RESEARCH ARTICLE

Near-optimal combination of disparity across a log-polar scaled visual field

Guido Maiello [1]*, **Manuela Chessa** [2], **Peter J. Bex** [3], **Fabio Solari** [2]

1 Department of Experimental Psychology, Justus Liebig University Giessen, Giessen, Hesse, Germany, **2** Department of Informatics, Bioengineering, Robotics and Systems Engineering, University of Genoa, Genoa, Italy, **3** Department of Psychology, Northeastern University, Boston, Massachusetts, United States of America

* guido_maiello@yahoo.it

Abstract

The human visual system is foveated: we can see fine spatial details in central vision, whereas resolution is poor in our peripheral visual field, and this loss of resolution follows an approximately logarithmic decrease. Additionally, our brain organizes visual input in polar coordinates. Therefore, the image projection occurring between retina and primary visual cortex can be mathematically described by the log-polar transform. Here, we test and model how this space-variant visual processing affects how we process binocular disparity, a key component of human depth perception. We observe that the fovea preferentially processes disparities at fine spatial scales, whereas the visual periphery is tuned for coarse spatial scales, in line with the naturally occurring distributions of depths and disparities in the real-world. We further show that the visual system integrates disparity information across the visual field, in a near-optimal fashion. We develop a foveated, log-polar model that mimics the processing of depth information in primary visual cortex and that can process disparity directly in the cortical domain representation. This model takes real images as input and recreates the observed topography of human disparity sensitivity. Our findings support the notion that our foveated, binocular visual system has been moulded by the statistics of our visual environment.

Author summary

We investigate how humans perceive depth from binocular disparity at different spatial scales and across different regions of the visual field. We show that small changes in disparity-defined depth are detected best in central vision, whereas peripheral vision best captures the coarser structure of the environment. We also demonstrate that depth information extracted from different regions of the visual field is combined into a unified depth percept. We then construct an image-computable model of disparity processing that takes into account how our brain organizes the visual input at our retinae. The model operates directly in cortical image space, and neatly accounts for human depth perception across the visual field.

Introduction

Humans employ binocular disparities, the differences between the views of the world seen by our two eyes, to determine the depth structure of the environment [1]. More specifically, stereoscopic depth perception relies on relative disparities, i.e. the differences in disparities between points at different depths in the world, which are independent of fixation depth [2]. Additional complexity in our estimate of the depth structure arises because spatial resolution is not uniform across the visual field. Instead, our visual system is space-variant: the foveae of both our eyes are sensitive to fine spatial detail, while vision in our periphery is increasingly coarse [3]. Therefore, when humans look at an object, the eyes are rotated so that the high-resolution foveae of both eyes are pointed at the same location on the surface of the object. The fixated object will extend into our binocular visual field by a distance proportional to the object's size, and over this area we will experience small stereoscopic depth changes, arising from relative retinal disparities due to the surface structure and slant or tilt of the fixated object. The world beyond the fixated object in our peripheral visual field will typically contain objects at a range of different depths. Consequently we will experience a greater magnitude and range of relative binocular disparities [4]. It has been proposed that the visual system may process disparity at different disparity spatial scales along separate channels [5], analogous to the channels selective for luminance differences at different luminance spatial frequencies [6]. Using a variety of paradigms to investigate both absolute and relative disparity processing, several authors have provided evidence for at least two [7–11] or more [12] disparity spatial channels for disparity processing, which in turn may rely on distinct sets of luminance spatial channels [13–16].

Given that our visual world contains small, fine disparities near the fovea and larger coarse disparities in our peripheral visual field, we might analogously expect sensitivity to disparity to vary across the visual field. Based on differences in experience during development, different regions of our visual field might therefore be expected to be optimized to process disparity at different spatial scales [17]. We test this hypothesis by measuring disparity sensitivity across the visual field of human participants. We employ annular pink noise stimuli embedded with disparity corrugations of different spatial scales and spanning rings of different retinal eccentricity. We hypothesize that as eccentricity increases from fovea to periphery, the tuning of depth sensitivity should shift from fine to coarse spatial scales. We also hypothesize that peak sensitivity to stereoscopic disparity should also decrease as eccentricity increases, following the general decrease in visual sensitivity observed in the visual periphery [18].

If indeed different visual field eccentricities preferentially process disparities at different spatial scales, then how does the visual system combine depth information processed throughout our visual field to recover the depth structure of the environment? If disparity information is integrated across different regions of the visual field, then sensitivity for full field stimuli should be better than for stimuli spanning smaller areas of the visual field. We test whether this integration process is optimal according to a maximum-likelihood estimation (MLE) principle [19–25].

Next, we construct a model. Prince and Rogers [17] were the first to suggest that disparity sensitivity across the visual field may be related to M-scaling (i.e. cortical magnification, the different number of cortical neurons that process information from different visual field locations). Gibaldi et al [26, 27] even suggest that the specific pattern of cortical magnification might be a consequence of how we visually explore the naturally occurring distribution of real-world depths. Therefore, we implement a simple, neurally-inspired model of disparity processing, in which we include a critical log-processing stage that mimics the transformation between retinal and cortical image space [28–30]. A unique advantage of this approach is that

disparity can be computed and analyzed directly in the cortical domain [31]. We have previously shown that this approach can account for motion processing throughout the visual field of human participants [32]. Here, we examine whether log-polar processing can also account for human disparity processing across the visual field.

Results

Fig 1a shows the pink noise stimuli we employed to psychophysically assessed disparity sensitivity in the central (red, 0-3 deg), mid peripheral (green, 3-9 deg), far peripheral (blue, 9-21 deg), and full (black, 0-21 deg) visual field of human observers. Noise stimuli were embedded with sinusoidal disparity corrugations of different spatial frequencies (Fig 1b, see detailed descriptions of stimuli and experimental procedures in the Materials and methods section).

Sensitivity to disparity corrugations varies with stimulus size and eccentricity

Fig 1c (bottom plot) shows the tuning of human disparity sensitivity across different regions of the visual field. Disparity sensitivity in the far periphery (blue curve) is tuned to depth
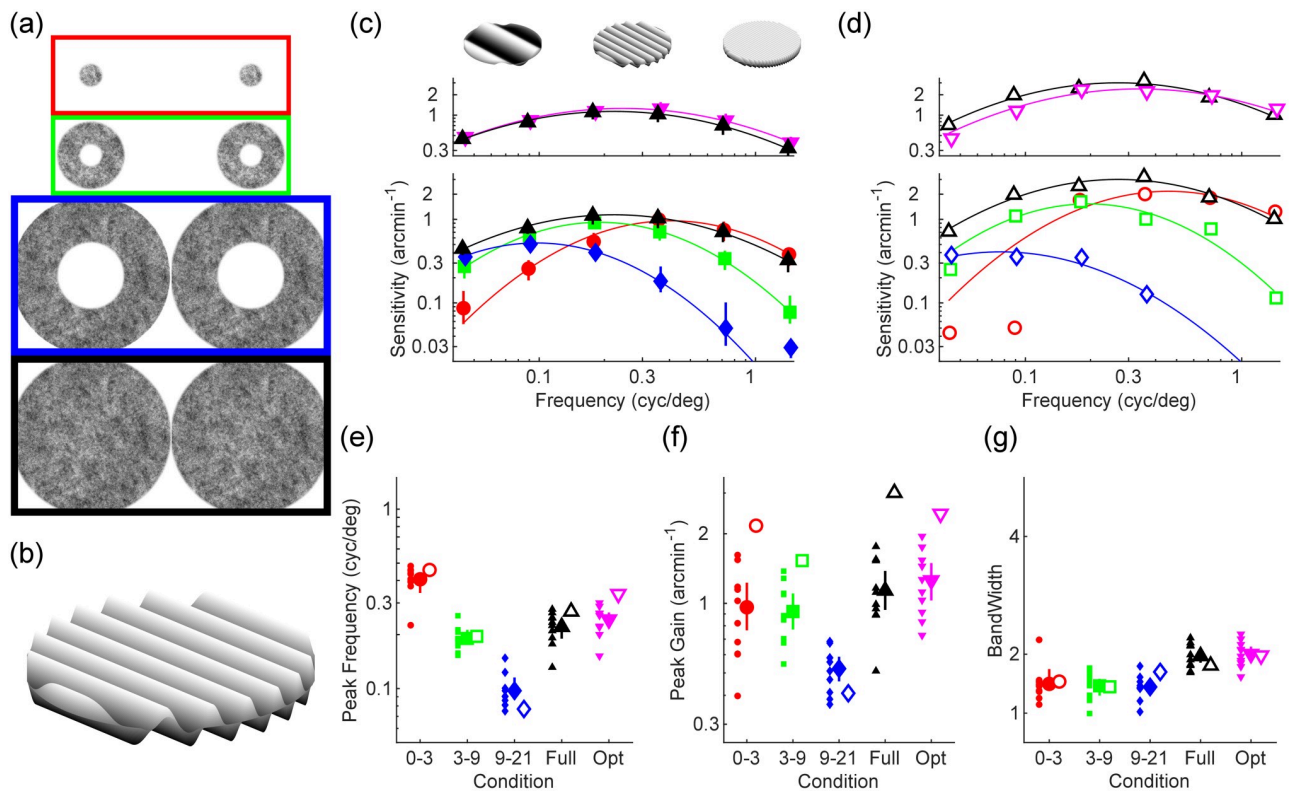

Fig 1. Disparity sensitivity across the visual field. (a) Participants and computational model were tested with annular pink noise stimuli spanning the foveal (red; 0-3 deg), mid (green; 3-9 deg), far (blue; 9-21 deg), and full (black, 0-21) visual field. (b) Noise stimuli were embedded with sinusoidal disparity corrugations. Cross-fuse stimuli in panel to view the disparity-defined corrugation. (c) In the bottom panel, human disparity sensitivity is plotted as a function of spatial frequency for stimuli spanning far (blue diamonds), mid (green squares), foveal (red circles), and full (black upwards pointing triangles) portions of the visual field. In the top panel, human disparity sensitivity for the full field stimuli is compared to MLE-optimal disparity sensitivities (magenta downwards pointing triangles). Continuous lines are best fitting log parabola functions passing through the data. (d) As in c, except for the computational model of disparity processing. (e-g) Peak frequency, peak gain, and bandwidth of the fitted log parabola model as a function of the portion of visual field tested, and for the MLE-optimal sensitivity. In all panels, filled markers represent human data, empty markers represent data from the computational model of disparity processing. Small markers are data from individual participants, large markers are the mean sensitivities across participants and error bars represent 95% bootstrapped confidence intervals.

variations at low spatial frequencies. Disparity sensitivity in the near periphery (green curve) is tuned to depth variations at mid spatial frequencies. Disparity sensitivity in the fovea (red curve) is tuned to depth variations at high spatial frequencies. Thus, the peak frequency of the disparity sensitivity curves shifts from high to low frequencies moving from the fovea to the peripheral visual field (Fig 1e, $F_{2,18} = 186.65$, $p = 9.2 \times 10^{-13}$). Peak sensitivity also decreases from the fovea to the peripheral visual field (Fig 1f, $F_{2,18} = 15.87$, $p = 1.1 \times 10^{-4}$), whereas the bandwidth of disparity tuning remains constant (Fig 1g, $F_{2,18} = 0.2$, $p = 0.82$).

Humans integrate disparity information across the visual field in a near-optimal fashion

Fig 1c (bottom plot) shows how disparity sensitivity for the full field stimuli (black) is the envelope of the disparity sensitivities estimated in the restricted visual field conditions. Additionally, Fig 1c (top plot) shows how disparity sensitivity for stimuli spanning the whole visual field (black) approaches the level of sensitivity predicted from the MLE-optimal combination of disparity sensitivity across the separate portions of the visual field (magenta, following [25], see Materials and methods section for precise mathematical formulation). While qualitatively similar, disparity tuning for the full field stimuli was statistically different from the MLE-optimal disparity tuning based on optimal integration of disparity across the retina. More specifically, disparity tuning for the full field stimuli exhibited lower peak frequency (Fig 1e, $t(9) = 3.95$, $p = 0.0033$) and lower peak gain (Fig 1f, $t(9) = 2.67$, $p = 0.026$) compared to the MLE-optimal disparity tuning, whereas bandwidth was not significantly different (Fig 1g, $t(9) = 0.53$, $p = 0.61$). Nevertheless, these differences amounted to a sub-optimal reduction in sensitivity of only 0.1 arcseconds, and a shift in tuning of only 0.02 cycles/degree.

A foveated model of disparity processing accounts for the patterns of human data

Fig 1d shows the spatial frequency tuning of disparity sensitivity in our log-polar computational model of disparity processing, tested with the same stimuli and procedure as the human observers (i.e. as if the model were an individual human participant). This pattern is strikingly similar to the patterns of disparity sensitivity across the visual field of human observers (Fig 1c), and the model shows a high level of agreement with the human data ($r = 0.91$; $p = 8.3 \times 10^{-10}$; $r^2 = 0.83$). Across experimental conditions, the estimates of peak frequency, peak gain and bandwidth for the computational model follow the same patterns as those of the human participants, and cover a similar range (compare filled and empty symbols in Fig 1e–1g).

Visual processing throughout the model

Fig 2 shows a scheme of the proposed model and how its different processing stages encode and decode visual information (a detailed description and precise mathematical formulation of the model is presented in the Materials and methods section). First, stereoscopic Cartesian images (Fig 2a) are mapped to the cortical domain (Fig 2b) using the log-polar transform. In the cortical domain, the coordinates of the transformed images represent log-scaled retinal eccentricity ξ and retinal angle η.

