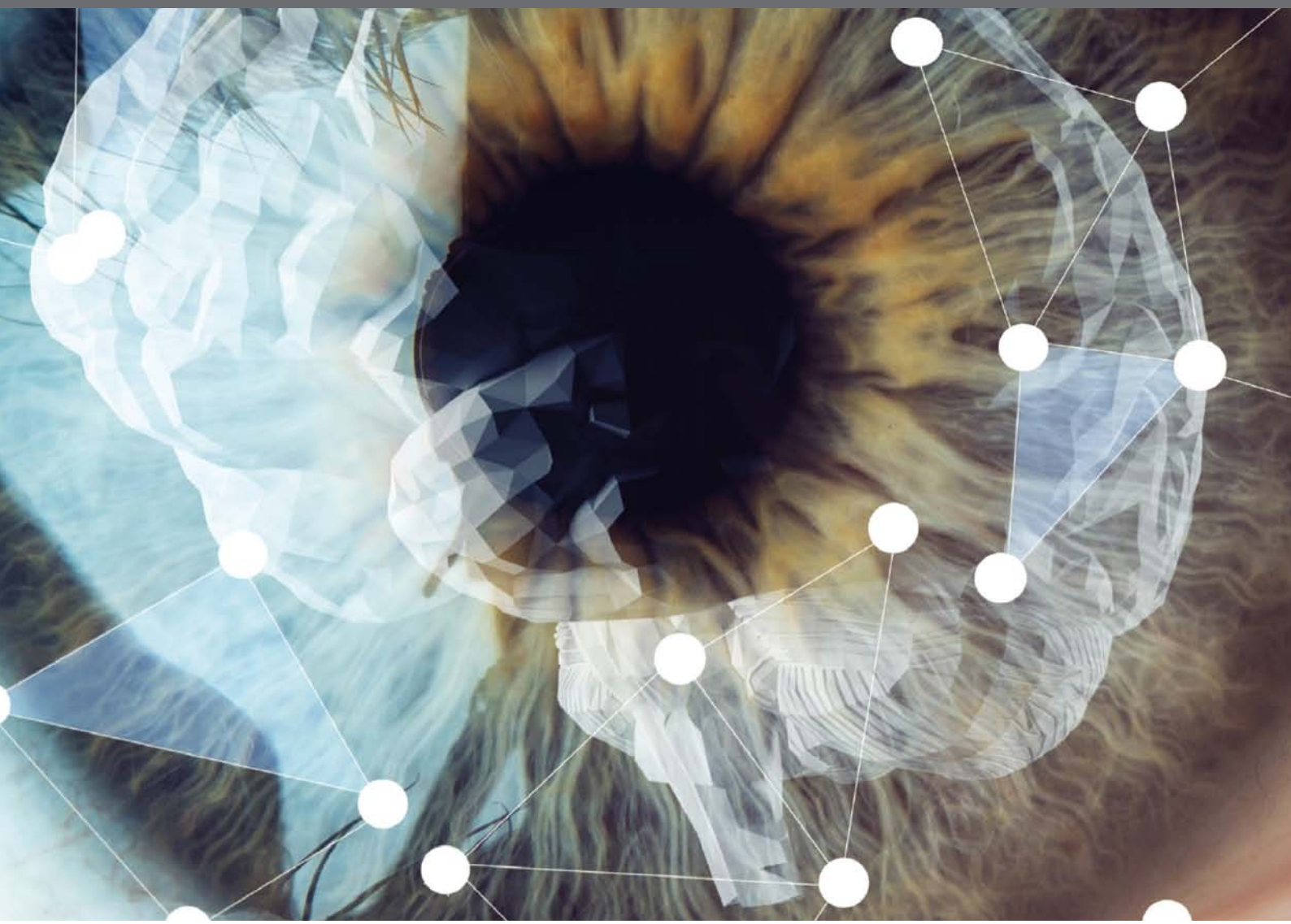
INTEGRATING VISUAL SYSTEM MECHANISMS, COMPUTATIONAL MODELS AND ALGORITHMS/TECHNOLOGIES

EDITED BY: Hedva Spitzer, Xavier Otazu and Hagit Hel-Or
PUBLISHED IN: Frontiers in Bioengineering and Biotechnology, Frontiers in Neuroscience, Frontiers in Computational Neuroscience and Frontiers in Psychology



frontiers Research Topics

frontiers

ISSN 1664-8714
ISBN 978-2-88963-510-8
DOI 10.3389/978-2-88963-510-8

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: researchtopics@frontiersin.org

INTEGRATING VISUAL SYSTEM MECHANISMS, COMPUTATIONAL MODELS AND ALGORITHMS/TECHNOLOGIES

Topic Editors:
Hedva Spitzer, Tel Aviv University, Israel
Xavier Otazu, Autonomous University of Barcelona, Spain
Hagit Hel-Or, University of Haifa, Israel

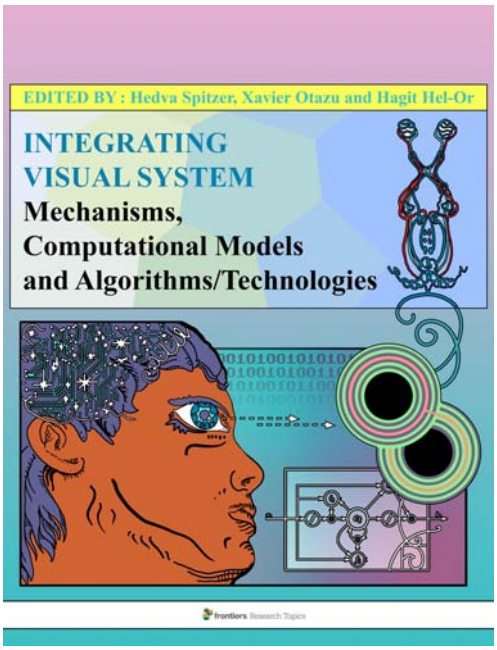


Illustration by Rony Griffit

Citation: Spitzer, H., Otazu, X., Hel-Or, H., eds. (2020). Integrating Visual System Mechanisms, Computational Models and Algorithms/Technologies. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88963-510-8

Table of Contents

frontiers in Bioengineering and Biotechnology

EDITORIAL
published: 22 January 2020
doi: 10.3389/fbioe.2019.00483

Editorial: Integrating Visual System Mechanisms, Computational Models and Algorithms/Technologies

Hedva Spitzer[1]*, *Xavier Otazu*[2] *and Hagit Hel-Or*[3]

[1] School of Electrical Engineering, Tel Aviv University, Tel Aviv, Israel, [2] Computer Science Department, Computer Vision Center, Autonomous University of Barcelona, Barcelona, Spain, [3] Department of Computer Science, University of Haifa, Haifa, Israel

Keywords: computational models, algorithms, technologies, visual system, mechanisms

Editorial on the Research Topic

Integrating Visual System Mechanisms, Computational Models and Algorithms/Technologies

The Research Topic on "Integrating Visual System Mechanisms, Computational Models and Algorithms/Technologies" collects novel studies that display a strong synergy between three entities: (1) the visual system from its various angles including physiological, psychophysical, and perceptual, (2) computational models whether descriptive or predictive, and (3) vision inspired algorithms and applications. The interaction between modeling and the various aspects of the visual system is expressed in the reciprocal contributions between the two. On one hand, visual mechanisms and neuronal units provide inspiration and basis for modeling approaches and their computational units within, and on the other hand, modeling provides novel insights and new understandings of the visual system mechanisms and its associated behaviors. Furthermore, computational models, and the underlying visual mechanisms, provide a basis for developing practical algorithms to perform image processing and image understanding.

The articles in this Research Topic present computational models of the visual system ranging from neuronal mechanisms, through visual mechanisms, to visual perceptual behavior and visual illusions. Modeling efforts take different computational approaches from building blocks that are inspired by mechanisms of the visual system, to a more global Gestalt approach that attempts to explain a phenomenon regardless of the underlying elements using functional, statistical, or learning approaches. Other articles develop applications ranging from visual system inspired measures such as image quality and image esthetics to applications such as classification and segmentation.

Several studies in this issue, present computational models of the visual system at the neuronal level, and some include feasible physiological components in the model. In Gonzalez and Tsotsos, the authors suggest a computational model of attention based on the adaptation mechanisms and selective tuning of the V4 neurons which is expressed in the neurons' firing rate during attentional tasks. Different computational models are tested, coinciding with different interpretations of the attention mechanism: (a) enhancing responses due to attention or (b) suppressing irrelevant signals. The authors follow a model of the second type and are able to predict the temporal profiles of neurons' firing rate, similar to those found electrophysiologically. Through their modeling, the authors show that high level vision processes can also be explained by low-level processes, namely, that selectively tuning a model of attention, can reprsoduce properties of neuron firing rates related to attention. In another article Banerjee et al., the authors propose a computational model, based on the extreme value theory, for the integration of two sensory modalities, namely, the olfactory input and visual sensitivity of zebrafish. The authors show that the neural signals (pattern and rate of neuronal firing) differ in their statistical fit when the signals are uni-modal (visual) or multi-modal (visual + olfaction). They further showed this by developing a Machine Learning based

OPEN ACCESS

Edited and reviewed by:
Richard D. Emes,
University of Nottingham,
United Kingdom

***Correspondence:**
Hedva Spitzer
hedva@eng.tau.ac.il

Specialty section:
This article was submitted to Bioinformatics and Computational Biology, a section of the journal Frontiers in Bioengineering and Biotechnology

Received: *21 November 2019*
Accepted: *27 December 2019*
Published: *22 January 2020*

Citation:
Spitzer H, Otazu X and Hel-Or H (2020) Editorial: Integrating Visual System Mechanisms, Computational Models and Algorithms/Technologies. Front. Bioeng. Biotechnol. 7:483. doi: 10.3389/fbioe.2019.00483

classifier that was able to successfully distinguish between these neural signals. This study forms a contribution to the intriguing area of interactions between different sensory modalities.

Two additional articles deal with the chromatic properties of the visual system as expressed in the retinal layer and cortical layers. In Barkan and Spitzer, a computational model is presented which suggests an explanation of the underlying visual mechanisms for compensating chromatic aberrations. The computational model takes into account the spatio-chromatic properties of the color-coded cells in the retina while taking into account the significance of the anatomical separation of the Konio and Parvo chromatic pathways in the visual system. Furthermore, the model predicts the enigmatic phenomenon of S-cone pattern reported by Shevell and Monnier. In a review article, by Patterson et al., the authors discuss the role of retinal midget RGC cells and cortical double opponent cells in the context of hue perception on one hand and spatial perception on the other. The authors present hypotheses that in some form are not in accord with those supported by some other models including that of Barkan and Spitzer mentioned above. As usual in Science, especially in neuroscience, conflicting results are always an interesting source for promotion of discussion and comparison of opposite/different ideas.

Another group of studies develop computational models in order to assist in understanding specific vision mechanisms. In Piu et al., the authors acquired experimental data and then performed statistical analysis on the data to obtain a representation of pupil size changes. They analyzed oscillatory dynamics of the pupil at rest by extracting features from the cross-recurrences of these oscillators as expressed in the power spectrum. The authors state that their novel analysis approach can form an adaptable diagnostic tool for identifying alertness and/or pathological status and thus might assist in clinical assessments of pathologies associated with the autonomous nervous system. In Reynaud and Hess, the authors analyze their previously measured dataset and assess the visual disparity sensitivity of subjects across different spatial frequencies. The computational factor in their study is the data analysis methods in which they applied inter-correlations and factor analysis on the data and found two spatial frequency channels for disparity sensitivity: one tuned to high spatial frequencies and one tuned to low spatial frequencies. The authors suggest that this tuning of disparity channels could be important in computer vision to design multi-scale stereo matching algorithms. In Marić and Domijan, binary attention maps are modeled using a recurrent competitive network with excitatory-inhibitory nodes. The model reproduces top-down mechanisms of attentions that enhance perceived saliency of low-level features. The model is based on an extension of previously suggested Winner Take All (WTA) choice models, and is inspired by neurological components such as dendritic non-linearity that act on the excitatory units and modulate synaptic transmission. The model integrates a large set of data in visual attention and successfully predicts several attentional effects including the ability to integrate information across space and time to form the intersection or union of two maps that are defined by different features.

Finally, a selection of articles uses computational models to predict and explain high level visual tasks, perceptual behavior, and visual phenomena. Some of these studies experiment with ambiguous stimuli and suggest explanations of visual system mechanisms that contribute to the stabilization of the visually perceived display content. The article Cohen-Duwek and Spitzer, models the Filling-In phenomenon and, specifically, the alternating effects in which the background of a stimulus may lead to two different types of perceived color: original or complementary color. The model successfully predicts both effects through a heat diffusion function that is triggered by both the chromatic edges of the stimulus and the achromatic *remaining* contours, in contrast to previous studies that use the edges as blockers for diffusion and not as triggers. In another article Cohen-Duwek and Spitzer, a computational model is presented that predicts spatial Filling-In effects such as the Watercolor illusion and the Cornsweet effects, that have several chromatic edges. The model is based on the heat diffusion equation where the scene gradients serve as heat sources. The model successfully predicts both the assimilative and non-assimilative watercolor effects, as well as additional Filling-In visual effects. The study thus supports the theory that a shared visual mechanism is responsible (or partly responsible) for the vast variety of the "conflicting" filling-in phenomena. Two articles studied motion integration using bi-stable moving visual stimuli that can induce two different percepts (e.g., coherent and transparent). In Li et al., a bi-stable moving visual stimuli of line segments was presented to participants and their individual biases were modeled using a Bayesian modeling approach indicating a preference for one of the two possible interpretations of the scene. The authors found that increasing density shows increasing bias in observers and that this effect is greater in regular patterns than in irregular patterns. The authors tested a number of Bayesian models and show that a motion segregation prior best explains the interaction of density and regularity observed in the collected experimental data. The authors suggest that bias is used by observers to stabilize visual perception of the world. In the article Liu et al., motion integration in normal observers was compared to integration by observers with Anisometropic Amblyopia, a neurodevelopmental disorder of the visual system. They showed that when the stimuli contrast is reduced, the control observers exhibit a change in percept patterns, but amblyopic eyes do not. Using Baysian modeling, the authors show that indeed contrast affects motion integration. Considering this together with the modeling outcomes, the authors suggest that there is a different motion coding mechanism in the amblyopic visual system. Finally, in Yankelovich and Spitzer, Boundary Completion was modeled, using a functional optimization approach in which there is no need to extract different image features. The model evaluates several possible interpretations of the input and assigns a cost to each. The interpretation with minimal cost is the model's output. The model successfully predicts real and illusory contours. Additionally, for ambiguous stimulus, the model is able to find multiple possible image interpretations, which are ranked according to the probability they are perceived.

