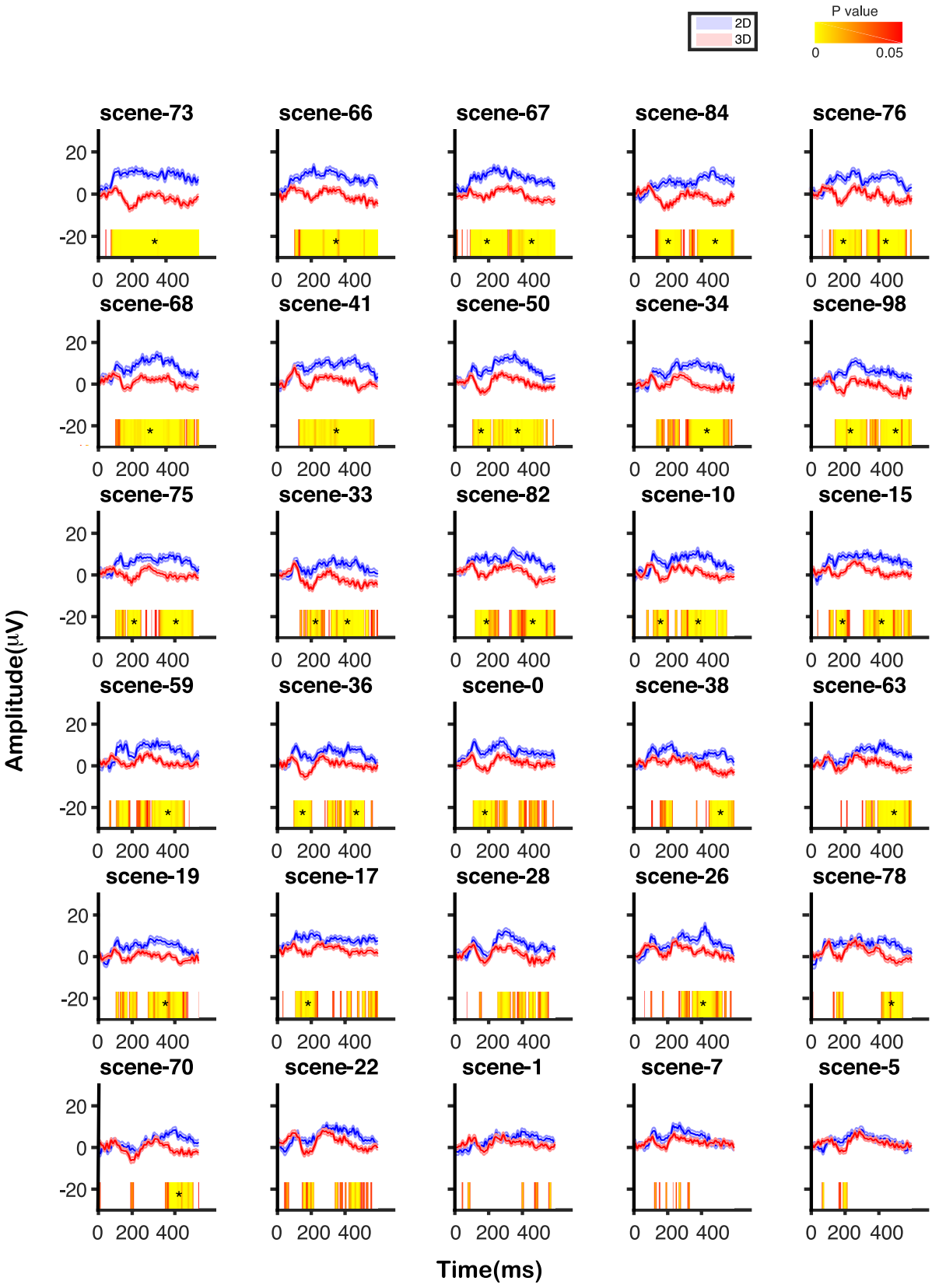

Figure 6. Individual scene differences in eliciting 2D and 3D neural responses. The pale blue and red lines plot the standard error of the mean. Sensor data were projected to the first RC space for each scene. The disparity response was quantified by the Euclidean distance between each 2D and 3D waveform, and scenes were ranked according to the magnitude of the distance and plotted in order from top left to bottom right. The RC waveforms were compared between the two conditions by waveform permutation test. The red/yellow bars on the horizontal axis indicate the time points at which a significant difference occurred, uncorrected for multiple comparisons. Those survived multiple comparison correction are reflected by the asterisks (n trials = 97, $p < 0.05$, corrected). Note that the order of the scenes presented in Figure 1 has been arranged to correspond to the order presented in Figure 6.