Next, the cortical representation of the input images is processed by a population of V1 binocular simple cell units, each unit characterized by a cortical receptive field size σ, a cortical preferred spatial orientation θ and a cortical preferred phase difference $\Delta\psi$ between the left- and right-eye components of a cell's receptive field (following the phase-shift model [33, 34]).
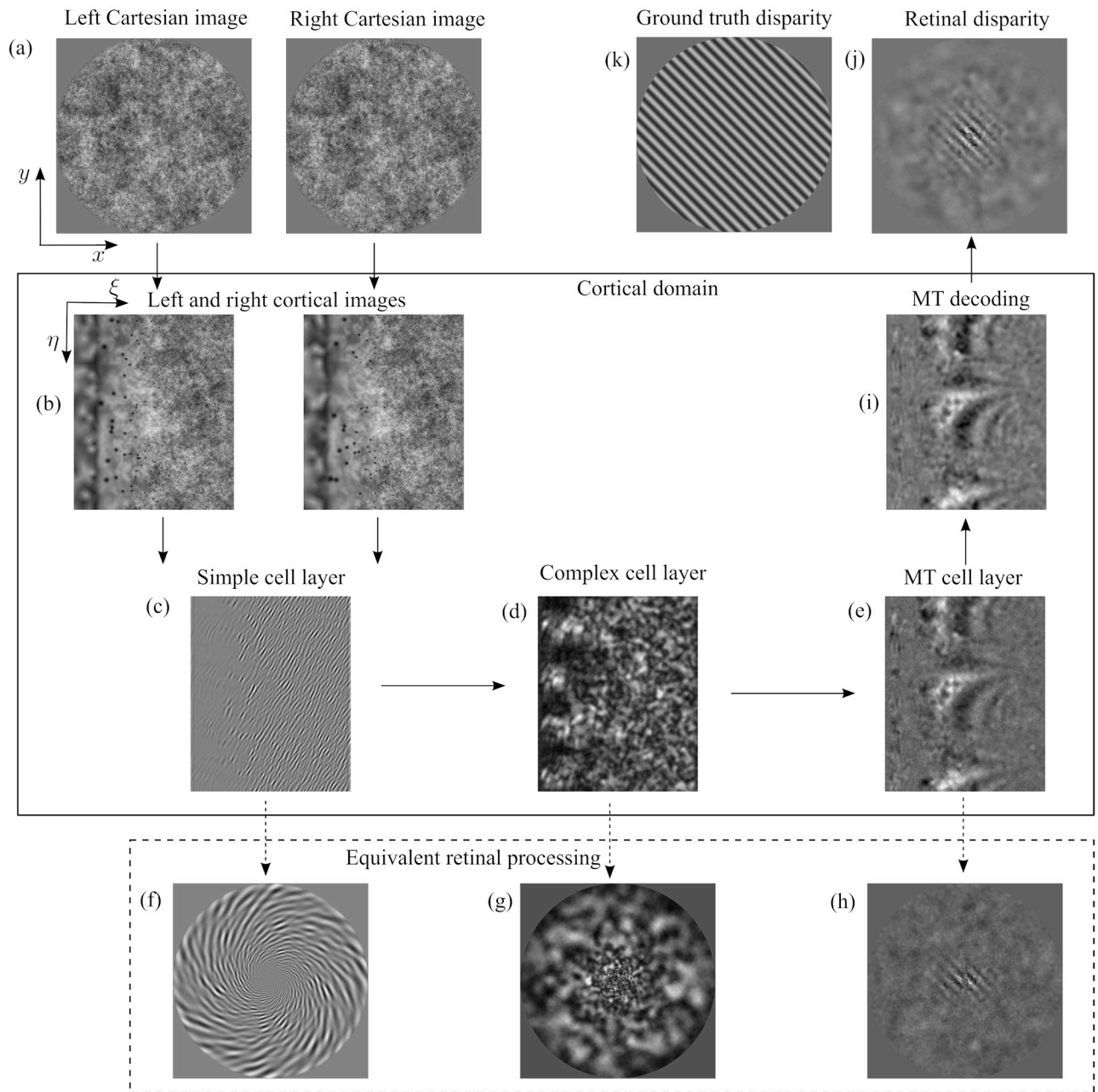
Fig 2. A scheme of the proposed model and of its encoding/decoding of the visual information. The left and right input stimuli (a) are mapped (Eq 3) to the corresponding cortical representation (b). These cortical images are the visual afferents to V1 layers (Eq 11): the activity of the simple cell layer (c) is non-linearly combined (Eq 13) to produce the complex cell layer (d) (for the sake of clarity we show an activity image for one set of tuning parameters, only). At this stage of the model the visual information is encoded in a distributed representation in the parameter space of V1 cells (i.e. spatial orientation θ, phase difference $\Delta\psi$ and spatial scale σ). Then, by pooling afferent V1 responses (Eq 14) the MT cell activity (e) shows a tuning to signal features (i.e. magnitude d and direction ϕ of disparity). The equivalent retinal processing is shown in (f-g-h), i.e. the cortical activity is back mapped to retinal space (only for visualization purposes, this representation is not computed or utilized by the model). (i) The MT activity is decoded (Eq 16) in order to estimate the disparity. (j) shows the estimated disparity map (in the retinal domain) for a disparity grating of 0.5 cycles/degree (k). This grating is optimized for the model's fovea, and the estimated disparity map is thus degraded by the log-polar mapping in the periphery.

https://doi.org/10.1371/journal.pcbi.1007699.g002

Fig 2c shows the output of one such V1 simple cell tuned for σ = 5.12 pixels, θ = 67.5 degrees, $\Delta\psi$ = 40.2 degrees. Note how the corresponding retinal processing (Fig 2f, obtained by applying the inverse log-polar mapping to Fig 2c) demonstrates the space-variant effects of the log-polar transform. A V1 unit tuned to a single cortical receptive field size and a single cortical orientation covers distinct orientations and receptive field sizes throughout the retinal domain.

Following the binocular energy model [33, 35, 36], quadrature pairs of binocular simple cells are combined to form the responses of V1 complex cell units. At this level the representation of visual information is distributed across the parameter space of V1 cells. This means that it is not possible to discern any information visibly related to the stereoscopic stimulus simply by looking at the cortical (Fig 2d) or retinal (Fig 2g) output of a single layer tuned to a specific parameter set. At the V1 level, cells are tuned to the component of the vector disparity orthogonal to the cell's spatial orientation tuning. This tuning behaviour is apparent when visualizing the model responses to visual stimuli of uniform disparity, such as the one we show in S1 Appendix.

Tuning to the vector disparity emerges at the MT level, where V1 complex cell responses are pooled across spatial and orientation domains, followed by a non-linearity. Fig 2e shows the response of an MT cell tuned to a specific cortical disparity. At this level, MT cells encode the magnitude d and direction ϕ of the stereoscopic stimulus. Thus, at MT level the representation of the visual information is distributed across d and ϕ parameter space. The equivalent retinal processing (Fig 2h) shows how this MT unit does indeed contain a partial representation of the disparity information embedded in the input images. By combining these partial disparity representations we can decode the MT activity in order to obtain a full estimate of cortical disparity (Fig 2i).

The estimated retinal disparity map shown in Fig 2j is obtained by backwards transforming the decoded cortical activity. Note the effect of the log-polar processing on the disparity corrugation. The input disparity corrugation (0.5 cycles/degree, Fig 2k) matches the frequency of the model's peak disparity sensitivity at the fovea. Therefore, the corrugation is primarily detectable in the model's fovea, and is degraded by the log-polar mapping towards the model's visual periphery.

A detailed description of the model processing for a uniform disparity stimulus is presented in S1 Appendix.

A comparative analysis of model parameters

The specific architecture and parameters selected for the proposed model (see Materials and methods) were derived from the literature or based on pilot work [37] where we compared model performance to normative data from Reynaud et al. [38]. To test how the model's ability to account for the human data is dependent upon the specific parameter values we chose, we present a comparative analysis of model performance and behaviour when varying key parameters and architecture.

A key component of the model is the retino-corical transform, which can be conveniently summarized into one parameter: the compression ratio (*CR*, see Materials and methods) of the cortical image with respect to the Cartesian one. The *CR* can therefore be equated to the strength of M-scaling between retina and cortex. Another key component of the model is the fact that processing occurs directly in cortical image space, and a primary determinant of cortical processing is the spatial support (or size) of the cortical receptive field. These two parameters together determine how visual processing varies from fine to coarse spatial scales moving from fovea to visual periphery. In addition to this the proposed model contains simulated

neural noise, since we've previously shown that the human visual system also contains internal noise [32]. Therefore another parameter of interest is the amount of neural noise that can be injected into the model before its agreement with human data begins to degrade.

Finally, a standard and computationally efficient approach to take into account the presence of distinct luminance spatial frequency channels in the visual cortex is to implement coarse-to-fine pyramidal processing [39–41], where every pyramid level processes a different spatial scale. The proposed model does not contain these distinct channels, since the log-polar spatial sampling acts as a "horizontal" multi-scale [42]. Nevertheless, distinct channels can be included alongside [39] or even replace log-polar spatial sampling, to test whether processing along distinct luminance channels leads to the observed human patterns of disparity frequency tuning.

The strength of M-scaling affects peak disparity sensitivity and disparity tuning bandwidth

Fig 3a shows that increasing or decreasing the model's CR degrades but does not destroy the model's agreement with human data. Across visual field conditions, varying the CR does not strongly affect disparity tuning in terms of the model's peak frequency (Fig 3b). Conversely, Fig 3c shows that increasing the CR decreases disparity sensitivity, whereas decreasing the CR
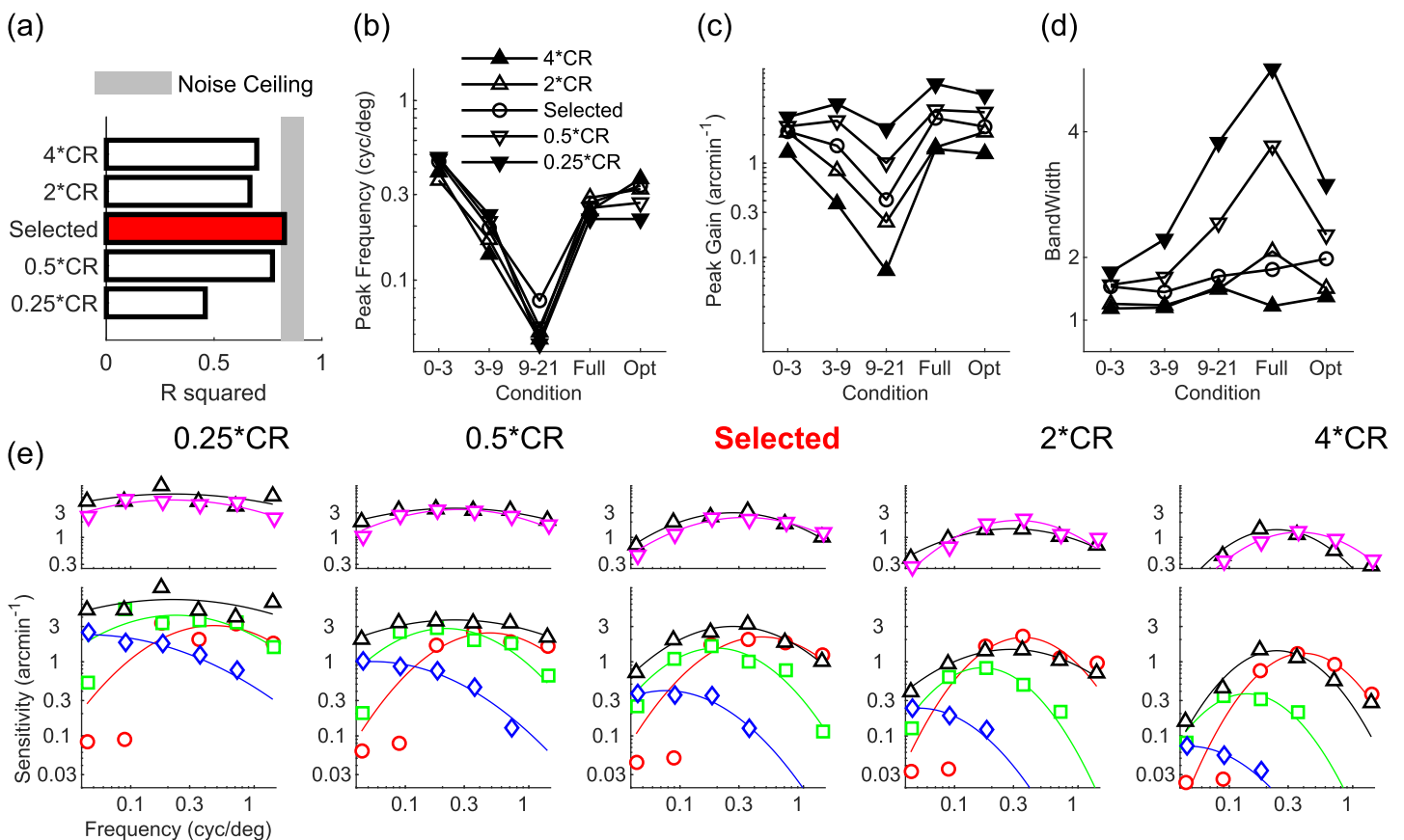
Fig 3. Effects of varying the model's compression ratio. (a) Agreement (R^2) between human data and models with smaller and larger CRs than the selected model (red bar). Shaded region represents the noise ceiling, an estimate of peak model performance (see Materials and methods). (b-d) Peak frequency, peak gain, and bandwidth of models with higher (upwards triangles) and lower (downwards triangles) CR than the selected model (circle), for all visual field conditions tested. (e) Disparity sensitivity plotted as a function of spatial frequency as in Fig 1d for all tested models of varying CR.