A different group of papers in this special issue, propose practical algorithms and applications that were inspired by elements of the Human Visual System, or include components that do so. In Tsitiridis et al., the authors attempt to develop a system to detect "Presentation Attacks" where a person's image is illegally reproduced and used to abuse a biometric system. The authors develop a biologically-inspired presentation attack detection model, based on features that mimic neurobiological processes in the human visual system. Machine learning tools are exploited to successfully predict whether incoming data is a spoofing-attack or is a legitimate image. In the article Paulun et al., a new system for dynamic visual recognition is introduced that combines bio-inspired sensor and hardware with a brain-like spiking neural network that mimics the layered structure and the retinotopic organization of the retina and visual cortex. Following training, the network showed a very high object classification accuracy. Finally, two papers in this group deal with image quality and esthetics. In Martinez-Garcia et al., the authors address the important question of biased or imbalanced datasets and their effect on quantitative modeling of the visual system. The authors show this in a specific case of layered retina-cortex models that learn to predict subjective quality ratings of images. They show that the database under-represents certain stimuli (such as cross-masking between different frequencies) and thus the model trained on this database does not generalize well. The authors show that by augmenting the database with synthetic examples, the model shows significant improvement in performance and generalization. The authors impress that naturalistic databases should be combined with artificial stimuli to improve model performance.

In the comprehensive review Brachmann and Redies, the authors describe the advances achieved by the Vision Science and the Computer Vision communities in the parallel fields of experimental visual aesthetics and computational visual aesthetics. The paper highlights the similarities between the types of features exploited for these tasks by both communities and the similarities between the quantitative tools used to analyze and define these features. The review covers models and algorithms that supply prediction of ratings, style, and artist identification as well as computational methods in art history of painting and photograph images. The review covers methods at both sensorial (low-level bottom-up) and cognitive levels (high-levels), including modern methods of deep learning. In addition, the review summarizes results from the field of experimental aesthetics and deal with several specific image properties. The authors show that a close interaction between computational and experimental approaches are fundamental to answering difficult questions.

In this special issue, we have collected a variety of articles that look at the intriguing cycle of: visual system, computational models, and applications. The studies show how computational models can explain the vision system from the neuronal level to the behavioral level providing understanding, and novel insights. On the other hand, the visual system provides ideas and inspiration for the computational units and driving rules of the models. The interaction cycle continues with the design of practical algorithms and applications in the field of computer vision, that arise from the computational models and the ensuing understanding of the visual system. Some of the papers in this collection, even succeeded in achieving algorithms that perform on par with state-of-the-art capabilities, due to the adoption of ideas from the visual systems. Other papers provide inspiration for future possible algorithms to accomplish different visual tasks.

Within this cycle of mutual contributions, we can learn some intriguing ideas and raise interesting questions.

A recurring notion is the idea of the visual system providing educated guesses on the visual scene, based on the visual input as well as on priors, and internal representations and computations. Multi-stable inputs in the 3D world, occluded and ambiguous scenes, allow several interpretations. However, these are processed by the visual system that considers the possible interpretations and produces an "educated guess" as the best explanation of the visual scene. Such a mechanism tends to lend stability and consistency to our visual world.

An interesting insight that has been previously established, is the importance of visual illusions as a basis for research on the visual system. As several of the articles in this issue have shown, illusions serve to mirror "errors" and "biases" of the visual system as well as provide a window into the visual system's mechanics via visual perception.

Finally, we note that several of the articles introduce the notion of aesthetics of the visual scene and raise the point that beyond a comprehensive review, a small step has been taken toward the famous philosophical-psychophysical problem also regarding to visual aesthetics through the discussion of originality and creativity.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Computational Neuroscience

ORIGINAL RESEARCH
published: 11 July 2017
doi: 10.3389/fncom.2017.00063

Characterization of Spatial Frequency Channels Underlying Disparity Sensitivity by Factor Analysis of Population Data

*Alexandre Reynaud * and Robert F. Hess*

McGill Vision Research, Department of Ophthalmology, McGill University, Montreal, QC, Canada

It has been suggested that at least two mechanisms mediate disparity processing, one for coarse and one for fine disparities. Here we analyze individual differences in our previously measured normative dataset on the disparity sensitivity as a function of spatial frequency of 61 observers to assess the tuning of the spatial frequency channels underlying disparity sensitivity for oblique corrugations (Reynaud et al., 2015). Inter-correlations and factor analysis of the population data revealed two spatial frequency channels for disparity sensitivity: one tuned to high spatial frequencies and one tuned to low spatial frequencies. Our results confirm that disparity is encoded by spatial frequency channels of different sensitivities tuned to different ranges of corrugation frequencies.

Keywords: disparity sensitivity, qDSF, binocular vision, stereopsis, individual differences, factor analysis

OPEN ACCESS

Edited by:
Hedva Spitzer,
Tel Aviv University, Israel

Reviewed by:
John E. Lewis,
University of Ottawa, Canada
Hagit Hel-Or,
University of Haifa, Israel
Ronen Segev,
Ben-Gurion University of the Negev, Beersheba, Israel

****Correspondence:***
Alexandre Reynaud
alexandre.reynaud@mail.mcgill.ca

Received: *15 February 2017*
Accepted: *28 June 2017*
Published: *11 July 2017*

Citation:
Reynaud A and Hess RF (2017) Characterization of Spatial Frequency Channels Underlying Disparity Sensitivity by Factor Analysis of Population Data. Front. Comput. Neurosci. 11:63. doi: 10.3389/fncom.2017.00063

INTRODUCTION

The visual system utilizes the displacement or disparity in the two images seen by the two eyes to compute the depth of objects. In terms of the underlying mechanisms, Pulliam (1982) first suggested that there were two global disparity mechanisms, one tuned to low spatial frequencies involving coarse disparities and one tuned to high spatial frequencies involving fine disparities. Yang and Blake (1991) also argued for only two spatial frequency channels for disparity processing and their model was later refined by Tyler et al. (1994). Additional evidence for two spatial frequency channels subserving disparity processing comes from the work of Norcia et al. (1985); Wilcox and Allison (2009); Witz et al. (2014). However, other studies suggest a multiple channels model (Julesz and Miller, 1975; Glennerster and Parker, 1997; Serrano-Pedraza et al., 2013).

Assessing the tuning of these channels has been of great importance for mechanistic models of stereo computer vision (Marr and Poggio, 1979; Nishihara, 1984; Quam, 1987; Rohaly and Wilson, 1993). These can be used to map different scales of matching in hierarchical structures (Nishihara, 1984; Quam, 1987) with, for instance, coarse-to-fine constraints (Rohaly and Wilson, 1993). In robotic vision, these tuning properties can be used to calibrate cameras (Tsai, 1986) and vergence algorithms (Piater et al., 1999; Lonini et al., 2013).

While most studies have used masking paradigms to characterize spatial frequency channels for stereopsis (Julesz and Miller, 1975; Yang and Blake, 1991; Shioiri et al., 1994; Tyler et al., 1994; Glennerster and Parker, 1997; Prince et al., 1998; Serrano-Pedraza et al., 2013), another possibility comes from factor analysis of population data (Read et al., 2016). The individual differences are then treated as systematic and meaningful, reflecting the true variability of underlying mechanisms rather than random noise (Peterzell, 2016). Identifying the sources of variability within the population will inform on the common processing mechanisms. Therefore, spatial and temporal

frequency channels can be characterized by analyzing individual differences and correlations. The rationale is that the correlation in detection thresholds for pairs of stimuli should be higher for stimuli detected by the same mechanism than for stimuli detected by different mechanisms (Owsley et al., 1983; Sekuler et al., 1984; Billock and Harding, 1996). Hence by looking at the inter-correlations between individuals' sensitivity at neighboring frequencies, one is able to determine the presence of frequency channels (Mayer et al., 1995; Billock and Harding, 1996; Peterzell and Teller, 2000; Simpson and McFadden, 2005; Rosli et al., 2009). Therefore, a factor analysis of the dataset consisting of a principal component analysis (PCA) and a rotation of the factors in order to determine a simple structure can characterize the tuning curves of the channels (Simpson and McFadden, 2005). Using factor analytics within the population sensitivities Peterzell and Teller (1996, 2000) assessed spatial frequency channels tuning for luminance and color contrast sensitivities. Here we use similar methods to analyze individual differences in our previously measured normative dataset on disparity sensitivity as a function of spatial frequency for oblique corrugations of 61 observers (**Figure 1**; Reynaud et al., 2015) in order to assess the spatial frequency tuning of the underlying disparity channels.

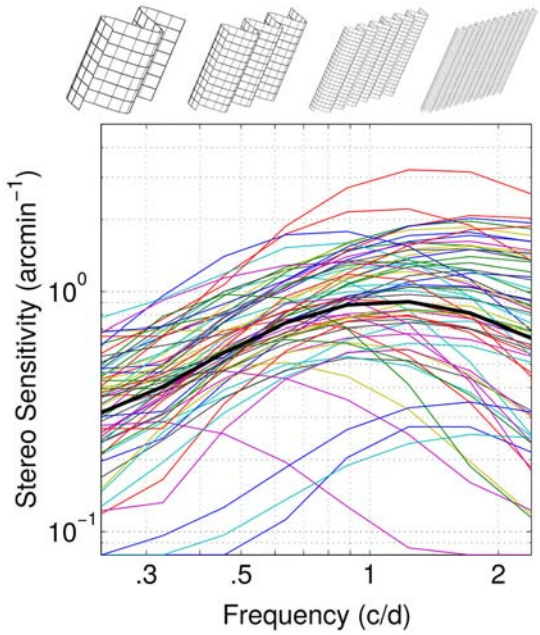


FIGURE 1 | Normative dataset. Disparity sensitivity as a function of spatial frequency is reported for 61 individual observers (thin color lines) and their average (thick black line). Sketches at the top illustrate the stimulus at different corrugations frequencies. Adapted with permission from Reynaud et al. (2015).

METHODS

In this paper, we analyze the normative dataset for the disparity sensitivity as a function of spatial frequency of 61 observers (25 males, 36 females, mean age 26 years, ±5.7 SD, with normal or corrected to normal-visual acuity) we measured previously using the quick Disparity Sensitivity Function (*qDSF*, Reynaud et al., 2015), a method adapted from the quick Contrast Sensitivity Function (*qCSF*, Lesmes et al., 2010).

The stimuli used in this dataset were stereograms composed of spatially filtered 2-D fractal noise carriers with oblique (45° or 135°) sinusoidal corrugations at 0.24, 0.33, 0.46, 0.64, 0.89, 1.23, 1.72, and 2.39 c/d. The spatial frequency of the carrier was 4 times the spatial frequency of the corrugation (see Reynaud et al., 2015). Disparity was modulated and the subjects' task was to identify the orientation of the corrugation in depth (45° or 135°) in a single-interval identification task to measure the disparity detection threshold. Stimuli were displayed on a passive wide 23″ 3D-Ready LED monitor ViewSonic V3D231, viewed with polarized 3D glasses at 70 cm, in a dim-lit room. Measured individual disparity sensitivity functions as a function of spatial frequency and their average are reproduced in **Figure 1**. Analysis was performed with Matlab R2016a (The MathWorks). The hierarchical clustering analysis was specifically performed with the statistics and machine learning toolboxes functions.

RESULTS

The average disparity sensitivity peaks are in the high spatial frequency range, around 1.2 c/d. However, we can observe a large variability in the individual sensitivities: some showing a low-pass, band-pass or high-pass profiles (**Figure 1**). Hence a factor analysis of these sensitivities might provide insight into the common mechanisms mediating them.

Figure 2 represents the scatterplot matrix of inter-correlations (Peterzell, 2016) for log-disparity sensitivity of all 61 observers. In each cell within the figure, the scatterplot represent the inter-correlation of the log-disparity sensitivity of all observers at one frequency (frequency indicated on the diagonal in the same row) as a function of their sensitivity at another frequency (frequency indicated on the diagonal in the same column) are depicted. For instance, in the bottom-left cell, the log-disparity sensitivity of each observer at 0.24 c/d is plotted pairwise against its log-disparity sensitivity at 2.39 c/d. Then the coefficient of determination R^2 between the two frequencies is computed. Two regions of high inter-correlations ($R^2 > 0.5$) at low spatial frequency (green) and high spatial frequency (blue) appear along the diagonal.

These two regions are supported by the hierarchical clustering analysis of the log-disparity sensitivity at all spatial frequencies. The pairwise distance between observations was calculated as one minus the sample linear correlation between observations and the hierarchical cluster tree was computed with the average distance. The resulting dendrogram is represented at the right of the inter-correlation matrix, with each spatial frequency being the leaves. Nevertheless, we can note that different distance measures and different linkage procedures can result in relatively different final clusters, some grouping the 3 lowest and 5 highest frequencies for instance. The two cluster branches whose linkage is less than the default 70% are represented in blue and green. As for the first qualitative approach, these two groups suggest the presence of two spatial frequency channels for disparity sensitivity, which might correspond to the coarse and fine disparity channels.