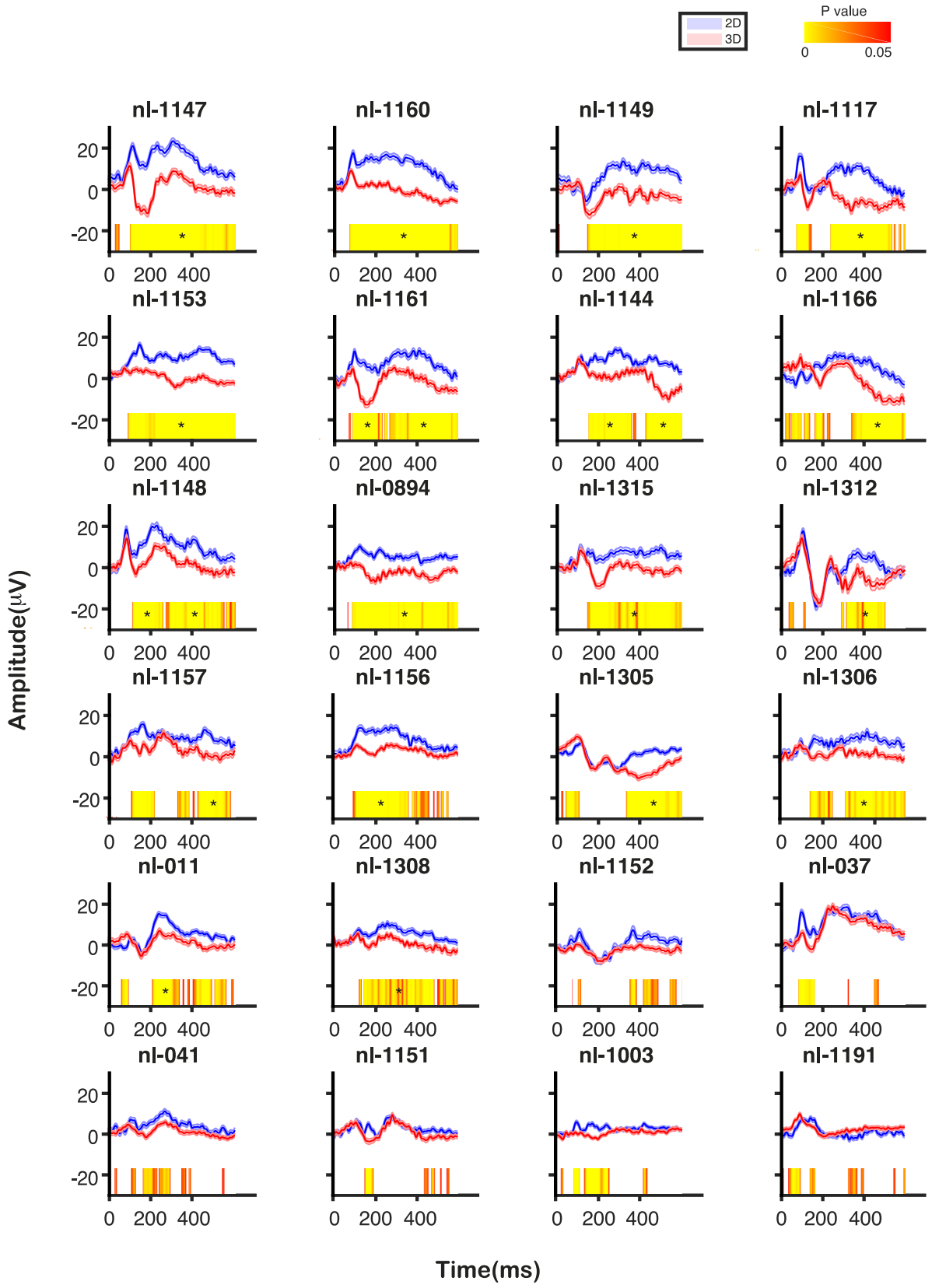

Figure 7. Individual subject differences in perceiving 3D natural scenes. The pale blue and red lines plot the standard error of the mean. Sensor data were projected to the first RC space for each individual. Individual differences in disparity response were quantified by the Euclidean distance between each 2D and 3D waveform. Participants' data were ranked according to the magnitude of the distances, and the waveforms are plotted in order from top left to bottom right. The RC waveforms were compared between the two conditions by waveform permutation test. The red/yellow bars on the horizontal axis indicate the time points at which a significant difference occurred, uncorrected for multiple comparisons. Those survived multiple comparison correction are reflected by the asterisks (n trials $= 120$, $p < 0.05$ corrected).

Figure 8. An example of a monocular image and its corresponding depth map. It can be observed that the two images appear to be similar in contours and structures.

heim & Lim, 1981). The evoked response after the onset of the first image in our trial sequence—a 2D scrambled image—from a uniform gray background reflects the population response to the low-level spectral features of a particular scene and a set of second-order statistics. The response after the onset of a 2D image from the baseline of its scrambled counterpart is driven by a combination of local contrast change and the introduction of higher-order structure.

The higher-order content of the natural scene is resolved no later than 125 ms, when the two waveforms of the dominant RC1 component diverge after an initial period of 50 ms of common activity. After this time point, the intact scene response is more sustained than the scrambled image response, suggesting continued processing of image structural content and monocular depth cues. Although no studies have directly compared the neural responses between scrambled and intact images of natural scenes, responses to other categories of natural images, such as faces, have shown the effect of high-level phase information on brain activation patterns. For example, Bieniek, Pernet, and Rousselet (2012) found that early event-related potentials (ERPs) to faces and objects are due to phase information, with almost no contribution from the amplitude spectrum. Similarly, Rousselet, Pernet, Bennett, and Sekuler (2008) manipulated phase information systematically along a continuum in a face discrimination task, and they found the mean ERP was modulated strongly by the level of integrity of the image phase information. Although a few other studies have found correlations between phase information and behavioral task performance (Baker, Yoonessi, & Arsenault, 2008; Emrith, Chantler, Green, Maloney, & Clarke, 2010; Joubert, Rousselet, Fabre-Thorpe, & Fize, 2009), we did not find such correlation in the current study, possibly because of the ceiling effect of the accuracy and reaction time.