https://doi.org/10.1371/journal.pcbi.1007699.g003
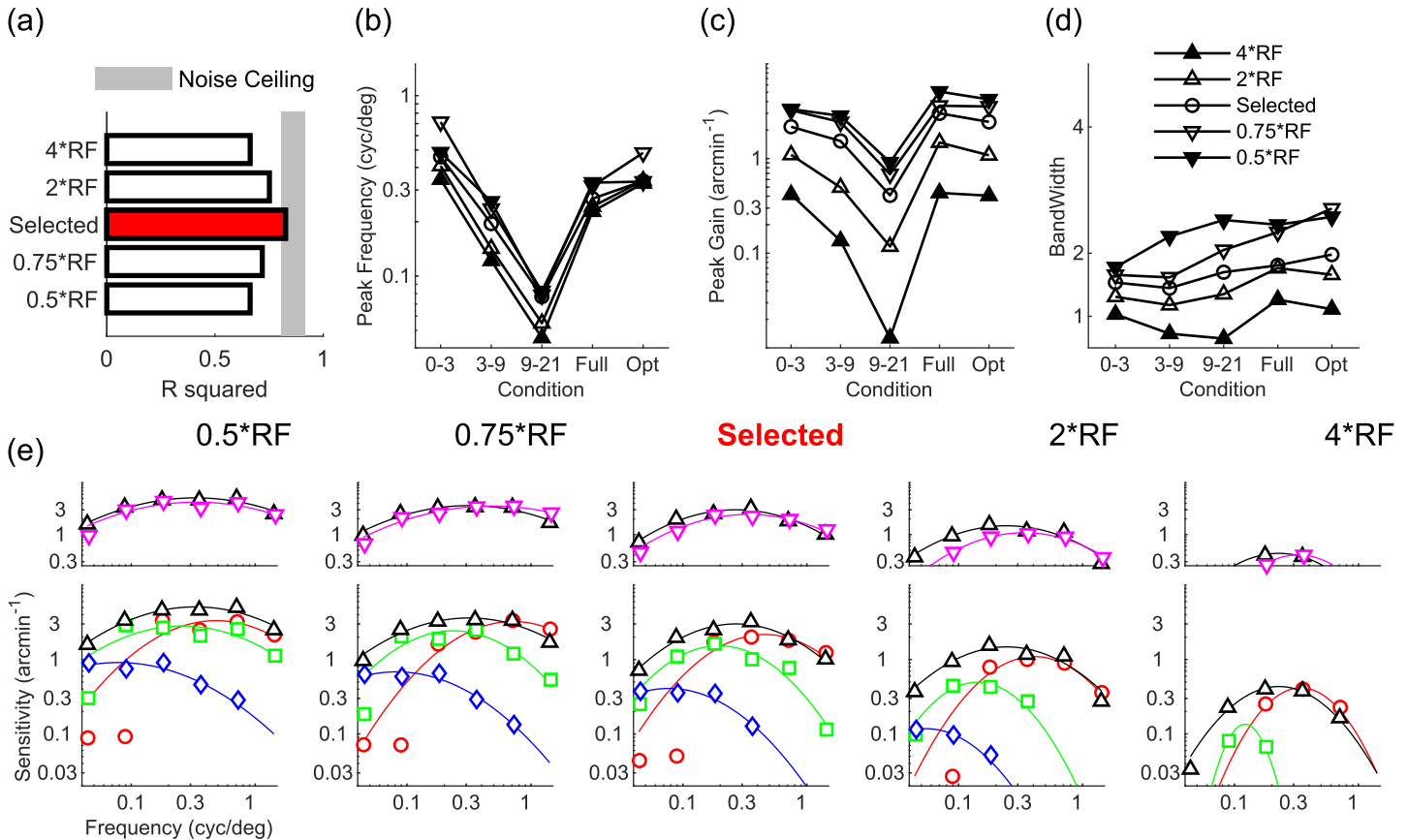
Fig 4. Effects of varying the model's cortical receptive field size. (a) Agreement (R^2) between human data and models with smaller and larger cortical receptive fields than the selected model (red bar). Shaded region represents the noise ceiling, an estimate of peak model performance (see Materials and methods). (b-d) Peak frequency, peak gain, and bandwidth of models with larger (upwards triangles) and smaller (downwards triangles) cortical receptive fields than the selected model (circle), for all visual field conditions tested. (e) Disparity sensitivity plotted as a function of spatial frequency as in Fig 1d for all tested models of varying cortical receptive field size.

increases sensitivity, and these effects are more pronounced in the periphery compared to the fovea. This is sensible, as the CR determines the rate of information loss moving into the visual periphery. Similarly therefore, disparity tuning narrows or widens when the CR is increased of decreased, and this effect is most pronounced in the visual periphery (Fig 3d). Fig 3e shows, for each tested CR value, the specific patterns of disparity sensitivity as a function of disparity corrugation spatial frequency and across visual field conditions.

Cortical receptive field size affects all aspects of disparity tuning

Fig 4a shows that increasing or decreasing the model's cortical receptive field size degrades but does not destroy the model's agreement with human data. Increasing or decreasing cortical receptive field size uniformly shifts tuning to lower or higher spatial frequencies respectively (Fig 4b). The model's overall disparity sensitivity decreases going from small to large cortical receptive field sizes (Fig 4c). Fig 4d also shows that disparity tuning narrows or widens with increasing and decreasing cortical receptive field sizes respectively, and this effect is more marked in the visual periphery. These shifts in frequency tuning (the specific patterns can be seen in Fig 4e) sensibly occur because smaller receptive fields better process high spatial frequencies.
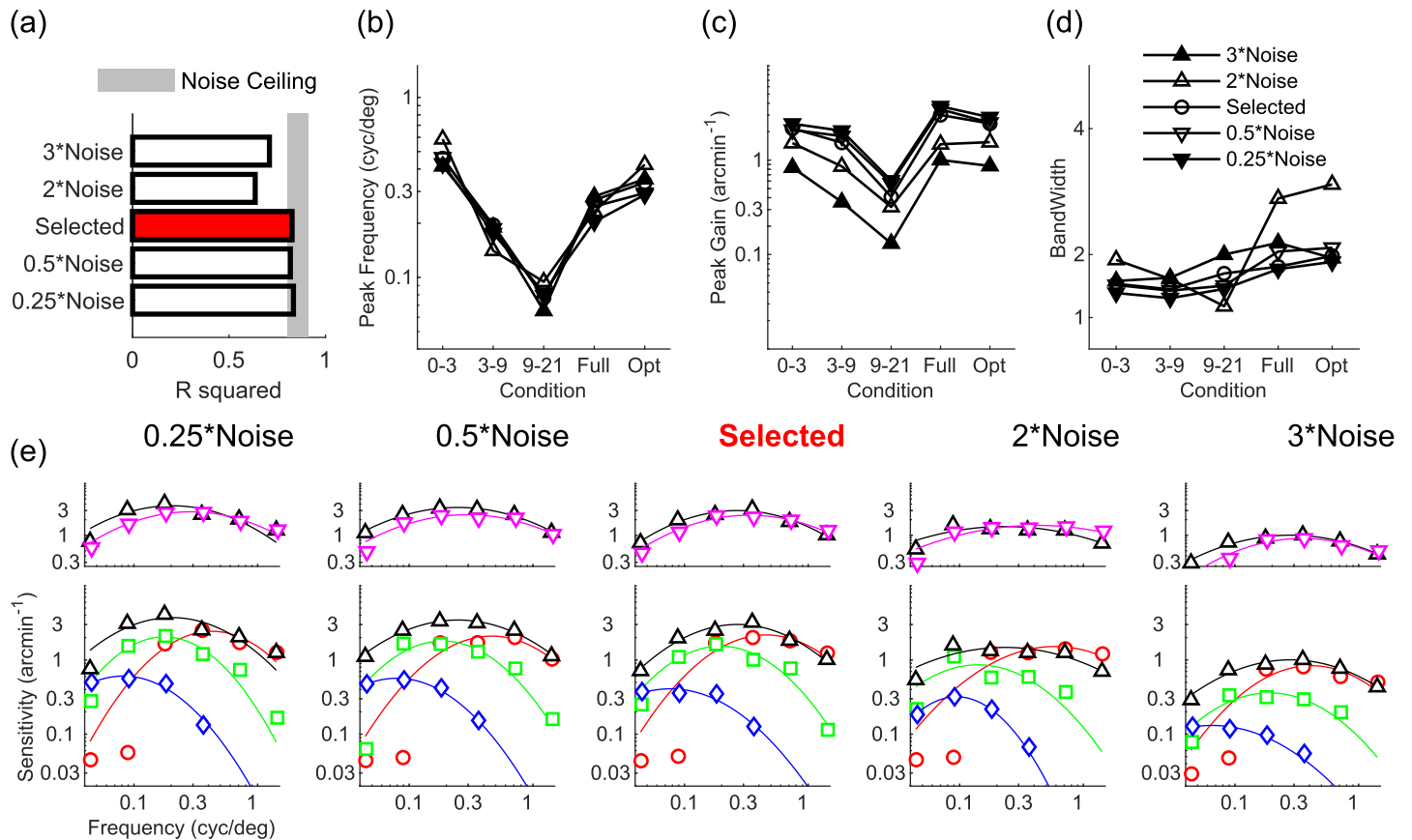
Fig 5. Effects of varying the model's simulated neural noise. (a) Agreement (R^2) between human data and models with smaller and larger simulated neural noise than the selected model (red bar). Shaded region represents the noise ceiling, an estimate of peak model performance (see Materials and methods). (b-d) Peak frequency, peak gain, and bandwidth of models with larger (upwards triangles) and smaller (downwards triangles) simulated neural noise than the selected model (circle), for all visual field conditions tested. (e) Disparity sensitivity plotted as a function of spatial frequency as in Fig 1d for all tested models of varying simulated neural noise.

https://doi.org/10.1371/journal.pcbi.1007699.g005

Simulated neural noise uniformly modulates disparity sensitivity

Fig 5a shows that decreasing the amount of simulated neural noise does not affect the model's agreement with human data, whereas increasing simulated noise degrades but does not destroy the model's agreement with human data. Increasing or decreasing simulated neural noise does not systematically affect the model's tuning frequency nor bandwidth (Fig 5b and 5d). The magnitude of simulated neural noise is instead inversely correlated with the model's peak sensitivity, independently of visual field location (Fig 5c). This uniform decrease in disparity sensitivity with simulated noise across spatial frequency and visual field conditions is evident in Fig 5e.

The log-polar stage of the computational model is crucial for replicating the patterns of human data

Fig 6a shows that adding spatial scales does not improve nor strongly degrade the model's agreement with human data. Conversely, a computational model without the log-polar processing stage (noLP) exhibits very low agreement with the human data, even if two distinct spatial scales are implemented (noLP/2S). Patterns of disparity tuning peak frequency (Fig 6b), peak gain (Fig 6c), and bandwidth (Fig 6d) are mostly unaffected by adding spatial scales on
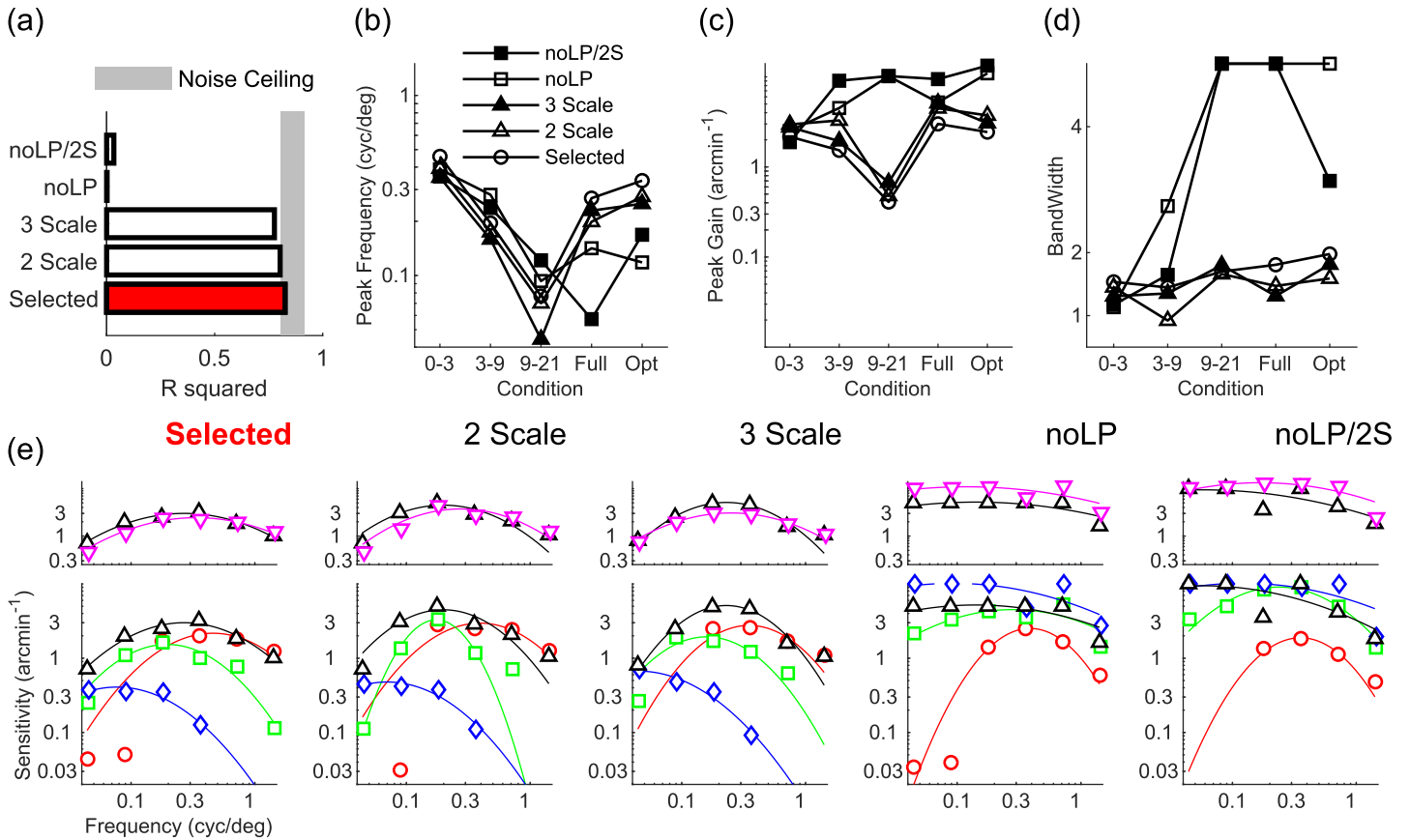
Fig 6. Effects of varying the model's architecture. (a) Agreement (R^2) between human data and models of varying architecture with respect to the selected one (red bar). Shaded region represents the noise ceiling, an estimate of peak model performance (see Materials and methods). (b-d) Peak frequency, peak gain, and bandwidth of models of varying architecture, for all visual field conditions tested. (e) Disparity sensitivity plotted as a function of spatial frequency as in Fig 1d for all tested models of of varying architecture.

https://doi.org/10.1371/journal.pcbi.1007699.g006

top of log-polar processing, contrary to what occurs when removing or replacing log-polar processing with pyramidal multi-scale processing. The two rightmost panels of Fig 6e in particular show how a computational model without the log-polar processing stage exhibits markedly different patterns of disparity sensitivity across the model's visual field. The non-log-polar models also show how stimulus configuration cannot account for the observed patterns of human data. Contrary to what occurs in humans, in the non-log-polar models performance is best in the far peripheral condition where the model can integrate disparity information across the largest image area. It is worth noting however that in all models *and humans* foveal disparity sensitivity falls off at the lowest spatial frequencies tested because the spatial extent of the foveal region cannot contain a full cycle of the disparity corrugation: central vision simply cannot process low spatial frequencies.