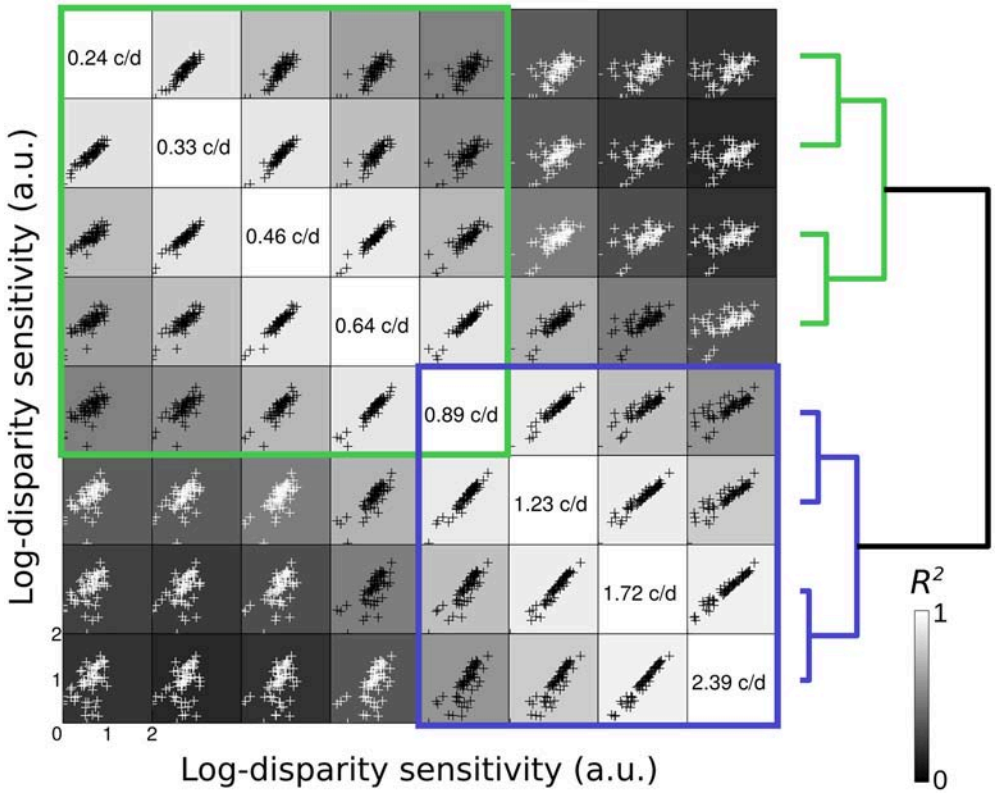


FIGURE 2 | Scatterplot matrix of inter-correlations. In each cell, the scatterplot represent the inter-correlation of the log-disparity sensitivity (arbitrary units) of all 61 observer at one frequency (frequency indicated on the diagonal in the same row) as a function of their sensitivity at another frequency (frequency indicated on the diagonal in the same column). The shade of the background in each cell indicates the value of the coefficient of determination R^2 between the two frequencies (from black = 0 to white = 1). Black datapoints indicate $R^2 > 0.5$ and white datapoints $R^2 < 0.5$. Blue and green squares highlight regions of high inter-correlations. On the right is represented the classification dendrogram of the spatial frequencies. The pairwise distance was calculated as one minus the sample linear correlation between observations and the hierarchical cluster tree was computed with the average distance.

In order to determine the precise tuning of these channels, we performed a factor analysis on the dataset. If we decompose the full dataset with a principal component analysis (PCA), we obtain the components shown in **Figure 3A**, with a percentage of explained variance (calculated from the eigenvalues of the PCA) associated with each component reported in the scree plot **Figure 3B**.

The first component has the shape of the average sensitivity (see **Figure 1**). The two first components (blue and green) explain more than 91% of the variance and the elbow of the scree plot occurs between the second and third components (**Figure 3B**). As we previously identified two regions of high inter-correlations and that this percentage of explained variance is considered enough to accurately describe the data (Simpson and McFadden, 2005), these two principal components were picked to describe the underlying disparity sensitivity channels. In order to make sense of them, these two principal components, or factors, were then rotated using a varimax orthogonal rotation to obtain a simple structure accounting for the channel tuning curves (Kaiser, 1958; Peterzell and Teller, 2000; Simpson and McFadden, 2005; Peterzell, 2016). These factors-tuning curves are reported in **Figure 3C**. The first factor peaks at the highest measured frequency 2.4 c/d and the second peaks around 0.65 c/d. They characterize the high and low spatial frequency channels identified by the inter-correlation analysis (respectively blue and green regions in **Figure 2**).

We wanted to test if the two channels we identified could in fact account for different classes within the population. In order to estimate the weights β of each of these factors in each individual sensitivity, we projected our dataset onto the basis defined by the two identified factors. The best linear unbiased estimator of β is obtained using the Moore-Penrose pseudo inverse X^+ (equation 1):

$$\beta = X^+ y \tag{1}$$

where y is the matrix of all individual sensitivities, X^+ is the Moore-Penrose pseudo inverse of the new basis matrix X whose two columns represent the two factors and β is a two-rows matrix in wihich each column contains the pair of weights associated to the two factors estimated for each subject (Friston et al., 1995; Woolrich et al., 2004; Reynaud et al., 2011).

The sensitivities $\hat{y}$ reconstructed solely from the linear combination of these two factors are plotted in **Figure 4A**

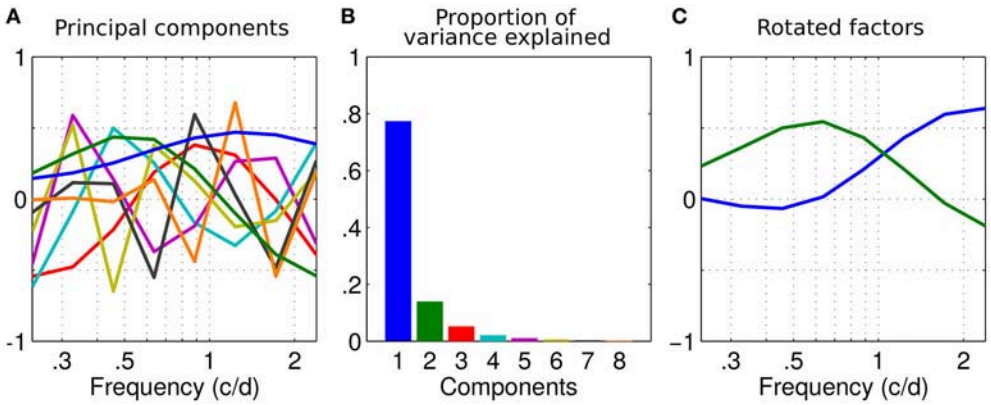


FIGURE 3 | Factor analysis. **(A)** Principal components of the dataset as a function of spatial frequency. Their order is indicated by colors in **(B)**. **(B)** Scree plot of the variance explained by each component of the principal component analysis (PCA) in **(A)**. **(C)** First two components rotated using a varimax rotation.

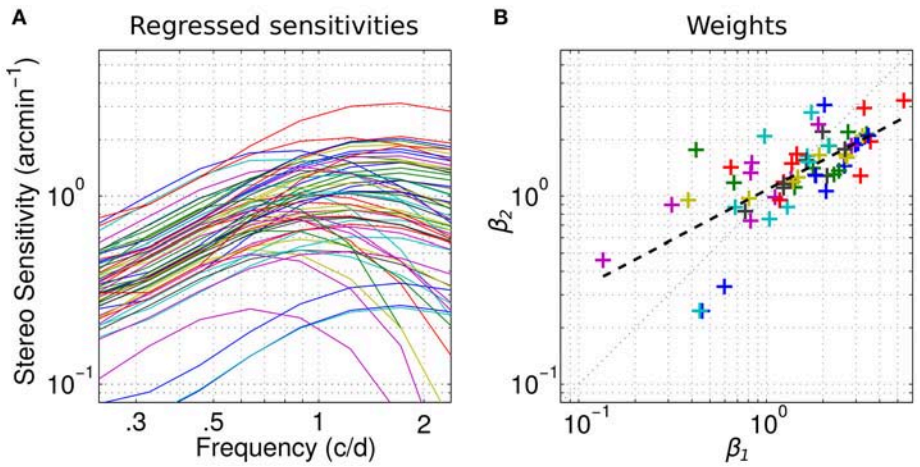


FIGURE 4 | Channels weights. **(A)** Individual sensitivities replotted using only the two channels factors. Same color-code as in **Figure 1**. **(B)** Scatterplot of the weights of the first factor β_1 vs. the weights of the second factor β_2 for all observers. Dashed line indicates linear regression on the log-values of the weights.

(Equation 2):

$$\hat{y} = \mathrm{X}\beta \tag{2}$$

We can see that they overall faithfully reproduce the original sensitivities except for the very low-pass profiles whose peaks shift to the right.

To determine whether these channels can account for different classes within the population, we report a scatterplot of the weights β_1 of the first factor vs. the weights β_2 of the second factor in **Figure 4B** for all observers. The mean weights for the first and second factor are, respectively, 1.76 and 1.48. As expected from the explained variance (**Figure 3B**), the weight of the first factor—the high-frequency channel—is greater than the weight of the second—the low frequency channel—in 70% of the cases. The distribution of these weights appears homogeneous and no clusters are revealed. However, the weights of the first factor seem to be relatively greater than the weights of the second in the high values range whereas it seems to be slightly the opposite in the low values range. This is further revealed by the slope of the linear regression between the log-values of the weights 0.53, which is inferior to 1 (dashed line). In fact, the correlation between the weight is very high (coefficient of determination $R^2 = 0.51$, $p < 0.0001$). Altogether, these observations suggest that the weight of the low and high spatial frequency channels co-vary: when the sensitivity is high for the low frequency channel, it is high for the high frequency channel too. But the high frequency channel contributes relatively more when the sensitivity is high and the low-frequency channel contributes relatively more when the sensitivity is low, in accordance with our previous observations (Reynaud et al., 2015).

DISCUSSION

The qDSF method assumes the sensitivity function follows the truncated log-parabola model and hence has a bell shape with a constant part, an increase to a peak and a drop-off

(Watson and Robson, 1981; Lesmes et al., 2010). We previously showed that this model can accurately represent the sensitivity function compared to non-constrained methods (Reynaud et al., 2015) and documents large differences in sensitivities within the population (see **Figure 1**). For different individuals, this function can peak at very different frequencies and can show low-pass, band-pass or high-pass profiles. The resultant variability in sensitivity across spatial frequency provides a rich dataset for inter-correlation analyses (Peterzell et al., 1995; Peterzell, 2016).

Because two regions of inter correlations were identified among the population in **Figure 1** and because 2 components accounted for more than 91% of the variance, our data could accurately be described by just 2 channels. However, the criterion to select the number of meaningful components in a PCA may vary. Popular selection methods such as a scree plot (Jackson, 1993) or the Random average under permutation analysis will indeed determine 2 components while some other methods will give less (the broken stick method gives 1 component) or more (the parallel analysis gives barely 3, the kaiser Guttman criterion which recommends eigenvalues >1 gives 3 too). Some methods such as the Bartlett tests even recommends all the 8 components which would not reduce the dimensionality of the data (Bartlett, 1950). A complete description of these methods can be found in Peres-Neto et al. (2005).

Hence, we cannot completely rule out the possibility of a single-channel or multiple-channels hypothesis. Serrano-Pedraza and Read reported a single channel mechanism specific to vertical corrugations (Serrano-Pedraza and Read, 2010, though see Witz et al., 2014). However, the large difference we can observe between the lowpass profile of sensitivity for some observers compared to the bandpass of other ones would indicate that more than one channel are involved. Several studies suggested a multiple-channels mechanism (Julesz and Miller, 1975; Schumer and Ganz, 1979; Cobo-Lewis and Yeh, 1994; Glennerster and Parker, 1997; Serrano-Pedraza et al., 2013) with a broad channel tuning of ∼2–3 octaves, comparable to our observations (Schumer and Ganz, 1979; Cobo-Lewis and Yeh, 1994). It is then possible that the 2 channels we observe are part of a multiple-channels system covering a wider range of spatial frequencies or could also overlap with intermediate channels continuously covering the spatial frequency range. Yang and Blake (1991) also observed two spatial frequency channels for disparity sensitivity using a masking paradigm. They described one channel centered around 3 c/d which could correspond to the high spatial frequency channel we observed and one centered around 5 c/d. However, their study and the present study didn't measure the same spatial frequency range which might explain why they didn't identify our low spatial frequency channel and why we didn't observe their high one.

The results of the present study suggests that there are two channels (**Figure 4B**), a low frequency channel that contributes to the detection of low corrugation frequencies and a more sensitive high frequency channel that contributes to the detection of high corrugation frequencies. We didn't observe any dichotomy based on these two channels within our population (Wilcox and Allison, 2009) which confirms the observations of most other population studies (Coutant and Westheimer, 1993; Bohr and Read, 2013; Bosten et al., 2015).

The implications of the assessment of the tuning of these disparity channels could be important in computer vision to design behaviorally relevant stereo matching algorithms. For instance, it could be used to tune the different layers of multi-scale algorithms (Rohaly and Wilson, 1993) or provide fine and coarse scales for algorithms processing in center and periphery, respectively, as stereopsis could be mediated by different mechanisms in central and peripheral vision (Wardle et al., 2012; Witz and Hess, 2013).

CONCLUSION

The analysis of the inter-correlations in the disparity sensitivity as a function of the spatial frequency, revealed two disparity channels. With a factor analysis of the population data, we determined that the first channel is tuned to high spatial frequencies (peaks at 2.4 c/d) and the second is tuned to low spatial frequencies (peaks at 0.65 c/d). We also observed that these two channels are well correlated with each other. Our results confirm that disparity is encoded by multiple spatial frequency channels that are of different sensitivities and subserve different ranges of corrugation frequencies.

AUTHOR CONTRIBUTIONS

AR and RH designed the research and wrote the manuscript. AR analyzed the data.

FUNDING

This work was supported by a Natural Sciences and Engineering Research Council of Canada grant (NSERC #46528) to RH.

ACKNOWLEDGMENTS

We thank the three reviewers for their helpful comments and suggestions. This work was supported by a Natural Sciences and Engineering Research Council of Canada grant (NSERC #46528) to RH.

REFERENCES

Bartlett, M. S. (1950). Tests of significance in factor analysis. *Br. J. Stat. Psychol.* 3, 77–85.

Billock, V. A., and Harding, T. H. (1996). Evidence of spatial and temporal channels in the correlational structure of human spatiotemporal contrast sensitivity. *J. Physiol.* 490(Pt 2), 509–517.

Bohr, I., and Read, J. C. A. (2013). Stereoacuity with frisby and revised FD2∼stereo tests. *PLoS ONE* 8:82999. doi: 10.1371/journal.pone.0082999

Bosten, J. M., Goodbourn, P. T., Lawrance-Owen, A. J., Bargary, G., Hogg, R. E., and Mollon, J. D. (2015). A population study of binocular function. *Vision Res.* 110(Pt A), 34–50. doi: 10.1016/j.visres.2015.02.017

Cobo-Lewis, A. B., and Yeh, Y. Y. (1994). Selectivity of cyclopean masking for the spatial frequency of binocular disparity modulation. *Vision Res.* 34, 607–620. doi: 10.1016/0042-6989(94)90016-7

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

Friston, K., Holmes, A., Worsley, K., Poline, J., Frith, C., and Frackowiak, R. (1995). Statistical parametric maps in functional imaging: a general linear approach *human brain. Mapping* 2, 189–210. doi: 10.1111/j.1475-1313.1993.tb00419.x

Glennerster, A., and Parker, A. (1997). Computing stereo channels from masking data. *Vision Res.* 37, 2143–2152.

Jackson, D. A. (1993). Stopping rules in principal components analysis: a comparison of heuristical and statistical approaches. *Ecology* 74, 2204–2214. doi: 10.2307/1939574

Julesz, B., and Miller, J. E. (1975). Independent spatial-frequency-tuned channels in binocular fusion and rivalry. *Perception* 4, 125–143. doi: 10.1068/p040125

Kaiser, H. F. (1958). The varimax criterion for analytic rotation in factor analysis. *Psychometrika* 23, 187–200.