Coefficients	Estimate	Standard error	*t* Value	Pr(>|*t*|)
(Intercept)	327.63	240.04	1.37	0.18
Response to scrambled scene (μV)	0.09	0.51	0.17	0.86
Response time (ms)	−0.11	0.08	−1.34	0.19
Response to sceneness (μV)	1.13	0.36	3.15	0.004**
Median of depth map	0.62	0.97	0.64	0.53
Standard deviation of depth map	−1.42	0.67	−2.12	0.044*

Table 1. Multiple regression analysis tablet for individual scene differences in terms of disparity response. *Notes*: Standard deviation of the depth maps and the "sceneness" are significant predictors of disparity response magnitude. Residual standard error: 25.22 on 24 *df*. Multiple R^2: 0.3702, adjusted R^2: 0.239. *F*-statistic: 2.822 on 5 and 24 *df*, *p*-value: 0.0384. * denotes $p < 0.05$; ** denotes $p < 0.005$.

Coefficients	Estimate	Standard error	*t* Value	Pr(>|*t*|)
(Intercept)	167.95	98.40	1.70	0.10
Age	−3.36	1.99	−1.69	0.11
Stereoacuity (arcsec)	−0.95	0.43	−2.24	0.038*
Response time (ms)	−0.003	0.02	−0.14	0.89
Response to scrambled scene (μV)	−0.14	0.23	−0.61	0.55
Response to sceneness (μV)	0.69	0.25	2.75	0.013*

Table 2. Multiple regression analysis output for individual subject differences in terms of disparity response. *Notes*: Stereoacuity and the "sceneness" are significant predictors of disparity response magnitude. Residual standard error: 37.16 on 18 *df*. Multiple R^2: 0.5775, adjusted R^2: 0.4602. *F*-statistic: 4.922 on 5 and 18 *df*, *p* value: 0.005166. * denotes $p < 0.05$.

Disparity structure modulates the evoked response

The differential visual evoked potential (VEP) responses between 2D and 3D natural scenes can be attributed to the responses related to the processing of binocular disparity cues because this is the only stimulus attribute that differs between conditions. The response to disparity measured here is broadly similar to that measured with dynamic random-dot stereograms (DRDS). Studies based on DRDS have reported an onset latency of typically about 100 ms (Fahle, Quenzer, Braun, & Spang, 2003; Lehmann & Julesz, 1978; Michel, Henggeler, & Lehmann, 1992; Neill & Fenelon, 1988; Regan & Spekreijse, 1970; Şahinoğlu, 2004), consistent with the onset time of 95 ms found in the current study. One of the earliest studies of evoked cortical responses to DRDS (Lehmann & Julesz, 1978) found a negative-going response in the hemisphere ipsilateral to the hemiretina of stimulation. Several subsequent studies confirmed this relative negativity associated with disparity-specific responses (Fahle et al., 2003; Julesz, Kropfl, & Petrig, 1980; Manning, Finlay, Dewis, & Dunlop, 1992; Skrandies, 2001). In agreement with these results, our results showed that although the global appearance of evoked brain activity is similar in both 2D and 3D conditions, the 3D waveform amplitude is significantly more negative for a sustained period. In addition, the scalp topography of RC1 for 3D responses is similar to that of previous EEG studies, being maximal at posterior occipital electrodes lying over early visual cortex (Lehmann & Julesz, 1978; Manning et al., 1992; Neill & Fenelon, 1988; Skrandies, 1991). It will be of interest in the future to use inverse modeling procedures to localize the sources derived from reliability components analysis. RCA produces multiple statistically defined sources that presumably correspond to a set of distributed electrical sources in cortex. However, at this point, there is no validated approach for using RC component topographies as input to source modeling procedures, and we are thus cautious in interpreting the anatomical site of their generation.

Depth structure and sceneness interact with disparity response of individual scenes

We found that different natural scenes can elicit different magnitudes of disparity-specific response, with some scenes eliciting large differences between the 2D and 3D versions (i.e., larger disparity response) and others eliciting small differences (Figure 6). Such variations are related to the depth structure of the image and to the response to the higher-order structure inherent in the scene—its "sceneness." With respect to depth structure, images with more homogeneous depth maps elicited larger responses. The homogeneity of the depth map may indicate the relative complexity of the image. Intuitively, depth structure of smoother images will be easier to perceive than that of images with complex and discontinuous features. This has been confirmed in the field of stereo-displays and image processing, where researchers have found that the perceived stereoscopic image quality can be increased by decreasing the standard deviation of its depth map by applying Gaussian filters (Alain, Tam, & Zhang, 2003; Fehn, 2003; Tam, Alain, Zhang, Martin, & Renaud, 2004). A preference for simpler depth structure may be a consequence of the very limited spatial resolution of the disparity system (Banks, Gepshtein, & Landy, 2004; Bradshaw & Rogers, 1999; Reynaud, Gao, & Hess, 2015; Tyler, 1974).