Discussion

Our human behavioural data demonstrate that different regions of the visual field preferentially process disparity at different disparity spatial scales. Our data broadly align with the shifts in spatial frequency tuning for depth reported by Prince and Rogers [17]. Furthermore, by approximately log scaling our stimuli, we show that the loss in peripheral sensitivity is not as

steep as that found with equally-sized annular stimuli that, unlike our stimuli, do not compensate for the change in sampling density across the visual field. Therefore, contrary to the common intuition that depth processing is best at the fovea, our results show that disparity sensitivity depends on both disparity spatial frequency and eccentricity. Disparity sensitivity to low and mid disparity spatial frequencies is higher in the far and near periphery respectively, than in the fovea.

This change in tuning across the visual field is remarkably similar to the change in naturally-occurring disparity statistics that have been reported for observers in natural indoor and outdoor environments [4, 26, 43]. We therefore speculate that the origin of this tuning may be related to the patterns of depth information the visual system has developed to process. Gibaldi and colleagues [26] were even able to correlate the empirical patterns of V1 receptive field size changes and cortical magnification [44] with the theoretical receptive field sizes required to cover the range of disparities experienced by participants actively exploring 3D visual scenes. Relatedly, we show that M-scaling and cortical receptive field size together determine the specific patterns of disparity tuning occurring throughout the visual field. This suggests the intriguing possibility that the specific structure of the retino-cortical transformation may arise, at least in part, from the disparity distributions experienced by humans as they actively visually explore the natural environment. Of course, our modeling shows that the connection between luminance and disparity scales is not a trivial one, as it depends upon the relationship between luminance scale tuning of binocular simple cells, receptive field size of binocular complex cells, and disparity frequency tuning of hypercyclopean channels. Additionally, cortical magnification most likely primarily arises from the necessity to trade-off acuity with sensitivity and the energy required to maintain high acuity vision, as it is observed across species, including animals that have no functional binocular vision or very small binocular overlap (e.g. mouse, rabbit) [45]. Given that disparity statistics are a function of both the arrangement of objects in the world and our own viewing parameters (i.e. interocular separation, viewing distance, visual exploration strategies), the mapping of disparity sensitivities onto a log-scaled representation of visual space may reflect the interdependent evolution of cortical magnification and the particular sets of disparities that humans have evolved to process.

We further observe that disparity information is integrated across the visual field in a near-optimal MLE fashion [19–25]. This finding informs how depth information at multiple scales is computed and combined across the visual field. Of course, in the natural environment, the perception of depth does not rely exclusively on binocular disparity, but is supported by several sources of visual information, such as linear perspective and motion parallax, that are combined into a unified depth percept [20]. These different cues likely have different reliability across different regions of the visual field. For example, defocus blur is a more variable cue to depth than disparity near the fovea [46], but disparity is more variable than blur away from fixation [47]. Here, we have only shown that within a single cue, binocular disparity, depth information is integrated near-optimally across different regions of the visual field. It remains to be seen whether depth information within and among different sources, such as blur, perspective and disparity, can be successfully or optimally integrated across the human visual field. It also remains unknown whether such integration would be weighted by the different patterns of reliability for different depth cues. Nevertheless, the possibility that multiple cues are integrated is supported by the observation that experiencing congruent blur and disparity information across the visual field facilitates binocular fusion compared with incongruent pairings [48].

The pattern of human disparity sensitivity that we observe is well captured by our biologically-motivated model of disparity processing that critically incorporates the log-polar retino-cortical transformation. It is generally accepted that our visual system processes disparity

along at least two [7–11] or more [12] channels that are selective for depth changes at different disparity spatial scales. These disparity spatial scales in turn may rely on distinct sets of luminance spatial channels [13–16] (as well as second-order channels [49–51]). A key insight provided by our work is that depth-selective channels emerge directly from the log-polar, retinocortical transform, since log-polar spatial sampling acts as a "sliding" multi-scale analysis, i.e. by design it processes different luminance (and consequently disparity) spatial scales at different image locations [31, 42].

By employing the log-polar transform, and thus a "sliding" multi-scale analysis, our model might help explain empirical observations beyond the ones tested in this work. For instance, not only does stereoacuity vary across the visual field, but also the upper disparity and binocular fusion limit (Panum's area) increase gradually with eccentricity [52–54]. Even though our model was not explicitly designed to account for this, since its receptive field sizes increase linearly towards the periphery, the model will gradually be able to estimate larger disparity ranges in its periphery. Additionally, since the model's receptive field density decreases with eccentricity (due to the log-polar sampling), the model's internal noise is effectively less averaged away in the periphery. Therefore, this may also explain the empirical observation that peripheral stereoacuity is limited by internal noise [55].

Materials and methods

Ethics statement

All methods were approved by the Internal Review Board of Northeastern University and adhered to the tenets of the Declaration of Helsinki. Informed consent was obtained from all human participants.

Disparity sensitivity in human and model observers

Participants. Author GM and nine naïve observers, (6 female, mean ±sd age: 24±6) participated in the study. All participants had normal or corrected to normal vision and normal stereo vision. Prior to testing, participants were screened using the Titmus stereopsis test and only participants with stereoacuity of 40 arcseconds or better were included in the study.

Apparatus. The experiment was programmed with the Psychophysics Toolbox Version 3 [56, 57] in Matlab (MathWorks). Stimuli were presented on an BenQ XL2720Z LCD monitor with a resolution of 1920 × 1080 pixels (display dot pitch 0.311 mm) at 120 Hz. The monitor was run from an NVidia Quadro K 420 graphics processing unit. Observers were seated in a dimly lit room, 45 cm in front of the monitor with their heads stabilized in a chin and forehead rest and wore active stereoscopic shutter-glasses (NVIDIA 3DVision) to control dichoptic stimulus presentation. The cross talk of the dichoptic system was 1% measured with a Spectrascan 6500 photometer.

Stimuli. Stimuli were 1/f pink noise stereograms presented on a uniformly gray background; examples for each experimental condition are shown in Fig 1a. The stimuli contained oblique (45 or 135 degrees) sinusoidal disparity corrugations of varying amplitude and spatial frequency, generated as in [38] (see also [58]). The stimuli were presented as disks or rings with 1 degree cosinusoidal edges. The central fixation target was a 0.25 degree black disk with 0.125 degree cosinusoidal edge. In pilot testing, we verified that it was not possible to perform the experiment without dichoptic stimulus presentation (i.e. the oblique sinusoidal corrugation did not generate visible compression and expansion artifacts in the pink noise patterns).

Procedure. Each trial, observers were presented with a black fixation dot on a uniformly gray background. As soon as the response from the previous trial had been recorded, the stimulus for the current trial was shown for 250 milliseconds. This was too brief a time for

observers to benefit from changes in fixation, since stimulus-driven saccade latencies are on average greater than 200 ms [59], saccade durations range from 20 to 200 ms [60], and visual sensitivity is reduced during and after a saccade [61, 62]. Once the stimulus had been extinguished, observers were required to indicate, via button press, whether the disparity corrugation was top-tilted leftwards or rightwards. Observers were given unlimited time to respond. The following trial commenced as soon as observers provided a response. Each trial, the amount of peak-to-trough disparity was under the control of a three-down, one-up staircase [63] that adjusted the disparity magnitude to a level that produced 79% correct responses.

Design. We measured how observer's disparity sensitivity (1/disparity threshold) varied, as a function of the spatial frequency of the sinusoidal disparity corrugation, throughout different portions of the visual field. We tested four visual field conditions. In the central visual field condition, stimuli were presented within a disk with a 3 degree radius centered at fixation. In the near and far peripheral visual field conditions, stimuli were presented within rings spanning 3-9 and 9-21 degrees into the visual periphery, respectively. Lastly, in the full visual field condition, stimuli were presented within a disk with a 21 degree radius, and thus spanned the full extent of the visual field tested in this study. In each condition, we measured disparity thresholds at six spatial frequencies: 0.04 0.09, 0.18, 0.35, 0.71, 1.41 cycles/degree. Thresholds were measured via 24 randomly interleaved staircases [63]. The raw data from 75 trials from each staircase were combined and fitted with a cumulative normal function by weighted least-squares regression (in which the data are weighted by their binomial standard deviation). Disparity discrimination thresholds were estimated from the 75% correct point of the psychometric function.

It is well known that disparity sensitivity varies lawfully as a function of spatial frequency following a bell-shaped function [64, 65]. This function is well described by a log-parabola model [38]. Therefore, we first converted disparity threshold estimates into disparity sensitivity (sensitivity = 1/threshold). Then, for each visual field condition, we fit the sensitivity data to a three-parameter log parabola Disparity Sensitivity Function (DSF) [38, 66] defined as:

$$DSF(f) = \log_{10}(\gamma_{max}) - \log_{10}(2)\left(\frac{\log_{10}(f) - \log_{10}(f_{max})}{\log_{10}(2\beta)/2}\right)^2 \tag{1}$$

where γ_{max} represents the peak gain (i.e. peak sensitivity), f_{max} is the peak frequency (i.e. the spatial frequency at which the peak gain occurs), and β is the bandwidth at half height (in octaves) of the function. The sensitivity data were fit to this equation, via least-squares regression, to obtain parameter estimates that could then be compared across experimental conditions.

Optimal integration model

It is unknown whether observers are able to combine binocular disparity information across different portions of the visual field. If this were the case, then the DSF estimated for the full visual field condition should be the envelope of the DSFs estimated in the restricted visual field conditions. We obtained an estimate of the upper bound of performance in the full visual field condition by designing an observer that optimally combines disparity information across the different portions of the visual field following a maximum-likelihood estimate (MLE) rule [25]. Let us assume each visual field region v can provide a disparity estimate $\hat{d}_v$, and that these estimates are corrupted by early, independent Gaussian noise with variance σ_v^2. If the Bayesian prior is uniform, then the maximum-likelihood disparity estimate across the full field is

$\hat{d}_{FF} = \sum_v w_v \hat{d}_v$, with $w_v = \frac{1/\sigma_v^2}{\sum_u 1/\sigma_u^2}$ and the variance of the full field estimate is $\sigma_{FF}^2 = \frac{1}{\sum_v 1/\sigma_v^2}$.

Adding the disparity estimates weighted by their normalized reciprocal variances produces the optimal, lowest-variance disparity estimate possible. Since thresholds are directly proportional to the standard deviation of the underlying estimator, according to the MLE method the disparity thresholds in the full-field condition should be lower (i.e. sensitivity should be higher) than in the restricted visual field conditions, following the rule:

$$
\frac{1}{T^2_{FF-Opt}} = \frac{1}{T^2_{0-3}} + \frac{1}{T^2_{3-9}} + \frac{1}{T^2_{9-21}} \Rightarrow
$$

$$
T_{FF-Opt} = \sqrt{\frac{T^2_{0-3} T^2_{3-9} T^2_{9-21}}{T^2_{0-3} T^2_{3-9} + T^2_{0-3} T^2_{9-21} + T^2_{3-9} T^2_{9-21}}}
$$

(2)

Therefore, we estimated the optimal disparity sensitivities as $1/T_{FF-Opt}$ at each tested spatial frequency. Then, we fit these optimal sensitivity data to the same DSF from Eq 1 to obtain DSF parameter estimates for an optimal integrator that could be compared to the DSF parameter estimates for the full field stimuli.

Statistical analyses

To test whether disparity sensitivity varied across the visual field of human observers, DSF parameter estimates from the restricted visual field conditions were analyzed using a one-way, within-subject Analysis of Variance (ANOVA). ANOVA normality assumptions were verified with Quantile-Quantile plots. Paired t-tests on the DSF parameter estimates were employed to test whether full field DSFs differed from MLE-optimal DSFs. To compare the computational model (described below) to human performance, we computed the square of the Pearson correlation r between the average human disparity sensitivity estimates and the model disparity sensitivity. To provide an estimate of peak model performance, we computed the correlation of each participant's disparity sensitivity estimates to the average of all other participants. We defined the squared, 95% bootstrapped confidence intervals of the mean between-participant correlation as the noise ceiling. Fisher's Z transformation was employed on the correlation values to ensure variance stabilization when computing confidence intervals of mean correlation [67]. If the model's agreement with human data were to fall within this noise ceiling, the model disparity sensitivity patterns would be essentially indistinguishable from those of a random human participant.

Foveated, image-computable model of disparity processing

We developed a biologically-inspired computational model that implements plausible neural processing stages underlying disparity computation in humans. The computational model mimics the dorsal visual pathway from the retinae to the middle temporal (MT) visual area [68, 69]. Critically, the model incorporates a biologically-plausible front end that approximates the space-variant sampling of the human retina. We hypothesized that this space-variant retinal sampling is responsible for the observed shifts in disparity tuning occurring across the visual field of human participants.

The computational model can be summarized as follows:

- a space variant front-end, i.e. the log-polar mapping that samples standard Cartesian stereo images;

- hierarchical neural processing layers for disparity estimation, based on V1 binocular energy complex cells and an MT distributed representation of disparity;

- a layer to take into account the optimal combination of disparity across annular regions of the visual field;

- a decoding layer in order to assess the encoded disparity into the cortical distributed representation.

Since the first processing stage is intended to mimic human retinal sampling, it consists of a log-polar transformation [28, 31] that maps standard Cartesian images onto a cortical image representation.

For disparity estimation we employ a feed-forward neural model that computes vector disparity [39]. This model can be directly applied on cortical images, since 2D vector disparity is computed without explicitly searching for image correspondences along epipolar lines. This allows us to discount the fact that straight lines in the Cartesian domain become curves in log-polar space [70], and this approach also does not require knowledge of the current pose of the stereo system (i.e. ocular vergence), even though in-principle this information could improve disparity estimation. Although disparities on the retina are predominantly horizontal, retino-cortical warping makes a vector representation of cortical disparity necessary. Fig 7 exemplifies this point: Even a simple horizontal (1D) disparity pattern is warped in the cortical domain. Therefore, to characterize properly a non vector (1D) Cartesian disparity pattern in cortical coordinates, a vector representation of cortical disparity is required.