Lesmes, L. A., Lu, Z.-L., Baek, J., and Albright, T. D. (2010). Bayesian adaptive estimation of the contrast sensitivity function: the quick CSF method. *J. Vision* 10, 17.1–17.21. doi: 10.1167/10.3.17

Lonini, L., Forestier, S., Teulière, C., Zhao, Y., Shi, B., and Triesch, J. (2013). Robust active binocular vision through intrinsically motivated learning. *Front. Neurorobot.* 7:20. doi: 10.3389/fnbot.2013.00020

Marr, D., and Poggio, T. (1979). A computational theory of human stereo vision. *Proc. R. Soc. Lond. B Biol. Sci.* 204, 301–328 doi: 10.1098/rspb.1979.0029

Mayer, M. J., Dougherty, R. F., and Hu, L.-T. (1995). A covariance structure analysis of flicker sensitivity. *Vision Res.* 35, 1575–1583.

Nishihara, H. K. (1984). Practical real-time imaging stereo matcher. *Opt. Eng. SPIE. Int. Soc. Opt. Eng.* 23.

Norcia, A. M., Sutter, E. E., and Tyler, C. W. (1985). Electrophysiological evidence for the existence of coarse and fine disparity mechanisms in human. *Vision Res.* 25, 1603–1611. doi: 10.1016/0042-6989(85)90130-0

Owsley, C., Sekuler, R., and Siemsen, D. (1983). Contrast sensitivity throughout adulthood. *Vision Res.* 23, 689–699. doi: 10.1016/0042-6989(83)90210-9

Peres-Neto, P. R., Jackson, D. A., and Somers, K. M. (2005). How many principal components? stopping rules for determining the number of non-trivial axes revisited. *Comput. Stat. Data Anal.* 49, 974–997. doi: 10.1016/j.csda.2004.06.015

Peterzell, D. H. (2016). Discovering sensory processes using individual differences: a review and factor analytic manifesto. *Electron. Imaging Human Vision Electron. Imaging* 11, 1–11. doi: 10.2352/ISSN.2470-1173.2016.16.HVEI-112

Peterzell, D. H., and Teller, D. Y. (1996). Individual differences in contrast sensitivity functions: the lowest spatial frequency channels. *Vision Res.* 36, 3077–3085. doi: 10.1016/0042-6989(96)00061-2

Peterzell, D. H., and Teller, D. Y. (2000). Spatial frequency tuned covariance channels for red–green and luminance-modulated gratings: psychophysical data from human adults. *Vision Res.* 40, 417–430. doi: 10.1016/S0042-6989(99)00187-X

Peterzell, D. H., Werner, J. S., and Kaplan, P. S. (1995). Individual differences in contrast sensitivity functions: longitudinal study of 4-, 6- and 8-month-old human infants. *Vision Res.* 35, 961–979. doi: 10.1016/0042-6989(94)00117-5

Piater, J. H., Grupen, R. A., and Ramamritham, K. (1999). "Learning real-time stereo vergence control," in *Proceedings of the 1999 IEEE International Symposium on Intelligent Control Intelligent Systems and Semiotics (Cat. No. 99CH37014)*, (Cambridge, MA), 272–277.

Prince, S. J. D., Eagle, R. A., and Rogers, B. J. (1998). Contrast masking reveals spatial-frequency channels in stereopsis. *Perception* 27, 1345–1355. doi: 10.1068/p271345

Pulliam, K. (1982). "Spatial frequency analysis of three-dimensional vision," in *Visual Simulation and Image Realism II, Proceedings of SPIE International Socity for Optical and Photonics*, ed K. Setty, (San Diego, CA).

Quam, L. H. (1987). "Readings in Computer Vision: Issues, Problems, Principles, and Paradigms," in *Hierarchical Warp Stereo*, eds M. A. Fischler and O. Firschein (San Francisco, CA: Morgan Kaufmann Publishers Inc.), 80–86.

Read., J., Serrano-Pedraza, I., Widdall, M., and Peterzell, D. (2016). Sensitivity to horizontal and vertical sine-wave corrugations defined by binocular disparity: factor analysis of individual differences reveals discrete processes with broad orientation and spatial frequency tuning. *J. Vision* 16:833. doi: 10.1167/16.12.833

Reynaud, A., Gao, Y., and Hess, R. F. (2015). A normative dataset on human global stereopsis using the quick Disparity Sensitivity Function (qDSF). *Vision Res.* 113 (Pt A), 97–103. doi: 10.1016/j.visres.2015.04.021

Reynaud, A., Takerkart, S., Masson, G. S., and Chavane, F. (2011). Linear model decomposition for voltage-sensitive dye imaging signals: application in awake behaving monkey. *Neuroimage* 54, 1196–1210. doi: 10.1016/j.neuroimage.2010.08.041

Rohaly, A. M., and Wilson, H. R. (1993). Nature of coarse-to-fine constraints on binocular fusion. *J. Opt. Soc.* 10, 2433–2441. doi: 10.1364/JOSAA.10.002433

Rosli, Y., Bedford, S. M., and Maddess, T. (2009). Low-spatial-frequency channels and the spatial frequency-doubling illusion. *Invest. Opthalmol. Visual Sci.* 50, 1956–1963. doi: 10.1167/iovs.08-1810

Schumer, R., and Ganz, L. (1979). Independent stereoscopic channels for different extents of spatial pooling. *Vision Res.* 19, 1303–1314. doi: 10.1016/0042-6989(79)90202-5

Sekuler, R., Wilson, H. R., and Owsley, C. (1984). Structural modeling of spatial vision. *Vision Res.* 24, 689–700.

Serrano-Pedraza, I., Brash, C., and Read, J. C. A. (2013). Testing the horizontal-vertical stereo anisotropy with the critical-band masking paradigm *J. Vision* 13, 15–15. doi: 10.1167/13.11.15

Serrano-Pedraza, I., and Read, J. C. A. (2010). Multiple channels for horizontal, but only one for vertical corrugations? A new look at the stereo anisotropy. *J. Vision* 10:10. doi: 10.1167/10.12.10

Shioiri, S., Hatori, T., Yaguchi, H., and Kubo, S. (1994). Spatial frequency channels for stereoscopic depth perception. *Opt. Rev.* 1, 311–313.

Simpson, W. A., and McFadden, S. M. (2005). Spatial frequency channels derived from individual differences. *Vision Res.* 45, 2723–2727. doi: 10.1016/j.visres.2005.01.015

Tsai, R. Y. (1986). "An efficient and accurate camera calibration technique for 3d machine vision," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition* (Miami Beach, FL), 364–374.

Tyler, C. W., Barghout, L., and Kontsevich, L. L. (1994). "Computational reconstruction of the mechanisms of human stereopsis," in *Computational Vision Based on Neurobiology, SPIE International Society Optical Engineering*, ed T. B. Lawton, (Park Grove, CA).

Wardle, S. G., Bex, P. J., Cass, J., and Alais, D. (2012). Stereoacuity in the periphery is limited by internal noise. *J. Vision* 12:12. doi: 10.1167/12.6.12

Watson, A. B., and Robson, J. G. (1981). Discrimination at threshold: labelled detectors in human vision. *Vision Res.* 21, 1115–1122. doi: 10.1016/0042-6989(81)90014-6

Wilcox, L. M., and Allison, R. S. (2009). Coarse-fine dichotomies in human stereopsis. *Vision Res.* 49, 2653–2665. doi: 10.1016/j.visres.2009.06.004

Witz, N., and Hess, R. F. (2013). Mechanisms underlying global stereopsis in fovea and periphery. *Vision Res.* 87, 10–21. doi: 10.1016/j.visres.2013.05.003

Witz, N., Zhou, J., and Hess, R. F. (2014). Similar mechanisms underlie the detection of horizontal and vertical disparity corrugations. *PLoS ONE* 9:e84846. doi: 10.1371/journal.pone.0084846

Woolrich, M. W., Behrens, T. E. J., and Smith, S. M. (2004). Constrained linear basis sets for HRF modelling using Variational Bayes. *Neuroimage* 21, 1748–1761. doi: 10.1016/j.neuroimage.2003.12.024

Yang, Y., and Blake, R. (1991). Spatial frequency tuning of human stereopsis. *Vision Res.* 31, 1177–1189 doi: 10.1016/j.neuroimage.2003.12.024

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Computational Neuroscience

REVIEW
published: 14 November 2017
doi: 10.3389/fncom.2017.00102

Computational and Experimental Approaches to Visual Aesthetics

*Anselm Brachmann and Christoph Redies**

Experimental Aesthetics Group, Institute of Anatomy, Jena University Hospital, School of Medicine, University of Jena, Jena, Germany

Aesthetics has been the subject of long-standing debates by philosophers and psychologists alike. In psychology, it is generally agreed that aesthetic experience results from an interaction between perception, cognition, and emotion. By experimental means, this triad has been studied in the field of *experimental aesthetics*, which aims to gain a better understanding of how aesthetic experience relates to fundamental principles of human visual perception and brain processes. Recently, researchers in computer vision have also gained interest in the topic, giving rise to the field of *computational aesthetics*. With computing hardware and methodology developing at a high pace, the modeling of perceptually relevant aspect of aesthetic stimuli has a huge potential. In this review, we present an overview of recent developments in computational aesthetics and how they relate to experimental studies. In the first part, we cover topics such as the prediction of ratings, style and artist identification as well as computational methods in art history, such as the detection of influences among artists or forgeries. We also describe currently used computational algorithms, such as classifiers and deep neural networks. In the second part, we summarize results from the field of experimental aesthetics and cover several isolated image properties that are believed to have a effect on the aesthetic appeal of visual stimuli. Their relation to each other and to findings from computational aesthetics are discussed. Moreover, we compare the strategies in the two fields of research and suggest that both fields would greatly profit from a joined research effort. We hope to encourage researchers from both disciplines to work more closely together in order to understand visual aesthetics from an integrated point of view.

Keywords: computational aesthetics, experimental aesthetics, visual preference, art history, artist identification, style identification, image features, statistical image properties

OPEN ACCESS

Edited by:
Xavier Otazu,
Universitat Autònoma de Barcelona, Spain

Reviewed by:
Qing Yun Wang,
Beihang University, China
Jesús Malo,
Universitat de València, Spain

****Correspondence:***
Christoph Redies
christoph.redies@med.uni-jena.de

Received: *30 May 2017*
Accepted: *30 October 2017*
Published: *14 November 2017*

Citation:
Brachmann A and Redies C (2017) Computational and Experimental Approaches to Visual Aesthetics. Front. Comput. Neurosci. 11:102. doi: 10.3389/fncom.2017.00102

1. INTRODUCTION

Dating back more than two thousand years ago, aesthetics has been the subject of debates by philosophers and other scholars alike. Defined by the Oxford Dictionary as "the philosophy of the beautiful or of art," "a system of principles for the appreciation of the beautiful," and "the distinctive underlying principles of a work of art or a genre" (OED, 2017), aesthetics represents a field of interest that has attracted researchers from diverse scientific disciplines, also outside of philosophy. In 1876, the founder of experimental aesthetics, Gustav Fechner, published his seminal book entitled "Vorschule der Ästhetik" (Fechner, 1876). He believed that the aesthetic appeal of physical objects manifests itself in stimulus properties that can be measured in an objective (formalistic) way. Specifically, he attempted to show that rectangles with an aspect ratio equal to the golden ratio

are more appealing to human observers than rectangles having other aspect ratios. Researcher have later raised concerns about the normative role in rectangular preferences (Green, 1995; McManus et al., 2010). Nevertheless, Fechner's scientific (objective) view of aesthetics provided the basis for the newly emerging field of *empirical aesthetics*. In this field, hypotheses regarding the perceived beauty of images, paintings or even every-day objects are proposed and tested experimentally for their validity. This stimulus-driven approach, called by Fechner *aesthetics from below*, was different from the aesthetics that was prevalent in Fechner's time and derived aesthetic principles from superordinate philosophical concepts *(aesthetics from above)* (Cupchik, 1986). Fechner is also credited for conceiving the field of *psychopysics*, which relates human perception to well-defined physical properties of stimuli. By applying this approach to aesthetics, he attempted to relate physical image properties to aesthetic perception in humans. The area of research that has taken up this idea in modern times is *experimental aesthetics*, a subfield of psychology.

Another discipline of natural science that studies aesthetics is *neuroaesthetics*, a subfield of brain research. In this field, modern imaging techniques, such as functional magnetic resonance imaging (fMRI), enable researcher to study the activation of brain regions when human observers view aesthetic stimuli (Cela-Conde et al., 2011; Chatterjee and Vartanian, 2014). This type of research has lead to a better understanding of what neural networks are involved in the human brain when we have an aesthetic experience. Research in neuroaesthetics is beyond the scope of the present review.

In recent years, aesthetics has also been studied using computational methods. In the field of computer science, *computational aesthetics*, a subfield of computer vision, has entered the field of aesthetics. In this area, there have been a variety of different studies on the aesthetics in digital images, for example, using digital reproductions of paintings. The birth of computational aesthetics is often attributed to Birkhoff's book "Aesthetic Measure" (Birkhoff, 1933), although the book does not mention the term itself (for an overview of the evolution of the term, see Greenfield, 2005). In a very mathematical way, Birkhoff proposed a formula for an aesthetic measure M, which is a function of O, order or reward by a positive tone of feeling, and C, complexity or a feeling of effort of attention. Stating that reward should be proportional to effort, Birkhoff concludes that $M = O/C$ best describes their relation.

A definition of computational aesthetics is given by Hoenig (2005), who describes it as "[...] the research of computational methods that can make applicable aesthetic decision in a similar fashion as humans can." To Hoenig, this definition emphasizes two major aspects: First, the use of computational methods, and second, their applicability to aesthetic decision making. More precisely, Galanter (2012) discusses how computational aesthetics is concerned with both, "the creation and evaluation of art using computers." He argues that the creation of art necessarily requires evaluation and gives the example of an artist, who, while learning about aesthetics and gathering experience, evaluates art created by others. When creating artworks himself, micro-evaluations help the artist guide his own creative process. Upon finishing his creation, the artist gains new insights about his art in a final evaluation of the created piece. Given the importance of the evaluation process, we will focus on it in the present review. As pointed out by Stork (2009a), the computational analysis of paintings has several advantages compared to an analysis carried out by human experts. For example, a computational analysis can pick up very subtle relationships that may escape the attention by human observers; moreover, computational methods are objective in nature and are potentially non-exhaustive in the amount of detail analyzed (e.g., every single brushstroke in a painting).