A novel result from our approach is our finding that the size of the disparity response depends on the size of the differential response to 2D versus scrambled images. Images in which the difference between scrambled and 2D response was larger also produce large disparity-specific responses. Past studies have reported that a higher-order image structure can interact with many perceptual phenomena (the term *interact* is used throughout the article for its literal meaning, not the statistical sense of interaction). For example, higher-order image statistics contribute importantly to boundary segmentation (Baker et al., 2008) as well as to detection of uniform photometric changes in natural images (Yoonessi, 2008). Strong preferences for images with natural phase spectra have been found during binocular rivalry, and such predominance could not be accounted for by the observer's bias toward recognizable features. (Baker & Graf, 2009). Here, we add disparity processing to that list of perceptual phenomena.

We discovered this linkage between monocular scene structure and 3D responses through a correlational approach based on simple image statistics and brain responses: We used depth map summary statistics and our neural sceneness metric to predict 2D versus 3D response differences. Different scenes contain different amounts of phase/higher-order structural information. It will be of interest to determine what image features drive sceneness and its relationship to 3D responses. Each image also had its own depth map. Simple visual inspection of the depth maps shows a strong relationship to the monocular scene structure (Figure 8). For example, the prominent tree in the natural scene displayed in Figure 8 is readily discernable in its corresponding depth map. This similarity suggests that the statistical linkage between 2D image structure and 3D responses may be driven via commonalities/consistencies between 2D and 3D cues for scene structure that are inherent in stereoscopic natural

images. Future studies could be designed to systematically vary these relationships to test this hypothesis. One thing worth mentioning is that the proportion of variability in the data set that is accounted for by our multiple regression model is only 23.9%. This suggests that although the associations are statistically significant, a substantial portion of the variance in the evoked response remains to be explained. How to best define perceptually relevant depth structure is an open question. We have used very simple metrics to summarize the depth structure in our images. In the future, it will be important to develop a more detailed understanding of the actual statistics of depth maps (Gibaldi, Canessa, & Sabatini, 2017; Hunter & Hibbard, 2015; Liu, Bovik, & Cormack, 2008) and natural scenes (Groen, Ghebreab, Prins, Lamme, & Scholte, 2013; Scholte, Ghebreab, Waldorp, Smeulders, & Lamme, 2009) so that better summary statistics can be developed for use in computational models of disparity processing that can make predictions on images (Didyk, Ritschel, Eisemann, Myszkowski, & Seidel, 2011; Read & Cumming, 2017).

Stereoacuity and sceneness interact with disparity response of individual subjects

We also found substantial individual differences in the magnitude of the response to disparity. It has previously been shown using behavioral measures that there are substantial individual differences in stereoscopic vision in adults who have excellent monocular visual acuity in each eye (Bosten et al., 2015; Howard, 1919; Richards, 1970). When analyzing the cortical responses of individuals, our multiple regression analysis showed that both stereoacuity and sceneness of the image contribute to individual variations in the disparity response. Individuals who have better stereoacuity and higher sensitivity to high-level scene information (i.e., larger difference in responses to intact 2D and scrambled images) tend to have larger disparity-specific responses. The relationship between stereoacuity and evoked response magnitude is weak but not surprising. Chao, Odom, and Karr (1988) measured participants' stereoacuity through the Titmus test and found a strong linear relationship between measured stereoacuity and VEP amplitude. Lower VEP amplitude has also been found to be associated with longer durations of disparity detection (Manning et al., 1992). What is somewhat surprising is that it was possible to measure a small but statistically reliable association between the VEP and behavior despite the restricted range of stereoacuity present in our participants. In the future, natural scene evoked responses may be useful in defining what drives individual differences, as this may be relevant both clinically and for applications in 3D display engineering (Patterson, 2015; Underwood, 1975; Wilmer, 2008).

Interpretation of the higher RC components

The neural system responding to disparity in natural images is not a simple one, and it cannot be entirely explained by a single dimension–reduced component. In addition to the first RC, the waveforms for the second and third components also differ between scrambled, 2D, and 3D conditions. These components account for ~30% and 11% of the trial-to-trial reliability, respectively. There are two features that differentiate the 3D responses of RC2 from that of RC1. First, their topographies are consistently different, with RC1 being distributed along posterior electrodes and RC2 being distributed dorsomedially. Second, the response to 3D images in RC1 is essentially a step function—there is an approximately constant relative negativity in the 3D response relative to the 2D response starting around 95 ms and lasting several hundred milliseconds more. By contrast, the RC2 response to 3D images is a ramp that steadily rises also from 95 ms (see the purples line in Figure 8). This response waveform bears little resemblance to the bi- or triphasic waveforms typical of sensory evoked potentials (Luck & Kappenman, 2011). This ramplike behavior is reminiscent of decision-related activity in other perceptual decision tasks (Dmochowski & Norcia, 2015; Donner, Siegel, Fries, & Engel, 2009; O'Connell, Dockree, & Kelly, 2012).