To mimic the near-optimal combination of disparity information across different portions of the visual field of human participants, we consider a simple pooling mechanism that combines neural activity across annular regions of the model's visual field.
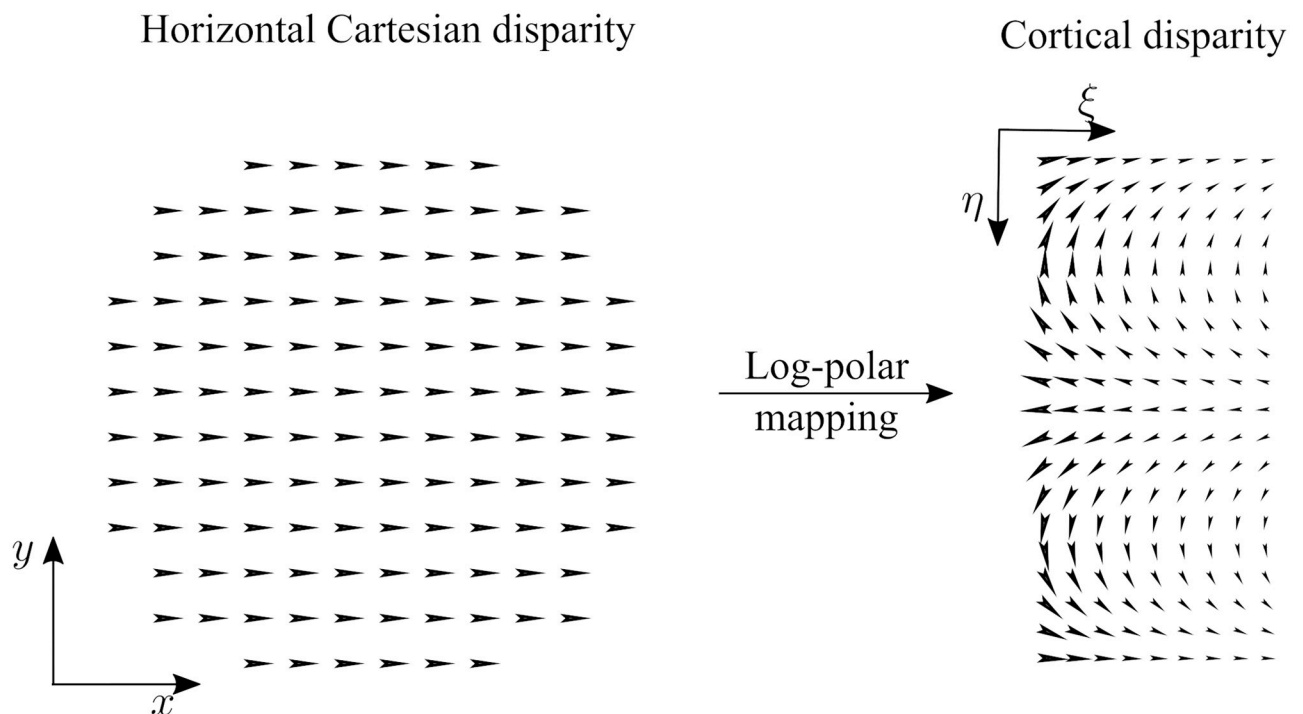
Fig 7. Retino-cortical disparity warping. (Left) A horizontal constant disparity map $d_x(x, y)$ that describes the horizontal shift between the left and right image in Cartesian domain can be considered as a vector disparity $\delta(x, y) = (d_x, 0)$, (i.e. horizontal vectors of constant magnitude) by considering the disparity map as the first component of the vector. (Right) The horizontal constant disparity vector field is warped in the cortical domain in a way that produces disparity vectors in several cortical directions, thus requiring a vector representation (see next Section for details about the log-polar mapping).

To compare the model to human disparity processing, we decode the model's distributed cortical activity and quantify the encoded disparity information. Even though this decoding stage is biologically plausible, we do not claim that it models the perceptual decision stage. We only employ this decoding stage to assess whether disparity estimation in the proposed model leads to patterns of disparity sensitivity similar to those measured in human participants.

Retino-cortical mapping.   To mimic the retino-cortical mapping of the primate visual system that provides a space-variant representation of the visual scene, we employ the central blind-spot model: each Cartesian image is transformed into its cortical representation through a log-polar transformation [28, 32, 71–73]. We chose this specific model with respect to other models in the literature (e.g. [74]) for several reasons: it captures the essential aspects of the retino-cortical mapping, it can be implemented efficiently, it provides a good preservation of image information [75, 76], and it allows us to provide an analytic description of cortical processing.

In the central blind-spot model, the mapping $\mathbf{T} : (x, y) \mapsto (\xi, \eta)$ from the Cartesian domain (x, y) to the cortical domain of coordinates (ξ, η) is described by the following equations:

$$\begin{cases} \xi = \log_a\left(\frac{\rho}{\rho_0}\right) \\ \eta = q\vartheta, \end{cases} \tag{3}$$

where a parameterizes the non-linearity of the mapping, q is related to the angular resolution, ρ_0 is the radius of the central blind spot, and $(\rho, \vartheta) = (\sqrt{x^2 + y^2}, \arctan(y/x))$ are the polar coordinates derived from the Cartesian ones. All points with $\rho < \rho_0$ are ignored (hence the central blind spot).

Discrete log-polar mapping.   Our aim was to test the model using the same experimental stimuli and procedures employed with human observers. Therefore, the log-polar transformation must be applied to digital images. Given a Cartesian image of $N_c \times N_r$ pixels, and defined $\rho_{max} = 0.5\min(N_c, N_r)$, we obtain an $R \times S$ (rings × sectors) discrete cortical image of coordinates (u, v) by taking:

$$\begin{cases} u = \lfloor \xi \rfloor \\ v = \lfloor \eta \rfloor, \end{cases} \tag{4}$$

where $\lfloor \cdot \rfloor$ denotes the integer part, $q = S/(2\pi)$, and the non-linearity of the mapping is $a = (\rho_{max}/\rho_0)^{1/R}$.

Fig 8 shows the log-polar pixels, which can be thought of as the log-polar receptive fields, in the Cartesian domain (Fig 8b) and in the cortical domain (Fig 8c): the Cartesian area (i.e. the log-polar pixel) that refers to a given cortical pixel defines the cortical pixel's receptive field. The non-linearity of the log-polar transformation can be described as follows: by referring to Fig 8b and 8c, a uniform (green) row of cortical units is mapped to a (green) sector of space variant receptive fields, and a vertical (cyan) column of cortical units is mapped to a (cyan) circular set of uniform receptive fields. By inverting Eq 3 the centers of the receptive fields can be computed, and these points present a non-uniform distribution throughout the retinal plane (see the yellow circles overlying the Cartesian images in Fig 8a). The magenta circular curve in Fig 8b, with radius $S/2\pi$, represents the locus where the size of log-polar pixels is equal to the size of Cartesian pixels. In particular, in the area inside the magenta circular curve (the fovea) a single Cartesian pixel contributes to many log-polar pixels (oversampling), whereas outside this region multiple Cartesian pixels will contribute to a single log-polar pixel. To avoid spatial aliasing due to the undersampling, we employ overlapping receptive fields. Specifically, we use
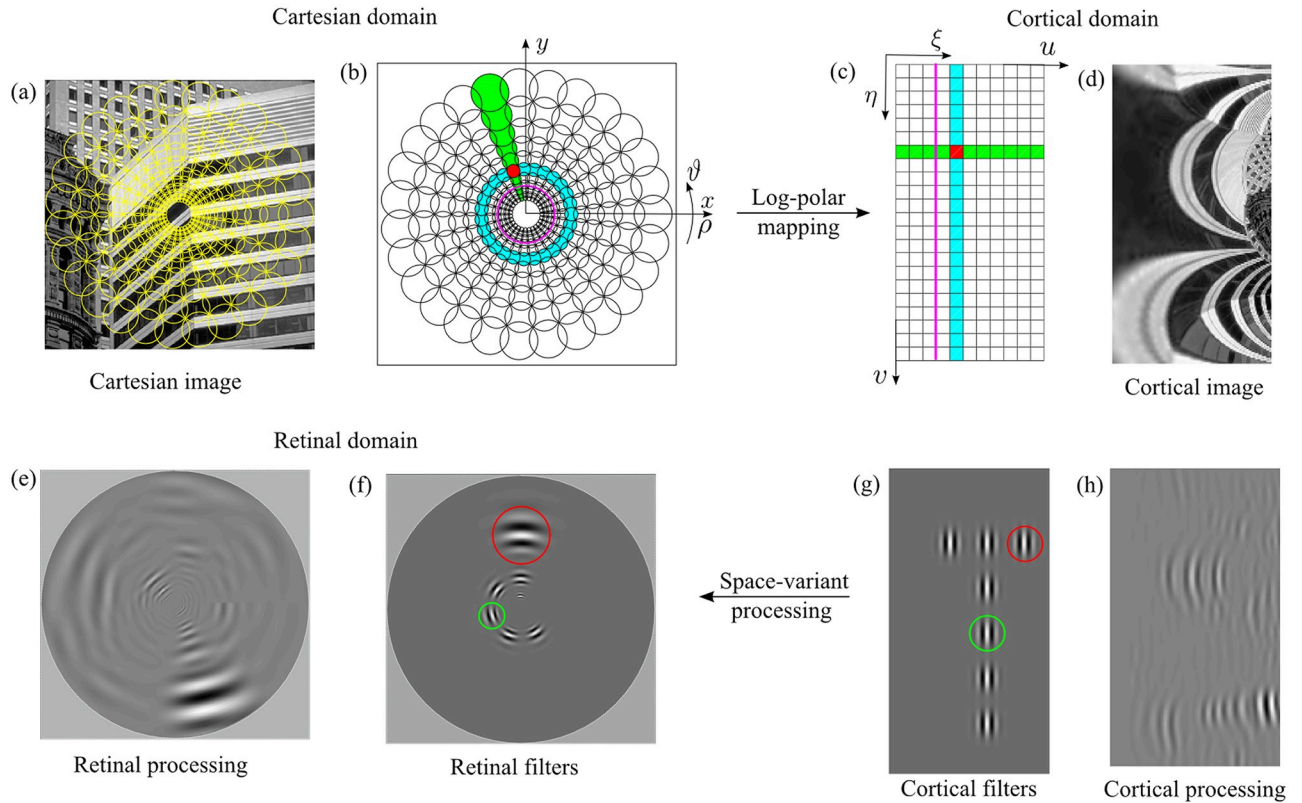
Fig 8. Log-polar mapping and cortical processing. (top) Log-polar mapping scheme for the central blind-spot model (Eq 3). (a) A standard Cartesian image with overlying log-polar pixels, the receptive fields (yellow circles). (b) Cartesian domain with the superposition of the circular overlapping log-polar receptive fields and (c) the corresponding cortical domain, where the squares denote the neural units. The green sector of receptive fields map to the horizontal row of (green) neural units and the cyan circle of receptive fields to a column of (cyan) neural units. The magenta circle delimits the oversampling (fovea) and undersampling areas (periphery). (d) The cortical representation of the standard Cartesian image. The cortical image is zoomed to improve the visualization. (bottom) A uniform processing in the cortical domain maps to a space-variant processing in the retinal domain. (a) The retinal space variant filtered image that is the backward mapping of the cortical uniform filtered image of subfigure (h). (f) The retinal filters that correspond to the filters in the cortical domain (g): a uniform filtering in the cortical domain results in a space-variant filtering operation in the retinal domain, where both the scale (red circle) and the orientation (green circle) of the filters vary. (h) The cortical filtered image obtained by applying the filter depicted in subfigure (g) on the cortical image shown in subfigure (d). The specific values of the log-polar parameters are: $R = 130$, $S = 203$, $\rho_o = 3$, $CR = 3.9$, $W_{max} = 4.8$. The spatial support of the filter is 31×31 cortical pixels.

https://doi.org/10.1371/journal.pcbi.1007699.g008

overlapping circular Gaussian receptive fields [77, 78], which are the most biologically plausible and optimally preserve image information [75]. An example of a transformation from Cartesian to cortical domain is shown in Fig 8a and 8d. The cortical image (Fig 8d) clearly demonstrates the non-linear effects of the log-polar mapping.

This discrete log-polar mapping provides a significant data reduction while preserving a large field of view and high resolution at the fovea [31, 79, 80]. To characterize the amount of data reduction provided by this transformation, we can can define the compression ratio (*CR*) of the cortical image with respect to the Cartesian one as:

$$CR = \frac{N_c \times N_r}{R \times S} \tag{5}$$

This compression ratio *CR* thus describes the data reduction occurring in the human visual system (that our computational model mimics), and will also affect the execution time of the simulated model.

The log-polar transformation models the space variant image resolution: the size of the receptive fields increases as a function of the eccentricity (the distance between the center of the receptive field and the fovea). We can define the relationship between the receptive field size (in particular, the maximum receptive field size W_{max}) and the parameters of the mapping as follows:

$$W_{max} = \rho_0 a^R (1 - a^{-1}).$$

(6)

Eq 6 provides a measure of the scale at which the periphery of the Cartesian image is processed. Moreover, the parameters of the log-polar mapping also influence the proportion of cortical units used to over-represent the fovea: we can define the percentage of the cortical area used to represent the fovea (χ). This can be derived from Eq 6 by setting the receptive field size to 1 and inverting the equation to find the corresponding u (see Eq 4), and by then dividing by the overall size of the modeled cortex R:

$$\chi = (1 - \log_a(\rho_0(a-1)))/R.$$

(7)

By exploiting Eqs 6 and 7 we can control the growth of the size of the receptive fields and the over-representation of the fovea in order to reproduce data from the literature on the size-to-eccentricity relationship [73, 81, 82].