The aim of the present review is to provide an overview of recent developments in the field of computational aesthetics and to point out its potential relevance for research in experimental aesthetics and vice versa. Our goal is to boost the awareness of researchers in experimental aesthetics for the wealth of data that computational aesthetics has generated in recent years. We would also like to inform scientists in computational aesthetics about some basic concepts and results from experimental aesthetics. Our review thus outlines a possible link between research on the objective (physical) properties of visual stimuli and experimental studies that take into account the subjective responses of humans to aesthetic stimuli, as originally proposed by Fechner. Specifically, we focus on the evaluation of visual images (photographs or digitally reproduced artworks) and the analysis of image properties. Important areas of research will be referenced and exemplary works will be presented, without striving for completeness. Topics include the prediction of ratings of photographs and paintings, the classification of images regarding their artist or style, computational methods for problems in art history, and, finally, the investigation of statistical properties of aesthetically pleasing images and artworks.

2. COMPUTATIONAL AESTHETICS: ALGORITHMS AND APPLICATIONS

Computational aesthetics is approached from different points of view. All articles reviewed here somehow deal with aesthetics in the form of photography and paintings and are motivated predominantly by producing applications and testing or improving algorithms. Accordingly, one of the tasks that is often pursued in computational aesthetics is to develop algorithms that allow to predict aesthetic ratings of photographs. Such algorithms have direct applications. For example, in online photo communities (for example Flickr, Photo.net, etc.), they can be used to select photographs of high aesthetic quality and discard snapshots that users would rate low. On a more commercial side, such systems are used for retrieving and licensing high-quality photographs from the internet for their use as stock photographs. Another possible application is to install such algorithms in industrial cameras and smartphones, which identify high-quality images in the split of a second. As we will show in the present article, there has been a tremendous success in building such systems.

The prediction of ratings is just one possible application among many, where computers can make decisions regarding

aesthetics. Computational methods have also been successfully applied to problems in art history, such as content analysis of paintings, forgery detection, or detection of a painter's influence. These applications will also be reviewed in the following sections.

2.1. Prediction of Ratings

One major trend in computational aesthetics is to predict ratings of image quality or aesthetic appeal. Possible applications of this technology are improved cameras, which automatically select the most appealing photos among many, optimization of advertisements for their aesthetic value, or even talent scouting in photo-sharing communities. In the early days of computational aesthetics, researcher followed the then popular practice to design features explicitly for a given task. In order to predict the aesthetic appeal of a given image, researchers determined in how far different photographic principles, like composition according to the rule of thirds or depth of field, were followed in images. They quantified these principles by expressing them numerically, either as binary or continuous values, called features. Features can be either local, describing only pixels or patches and their immediate neighborhood, or they can be global and describe properties of the image as a whole. Global features seem especially suitable to describe artistic photographs or artworks because concepts such as artistic composition refer to the relation between pictorial elements across the image. Another difference can be made concerning the level of abstraction: Low-level features describe basic features, such as colors and edges, while high-level features can describe more abstract image content. The features can then be used to train a classifier on a dataset of images so that it can learn to predict ratings given by humans. This goal is achieved by mathematically describing the relation between the subjective scores and the feature set. Popular choices for classifiers are, for example, Bayes classifiers, Decision Trees, or Support Vector Machines (SVMs). This approach will be presented in more detail in section 2.1.1. In recent years, computational aesthetics has gone from designing features by hand to using generic features that have been developed for other purposes in computer vision. This development has reached a pinnacle with the development and widespread use of Deep Neural Networks. Approaches using generic features will be discussed in section 2.1.2.

2.1.1. Hand-Crafted Image Features

One of the first attempts to measure aesthetics in an image was published by Tong et al. (2004), who proposed a method to distinguish between photographs taken by professional photographers and photographs taken by non-expert (home) users. They used a set of low-level features that describe blur, contrast, colorfulness and saliency, and combined it with general purpose low-level features that capture texture, shape and energy in the frequency spectrum, by using difference-edge histograms. In total, they proposed 21 different features which added up to 846 dimensions. After reducing the dimensionality, they reported classification results comparing Boosting, an SVM and a Bayesian classifier, which performed best.

Using another set of low-level features, Datta et al. (2006) build a classifier for distinguishing images of high aesthetic appeal from other images, as rated by the community of the popular photo-sharing website Photo.net. Overall, the authors collected 3,581 different images and split them into two classes according to their aesthetic rating by the users of the site (low and high rating). They explicitly stated that their goal was not to build the best-performing classifier, but rather to be able to draw conclusions from the best performing features. Their choice of features was based on common intuition, rules of thumb in photography and trends that they observed for the ratings of the collected images. In total, they proposed a set of 56 different features, containing basic ones, such as colorfulness, saturation, hue, size and aspect ratio, as well as adherence to the rule of thirds. The features were selected as follows: First, the authors used a one-dimensional SVM to find the features with the most discriminative power and selected the top 30. Starting with an empty features set, they then iteratively added those features that improved the classification the most. As a result, they found that average hue, average pixel intensity as well as a saturation-based rule of thirds measure contributed the most to the aesthetic value of an image, as rated by human observers.

Ke et al. (2006) designed a system to distinguish between high-quality professional photographs and low-quality snapshots. They reference the work of Tong et al. (2004) but criticize their black-box approach, which prevents them from gaining any insight into why some photos are better than others, although the system by Tong and colleagues performed well for the task. Ke et al. (2006) therefore chose an approach similar to the one by Datta et al. (2006) and designed a set of features that capture image quality. They based their choice of the features on interviews conducted with photographers. Their feature set contained the spatial distribution of edges, color distribution, hue count and blur as well as contrast and brightness. For classification, they used a naive Bayes classifier and tested their system on images that were downloaded from a photo contest website. The blur feature turned out to be the most discriminative metric.

Luo and Tang (2008) extracted very simple features that captured lighting, simplicity, composition or color harmony, based on the subject region and the background of an image. They reported an improvement of classification upon Datta et al. (2006) and Ke et al. (2006) and contributed this success to the distinction of foreground and background, while the previous methods computed their features on the image as a whole.

Besides focusing on low-level features as provided by Ke et al. (2006) and Dhar et al. (2011) also integrate high-level attributes in their system in order to predict aesthetic value and interestingness. According to the authors, high-level attributes define characteristics of images as humans would describe them, and can be classified into compositional attributes (like the rule of thirds), content attributes (like the presence of people) and sky illumination attributes. Dhar et al. (2011) reported improved performance compared to the approach by Ke et al. (2006).

Although the general focus of aesthetic quality assessment in computational aesthetics is on the prediction of ratings of photographs, a few researchers have also proposed methods for quality assessment of paintings. Li and Chen (2009), for example, propose a total of 40 features that capture color, brightness and

compositional characteristics of a paintings. Using these features, they use a Bayes classifier as well as AdaBoost on a binary task to predict whether a painting received high or low rating scores. In their work, they provide a detailed discussion of the importance of the individual features.

What all these approaches have in common is that a combination of multiple features is used to predict aesthetic ratings. While this has proven successful for automated aesthetic decision making, there are a number of problems that preclude a deeper understanding of the role of individual features in these decisions. First, because the features are not necessarily independent of each other, it would require more sophisticated statistical methods to extract the influence of each of them. Second, the experimental conditions, under which ratings are obtained in most of the above-mentioned studies, are unknown, unspecified or variable (for example, with regard to the size of the stimuli on the retina, the brightness of the stimuli, contrast settings of the monitors, background illumination, sequence of stimulus presentation etc.,). Third, the rating by users of internet platforms often remain anonymous which precludes any specification of their personal characteristics (sex, age, cultural background etc.,). All these factors might influence the results or introduce artifacts.

In experimental aesthetics, some of the features used in the above combinatorial approaches have been isolated and studied in psychological experiments under well-defined experimental conditions (for a survey of such studies, see section 3).

2.1.2. Generic Image Features

Generic image features are features that are not explicitly designed for the prediction of image aesthetics, but rather for other popular research topics in computer vision, like object detection and classification, scene understanding, or image retrieval. An example of such features are the SIFT descriptors (scale-invariant feature transform; Lowe, 2004), which were originally designed for feature matching and image stitching. SIFT encodes edge orientations in gray-scale images as a vector (for more recent image descriptors, see Canclini et al., 2013).

The first study to model aesthetic ratings based on generic image features was published by Marchesotti et al. (2011). They used SIFT descriptors together with a color descriptor, motivated by the assumption that aesthetic properties, such as the presence of sharp edges or the saturation of colors, can be described implicitly by these kind of features. The authors chose a Bag-Of-Visual-Words and a Fisher-Vector representation in order to represent prototypical patches for aesthetic and non-aesthetic photographs. As a result, they reported an improvement in classification rates for high-quality and low-quality images, compared to the methods by Datta et al. (2006) and Ke et al. (2006) who used hand-crafted features (see section 2.1.1). While hand-crafted features allow to quantify which feature contributes the most to an aesthetic rating, this interpretability is lost with generic features. Here, conclusions can only be drawn by a comparison of the images that are rated high or low by the model because the features of the model are not deliberately designed to capture known properties of aesthetics, but they rather hide their relation to them. For example, Marchesotti et al. report that all blurry and low-resolution images were rated low in his model, whereas images that displayed foreground objects with sharp edges on out-of-focus backgrounds were rated highly. Moreover, highly-rated images had a dominant color or used complementary colors in their palette; if too many colors were present, images received low scores in general. On the same dataset, Murray (2012) used a low-level contrast model that was originally developed for saliency estimation and showed that it can also be applied to predict aesthetic preferences.

In recent years, deep learning models, in particular Convolutional Neural Networks (CNNs), have started to conquer many subareas in the field of computer vision and artificial intelligence. Although the basic idea of CNNs has already been proposed more than three decades ago (Fukushima, 1980; Lecun and Bengio, 1995), only recently, progress in computing technologies and the availability of huge datasets for training have helped to restore the interest in using CNNs for image processing (Krizhevsky et al., 2012; Simonyan and Zisserman, 2014; He et al., 2015; Huang et al., 2016). CNNs learn a hierarchy of filters, which are applied to an input image in order to extract meaningful information from the input. The training is done using backpropagation, a supervised training algorithm, in which the current output of a network is compared to a desired output. Filter parameters of the network are changed according to their contribution to the current error. When used on a large training set of images, CNNs tend to learn features that resemble Gabor-like edge detectors and color-opponent filters at lower layers of the CNNs. These features are akin to neural responses in the early mammalian visual system. On higher layers of the CNNs, features capture more abstract image content by integrating the lower-layer features (Yosinski et al., 2015). Different open-source implementations exist, which also include a variety of models that were pretrained for object or scene recognition. Their availability enables researchers to either retrain networks that already work well for recognition tasks (a process called fine-tuning), or to use features from pretrained models without any further modification.

CNNs have been applied to the task of rating image aesthetics. Lu et al. (2015) trained a two-column deep neural network simultaneously on global and local views of photographs in order to predict their aesthetic rating class (high or low). The authors motivated their architecture by the observation that the aesthetics of an image is influenced by local cues, such as sharpness, as well as global cues, which capture compositional aspects. They evaluated different cropping strategies for the local image view and report a higher accuracy in the prediction of image aesthetics than reported for previous approaches on the same dataset (Murray et al., 2012).

Dong et al. (2015) applied the AlexNet architecture presented by Krizhevsky et al. (2012), which was trained on 1.2 million images to discriminate between 1,000 different object categories. They used the features of the top convolutional layer, which are computed on the entire image, as well as on five local crops, and trained an SVM on the concatenated features. They improved upon the results by Marchesotti et al. (2011) by a margin of about 10%. Interestingly, their approach did not explicitly use features trained in the context of an aesthetic evaluation, but rather for

object recognition, so that the decision whether an image was rated as highly aesthetic or not seemed to rely more on image content than on image form.

Denzler et al. (2016) proposed to use CNNs as model of perception for research in aesthetics. They trained the AlexNet model (Krizhevsky et al., 2012) on different datasets to experimentally evaluate how well pre-learned features of different layers are suited to distinguish art from non-art images using an SVM classifier. They report the highest discriminatory power with a Network trained on the ImageNet dataset, which outperforms a network solely trained on natural scenes.

Kao et al. (2016) proposed a multi-task learning approach, in which a CNN was trained to simultaneously assign semantic and aesthetic labels. They explored different network architectures and showed that a network trained to recognize semantic labels in addition to the aesthetic class outperforms a network trained solely to recognize the aesthetic class of an image. This finding is compatible with the role of both content and form in psychological models of aesthetic experience (see below).

Nowadays, deep neural networks have largely replaced the conventional approach of designing features deliberately in order to reflect aesthetic concepts that derive from human intuition. They outperform the conventional approach easily and have a number of additional advantages: (1) Deep neural networks learn features that are important for aesthetic evaluations automatically, provided that a dataset is big enough. (2) They can combine local image properties, such as sharpness or blur, with global properties, such as composition or color harmony. (3) They can even take into account abstract features, such as image content, without the explicit design of such features by humans. (4) Last but not least, deep neural networks are able to learn image properties that humans may not even be aware of. Such properties include unspecified compositional rules that are employed intuitively by photographers and painters (Bell, 1914; Arnheim, 1954; Redies, 2007, 2015).

While deep learning models are state-of-the-art in aesthetic image evaluation, their success comes at a cost. At present, the understanding of deep features and how they work in object or aesthetic recognition lacks behind. Although there have been attempts to analyze what deep neural networks actually encode at higher layers (Yosinski et al., 2015), we are far from understanding the success of deep learning in any significant detail. For applications in aesthetic image evaluation, it may be sufficient to simply build systems that closely match human perception in deciding whether an image is considered to be beautiful. However, for researchers who want to learn more about aesthetics *per se*, the limitations of deep learning models are particularly obvious. With handcrafted features, it is easy to draw conclusion about which features contribute to the aesthetic value of an image. Deep neural networks and generic features basically represent a black-box approach that lacks this kind of interpretability. Nevertheless, if we can develop tools to understand deep representations in the future, the drawback of deep learning approaches may eventually turn out into an asset for understanding aesthetics. Such a more profound understanding would also require that deep learning be better explainable in terms of actual neural mechanisms. Although some recent studies lead in this direction (for example, see Brachmann et al., 2017), an abundance of questions remains.