Ramping activity in these tasks is frequently modeled as an evidence accumulation process (Gold & Shadlen, 2007; Hanes & Schall, 1996; Smith & Ratcliff, 2004; Smith & Vickers, 1988). It is possible that the ramplike behavior in RC2 reflects a process in which a steplike 3D evidence signal is being integrated by the RC2 generator. It is unlikely that this activity is simply related to response choice generation or motor planning, given that the motor responses themselves were made much later in the trial. This activity would then need to be stored in memory for possible use in the comparison with the image in the second interval and the subsequent motor response. Simple tests of this model could involve using a different set of task instructions in which the decision is on a variable that is orthogonal to disparity. This would show whether the ramp is task related or simply a passive process related to disparity processing.

RC1 and RC2 also differ in their pattern of sensitivity to the three levels of images structure: scrambled, intact 2D, and intact 3D. For convenience, the three response waveforms (scrambled image onset, 2D natural image onset, and 3D natural image) are plotted together in Figure 9 for each RC component.

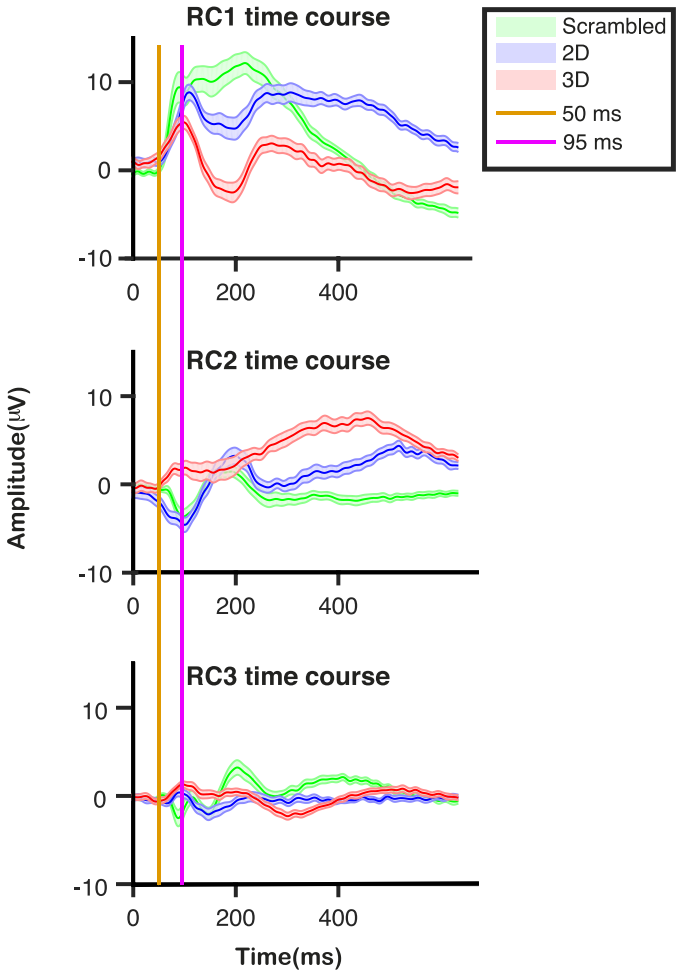

Figure 9. Waveforms of the three reliable components for scrambled natural scenes (green), intact 2D natural scenes (blue), and 3D natural scenes (red). The pale blue, green, and red lines plot the standard error of the mean.

For RC1, the structure inherent to intact natural scenes and that of 3D versus 2D scenes becomes resolved no later than 95 ms. Prior to 95 ms, there is a period of common activity, starting at 50 ms, that does not measurably differentiate scene content. After this time point, the intact scene responses are each more sustained than the scrambled image response. For RC2, 3D intact image responses diverge from the 2D and scrambled image response around 50 ms. Under the accumulator model of RC2 discussed above, the differential 3D activity needs to start around or before 50 ms. This may be possible, but our measurements of RC1 may not be sufficiently precise to resolve this activity from the 2D activity on the rising slope of RC1 between 50 and 95 ms.

In conclusion, the current study examined the topography, strength, and temporal dynamics of brain responses evoked by 2D and 3D natural scenes. The disparity-evoked response to natural scene stimuli is to first order similar to that from RDS, comprising sustained relative negativity of the dominant response component, RC1. At a finer grain of analysis, we found that depth structure contributes to scene-level variations and that stereoacuity contributes to individual differences in the disparity-specific response. Importantly, variation in the response to high-order scene statistics contributes to both scene-level and individual-level differences in the disparity-specific response. Through the use of RCA, we found multiple underlying sources' sensitivity to 2D and 3D structure in natural scenes, with RC1 having properties consistent with sensory encoding and RC2 having properties more consistent with decoding this information for task performance.

Keywords: disparity, natural scenes, individual differences

Acknowledgments

Supported by EY018875-04 from the National Institutes of Health.

Commercial relationships: none.
Corresponding author: Yiran Duan.
Email: yirand@gmail.com.
Address: Department of Psychology, Stanford University, Stanford, CA, USA.