Cortical processing. In the human visual system, visual processing is performed by networks of units (cells) described by their receptive fields. This neural network can be approximated by sets of filter banks whose responses to visual stimuli mimic those of neurons throughout the human visual system. The proposed model for disparity estimation could therefore embed the processing of V1 binocular simple units directly into the log-polar receptive fields. Specifically, the log-polar transform could be modified by using, as receptive fields, filters that perform V1-like feature extraction. However, to minimize the model's computational load, we can consider that filter banks embedded in the log-polar transform can be "implemented" as a filtering process applied directly to the cortical image [31, 83]. We can demonstrate that the extraction of visual features can be carried out directly in the cortical domain by using solutions developed for the Cartesian domain without any modifications. To do so, in the following we analyze the relationships between the different parameters of a discrete log-polar mapping and of a bank of multi-scale and multi-orientation band-pass filters [84].

To maintain equivalence between Cartesian and cortical visual processing, the discreet log polar mapping should provide an isotropic sampling of Cartesian coordinates. To avoid anisotropy, circular sampling must be (locally) equal to radial sampling, since the cortical space consists of a uniform network of neural units. Sampling points can be derived by considering the inverse of the cortical mapping $\mathbf{T}$ (Eq 3). Specifically, the circular sampling interval is $(2\pi/S)\rho_0 a^{u-1}$ and the radial sampling interval is $\rho_0 a^{u-1}(a-1)$. To maintain isotropic sampling these sampling intervals must be equal, therefore the relationship between rings and sectors of the log-polar mapping must follow the rule:

$$S = 2\pi(\rho_{max}/\rho_0)^{-1/R}.$$

(8)

From a geometric point of view, the optimal relationship between R and S, expressed by Eq 8, is the one that optimizes the log-polar pixel aspect ratio making it as close as possible to 1.

The receptive fields of V1 simple cells are classically modeled as band-pass filters [85], thus we define the following complex-valued Gabor filter [86]:

$$g(\xi, \eta, \theta, \sigma, \psi) = A e^{(-(\xi^2+\eta^2)/2\sigma^2)} e^{j2\pi(f_s\cos(\theta)\xi + f_s\sin(\theta)\eta + \psi)},$$

(9)

where σ defines the spatial scale, f_s the peak spatial frequency, and ψ is the phase of the sinusoidal modulation. By considering filters that are normalized by their energy, we have $A = (\sqrt{\pi}\sigma)^{-1}$.

In order to process the cortically-transformed images, it is necessary to characterize the filters, defined in the Cartesian domain, with respect to the cortical domain, i.e. to map the filters into the cortical domain, thus obtaining $g(x(\xi, \eta), y(\xi, \eta), \theta, \sigma, \psi)$. As a consequence of the nonlinearity of the log-polar mapping, the mapped filters are distorted [87, 88], thus a filtering operation directly in the cortical domain could introduce undesired distortions in the outputs. Here, we show that under specific conditions these distortions can be kept to a minimum: under these assumptions, it is possible to directly work in the cortical domain, by considering spatial filters sampled in log-polar coordinates $g(\xi, \eta, \theta, \sigma, \psi)$.

At a global level (e.g. see Fig 8d) log-polar transformed images exhibit large distortions. However, we can consider what occurs at a more local level, at the scale of the receptive field of a single Gabor filter. First, we consider that the log-polar mapping can be expressed in terms of general coordinates transformation [89], thus the Jacobian matrix of the coordinates transformation allows us to describe how the receptive field locally changes. Specifically, the scalar coefficient $\rho_0\, a^{\xi}\ln(a)$ represents the scale factor of the log-polar vector, and the matrix describes the rotation η due to the mapping. Fig 8g shows a set of cortical filters and Fig 8f their retinal counterpart (i.e. the inverse log-polar transform): the red circle highlights the scale factor (i.e. the spatial support) of the filter and the green one its rotation. It is worth to note that the column of equally oriented filters in the cortical domain maps on a circle of filters in the retinal domain and each retinal filter is also at a different orientation. Specifically, vertically-oriented filters on the cortex correspond to azimuthally/tangentially-oriented filters on the retina; horizontally-oriented filters on the cortex correspond to radially-oriented filters on the retina.

Next, we want to analyze how the distortion affects the receptive field shape as a function of the distance from its center $p_0 = (\xi_0, \eta_0)$: we can consider that the ratio $g(x(\xi, \eta), y(\xi, \eta), \theta, \sigma, \psi)/g(\xi, \eta, \theta, \sigma, \psi)$ around a given point should be equal to 1. Since the filter $g(\cdot)$ is an exponential function, we can evaluate the difference $h(\cdot)$ between their arguments. We can approximate such a difference by using a Taylor expansion of a multi-variable function:

$$h(p) \approx h(p_0) + (p - p_0)^T \nabla h(p_0) + 0.5(p - p_0)^T H(p_0)(p - p_0), \qquad (10)$$

where $(\cdot)^T$ denotes the transpose, and $H(\cdot)$ the Hessian matrix. In the following we only focus on the terms that are relevant to describe how the distortion affects the receptive field shape: essentially, this depends on the partial derivatives of $(x(\xi, \eta), y(\xi, \eta))$ that constitute the gradient and the Hessian of $h(\cdot)$. The first order term takes into account how the mapping depends on the spatial position of the receptive field center. Indeed, the gradient has terms that are in common with the Jacobian matrix of the coordinates transformation, thus it describes the scale factor and the rotation of the receptive field as a function of the position p_0. The approximation error can be expressed by the second order term of the Taylor expansion: thus, there is an error that increases as a quadratic function of the distance $p - p_0$ (i.e. from the receptive field center), and an error that depends on the Hessian matrix that is related to the log-polar parameters. For instance, the mixed partial derivative of $x(\xi, \eta)$ is $\rho_0\ln(a)a^{\xi}\sin(\eta)$, thus we can consider that the error related to the log-polar parameters is proportional to $\rho_0\ln(a) = (\rho_0/R)\ln(\rho_{max}/\rho_0)$. It increases as a function of ρ_0 (given a fixed ρ_{max}) and decreases as R increases, which in turn decreases the compression ratio (Eq 5). Fig 8f and 8g shows that such distortions can be negligible, though the spatial support of the displayed filters is large for sake of visualization. Fig 8h shows the cortical image (Fig 8d) filtered by the filter that is drawn in different

cortical positions in Fig 8g. In Fig 8e the retinal (i.e. space variant) processing is shown, which is obtained through the inverse log-polar mapping of Fig 8h.

Cortical computational model of disparity estimation.   We consider a pair of (grayscale) cortical images $I^L(p)$ and $I^R(p)$, for all positions $p = (\xi, \eta)$ that are the cortical representations of an input stereo pair of Cartesian images. Our goal is to define a computational model that is able to encode in its cortical activity the information related to the disparity present in the Cartesian images. The cortical images are a warped version of the Cartesian images. The representation of disparity is a vector quantity. We thus define the disparity map $\delta(p) = (d_\xi, d_\eta)(p)$ as the difference between the pair of cortical images at each position p. To compute this cortical disparity map, the proposed model is composed of several processing stages.

V1 binocular energy computation and normalization.   In the proposed model we consider two sub-populations of neurons at the V1 level: binocular simple cells and complex cells. V1 simple cells are characterized by a preferred spatial orientation θ and a preferred phase difference $\Delta\psi$ between the left- and right-eye components of a cell's receptive field. We model the receptive fields of V1 simple cells as Gabor filters (see Eq 9). The spatial support of the filters is defined as a function of their spatial radial peak frequency f_s and bandwidth B: $\sigma = f_s^{-1}(2^B + 1)(2^B - 1)^{-1}$. We consider one standard deviation of the amplitude spectrum as the cut-off frequency.

Following the phase-shift model [33, 34], we define the receptive fields of the binocular simple cell as $S^L(p, \theta, \sigma, \psi^L) = \Re[g^L(p, \theta, \sigma, \psi^L)]$ and $S^R(p, \theta, \sigma, \psi^R) = \Re[g^R(p, \theta, \sigma, \psi^R)]$. These receptive fields are centered at the same position in the left- and right-eye images, and have a binocular phase difference $\Delta\psi = \psi^L - \psi^R$. For each spatial orientation, a set of K binocular phase differences are chosen to obtain tuning to different disparities: $d = \Delta\psi/f_s$.

We define the response of binocular simple cells as

$$R(p, \theta, \sigma, \Delta\psi) = (S^L(\cdot, \theta, \sigma, \psi^L) \overset{(\xi,\eta)}{*} I^L)(p) + (S^R(\cdot, \theta, \sigma, \psi^R) \overset{(\xi,\eta)}{*} I^R)(p). \qquad (11)$$

We can compute the response $R_q(p, \theta, \sigma, \Delta\psi)$ of a quadrature binocular simple cell by using the imaginary part of the Gabor filters.

The response of a complex cell is described by the binocular energy (the sum of the squared responses of a quadrature pair of binocular simple cells) [33, 35, 36]:

$$E(p, \theta, \sigma, d) = R(p, \theta, \sigma, \Delta\psi)^2 + R_q(p, \theta, \sigma, \Delta\psi)^2, \qquad (12)$$

by considering that $d = \Delta\psi/f_s$. By taking into account the extensions of the binocular energy model proposed in [90, 91], we apply a static non-linearity to the complex cell response described in Eq 12.

The response of the V1 layer of our model, when considering a finite set of orientations $\theta = \theta_1 \ldots \theta_N$, can be defined, through a divisive normalization to remove confounds due to variations in the local amount of contrast [92, 93], as

$$E^{V1}(p, \theta, d) = \frac{E(p, \theta, \sigma, d)^{0.5}}{\sum_{i=1}^{N} E(p, \theta_i, \sigma, d)^{0.5} + \varepsilon}, \qquad (13)$$

where $0 < \varepsilon \ll 1$ is a small constant to avoid dividing by zero in regions where no binocular energy is computed (i.e. no texture is present). For simplicity we omit from the notation the spatial scale σ. At this level, V1 responses are tuned to the spatial orientation and magnitude of the stimulus. The model neurons are tuned to disparity orthogonal to their orientation on the cortex; e.g. a horizontally-oriented cortical RF is tuned to the radial component of retinal disparity. It's important to recognise that the tuning is to 1D disparity—a cell will respond

strongly if the component of stimulus disparity along its preferred direction matches the magnitude of disparity that the cell is tuned to, regardless of stimulus disparity in the orthogonal direction.

In order to mimic natural neural activity, we consider that neural noise is present [90]. We model this neural noise as: $E^{V1}(p, \theta, d) = E^{V1}(p, \theta, d) + n_{V1}(p)$. The noise is uniformly distributed and its value is a fraction of the local average neural activity.

MT cells response.   Orientation-independent disparity tuning is obtained at the MT level of the model by pooling afferent V1 responses in the spatial and orientation domains, followed by a non-linearity [39, 94].

The responses of an MT cell, tuned to the magnitude d and direction ϕ of the vector disparity δ, can be expressed as follows:

$$E^{MT}(p, \phi, d) = F\left(\lambda \sum_{i=1}^{N} w_\phi(\theta_i)(G_{\sigma_{pool}} \overset{\xi,\eta}{*} E^{V1}(\cdot, \theta_i, d))(p)\right), \tag{14}$$

where $G_{\sigma_{pool}}$ denotes a Gaussian kernel (standard deviation σ_{pool}) for the spatial pooling, $F(s) = exp(s)$ is a static non-linearity, specifically an exponential function [39, 92], λ is the gain of the non-linearity, and w_ϕ represents the MT linear weights that give origin to the MT tuning. Spatial pooling accounts for the fact that MT receptive fields are larger than V1 receptive fields, and has the effect of improving the accuracy of disparity estimation [39]. The static non-linearity is employed since linear models fail to account for the response patterns of MT cells, whereas an exponential nonlinearity provides a good description of the MT firing patterns [92] and improves the accuracy of disparity estimation [39].

Similarly to what occurs at the V1 layer, we model neural noise at the MT level as: $E^{MT}(p, \phi, d) = E^{MT}(p, \phi, d) + n_{MT}(p)$.

Experimental evidence suggests that w_ϕ is a smooth function with central excitation and lateral inhibition. Therefore, by considering the MT linear weights shown in [92], we define $w_\phi(\theta)$ as

$$w_\phi(\theta) = \cos(\phi - \theta) \quad \phi \in [0, 2\pi]. \tag{15}$$

Vector disparity is thus encoded as a distributed representation through a population of MT neurons that span over the 2-D disparity space with a preferred set of tuning directions ($\phi = \phi_1 \ldots \phi_P$) in $[0, 2\pi]$ and tuning magnitudes ($d = d_1 \ldots d_K$). Thus, this processing stage contributes to represent the disparity stimulus in terms of its parameters, i.e. directions and magnitude, with respect to the V1 representation of the stimulus that is described in terms of the cells' parameters.

Such a representation mimics the neural distributed representation of information. However, from a computational point of view, cosine functions shifted over various orientations (see Eq 15) are described by the linear combination of an orthonormal basis (i.e., sine and cosine functions). Thus, all the V1 afferent information can be encoded by a population of MT neurons tuned to the directions $\phi = 0$ and $\phi = \pi/2$, only, with varying tuning magnitudes (see Eq 14).

This observation may help account for the larger selectivity for horizontal disparity reported in the literature [95–97]. Since a neural population tuned to two directions (at an angular difference of $\phi = \pi/2$) can encode the full vector disparity, a neural population of MT units tuned to a retinal disparity range slightly larger than $[-\pi/4, \pi/4]$ is able to recover the full vector disparity, i.e. a population of MT cells tuned around the horizontal axis might account also for the selectivity to vertical disparity [39].

Our model implementation however does not incorporate this anisotropy, nor does it account for the fact that the anisotropy between horizontal and vertical disparity tuning has been found already at the V1 level [98]. Indeed, our model is not meant to incorporate all known properties of V1 (such as the differences in crossed/uncrossed disparity tuning across upper and lower visual field [43]). However, we highlight how the vertical/horizontal anisotropy may arise at the MT layer, since this is where we have orientation-independent disparity tuning and is therefore where we can first explicitly estimate vector disparity.

Multi-scale analysis. A standard approach to handle multi-scale analysis is to adopt the following steps [39]: (i) a pyramidal decomposition with L levels [40] and (ii) a coarse-to-fine refinement [41]. This is a computationally efficient way to take into account the presence of different spatial frequency channels in the visual cortex and of large range of disparities and spatial frequencies in the real visual signal.