2.2. Other Classifications of Images

Besides the prediction of visual preference, there has been another trend in computational aesthetics, which tends to be more focused on artworks than on photography. In this trend, images are not classified according to their aesthetic appeal, but with respect to the correct identification of the painter or the artistic style, an undertaking which is usually performed by art experts. From a methodological point of view, the identification of painter and style are related tasks that often go hand in hand. However, in the early days of computational aesthetics, the identification of the artist who created a given painting (Cezanne, Vermeer, Rembrandt, etc.,) was more popular. More recently, there seems to be a shift to the prediction of the style (Realism, Impressionism, Cubism, etc.,), as works from more and more art collections become digitized and available on the web. These open-source collections enable researchers to easily collect the huge number of images that are needed in order to train and test algorithms. Possible applications for such methods are recommender systems for online art markets or the more precise description of the stylistic singularities of particular artists.

2.2.1. Artist Identification

Using a Naive Bayes Classifier, Keren (2002) computed Discrete Cosine Transform (DCT) coefficients on an image and identified the painters of art images (Rembrandt, van Gogh, Picasso, Magritte, Dali) by using a voting scheme, where each 9×9 block of an image is assigned the style of an artist. A majority voting for an image yielded the final result and the authors reported an accuracy of 86% for choosing the correct painter. Widjaja et al. (2003) focused on nude paintings and used color of skin in order to identify the artist. They trained an SVM on color profiles of patches extracted from images of four different painters (Rubens, Michelangelo, Ingres, and Botticelli) and reported a rate of correct identifications of 85%. Li and Wang (2004) proposed a system for artist identification based on wavelets and a Multiresolution Hidden Markov Model and tested their approach on a dataset of grayscale Chinese ink images that contained works by five different Chinese artists. Besides the classification of paintings regarding their artist, they found that their modeling approach can also be used as a measure of similarity. To recognize the artist of an image, Lombardi (2005) proposed a system that used a set of low-level features for intensity, edge information, spatial frequency information, as well as a new feature that captured color. Shen (2009) combined a set of global visual features (color, textures, shape) and local visual features (Gabor wavelets) and reported an identification accuracy of 69.7% when distinguishing 25 classical Western painters in a dataset that included Caravaggio, Rubens, Vermeer, and van Gogh. For classification, they used an RBF neural network. Khan et al. (2010) automatically predicted painters (Ingres, Matisse, Monet, Picasso, Rembrandt, Rubens, Titian and van Gogh) by using a Bag-of-Visual-Words approach. They computed SIFT descriptors, as well as color name descriptors and trained an SVM on a dataset which consisted of 40 images each of the

eight artist (320 images total). They report an accuracy of 62% for the combination of color and shape features. Condorovici et al. (2013) used a dataset of 1,896 paintings by 15 different artist (including Pollock, Rembrandt, Cezanne, and Magritte), from which they extracted low-level features like an RGB color histogram and edge information by Gabor filters. The authors experimented with eight different classifiers, among which multi-class logistic regression yielded the best results. Cetinic and Grgic (2013) extracted three types of features, namely image-intensity statistics, color-based features, and texture-based features and used a multi-layer perceptron with one hidden layer; they reported a 75.3% accuracy of identifying the correct one among 20 painters.

Overall, it is difficult to compare the performance of the different methods for artist identification because a common database, on which results could be reported and compared to others, is lacking to date. Condorovici et al. (2013) addressed this problem by comparing different methods to their guessing baseline. However, this approach may give an advantage to researcher who select painters who are more diverging to begin with. For example, it may be harder to distinguish an impressionist painting by Claude Monet from one by Paul Cezanne, than to distinguish an abstract drip painting by Jackson Pollock from a surrealist painting by René Magritte.

In summary, the most popular choices for features that are used for the classifiers include a measure to capture texture or spatial frequency, edge histograms for shape detection and histograms for color analysis; all these features are low-level and do not describe image content.

More recently, classification studies in other areas of research no longer rely on one classifier, but report results for a set of different classifiers that are studied in parallel. A popular choice for this type of analysis is the Weka data mining software (Hall et al., 2009).

2.2.2. Style Prediction

To predict art styles in various sets of artworks, different approaches have been used. Gunsel et al. (2005) trained an SVM classifier in order to discriminate among five painting styles (Classicism, Impressionism, Cubism, Expressionism, and Surrealism) as well as between twelve different painters. They proposed a system that computes a 6-dimensional vector of low-level features including brightness and gradient information of an image as well as statistics of the gray-level histogram. This system allows a user to query the system for similar paintings of unknown style. For painter and art movement classification, the authors report a high accuracy with a low number false positive results. A different approach was taken by Jiang et al. (2006) who designed a way to retrieve traditional Chinese paintings and then classify them into one of the two styles, *Gongbi* (traditional Chinese realistic painting) or *Xieyi* (freehand style). For this task, they used low-level features, which captured color, texture and edges. With a classifier that combined a decision tree and SVMs, they obtained accuracies that are suitable for practical purposes.

Wallraven et al. (2009) asked participants to group images from 11 different art periods (e.g., Gothic, Renaissance, Classicism, Surrealism and Postmodern Art) and different artists into self-selected categories. The resulting categories of artworks corresponded well with the canonical art periods. The authors then computed several low-level features of the images (e. g. raw pixel values, color histograms, frequency, or a GIST descriptor; Oliva and Torralba, 2006) and tested how well the features described the clustering into different art periods. The authors found a low correlation between their set of low-level features and the grouping into art periods and concluded that humans rely more on higher-layer properties. Siddiquie et al. (2009) used multiple kernel learning in their approach and chose texture, histograms of gradient orientations (HOGs), color, and saliency as their features to discriminate between seven different styles (Abstract Expressionism, Baroque, Cubism, Graffiti, Impressionism and Rennaissance). Zujovic et al. (2009) chose five different genres (Abstract Expressionism, Cubism, Impressionism, Pop Art, and Realism). As features, they used steerable filters as well as edge information extracted by a canny edge detector. For color, they calculated HSV histograms and used their bins as features. The classification was done with several different classifiers and the authors reported a best overall accuracy of 69.1% for the AdaBoost classifier. Shamir et al. (2010) classified paintings of nine artists of different genres (Impressionism, Surrealism and Abstract Expressionism) and reached an accuracy of 91.0% in style classification by using a set of features that contained frequency statistics, edge information and color information. Čuljak et al. (2011) focused on texture and color features, stating that such features are closely related to the way humans perceive artworks. As genres, they chose Realism, Impressionism, Cubism, Fauvism, Pointillism and Naïve Art. They tested a range of classifiers and reported best results for an SVM, reaching 60.2% accuracy. Ivanova et al. (2012) used various MPEG-7 descriptors in order to distinguish different art styles. In their experiment, they noted that color features were better suited than texture features for distinguishing between art styles and artists. Condorovici et al. (2015) reported that key to a better accuracy in style discrimination is to let features be inspired by human perception. Accordingly, they used luminance and features that detected shape, texture, edges and color. A total of eight genres was selected for style classification in their study. Like other authors, they tested a set of classifiers and reached best results with an SVM, outperforming their predecessors.

While all articles mentioned above used low-level features, which capture formal aspects of paintings, results from Arora and Elgammal (2012) first indicated that semantic features are also important for style classification. The author compared different features and reported the best results for an SVM trained on classeme feature vectors (Torresani et al., 2010), which represent an image as combined classification scores for many weak classifiers that were trained on low-level descriptors.

Beginning with the work of Krizhevsky et al. (2012) and due to the renewed interest in deep neural networks, these models have also been applied to style prediction. Karayev et al. (2013) used a relatively large dataset of 100K images together with color features, GIST descriptors, saliency, meta-class features (Bergamo and Torresani, 2012) for image content, as well as DeCAF features (Donahue et al., 2014), which are activations of higher layers of CNNs that encode image content rather than

image form. They additionally trained a classifier for content features on the categories of animals, vehicles, indoor objects and people. For 25 different painting styles, they reached a mean accuracy of 47.3% with all features in combination. Other than painting style, they also reported results for photographic styles in their article. One of their main conclusions is that style is highly dependent on content. Another approach that also relied on DeCAF features can be found in Bar et al. (2014). These authors reported that a combination of DeCAF features and PiCoDes features (Bergamo et al., 2011), a binary descriptor, which incorporates several low-level descriptors, shows the best performance in style recognition.

Saleh and Elgammal (2015) used the object labels that were produced by the networks proposed in (Krizhevsky et al., 2012) as a feature to discriminate the artist, the style and the genre of roughly 80K paintings. They concluded that classemes (Torresani et al., 2010) are the best way to represent artist, genre, and style-specific properties for discrimination. Tan et al. (2016) conducted several experiments regarding painting style, genre, and artist discrimination and used the architecture proposed by Krizhevsky et al. (2012). They fine-tuned a model that was trained on the ImageNet (Deng et al., 2009) dataset for object recognition, trained a model from scratch, and also tested SVM classifiers on deep features. Interestingly, the fine-tuned model yielded the best results in all tasks and even outperformed the model that was trained from scratch.

Painter and style prediction go hand in hand. In the early days, hand-crafted features that captured the same type of image properties were equally suitable for both tasks. With more and more image data becoming available for training, style prediction can now be trained and tested on exceedingly large sets of images and collections of style categories can be expanded with ease. For painter identification, this is not necessarily the case because, for most artists, only a relatively limited number of paintings are available for training deep networks. As another complicating factor, many artists changed their style during their lifetime. For example, several abstract artists started their career with realistic paintings (for example, Wassily Kandinsky, Piet Mondrian, and Jackson Pollock). As a result, training deep neural networks for painter identification will likely remain more difficult than for style prediction.

For style prediction, the availability of huge collections of digitized artworks will open new possibilities for researchers who will use machine learning methods in the future. For example, popular and widely used datasets of paintings, such as the databases of the Google Art Project and WikiArt (formerly WikiPaintings), contains several thousands of annotated artworks.

As outlined for rating prediction (section 2.1), deep features are getting more and more popular for style prediction and increasingly replace hand-crafted features because they are capable of representing semantic information also. For example, Chiaroscuro style paintings often depict indoor scenes and people, while Impressionist paintings frequently display landscapes. Therefore, deep features do well on style prediction and prove to be more powerful than low-level features that focus on image form only. On the other hand, as with the prediction of ratings, interpretability is not as high as it has been with purposely designed features.

Although the vast area of computer-generated artistic images is beyond the scope of the present review, we would like to point out that deep models have boosted recent developments in this area that harbor a large potential for understanding aesthetics. Gatys et al. (2016) proposed an algorithm that can transfer the style of any image to another, by matching the statistics of the gram matrix of lower-layer features, as well as image content that is represented at higher layers. They demonstrated that arbitrary images can be redrawn in the style of famous paintings from Van Gogh or Picasso. More recent generative models (Generative Adversarial Networks [GANs]; Goodfellow et al., 2014) are even capable of matching the style of entire collections of artworks, as shown by Zhu et al. (2017), who used collections of paintings by Monet, Cezanne and Van Gogh to redraw landscape photographs to match the respective painter's style. While GANs are advanced methods that originate in Machine Learning, other methods like the approach by Malo and Simoncelli (2015) focus more on using physiologically plausible architectures to generate images with similar textures. This latter approach is likely to have more explanatory power because it makes use of mathematical tools that are more directly related to findings from vision science.

2.3. Other Applications

In the previous sections, we described computational methods to predict ratings and to discriminate between paintings by different artists and art styles. Most of these methods rely of the perceptual distinctness of different types of artworks. However, art has also been studied from other perspectives. In the present section, we review computational methods that can provide useful help in solving questions relevant to art history as well as art forgery detection. Some of these methods aim to discriminate rather subtle differences between artworks that may not even be apparent to the human eye.

For a review on earlier methods, see Stork (2009a). A more recent overview is given in Spratt and Elgammal (2014), who list different applications and publications of computational methods for art analysis, including semantic annotation of artworks, ordering of paintings by creation date, or the detection of similarities in paintings and artists in order to reveal mutual influences between artists.

2.3.1. Art History

Among the methods that address art historical questions, we can discern two areas of interest. First, some researchers have developed computational methods to study artistic technique. Second, the influence of a painter on the style of other artists has been studied.

Criminisi et al. (2002) developed methods for investigating the perspective and the reconstruction of the 3-dimensional space from realistic paintings. This information can help art historians to answer spatial questions like, for example, to determine the height of people or objects that are depicted in paintings. In another study, Criminisi and Stork (2004) analyzed inaccuracies in the perspective cues in a painting by Jan van Eyck and demonstrated that is it unlikely that the painter used

optical aids like mirrors during the creation of the painting "Portrait of Arnolfini and his wife." Stork and Johnson (2006) applied a technique that was originally designed for detection of tampering in photographs, in order to localize light sources in paintings. They presented such an analysis for Georges de La Tour's painting "Christ in the carpenter's studio." Based on their findings, they rebutted the claim that the light source of the depicted scene lays outside the painting, which could have been an indication of the use of optical aids as well. Papaodysseus et al. (2006) investigated the use of stencils in late Bronze Age wall paintings by applying a Hough Transform (a method for finding instances of mathematically defined shapes in images), and identified a set of stencils that were likely used during creation of the wall paintings. Kim et al. (2014) propose statistical measures to quantify the usage of individual colors, their variety in a painting, and the roughness of the brightness of a painting and report significant differences for different art periods. Berezhnoy et al. (2005) studied color and texture features in paintings by van Gogh. They confirmed that the painter increasingly made use of opponent colors later in his lifetime. Later, Berezhnoy et al. (2009) proposed a method for aiding art experts in automatically extracting the orientations of brushstrokes in a painting.

The study of a painter's influence on other artists, which can be investigated by detecting similarities between images, is a popular topic of research in computational aesthetics. Bressan et al. (2008) used SIFT features and local color statistics to compute similarities between images based on a Fisher Kernel representation of the images. Shamir and Tarakhovsky (2012) used a set of 4,027 features that represented many different aspects of visual appearance (e.g., shape, texture, color) and computed a phylogeny, which shows distinct clusters for classic artists like Vermeer or Rembrandt and for modern artists like Jackson Pollock, Marc Rothko, or Wassily Kandinsky. Wang and Takatsuka (2012) extracted color and composition features, which allowed them to classify Renaissance, Impressionist and Postimpressionist paintings. Furthermore, they applied hierarchical clustering in order to identify relationships among artists and demonstrated that they can detect influences of preceding art periods on Picasso's works. Abe et al. (2013) proposed a framework for determining artistic influences based on the semantics of images. By using classeme features to compute distances between images (Torresani et al., 2010), they succeeded in identifying novel cases where one artist influenced another, which had not been considered by art historians before. Elgammal and Saleh (2015) approached the problem of assessing creativity in terms of the originality of an artwork and represented influences and originality as a graph. Relying on classemes for subject matter and GIST features for compositional aspects, they computed a creativity score for each painting in comparison to contemporary artworks.