References

Alain, G., Tam, W., & Zhang, L. (2003). Improving stereoscopic image quality of pictures generated from depth maps. Communications Research Center Canada, Internal CRC report, Ottawa, April 2003.

Appelbaum, L. G., Wade, A. R., Vildavski, V. Y., Pettet, M. W., & Norcia, A. M. (2006). Cue-invariant networks for figure and background processing in human visual cortex. *Journal of Neuroscience*, *26*, 11695–11708.

Backus, B. T., Fleet, D. J., Parker, A. J., & Heeger, D. J. (2001). Human cortical activity correlates with stereoscopic depth perception. *Journal of Neurophysiology*, *86*, 2054–2068.

Baker, C., & Graf, E. W. (2009). Natural images dominate in binocular rivalry. *Proceedings of the National Academy of Sciences, USA*, *106*, 5436–5441.

Baker, C., Yoonessi, A., & Arsenault, E. (2008). Texture segmentation in natural images: Contribution of higher-order image statistics to psycho-

physical performance. *Journal of Vision*, *8*(6):350, https://doi.org/10.1167/8.6.350. [Abstract]

Banks, M. S., Gepshtein, S., & Landy, M. S. (2004). Why is spatial stereoresolution so low? *Journal of Neuroscience*, *24*, 2077–2089.

Bieniek, M. M., Pernet, C. R., & Rousselet, G. A. (2012). Early ERPs to faces and objects are driven by phase, not amplitude spectrum information: Evidence from parametric, test-retest, single-subject analyses. *Journal of Vision*, *12*(13):12, 1–24, https://doi.org/10.1167/12.13.12. [PubMed] [Article]

Blair, R. C., & Karniski, W. (1993). An alternative method for significance testing of waveform difference potentials. *Psychophysiology*, *30*, 518–524.

Bosten, J., Goodbourn, P., Lawrance-Owen, A., Bargary, G., Hogg, R., & Mollon, J. (2015). A population study of binocular function. *Vision Research*, *110*, 34–50.

Bradshaw, M. F., & Rogers, B. J. (1999). Sensitivity to horizontal and vertical corrugations defined by binocular disparity. *Vision Research*, *39*, 3049–3056.

Burge, J., McCann, B. C., & Geisler, W. S. (2016). Estimating 3D tilt from local image cues in natural scenes. *Journal of Vision*, *16*(13):2, 1–25, https://doi.org/10.1167/16.13.2. [PubMed] [Article]

Carandini, M., Demb, J. B., Mante, V., Tolhurst, D. J., Dan, Y., Olshausen, B. A., ... Rust, N. C. (2005). Do we know what the early visual system does? *Journal of Neuroscience*, *25*, 10577–10597.

Chao, G.-M., Odom, J. V., & Karr, D. (1988). Dynamic stereoacuity: A comparison of electrophysiological and psychophysical responses in normal and stereoblind observers. *Documenta Ophthalmologica*, *70*, 45–58.

Coutant, B. E., & Westheimer, G. (1993). Population distribution of stereoscopic ability. *Ophthalmic and Physiological Optics*, *13*, 3–7.

David, S. V., Vinje, W. E., & Gallant, J. L. (2004). Natural stimulus statistics alter the receptive field structure of v1 neurons. *Journal of Neuroscience*, *24*, 6991–7006.

Didyk, P., Ritschel, T., Eisemann, E., Myszkowski, K., & Seidel, H.-P. (2011, August). A perceptual model for disparity. In *ACM Transactions on Graphics (TOG)* (Vol. 30, No. 4, p. 96). Chicago: ACM.

Dmochowski, J. P., Greaves, A. S., & Norcia, A. M. (2015). Maximally reliable spatial filtering of steady state visual evoked potentials. *Neuroimage*, *109*, 63–72.

Dmochowski, J. P., & Norcia, A. M. (2015). Cortical components of reaction-time during perceptual decisions in humans. *PLoS One*, *10*, e0143339.

Dmochowski, J. P., Sajda, P., Dias, J., & Parra, L. C. (2012). Correlated components of ongoing EEG point to emotionally laden attention—A possible marker of engagement? *Frontiers in Human Neuroscience*, *6*, 112.

Dodgson, N. A. (2004, May). Variation and extrema of human interpupillary distance. In *Stereoscopic Displays and Virtual Reality Systems XI* (Vol. 5291, pp. 36–47). International Society for Optics and Photonics.

Donner, T. H., Siegel, M., Fries, P., & Engel, A. K. (2009). Buildup of choice-predictive activity in human motor cortex during perceptual decision making. *Current Biology*, *19*, 1581–1585.

Emrith, K., Chantler, M., Green, P., Maloney, L., & Clarke, A. (2010). Measuring perceived differences in surface texture due to changes in higher order statistics. *Journal of the Optical Society of America A*, *27*, 1232–1244.

Fahle, M., Quenzer, T., Braun, C., & Spang, K. (2003). Feature-specific electrophysiological correlates of texture segregation. *Vision Research*, *43*, 7–19.

Fehn, C. (2003, September). A 3D-TV approach using depth-image-based rendering (DIBR). In *Proceedings of VIIP* (Vol. 3, No. 3).

Felsen, G., & Dan, Y. (2005). A natural approach to studying vision. *Nature Neuroscience*, *8*, 1643–1646.