However, our model implements a log-polar mapping, thus its space variance, i.e. the linear increase of the filter size with respect to the eccentricity, can be exploited to efficiently implement a multi-scale analysis. Specifically, a pyramidal approach can be considered as a "vertical" multi-scale (the variation of the filter size at a single location), whereas the log-polar spatial sampling acts as an "horizontal" multi-scale (the variation of the filter size across different location [42]). The "vertical" multi-scale is also addressed in the literature as "cortical pyramids".

Cue combination across the visual field. Human observers and model were tested with annular stimuli spanning sub-portions of the visual field, as well as with full field stimuli spanning the whole region of the visual field visual within a 21 degree radius. When considering the responses of the model to the foveal, mid-peripheral, and far-peripheral stimuli, only the neural units corresponding to the stimulated field regions exhibited any neural activity (as described by Eq 14) and contributed to the model output. When analyzing the responses of the model to the full-filed stimuli, we pooled the neural activities of the distinct MT populations across the three considered annular regions.

Decoding. To assess whether the proposed computational model is able to effectively encode information about the features of the visual signal, and whether the model DSF is similar to the DSF of human observers, we decode the population responses of the MT neurons [90], which encode the disparity stimulus parameters in their distributed representation. The population responses of the MT neurons essentially highlight the most probable disparity values. We adopt a linear combination approach to decode the MT population response as in [39, 99, 100]:

$$\begin{cases} d_\xi(p) = \sum_{i=1}^{K} d_i E^{MT}(p, 0, d_i), \\ d_\eta(p) = \sum_{i=1}^{K} d_i E^{MT}(p, \pi/2, d_i). \end{cases} \tag{16}$$

Note that when considering P tuning directions ($\phi_1 \ldots \phi_P$), Eq 16 would normally contain a $2/P$ normalization term (see [39] for how this term is derived). Here we consider only 2 tuning directions, thus $P = 2$ and the normalization term is 1.

Next, we backwards transform into the retinal domain the disparity map described by Eq 16. To easily detect whether the disparity corrugation is top-tilted leftwards or rightwards, we apply the Fourier transform to the retinal disparity map and check the position of the peak of its magnitude.

Simulation parameters. The simulation parameters selected to obtain the results presented in Fig 1 were adapted from the simulation parameters reported in [39], which were originally tuned to perform on computer vision benchmarks [101–104]. Since the proposed

algorithm is meant to model human stereo vision, not compete on computer vision benchmarks, we modified the simulation parameters to reflect the known properties of the human visual system. Most parameter choices were derived from the literature, and the rest were selected based on pilot work [37] where we compared model performance to the normative data from Reynaud et al. [38]. The most notable differences between the current model and the one presented in [39] are:

- The foveated architecture and the related cortical processing that were not present in [39]: the log-polar paradigm, employed in the proposed computational model, is crucial for replicating the patterns of human data

- The algorithm presented in [39] did not contain neural noise, which is instead present in the human visual system [32] and was thus incorporated into the current model

- In [39] a multi-scale approach was adopted with 11 sub-octave scales in order to recover a large range of disparities (common in computer vision) by using Gabor filters with peak frequency of 0.26 cycles/pixel. However, in the current model, only 1 scale was employed, since as we've noted, the log-polar spatial sampling acts as a "sliding" multi-scale

  The specific model parameters employed here were:

- $\mathcal{D} = \pm 1.52$ pixels, the cortical disparity range to which the neural units are sensitive (this range is constrained by the spatial peak frequency f_s of the filters). Note that the retinal disparity range increases linearly (with receptive field size) across the model's visual field, from ± 0.43 arcmin at the fovea to ± 25 arcmin in the model's periphery.

- $K = 5$, the sampling of the disparity range, i.e. the number of neural units for a given spatial orientation θ.

- the V1 static non-linearity is a power function with exponent 0.5.

- $\sigma_{pool} = 3.66$ pixels, the spatial pooling of V1 responses (its standard deviation).

- $\lambda = 0.65$, the gain of the exponential static non-linearity at the MT level.

- $N = 12$, the number of spatial orientations, i.e. the number of neural units that sample the spatial orientation θ.

- the neural noise is set to 34% and 18% of the local average neural activity at the V1 and MT levels, respectively.

- $f_s = 0.13$ cycles/pixel, the radial peak frequency of the Gabor filters.

- $\sigma = 5.12$ pixels, the standard deviation of the Gabor filters.

- the Gabor filters are zero-mean.

- $R = 318$, the number of rings of the log-polar mapping.

- $\rho_0 = 9$ pixels, the radius of the central blind spot.

- $CR = 6.4$, the compression ratio of the cortical image compared to the Cartesian image.

Supporting information

S1 Appendix. Visual processing throughout the model for a uniform disparity stimulus. (PDF)

Acknowledgments

Author Contributions

Conceptualization: Guido Maiello, Manuela Chessa, Peter J. Bex, Fabio Solari.

Data curation: Guido Maiello, Peter J. Bex.

Formal analysis: Guido Maiello, Manuela Chessa, Fabio Solari.

Funding acquisition: Guido Maiello, Peter J. Bex.

Investigation: Guido Maiello.

Methodology: Guido Maiello, Manuela Chessa, Peter J. Bex, Fabio Solari.

Project administration: Guido Maiello.

Software: Guido Maiello, Manuela Chessa, Fabio Solari.

Validation: Guido Maiello, Manuela Chessa, Peter J. Bex, Fabio Solari.

Visualization: Guido Maiello, Manuela Chessa, Fabio Solari.

Writing – original draft: Guido Maiello, Manuela Chessa, Fabio Solari.

Writing – review & editing: Guido Maiello, Manuela Chessa, Peter J. Bex, Fabio Solari.

References

1. Wheatstone C. On some remarkable, and hitherto unobserved, Phenomena of Binocular Vision. Philosophical Transactions of the Royal Society of London. 1938; 128:371–394.

2. Westheimer G. Cooperative neural processes involved in stereoscopic acuity. Experimental Brain Research. 1979; 36(3):585–597. https://doi.org/10.1007/bf00238525 PMID: 477784

3. Aubert HR, Foerster CFR. Beiträge zur Kenntniss des indirecten Sehens. (I). Untersuchungen über den Raumsinn der Retina. Archiv für Ophthalmologie. 1857; 3:1–37.

4. Liu Y, Bovik AC, Cormack LK. Disparity statistics in natural scenes. Journal of Vision. 2008; 8(11):19–19. https://doi.org/10.1167/8.11.19 PMID: 18831613

5. Pulliam K. Spatial frequency analysis of three-dimensional vision. In: Visual Simulation and Image Realism II. vol. 303; 1982. p. 71–78. https://doi.org/10.1117/12.932660

6. Campbell FW, Robson J. Application of Fourier analysis to the visibility of gratings. The Journal of Physiology. 1968; 197(3):551–566. https://doi.org/10.1113/jphysiol.1968.sp008574 PMID: 5666169

7. Norcia AM, Suiter EE, Tyler CW. Electrophysiological evidence for the existence of coarse and fine disparity mechanisms in human. Vision Research. 1985; 25(11):1603–1611. https://doi.org/10.1016/0042-6989(85)90130-0 PMID: 3832583

8. Yang Y, Blake R. Spatial frequency tuning of human stereopsis. Vision Research. 1991; 31(7):1176–1189. https://doi.org/10.1016/0042-6989(91)90043-5

9. Tyler CW, Barghout L, Kontsevich LL. Computational reconstruction of the mechanisms of human stereopsis. In: Computational Vision Based on Neurobiology. vol. 2054; 1994. p. 52–69. https://doi.org/10.1117/12.171143

10. Wilcox LM, Allison RS. Coarse-fine dichotomies in human stereopsis. Vision Research. 2009; 49(22):2653–2665. https://doi.org/10.1016/j.visres.2009.06.004 PMID: 19520102

11. Reynaud A, Hess RF. Characterization of spatial frequency channels underlying disparity sensitivity by factor analysis of population data. Frontiers in Computational Neuroscience. 2017; 11:63. https://doi.org/10.3389/fncom.2017.00063 PMID: 28744211

12. Serrano-Pedraza I, Read JC. Multiple channels for horizontal, but only one for vertical corrugations? A new look at the stereo anisotropy. Journal of Vision. 2010; 10(12):10–10. https://doi.org/10.1167/10.12.10 PMID: 21047742

13. Julesz B, Miller JE. Independent spatial-frequency-tuned channels in binocular fusion and rivalry. Perception. 1975; 4(2):125–143. https://doi.org/10.1068/p040125

14. Glennerster A, Parker A. Computing stereo channels from masking data. Vision Research. 1997; 37 (15):2143–2152. https://doi.org/10.1016/s0042-6989(97)00036-9 PMID: 9327061

15. Witz N, Hess RF. Mechanisms underlying global stereopsis in fovea and periphery. Vision Research. 2013; 87:10–21. https://doi.org/10.1016/j.visres.2013.05.003 PMID: 23680486

16. Witz N, Zhou J, Hess RF. Similar mechanisms underlie the detection of horizontal and vertical disparity corrugations. PLoS ONE. 2014; 9(1):e84846. https://doi.org/10.1371/journal.pone.0084846 PMID: 24404193

17. Prince SJ, Rogers BJ. Sensitivity to disparity corrugations in peripheral vision. Vision Research. 1998; 38(17):2533–2537. https://doi.org/10.1016/s0042-6989(98)00118-7 PMID: 12116701

18. Virsu V, Rovamo J. Visual resolution, contrast sensitivity, and the cortical magnification factor. Experimental Brain Research. 1979; 37(3):475–494. https://doi.org/10.1007/bf00236818 PMID: 520438

19. Blake A, Bülthoff HH, Sheinberg D. Shape from texture: Ideal observers and human psychophysics. Vision Research. 1993; 33(12):1723–1737. https://doi.org/10.1016/0042-6989(93)90037-w PMID: 8236859

20. Landy MS, Maloney LT, Johnston EB, Young M. Measurement and modeling of depth cue combination: In defense of weak fusion. Vision Research. 1995; 35(3):389–412. https://doi.org/10.1016/0042-6989(94)00176-m PMID: 7892735

21. Knill DC. Discrimination of planar surface slant from texture: human and ideal observers compared. Vision Research. 1998; 38(11):1683–1711. https://doi.org/10.1016/s0042-6989(97)00325-8 PMID: 9747503

22. Backus BT, Banks MS. Estimator reliability and distance scaling in stereoscopic slant perception. Perception. 1999; 28(2):217–242. https://doi.org/10.1068/p2753 PMID: 10615462

23. van Beers RJ, Sittig AC, Gon JJDvd. Integration of proprioceptive and visual position-information: An experimentally supported model. Journal of Neurophysiology. 1999; 81(3):1355–1364. https://doi.org/10.1152/jn.1999.81.3.1355 PMID: 10085361

24. Schrater PR, Kersten D. How optimal depth cue integration depends on the task. International Journal of Computer Vision. 2000; 40(1):71–89. https://doi.org/10.1023/A:1026557704054

25. Ernst MO, Banks MS. Humans integrate visual and haptic information in a statistically optimal fashion. Nature. 2002; 415(6870):429. https://doi.org/10.1038/415429a PMID: 11807554

26. Gibaldi A, Canessa A, Sabatini SP. The active side of stereopsis: Fixation strategy and adaptation to natural environments. Scientific Reports. 2017; 7:44800. https://doi.org/10.1038/srep44800 PMID: 28317909

27. Gibaldi A, Banks MS. Binocular Eye Movements Are Adapted to the Natural Environment. Journal of Neuroscience. 2019; 39(15):2877–2888. https://doi.org/10.1523/JNEUROSCI.2591-18.2018 PMID: 30733219

28. Schwartz EL. Spatial Mapping in the Primate Sensory Projection: Analytic Structure and Relevance to Perception. Biological Cybernetics. 1977; 25:181–194. https://doi.org/10.1007/bf01885636 PMID: 843541

29. Tootell RB, Silverman MS, Switkes E, De Valois RL. Deoxyglucose analysis of retinotopic organization in primate striate cortex. Science. 1982; 218(4575):902–904.

30. Traver VJ, Bernardino A. A review of log-polar imaging for visual perception in robotics. Robotics and Autonomous Systems. 2010; 58(4):378–398. https://doi.org/10.1016/j.robot.2009.10.002

31. Solari F, Chessa M, Sabatini SP. Design strategies for direct multi-scale and multi-orientation feature extraction in the log-polar domain. Pattern Recognition Letters. 2012; 33(1):41–51. https://doi.org/10.1016/j.patrec.2011.09.021

32. Chessa M, Maiello G, Bex PJ, Solari F. A space-variant model for motion interpretation across the visual field. Journal of Vision. 2016; 16(2):12. https://doi.org/10.1167/16.2.12 PMID: 27580091

33. Fleet DJ, Wagner H, Heeger DJ. Neural encoding of binocular disparity: energy models, position shifts and phase shifts. Vision research. 1996; 36(12):1839–1857. https://doi.org/10.1016/0042-6989(95)00313-4 PMID: 8759452

34. Qian N, Zhu Y. Physiological computation of binocular disparity. Vision research. 1997; 37(13):1811–1827. https://doi.org/10.1016/s0042-6989(96)00331-8 PMID: 9274767

35. Ohzawa I, DeAngelis GC, Freeman RD. Stereoscopic depth discrimination in the visual cortex: neurons ideally suited as disparity detectors. Science. 1990; 249(4972):1037–1041.

36. Allenmark F, Read JCA. Spatial Stereoresolution for Depth Corrugations May Be Set in Primary Visual Cortex. PLOS Computational Biology. 2011; 7(8):1–14. https://doi.org/10.1371/journal.pcbi.1002142

37. Maiello G, Chessa M, Bex PJ, Solari F. Can Neuromorphic Computer Vision Inform Vision Science? Disparity Estimation as a Case Study. In: Computational and Mathematical Models in Vision (MOD-VIS); 2016.