2.3.2. Forgery Detection

Another example where computational methods can help art historians is in the detection of forgeries, which is a problem closely related to artist identification. In artist identification, the works of an artist are identified among many others that usually possess rather different characteristics, which are often obvious even to laymen. However, when detecting forgeries, any differences may no longer be as easy to spot so that the task may be difficult even for art experts. Both approaches aim at identifying unique features of an artist, but an algorithm, which works well for artist identification, may not work as well for authentication and vice versa.

For example, Lyu et al. (2004) performed a wavelet decomposition of eight works attributed to the Renaissance painter Pieter Bruegel the Elder and five imitations of his work. From the wavelet statistics, they extracted a feature vector for subimages of each image and performed authentication by measuring distances between these high-dimensional points. They found that imitations of Bruegel's works differ significantly from authentic paintings. In another application of their technique, they solved the problem of "many hands." Here, art historians are interested in how many different painters contributed to one particular painting. Using their method, they were able to identify at least four different painters for face depictions in an image attributed to Pietro Perugino, a notion that is shared by art historians. Polatkan et al. (2009) introduced a new dataset of images that included originals and purposely copied paintings. Using the parameters of a Hidden Markov Model trained on wavelet coefficients, they succeeded in discriminating the copies from the originals. Li et al. (2012) studied the brushstrokes of paintings by Vincent van Gogh and used them for comparison with contemporaries and forgeries, as well as for dating different periods of van Gogh's work. Johnson et al. (2008) summarize different approaches by three research groups for discriminating between 82 original van Gogh paintings, 6 non-original works, and 13 paintings of questionable authorship. All approaches are based on a wavelet decomposition of the images.

The work of American painter Jackson Pollock has received particular interest from the scientific community. Taylor et al. (1999) performed a fractal analysis of the artist's drip paintings and found that the fractal dimension, computed using a box-counting approach, increased over the artist's lifetime. The authors suggested that this method could be used for authenticating or dating individual works by the artist. Taylor's approach was criticized by Jones-Smith and Mathur (2006), who showed that they could easily generate images that had the same fractal properties albeit not being similar to Pollock's paintings in their aesthetic value. Stork (2009b) later defended Taylor and colleagues and argued that, while one feature in isolation may not be sufficient for the analysis, a combination of multiple fractal measures can provide useful information. Shamir (2015) used a set of features from biological image analysis (Shamir et al., 2008) and reported an accuracy of 93.0% in discriminating between original and non-original drip paintings.

Hughes et al. (2010) applied a sparse coding scheme in order to compare authentic Bruegel paintings with works by imitators. They demonstrated that their technique can be used to discriminate between authentic and non-authentic Bruegel drawings. Olshausen and DeWeese (2010) suggested that the methods of detecting forgeries brought forward by Hughes et al. (2010) could be useful not only in learning styles of particular artists but also for using these statistics to generate novel images.

Montagner et al. (2016) proposed a system for forgery detection of paintings by the Portuguese painter Amadeo Souza-Cardoso. In their approach, they combined a brushstroke analysis using SIFT features on RGB images and an analysis of the pigments in the painting by hyperspectral imaging. Using a dataset of 12 images, among which one was not painted by the artist, they successfully determined the authenticity of the original paintings.

In summary, computational methods can provide support for art historians who study individual paintings or artists. Computational methods have aided art historians in multiple ways, for example by enabling them to detect the use of practical aids like stencils or projectors in the creation of an artwork. Furthermore, telling forgeries from originals as well as the dating of an artist's work can be improved with the help of algorithmic approaches. Other applications are the exploration of hitherto unknown influences between artists.

3. EXPERIMENTAL AESTHETICS: INVESTIGATION OF SPECIFIC IMAGE PROPERTIES

In experimental aesthetics, researchers are not primarily interested in reaching automatic decisions that mimic human aesthetic judgments. Rather, the goal is to find out on what grounds aesthetic judgement are made by human observers and what their biological basis and evolutionary purpose might be. In other words, applications are not the focus of research, but rather a better understanding of aesthetic experience (Berlyne, 1974; Cela-Conde et al., 2011; Chatterjee and Vartanian, 2014; Shimamura, 2014). Before proceeding to concrete examples, we will briefly review some key concepts in experimental aesthetic research.

3.1. Basic Concepts in Experimental Aesthetics

It is generally agreed that aesthetic experience is a highly complex phenomenon and involves at least three key domains (perception, cognition and emotion), which are realized at multiple levels of human social organization (universal, cultural and individual) (Jacobsen, 2006; Marković, 2012; Chatterjee and Vartanian, 2014; Redies, 2015).

To a large extent, perception represents bottom-up processing of visual information. Perceptual mechanisms are thought to be universal among humans and are likely to have their origin in the evolution of the human visual system. Whereas it is self-evident that any information associated with a visual stimulus must be processed by the visual system in order to be perceived, it is still a matter of debate whether there are specific mechanisms that mediate the perception of aesthetic (or beautiful) stimuli at lower or mid-levels of visual processing.

On the one hand, it has been demonstrated that visually pleasing images are associated with specific image features that can be measured by objective means. Because artworks of different styles, cultures and artists differ in their content, these common image properties reflect formal characteristics of images (*significant form*; Bell, 1914). Possibly, these stimulus properties elicit a particular state of neural activity in the visual system (*resonance*; Taylor et al., 2005; Redies et al., 2007b) or induce the activation of a specific (beauty-responsive) neural mechanism in receptive individuals (Redies, 2015). This specific activation can be thought of as the correlate of visual preference or, more specifically, of the perception of beauty in images.

On the other hand, it has been argued by some modern philosophers, art critics, psychologists and neuroscientists that any visual stimulus can elicit an aesthetic experience, as long as it is presented in an appropriate cultural context. Followers of this cognitive hypothesis often reject the notion that there are objective and universal stimulus properties that characterize aesthetic stimuli. Instead, they emphasize the role of the art-historical context of artworks, the intentions of the artists, conceptual issues, the expertise of the beholder, the status of the artwork and other culturally determined factors (Danto, 1981; Leder et al., 2004; Zeki, 2013; Gopnik, 2014). These factors are, by definition, not universal and do not persist over time, because cultural conditions change perpetually; they reflect cognitive (predominantly top-down) mechanisms in the human brain and relate more to the content and context of artworks than to their form. However, perceptual (sensory) and cognitive factors are not mutually exclusive in aesthetic appreciation; several researchers have included combinations of both types of factors in their models of aesthetic experience (for example, see Jacobsen, 2006; Locher et al., 2007; Marković, 2012; Chatterjee and Vartanian, 2014; Kozbelt and Kaufman, 2014; Shimamura, 2014; Redies, 2015).

Individual experiences also play an important role in aesthetic experience, both in terms of short-term adaptation to the beauty of visual stimuli and in long-term processes, such as familiarization and the acquisition of knowledge about art. Interestingly, interindividual differences have been found even in the preference for basic stimulus properties, such as stimulus complexity (Bies et al., 2016a; Güçlütürk et al., 2016; Lyssenko et al., 2016; Spehar et al., 2016), color (Mallon et al., 2014; Palmer et al., 2016), or the preference for the aspect ratio of rectangles (McManus et al., 2010). Last but not least, the emotions of the beholder also play an important role in aesthetic appreciation (Leder et al., 2004, 2014; Silvia, 2005, 2014).

Against this background of concepts in experimental aesthetics, it is clear the identification of objective image properties in computational aesthetics can provide an important basis for the understanding of aesthetic perception. Indeed, the notion that aesthetic stimuli are endowed by objectively measurable properties that can be universally recognized and are preferred by humans across cultures seems implicit in many studies in computational aesthetics. However, the knowledge about other factors that depend on the cultural context of individual artworks, on the intentions of the artists and on the cognitive and emotional state of the beholder should make us cautious when confronted with claims that particular image properties are universally preferred across individuals, groups of people or cultures.

A major research topic of experimental aesthetics is the investigation of the specific properties of artworks. This research allows us to gain insight into how aesthetic perception is linked

to human vision and contributes to our knowledge on how we perceive the world (Graham and Redies, 2010). In the field of experimental aesthetics, researchers have studied a wide variety of aesthetic experiences, ranging from deeply moving emotions elicited when viewing famous artworks in a prestigious museum, to aesthetic ratings of artworks in a laboratory setting, and to visual preferences for simple artificial patterns displayed on a computer screen. This wide range of aesthetic experiences brings up two issues. First, beyond statistical image properties, cultural, social and psychological factors play an important role in aesthetic experience. Undoubtedly, these factors interact with image properties that characterize artworks. Second, the role of specific image properties may depend on the type (or the intensity) of the aesthetic experience studied. For example, if an image property plays a role in aesthetic preference of simple, computer-generated patterns in a laboratory experiment, the same property may not necessarily influence the aesthetic appreciation of high-quality artworks in a museum (or the classification of photographs in a computational study). With these caveats in mind, we will describe several image properties that have been associated with aesthetic experience in the following sections. Again, we do not strive for completeness, but rather review selected examples that seem particularly instructive, with a focus on artworks and photographs.

3.2. Luminance and Color Statistics

The distribution of luminance, color and contrast belong to the low-level image properties that can affect the preference ratings of photographs. For example, Graham and Field (2008) showed that luminance statistics differ between artworks and natural scenes, as do their optical properties. By manipulating luminance statistics in a variety of natural images, including artistic photographs of landscapes, Graham et al. (2016) found that humans prefer images of low skewness (i.e., the third statistical moment) of their luminance distribution, with roughly equal proportions of light and dark in the images. Indeed, artworks tend to have lower-skew luminance histograms than photographs of real scenes across cultures and time periods (Graham and Field, 2007). The authors argue that artists use a non-linear compression to obtain low skewness in their paintings because images with this property can be more efficiently processed by the visual system.

Color is a feature that has been frequently used in classifiers in the field of computational aesthetics (see section 2.1.1). Although it is clear that color contributes much to aesthetics of visual art, there have been relatively few studies on color in experimental aesthetics. For example, by manipulating color statistics of Renaissance paintings, Pinto et al. (2006) studied lighting conditions that viewers consider optimal; they found that human observers generally prefer illumination conditions that yield increased chromatic diversity. Palmer and Schloss (2010) studied human aesthetic preferences for color, using simple visual stimuli. In their ecological valence theory, they suggest that color preferences arise from the affective responses to color-associated objects. In other words, people like colors that are associated with objects they like. In how far these results generalize to artworks remains unclear. Mallon et al. (2014) observed that participants preferred specific combinations of color measures in abstract artworks and that this aesthetic preference is subject to short-term visual adaptation.

In the field of computational aesthetics, Leykin and Cutzu (2003) compared the occurrence of color and luminance intensity edges in paintings and photographs of real scenes. Their results indicated that, in paintings, there are significantly more color-only edges than in photographs of real scenes. Moreover, color edges and intensity edges tend to coincide less frequently in paintings than in photographs of real scenes. Cutzu et al. (2005) build a classifier that combined color, edge and texture properties and distinguished artworks and photographs with 90% accuracy.

Aragón et al. (2008) studied the distribution of luminance in Vincent van Gogh's "Starry Night" and other paintings by the artist. Interestingly, the distribution of luminance fluctuations in some of these images resembled the mathematical distribution of fluid turbulence, as described by the Russian mathematician Andrei Kolmogorov. The authors speculated that the painter might have unwittingly introduced this property in order to produce a special feeling of unease and motion.

3.3. Complexity

Complexity relates the subjective impression of how many pictorial elements are contained in a visual stimulus. This property has been studied extensively, both in computational aesthetics and in psychological experiments. Complexity has been captured by a multitude of statistical measures, such as the number of visual elements in an image (Birkhoff, 1933), the fractal dimension (Mureika, 2005; Taylor et al., 2011), GIF compression (Forsythe et al., 2011), overall luminance gradient strength (Braun et al., 2013), or edge density (Redies et al., 2017).

In his seminal work on aesthetics, Berlyne (1974) suggested that images with an intermediate degree of complexity are preferred by humans over images of low or high complexity. His interpretation of the inverted u-shaped relation between beauty and complexity was that preference and interest increase steadily with visual complexity until a maximal level of affective appraisal is reached. With a further increase in complexity, appraisal decreases again because of decreasing preference. Others have argued that humans prefer an intermediate visual complexity because our ancestors lived in a savanna-type landscape of similar complexity (for a review, see Forsythe et al., 2011). The relationship between liking and stimulus complexity is subject to considerable interindividual variability, at least for artificial images (Jacobsen and Höfel, 2002). By automatically clustering the participants, Güçlütürk et al. (2016) described that, for one group of participants, liking decreased as stimuli became more complex, while another group exhibited the opposite pattern of preference (i.e., higher liking for more complex stimuli). Bies et al. (2016a) obtained similar results by investigating preference ratings for exact (mathematical) fractal patterns. They also described that their measure of complexity (fractal dimension) interacted with symmetry and recursion of their stimuli.

Rigau et al. (2008) took Birkhoff's aforementioned idea of aesthetics being a trade off between order and complexity, and proposed different global measures based on principles from

information theory and Kolmogorov complexity. The authors applied these measures to nine paintings by van Gogh, Seurat, and Mondrian.

3.4. Symmetry, Balance and The Rule of Thirds

Symmetry is a well-established property that plays a prominent role in the perception of many natural and artificial patterns. Symmetry can be perceived at a glance and can affect visual detection, attention, eye movements and physiological arousal (Locher and Nodine, 1989). Not surprisingly, several studies have demonstrated that symmetry is involved also in aesthetic perception. A particularly well-known example is the perception of attractiveness of human faces (Grammer and Thornhill, 1994). In simple geometrical (graphic) and ornamental patterns, symmetry was shown to have a high correlation with aesthetic judgements (Jacobsen and Höfel, 2002; Westphal-Fitch et al., 2013; Rampone et al., 2016; al Rifaie et al., 2017). However, the role of symmetry in photography and artworks seems less clear. The visitor to any art museum will readily realize that simple types of geometrical symmetry (reflectional, translational or rotational) are not general principles of composition in traditional visual art, although symmetry can attract attention if present in a painting (Locher and Nodine, 1989). Accordingly, studies that link symmetry to the aesthetic appreciation of artworks are infrequent (Osborne, 1986). It has therefore been suggested that the link between symmetry and attractiveness/beauty is domain-specific (Little, 2014).