Felsen, G., Touryan, J., Han, F., & Dan, Y. (2005). Cortical sensitivity to visual features in natural scenes. *PLoS Biology*, *3*, e342.

Field, D. J. (1987). Relations between the statistics of natural images and the response properties of cortical cells. *Journal of the Optical Society of America A*, *4*, 2379–2394.

Fischmeister, F. P. S., & Bauer, H. (2006). Neural correlates of monocular and binocular depth cues based on natural images: A LORETA analysis. *Vision Research*, *46*, 3373–3380.

Gaebler, M., Biessmann, F., Lamke, J.-P., Müller, K.-R., Walter, H., & Hetzer, S. (2014). Stereoscopic depth increases intersubject correlations of brain networks. *NeuroImage*, *100*, 427–434.

Geisler, W. S., & Diehl, R. L. (2002). Bayesian natural selection and the evolution of perceptual systems. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *357*, 419–448.

Gibaldi, A., Canessa, A., & Sabatini, S. P. (2017). The active side of stereopsis: Fixation strategy and adaptation to natural environments. *Scientific Reports*, *7*, 44800.

Gold, J. I., & Shadlen, M. N. (2007). The neural basis of decision making. *Annual Review of Neuroscience*, *30*, 535–574.

Groen, I. I., Ghebreab, S., Prins, H., Lamme, V. A., & Scholte, H. S. (2013). From image statistics to scene gist: Evoked neural activity reveals transition from low-level natural image structure to scene category. *Journal of Neuroscience*, *33*, 18814–18824.

Hanes, D. P., & Schall, J. D. (1996). Neural control of voluntary movement initiation. *Science*, *274*, 427.

Haufe, S., Meinecke, F., Görgen, K., Dähne, S., Haynes, J.-D., Blankertz, B., Bießmann, F. (2014). On the interpretation of weight vectors of linear models in multivariate neuroimaging. *Neuroimage*, *87*, 96–110.

Hoffman, D. M., Girshick, A. R., Akeley, K., & Banks, M. S. (2008). Vergence–accommodation conflicts hinder visual performance and cause visual fatigue. *Journal of Vision*, *8*(3):33, 1–30, https://doi.org/10.1167/8.3.33. [PubMed] [Article]

Howard, C. H. J. (1919). A test for the judgment of distance. *American Journal of Ophthalmology*, *2*, 656–675.

Hunter, D. W., & Hibbard, P. B. (2015). Distribution of independent components of binocular natural images. *Journal of Vision*, *15*(13):6, 1–31, https://doi.org/10.1167/15.13.6. [PubMed] [Article]

Joubert, O. R., Rousselet, G. A., Fabre-Thorpe, M., & Fize, D. (2009). Rapid visual categorization of natural scene contexts with equalized amplitude spectrum and increasing phase noise. *Journal of Vision*, *9*(1):2, 1–16, https://doi.org/10.1167/9.1.2. [PubMed] [Article]

Julesz, B., Kropfl, W., & Petrig, B. (1980). Large evoked potentials to dynamic random-dot correlograms and stereograms permit quick determination of stereopsis. *Proceedings of the National Academy of Sciences, USA*, *77*, 2348–2351.

Lehmann, D., & Julesz, B. (1978). Lateralized cortical potentials evoked in humans by dynamic random-dot stereograms. *Vision Research*, *18*, 1265–1271.

Liu, Y., Bovik, A. C., & Cormack, L. K. (2008). Disparity statistics in natural scenes. *Journal of Vision*, *8*(11):19, 1–14, https://doi.org/10.1167/8.11.19. [PubMed] [Article]

Luck, S. J., & Kappenman, E. S. (2011). *The Oxford handbook of event-related potential components*. Oxford, UK: Oxford University Press.

Manning, M. L., Finlay, D. C., Dewis, S. A., & Dunlop, D. B. (1992). Detection duration thresholds and evoked potential measures of stereosensitivity. *Documenta Ophthalmologica*, *79*, 161–175.

McCann, B. C. (2015). *Naturalistic depth perception* (Unpublished doctoral dissertation).

McKee, S. P., & Taylor, D. G. (2010). The precision of binocular and monocular depth judgments in natural settings. *Journal of Vision*, *10*(10):5, 1–13, https://doi.org/10.1167/10.10.5. [PubMed] [Article]

Michel, C. M., Henggeler, B., & Lehmann, D. (1992). 42-channel potential map series to visual contrast and stereo stimuli: Perceptual and cognitive event-related segments. *International Journal of Psychophysiology*, *12*, 133–145.

Morgan, M., Ross, J., & Hayes, A. (1991). The relative importance of local phase and local amplitude in patchwise image reconstruction. *Biological Cybernetics*, *65*, 113–119.

Neill, R., & Fenelon, B. (1988). Scalp response topography to dynamic random dot stereograms. *Electroencephalography and Clinical Neurophysiology*, *69*, 209–217.

O'Connell, R. G., Dockree, P. M., & Kelly, S. P. (2012). A supramodal accumulation-to-bound signal that determines perceptual decisions in humans. *Nature Neuroscience*, *15*, 1729–1735.