38. Reynaud A, Gao Y, Hess RF. A normative dataset on human global stereopsis using the quick Disparity Sensitivity Function (qDSF). Vision Research. 2015; 113:97–103. https://doi.org/10.1016/j.visres.2015.04.021 PMID: 26028556

39. Chessa M, Solari F. A Computational Model for the Neural Representation and Estimation of the Binocular Vector Disparity from Convergent Stereo Image Pairs. International Journal of Neural Systems. 2019; 29(05):1850029. https://doi.org/10.1142/S0129065718500296 PMID: 30045646

40. Bergen J R Adelson CHA E H, Burt PJ, Ogden JM. Pyramid methods in image processing. RCA Engineer. 1984; 29:33–41.

41. Simoncelli EP. Course-to-fine Estimation of Visual Motion. In: IEEE Eighth Workshop on Image and Multidimensional Signal Processing; 1993.

42. Bonmassar G, Schwartz EL. Space-Variant Fourier Analysis: The Exponential Chirp Transform. IEEE Trans Pattern Anal Mach Intell. 1997; 19(10):1080–1089. https://doi.org/10.1109/34.625108

43. Sprague WW, Cooper EA, Tošić I, Banks MS. Stereopsis is adaptive for the natural environment. Science Advances. 2015; 1(4). https://doi.org/10.1126/sciadv.1400254

44. Harvey BM, Dumoulin SO. The relationship between cortical magnification factor and population receptive field size in human visual cortex: constancies in cortical architecture. Journal of Neuroscience. 2011; 31(38):13604–13612. https://doi.org/10.1523/JNEUROSCI.2572-11.2011 PMID: 21940451

45. Land MF, Nilsson DE. Animal Eyes. Oxford University Press; 2012.

46. Maiello G, Chessa M, Solari F, Bex PJ. The (in) effectiveness of simulated blur for depth perception in naturalistic images. PLoS ONE. 2015; 10(10):e0140230. https://doi.org/10.1371/journal.pone.0140230 PMID: 26447793

47. Held RT, Cooper EA, Banks MS. Blur and disparity are complementary cues to depth. Current Biology. 2012; 22(5):426–431. https://doi.org/10.1016/j.cub.2012.01.033 PMID: 22326024

48. Maiello G, Chessa M, Solari F, Bex PJ. Simulated disparity and peripheral blur interact during binocular fusion. Journal of Vision. 2014; 14(8):13–13. https://doi.org/10.1167/14.8.13 PMID: 25034260

49. Hibbard PB, Goutcher R, Hunter DW. Encoding and estimation of first-and second-order binocular disparity in natural images. Vision research. 2016; 120:108–120. https://doi.org/10.1016/j.visres.2015.10.016 PMID: 26731646

50. Tanaka H, Ohzawa I. Neural basis for stereopsis from second-order contrast cues. Journal of Neuroscience. 2006; 26(16):4370–4382. https://doi.org/10.1523/JNEUROSCI.4379-05.2006 PMID: 16624957

51. Schor CM, Edwards M, Pope DR. Spatial-frequency and contrast tuning of the transient-stereopsis system. Vision research. 1998; 38(20):3057–3068. https://doi.org/10.1016/s0042-6989(97)00467-7 PMID: 9893815

52. Ogle KN. On the limits of stereoscopic vision. Journal of Experimental Psychology. 1952; 44(4):253. https://doi.org/10.1037/h0057643 PMID: 13000066

53. Schor C, Wood I, Ogawa J. Binocular sensory fusion is limited by spatial resolution. Vision Research. 1984; 24(7):661–665. https://doi.org/10.1016/0042-6989(84)90207-4 PMID: 6464360

54. Ghahghaei S, McKee S, Verghese P. The upper disparity limit increases gradually with eccentricity. Journal of Vision. 2019; 19(11):3–3. https://doi.org/10.1167/19.11.3 PMID: 31480075

55. Wardle SG, Bex PJ, Cass J, Alais D. Stereoacuity in the periphery is limited by internal noise. Journal of vision. 2012; 12(6):12–12. https://doi.org/10.1167/12.6.12 PMID: 22685339

56. Brainard DH. The Psychophysics Toolbox. Spatial Vision. 1997; 10:433–436. https://doi.org/10.1163/156856897X00357 PMID: 9176952

57. Pelli DG. The VideoToolbox software for visual psychophysics: Transforming numbers into movies. Spatial Vision. 1997; 10:437–442. https://doi.org/10.1163/156856897X00366 PMID: 9176953

58. Georgeson MA, Yates TA, Schofield AJ. Discriminating depth in corrugated stereo surfaces: Facilitation by a pedestal is explained by removal of uncertainty. Vision Research. 2008; 48(21):2321–2328. https://doi.org/10.1016/j.visres.2008.07.009 PMID: 18682260

59. Yang Q, Bucci MP, Kapoula Z. The latency of saccades, vergence, and combined eye movements in children and in adults. Investigative Ophthalmology & Visual Science. 2002; 43(9):2939–2949.

60. Baloh RW, Sills AW, Kumley WE, Honrubia V. Quantitative measurement of saccade amplitude, duration, and velocity. Neurology. 1975; 25(11):1065–1065. https://doi.org/10.1212/wnl.25.11.1065 PMID: 1237825

61. Volkmann FC. Vision during voluntary saccadic eye movements. Journal of the Optical Society of America. 1962; 52(5):571–578. https://doi.org/10.1364/josa.52.000571 PMID: 13926602

62. Dorr M, Bex PJ. Peri-saccadic natural vision. Journal of Neuroscience. 2013; 33(3):1211–1217. https://doi.org/10.1523/JNEUROSCI.4344-12.2013 PMID: 23325257

63. Wetherill G, Levitt H. Sequential estimation of points on a psychometric function. British Journal of Mathematical and Statistical Psychology. 1965; 18(1):1–10. https://doi.org/10.1111/j.2044-8317.1965.tb00689.x PMID: 14324842

64. Tyler CW. Spatial organization of binocular disparity sensitivity. Vision Research. 1975; 15(5):583–590. https://doi.org/10.1016/0042-6989(75)90306-5 PMID: 1136171

65. Bradshaw MF, Rogers BJ. Sensitivity to horizontal and vertical corrugations defined by binocular disparity. Vision Research. 1999; 39(18):3049–3056. https://doi.org/10.1016/s0042-6989(99)00015-2 PMID: 10664803

66. Lesmes LA, Lu ZL, Baek J, Albright TD. Bayesian adaptive estimation of the contrast sensitivity function: The quick CSF method. Journal of Vision. 2010; 10(3):17–17. https://doi.org/10.1167/10.3.17 PMID: 20377294

67. Fisher RA. Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population. Biometrika. 1915; 10(4):507–521. https://doi.org/10.2307/2331838

68. Goodale MA, Westwood DA. An evolving view of duplex vision: separate but interacting cortical pathways for perception and action. Current Opinion in Neurobiology. 2004; 14(2):203–211. https://doi.org/10.1016/j.conb.2004.03.002 PMID: 15082326

69. Nguyenkim JD, DeAngelis GC. Disparity-based coding of three-dimensional surface orientation by macaque middle temporal neurons. Journal of Neuroscience. 2003; 23(18):7117–7128. https://doi.org/10.1523/JNEUROSCI.23-18-07117.2003 PMID: 12904472

70. Schindler K. Geometry and construction of straight lines in log-polar images. Computer Vision and Image Understanding. 2006; 103(3):196–207. https://doi.org/10.1016/j.cviu.2005.07.012

71. Traver VJ, Pla F. Log-polar mapping template design: From task-level requirements to geometry parameters. Image Vision Computing. 2008; 26(10):1354–1370. https://doi.org/10.1016/j.imavis.2007.11.009

72. Solari F, Chessa M, Sabatini SP. An integrated neuromimetic architecture for direct motion interpretation in the log-polar domain. Computer Vision and Image Understanding. 2014; 125:37–54. https://doi.org/10.1016/j.cviu.2014.02.012

73. Wilkinson MO, Anderson RS, Bradley A, Thibos LN. Neural bandwidth of veridical perception across the visual field. Journal of vision. 2016; 16(2):1–1. https://doi.org/10.1167/16.2.1 PMID: 26824638

74. Schira MM, Tyler CW, Spehar B, Breakspear M. Modeling Magnification and Anisotropy in the Primate Foveal Confluence. PLOS Computational Biology. 2010; 6:1–10.

75. Chessa M, Sabatini SP, Solari F, Tatti F. A Quantitative Comparison of Speed and Reliability for Log-Polar Mapping Techniques. In: Crowley J, Draper B, Thonnat M, editors. Computer Vision Systems. vol. 6962 of Lecture Notes in Computer Science; 2011. p. 41–50.

76. Lungarella M, Sporns O. Mapping Information Flow in Sensorimotor Networks. PLOS Computational Biology. 2006; 2:1–12. https://doi.org/10.1371/journal.pcbi.0020144

77. Bolduc M, Levine MD. A Real-Time Foveated Sensor with Overlapping Receptive Fields. Real-Time Imaging. 1997; 3(3):195–212. https://doi.org/10.1006/rtim.1996.0056

78. Pamplona D, Bernardino A. Smooth Foveal Vision with Gaussian Receptive Fields. In: 9th IEEE-RAS International Conference on Humanoid Robots; 2009.

79. Traver VJ, Bernardino A. A review of log-polar imaging for visual perception in robotics. Robotics and Autonomous Systems. 2010; 58(4):378–398. https://doi.org/10.1016/j.robot.2009.10.002

80. Berton F, Sandini G, Metta G. Anthropomorphic visual sensors. In: Encyclopedia of Sensors. American Scientific Publishers; 2006. p. 1–16.

81. Freeman J, Simoncelli EP. Metamers of the ventral stream. Nature Neuroscience. 2011; 14(9):1195–1201. https://doi.org/10.1038/nn.2889 PMID: 21841776

82. Wurbs J, Mingolla E, Yazdanbakhsh A. Modeling a space-variant cortical representation for apparent motion. Journal of Vision. 2013; 13(10):2. https://doi.org/10.1167/13.10.2 PMID: 23922444

83. Chessa M, Solari F. Local feature extraction in log-polar images. In: International Conference on Image Analysis and Processing. Springer; 2015. p. 410–420.

84. Granlund GH, Knutsson H. Signal Processing for Computer Vision. Kluwer Academic Publishers, Dordrecht; 1995.

85. Marčelja S. Mathematical description of the responses of simple cortical cells. JOSA. 1980; 70(11):1297–1300. https://doi.org/10.1364/JOSA.70.001297

86. Daugman JG. Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. JOSA A. 1985; 2(7):1160–1169. https://doi.org/10.1364/JOSAA.2.001160

87. Mallot HA, Seelen W, Giannakopoulos F. Neural mapping and space-variant image processing. Neural Networks. 1990; 3(3):245–263. https://doi.org/10.1016/0893-6080(90)90069-W

88. Wallace AM, McLaren DJ. Gradient detection in discrete log-polar images. Pattern Recognition Letters. 2003; 24(14):2463–2470. https://doi.org/10.1016/S0167-8655(03)00091-6

89. Chan Man Fong CF, Kee D, Kaloni PN. Advanced Mathematics For Applied And Pure Sciences. CRC Press; 1997.

90. Henriksen S, Cumming BG, Read JC. A single mechanism can account for human perception of depth in mixed correlation random dot stereograms. PLoS computational biology. 2016; 12(5):e1004906. https://doi.org/10.1371/journal.pcbi.1004906 PMID: 27196696

91. Nishimoto S, Gallant JL. A three-dimensional spatiotemporal receptive field model explains responses of area MT neurons to naturalistic movies. The Journal of Neuroscience. 2011; 31(41):14551–14564. https://doi.org/10.1523/JNEUROSCI.6801-10.2011 PMID: 21994372

92. Rust NC, Mante V, Simoncelli EP, Movshon JA. How MT cells analyze the motion of visual patterns. Nature Neuroscience. 2006; 9:1421–1431. https://doi.org/10.1038/nn1786 PMID: 17041595

93. Heeger DJ. Normalization of cell responses in cat striate cortex. Visual neuroscience. 1992; 9 (02):181–197. https://doi.org/10.1017/s0952523800009640 PMID: 1504027

94. Born RT, Bradley DC. Structure and function of visual area MT. Annu Rev Neurosci. 2005; 28:157–189. https://doi.org/10.1146/annurev.neuro.26.041002.131052 PMID: 16022593

95. Read JC. Vertical binocular disparity is encoded implicitly within a model neuronal population tuned to horizontal disparity and orientation. PLoS computational biology. 2010; 6(4):e1000754. https://doi.org/10.1371/journal.pcbi.1000754 PMID: 20421992

96. Serrano-Pedraza I, Read JCA. Stereo vision requires an explicit encoding of vertical disparity. Journal of Vision. 2009; 9(4):3. https://doi.org/10.1167/9.4.3 PMID: 19757912

97. Read JCA, Cumming BG. Does depth perception require vertical-disparity detectors? Journal of Vision. 2006; 6(12):1. https://doi.org/10.1167/6.12.1

98. Cumming B. An unexpected specialization for horizontal disparity in primate primary visual cortex. Nature. 2002; 418(6898):633–636. https://doi.org/10.1038/nature00909 PMID: 12167860

99. Pouget A, Zhang K, Deneve S, Latham PE. Statistically efficient estimation using population coding. Neural Computation. 1998; 10(2):373–401. https://doi.org/10.1162/089976698300017809 PMID: 9472487

100. Rad KR, Paninski L. Information Rates and Optimal Decoding in Large Neural Populations. In: Shawe-Taylor J, Zemel RS, Bartlett PL, Pereira FCN, Weinberger KQ, editors. NIPS; 2011. p. 846–854.

101. Scharstein D, Szeliski R. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. International Journal of Computer Vision. 2002; 47(1-3):7–42. https://doi.org/10.1023/A:1014573219977

102. Scharstein D, Szeliski R. High-accuracy stereo depth maps using structured light. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2003). vol. 1; 2003. p. 195–202.

103. Chessa M, Solari F, Sabatini SP. A Virtual Reality Simulator for Active Stereo Vision Systems. In: Proceedings of the Fourth International Conference on Computer Vision Theory and Applications (VISAPP 2009). vol. 2; 2009. p. 444–449.

104. Canessa A, Gibaldi A, Chessa M, Fato M, Solari F, Sabatini SP. A dataset of stereoscopic images and ground-truth disparity mimicking human fixations in peripersonal space. Scientific Data. 2017; 4:170034. https://doi.org/10.1038/sdata.2017.34 PMID: 28350382