Ogawa, A., & Macaluso, E. (2015). Orienting of visuospatial attention in complex 3D space: Search and detection. *Human Brain Mapping*, *36*, 2231–2247.

Oppenheim, A. V., & Lim, J. S. (1981). The importance of phase in signals. *Proceedings of the IEEE*, *69*, 529–541.

Parker, A. J. (2007). Binocular depth perception and the cerebral cortex. *Nature Reviews Neuroscience*, *8*, 379–391.

Parra, L. C., Spence, C. D., Gerson, A. D., & Sajda, P. (2005). Recipes for the linear analysis of EEG. *Neuroimage*, *28*, 326–341.

Patterson, R. E. (2015). *Human factors of stereoscopic 3D displays*. London: Springer.

Patterson, R., & Martin, W. L. (1992). Human stereopsis. *Human Factors*, *34*, 669–692.

Portilla, J., & Simoncelli, E. P. (2000). A parametric texture model based on joint statistics of complex wavelet coefficients. *International Journal of Computer Vision*, *40*, 49–70.

Read, J. C., & Cumming, B. G. (2017). Visual perception: Neural networks for stereopsis. *Current Biology*, *27*, R594–R596.

Regan, D., & Spekreijse, H. (1970). Electrophysiological correlate of binocular depth perception in man. *Nature*, *225*, 92–94.

Reynaud, A., Gao, Y., & Hess, R. F. (2015). A normative dataset on human global stereopsis using the quick Disparity Sensitivity Function (qDSF). *Vision Research*, *113*, 97–103.

Richards, W. (1970). Stereopsis and stereoblindness. *Experimental Brain Research*, *10*, 380–388.

Rousselet, G. A., Pernet, C. R., Bennett, P. J., & Sekuler, A. B. (2008). Parametric study of EEG sensitivity to phase noise during face processing. *BMC Neuroscience*, *9*, 98.

Şahinoğlu, B. (2004). Depth-related visually evoked potentials by dynamic random-dot stereograms in humans: Negative correlation between the peaks elicited by convergent and divergent disparities. *European Journal of Applied Physiology*, *91*, 689–697.

Scholte, H. S., Ghebreab, S., Waldorp, L., Smeulders, A. W., & Lamme, V. A. (2009). Brain responses strongly correlate with Weibull image statistics when processing natural images. *Journal of Vision*, *9*(4):29, 1–15, https://doi.org/10.1167/9.4.29. [PubMed] [Article]

Simoncelli, E. P., & Olshausen, B. A. (2001). Natural image statistics and neural representation. *Annual Review of Neuroscience*, *24*, 1193–1216.

Skrandies, W. (1991). Contrast and stereoscopic visual stimuli yield lateralized scalp potential fields associated with different neural generators. *Electroencephalography and Clinical Neurophysiology*, *78*, 274–283.

Skrandies, W. (2001). The processing of stereoscopic information in human visual cortex: Psychophysical and electrophysiological evidence. *Clinical Electroencephalography*, *32*, 152–159.

Smith, P. L., & Ratcliff, R. (2004). Psychology and neurobiology of simple decisions. *Trends in Neurosciences*, *27*, 161–168.

Smith, P. L., & Vickers, D. (1988). The accumulator model of two-choice discrimination. *Journal of Mathematical Psychology*, *32*, 135–168.

Snow, J. C., Pettypiece, C. E., McAdam, T. D., McLean, A. D., Stroman, P. W., Goodale, M. A., & Culham, J. C. (2011). Bringing the real world into the fMRI scanner: Repetition effects for pictures versus real objects. *Scientific Reports*, *1*, 130.

Tam, W. J., Alain, G., Zhang, L., Martin, T., & Renaud, R. (2004, October). Smoothing depth maps for improved steroscopic image quality. In *Three-Dimensional TV, Video, and Display III* (Vol. 5599, pp. 162–173). International Society for Optics and Photonics.

Tanaka, K. (1996). Inferotemporal cortex and object vision. *Annual Review of Neuroscience*, *19*, 109–139.

Tyler, C. W. (1974). Depth perception in disparity gratings. *Nature*, *251*, 140.

Underwood, B. J. (1975). Individual differences as a crucible in theory construction. *American Psychologist*, *30*, 128.

Welchman, A. E. (2016). The human brain in depth: how we see in 3D. *Annual Review of Vision Science*, *2*, 2.6.1–2.6.32.

Wheatstone, C. (1838). Contributions to the physiology of vision. Part the first. On some remarkable, and hitherto unobserved, phenomena of binocular vision. *Philosophical Transactions of the Royal Society of London*, 371–394.

Wilmer, J. B. (2008). How to use individual differences to isolate functional organization, biology, and utility of visual functions; with illustrative proposals for stereopsis. *Spatial Vision*, *21*, 561–579.

Yoonessi, A. (2008). Comparison of sensitivity to color changes in natural and phase-scrambled scenes. *Journal of the Optical Society of America A*, *25*, 676–684.

Zaroff, C. M., Knutelska, M., & Frumkes, T. E. (2003). Variation in stereoacuity: Normative description, fixation disparity, and the roles of aging and gender. *Investigative Ophthalmology & Visual Science*, *44*, 891–900.