The century-old concept of pictorial balance is related to symmetry, but on a more complex level. Unlike symmetry, it is considered to be an important and universal factor that contributes to the aesthetic appreciation of most types of images, including abstract visual patterns, photographs and artworks (McManus et al., 1985; Gershoni and Hochstein, 2011; Jahanian et al., 2015). According to Arnheim's Gestalt theory of visual balance (Arnheim, 1954), an image is balanced if the center of the displayed attractions is placed on any of the major axes of the image (vertical, horizontal and diagonal). There are different ways to measure balance. For example, in their study on Arnheim's theory, McManus et al. (2011a) used a physicalist approach and measured the center-of-mass of the luminance values in images. They considered an image more balanced if the center-of-mass was closer to the geometrical center of an image. Overall, the authors did not find evidence to support Arnheim's theory when they compared art photographs to photographs that were randomly taken, or when they studied simple geometrical figures. Jahanian et al. (2015) took another approach and modeled pictorial balance in terms of the visual weight of several low-level visual features that are used to calculate visual saliency. In a large set of 120,000 images that were rated highly, the saliency-based image hotspots aligned with Arnheim's axes, thus confirming his theory. A similar difference was obtained in a study on photographic cropping. The details of photographs that were preferred during cropping showed a more balanced saliency distribution than the details that were avoided during cropping (Abeln et al., 2016); no such difference was observed for luminance-based balance McManus et al. (2011b). Some of the computer algorithms that predict ratings of photographs and artworks (see section 2.1.1) incorporate measures of pictorial balance in their calculations (for example, see Ke et al., 2006; Li and Chen, 2009).

The rule of thirds, which is a principle of composition avidly followed in photography, seems to contradict the notion that the major axis of an image play a significant role in balance; it stipulates that salient compositional elements are to be placed close to one of the third lines of the image in order for images to be aesthetically pleasing. The rule of thirds has been used in many computational methods to predict ratings of photographs and artworks (for example, see Datta et al., 2006; Luo and Tang, 2008; Li and Chen, 2009). However, experimental studies did not confirm the significance of this rule in high-quality photographs (Amirshahi et al., 2014a) or "selfie" photographs (Bruno et al., 2014).

3.5. Fourier Spectral Properties

Graham and Field (2007) and Redies et al. (2007b) compared the Fourier spectral properties of natural scenes and images of Western artworks. They found that both types of stimuli share a scale-invariant amplitude (or power) frequency spectrum and both have a similar slope in log-log plots. Similar results were obtained for artworks of East Asian provenance (Graham and Field, 2008) and for other visual stimuli that were created to please the human eye, such as cartoons, comics and mangas (Koch et al., 2010). In contrast, several types of non-art images, such as photographs of simple objects and plants, do not possess this property (Redies et al., 2007b). Notably, photographs of faces portraits have steeper slopes of the log-log plots than human portraits drawn by artists (Redies et al., 2007a). Mather (2014) compared the spectral slopes of 31 artworks with those of closely matching photographs. He found that artists compress the spectral slopes of their works to a relatively narrow range compared to the slopes of the photographs and proposed that the artist's visual system plays a central role in adjusting the spectral slope of artworks. Humans observers tend to prefer artificial, random-phase patterns with Fourier properties similar to natural scenes (Menzel et al., 2015), but exhibit significant interindividual differences in this preference (Spehar et al., 2016). Moreover, the visual preference for these synthetic noise images correlated well with the discrimination sensitivity of the observers for different amplitude spectra of the images (Spehar et al., 2016).

Interestingly, the amplitude spectrum of many uncomfortable visual stimuli contains an excessive energy at medium spatial frequencies and thereby deviates from the linear spectral properties of natural scenes and images of artworks that are perceived as pleasant (Fernandez and Wilkins, 2008; O'Hare and Hibbard, 2011). The Fourier spectral slope of images correlates with measures of image complexity (Table S1 in Redies et al., 2017), in particular with the fractal dimension (Bies et al., 2016b). A shallower slope indicates more power in the high-frequency part of the spectrum; consequently, the images show more fine detail and thus higher complexity.

Schweinhart and Essock (2013) analyzed the Fourier spectral properties in landscape paintings that were produced by a group of local artists, and compared them to photographs of the scenes, which the artists had painted. They asked whether the well-known oblique effect can be observed in paintings. The oblique effect refers to the fact that, in our natural environment, cardinal (horizontal and vertical) edge orientations are more prominent than oblique orientations. In the Fourier domain, this difference translates into stronger amplitudes for cardinal vs. oblique orientations. In the natural environment, this effect is observed only for the lowest spatial frequencies but not for high spatial frequencies. However, the artists implemented the oblique effect also at high spatial frequencies, thus overregulating this image property in their works.

3.6. Fractals and Self-similarity

The work of the abstract expressionist artist Jackson Pollock (1912–1956) has received particular interest from the scientific community. Taylor performed a fractal analysis of the artist's drip paintings using a box-counting approach and found that Pollock's paintings are not chaotic but possess a fractal structure (Taylor, 2002). This surprising finding prompted a series of investigations of human responses to fractals, which are not only prevalent in nature but can also be found in geometric and mathematical patterns produced by humans. The studies included behavioral investigations, studies of physiological responses, eye tracking and brain imaging studies (Taylor et al., 2011; Taylor and Spehar, 2016). Converging evidence from these studies indicate that both natural and artificial fractals of mid-range complexity (as measured by the fractal dimension) elicit favorable physiological responses and are thus preferred by human observers (see also section 3.3). Fractals have even been shown to reduce stress levels in the observers (Taylor, 2006) and it has been suggested that the beneficial effect of fractal patterns can enhance architecture and our urban environment (Joye, 2007). However, as already observed by Aks and Sprott in their seminal study on chaotic visual patterns (Aks and Sprott, 1996), there are large interindividual differences in human responses to fractals and their complexity (see section 3.3). Interestingly, Pollock created fractal structure in his artworks long before fractal geometry was described and studied in detail in the 1970ies (Mandelbrot and Pignoni, 1983); he must have followed this principle intuitively and without explicit cognitive control. As noted by Alvarez-Ramirez et al. (2008), the finding that Pollock's drip paintings possess fractal structure is closely related to its scale-invariant spectral properties (see section 3.5).

The fractal-like structure of artworks was studied also by Amirshahi et al. (2012) who derived a measure for self-similarity in images, based on a Pyramid Histogram of Oriented Gradients (PHOG) representation of images (Bosch et al., 2007). In this approach, images are self-similar if the Histograms of Oriented Gradients (HOGs) of parts of an image resemble the HOG of the entire image. Redies et al. (2012) applied this measure to different image categories, ranging from natural scenes to man-made stimuli and artworks, including a large and diverse sets of traditional paintings of Western provenance (Amirshahi et al., 2014b). For artworks and most natural patterns, Redies and colleagues reported an intermediate to high self-similarity, whereas other patterns, such as images of simple objects, faces of buildings, were less self-similar.

Both lines of evidence suggest that traditional artworks share specific stimulus properties with our natural environment. Our visual system has adapted to these properties in evolution so that it can process them with a sparse (efficient) code in order to save computational and metabolic resources (Simoncelli and Olshausen, 2001). It has therefore been suggested that artworks are created so that they can be processed efficiently/sparsely by the human visual system (Redies, 2007; Renoult et al., 2016). The concept of sparse coding is familiar also to researchers in computer vision (Mairal et al., 2014). Akin to the efficient coding hypothesis is the idea that artworks can be processed fluently and therefore evoke a pleasant feeling in human observers (Reber et al., 2004). The fluency concept has its origin in the field of psychology; the underlying neuronal mechanism and possible coding strategies in the human brain remain unspecified to date.

3.7. Regularities in the Orientation of Luminance Gradients, Edges, and Lines

In a study on large subsets of traditional Western artworks, histograms of oriented gradients (HOGs; see section 3.6) were found to possess a surprising regularity (Redies et al., 2012; Braun et al., 2013): Artworks possess a relatively uniform spectrum of luminance gradient (edge) orientations. This result implies that all edge orientations in the artworks tend to be similarly prominent. In other words, anisotropy of edge orientations is low in artworks. Other types of images with low anisotropy can be found in nature (for example, large vista scenes and images of plants, lichen growth patterns, branches and clouds; Redies et al., 2012). Anisotropy is larger in images of simple objects, including faces, and other man-made patterns, such as advertisements, building facades and urban scenes, due to the relative prominence of single or a few orientations. For example, horizontal and vertical orientations predominate in images of building facades.

The finding of low anisotropy of edge orientations in artworks was recently confirmed and extended by Redies et al. (2017), who studied edge orientations in different categories of images, including traditional artworks of different cultural provenance (Western, Islamic and East Asian). They showed that the art images possess a more uniform histogram of edge orientations across cultures than many non-art types of images, in particular, photographs of man-made objects and scenes. This result mirrors the low anisotropy found in artworks (see above). In addition, by pairwise comparison of edge orientations across each image, Redies and colleagues found that edge orientations are independent of each other across art images, except for edge pairs at short distances, which tend to be collinear. In other words, the edge orientation at one position of an image does not allow predicting the orientations of distant edges at other positions in the same image. Similar statistical regularities of edge orientations are observed in some natural images, such as lichen growth patterns. This property is independent of cultural provenance, artistic genre or technique, or image content of the artworks studied. The authors speculated that this regularity

might relate to the notion of "good composition" (Arnheim, 1954) or "visual rightness" (Locher et al., 1999), which has been advanced for traditional artworks.

Another regularity with respect to the perception of contours is that smoothly curved lines and objects are generally preferred over sharply angular ones (Gómez-Puerto et al., 2015). Interestingly, humans share this preferences not only across cultures but also with great apes (Munar et al., 2015). As a possible explanation, Bar and Neta (2006) proposed that sharp transitions in contour convey a sense of threat in the observer and are therefore disliked. However, Bertamini et al. (2016) questioned this notion and provided experimental evidence that humans prefer curvature due to its intrinsic characteristics and not because they reject the threat potential of angular contours.

4. CONCLUSION AND OUTLOOK

In recent years, computer vision has successfully contributed computational methods to the evaluation of photographs and digitally reproduced artworks. In the present work, we discussed recent progress in this field, which has become known as computational aesthetics. Specifically, we reviewed methods that were developed to predict the aesthetic rating of photographs and artworks by computational approaches. For artworks, we provided an overview on applications of computational algorithms to artist identification, style prediction, art historical questions, and forgery detection.

In general, researchers in the computer vision community tend to measure success by comparing different methods regarding their accuracy of classification or prediction. When using the same database, systems can easily be compared and finding the best working approach is straightforward. However, with recent advances in technology, algorithmic and larger datasets, the best-performing classifiers have become black boxes and their discrimination boundaries are no longer obvious. From an application standpoint of view, this is not necessarily a limitation. For example, such systems can be readily deployed in image processing pipelines to identify images of high vs. low aesthetic value. While early methods where restricted to the formal aspects of a scene, more advanced methods, like Deep Neural Networks, can take into account the content of images as well. It was shown that the inclusion of content results in major improvements, because different stylistic elements come along with different content matter. For example, bright colors are usually more pronounced in pleasant images that depict fresh fruits than in gloomy images of street scenes at night. Such combinatorial information can improve classification results.

Lately, computational methods have gained increasing popularity also in the field of experimental aesthetics, an area of research that has a long tradition as a branch of psychology and, more recently, of neuroscience. In experimental aesthetics, the focus is not on improving algorithms for rating prediction systems or identifying artists or artistic styles, but rather on gaining a better understanding of what specific stimulus properties induce human observers to reach judgements on beauty and to have an aesthetic experience. For example, as discussed in section 3, converging evidence suggests that some global image properties that also characterize natural scenes can be found in large subsets of traditional artworks.

With recent developments in Deep Learning, it has become harder to share knowledge between computational aesthetics and experimental aesthetics. In the early days, insights from the active field of experimental aesthetics provided a wealth of knowledge, also for computational aesthetics. This knowledge resulted in the development of computational algorithms based on handcrafted features, which were known (or suspected) to contribute to the aesthetic appeal of an image. During this time, empirical aesthetics also profited greatly from the computational methods because, for the first time, very large datasets of images could be analyzed, rather than the small number of images that are usually tested in psychological experiments with human observers. However, with Deep Learning, it has became harder for empirical aesthetics to catch up with the computational approaches. Deep Learning models basically represent black boxes, which prevent insight into what features they learn and how they use them to evaluate the aesthetic quality of images, which is the main motivation for empirical aesthetics. In future work, it will therefore be essential to gain a better understanding and interpretability of the decision boundaries that the computational models draw, in order to identify concrete properties of human aesthetic preference. Moreover, recent generative models from computer vision (Gatys et al., 2016) are capable of producing synthetic images that match the style of famous painters, and are no longer discriminative only. This generative approach may provide researchers with well-controlled stimuli for testing human observers in experimental aesthetics.

In conclusion, much can be learned if the two areas of aesthetic research can be recombined, taking advantage of the methodological advances in computational aesthetics and the identification of perceptual mechanisms in experimental aesthetics. As an example, we recently investigated the variability of CNN feature responses to traditional artworks and non-art images and found that the two categories of images can be separated by a classifier that is based on only two variance values (Brachmann et al., 2017). However, results for some styles of (post-)modern and contemporary art clearly deviated from traditional art. The investigation of differences between art styles may therefore be of particular interest in the future, not only in computational aesthetics but also in experimental aesthetics. Moreover, in view of the interindividual differences in aesthetic preferences (see section 3.1), cultural diversity will be an important issue in future research.

AUTHOR CONTRIBUTIONS

AB and CR conceived this review, carried out the literature search and wrote the manuscript.

FUNDING

This work was supported by funds from the Institute of Anatomy, Jena University Hospital.

REFERENCES

Abe, K., Saleh, B., and Elgammal, A. (2013). "An early framework for determining artistic influence," in *International Conference on Image Analysis and Processing* (Berlin: Springer), 198–207.

Abeln, J., Fresz, L., Amirshahi, S. A., McManus, I. C., Koch, M., Kreysa, H., et al. (2016). Preference for well-balanced saliency in details cropped from photographs. *Front. Hum. Neurosci.* 9:704. doi: 10.3389/fnhum.2015.00704

Aks, D. J., and Sprott, J. C. (1996). Quantifying aesthetic preference for chaotic patterns. *Emp. Stud. Arts* 14, 1–16.

al Rifaie, M. M., Ursyn, A., Zimmer, R., and Javid, M. A. J. (2017). "On symmetry, aesthetics and quantifying symmetrical complexity," in *International Conference on Evolutionary and Biologically Inspired Music and Art* (Cham: Springer), 17–32.

Alvarez-Ramirez, J., Ibarra-Valdez, C., Rodriguez, E., and Dagdug, L. (2008). 1/f-Noise structures in Pollocks's drip paintings. *Phys. A Stat. Mechan. Applic.* 387, 281–295. doi: 10.1016/j.physa.2007.08.047

Amirshahi, S. A., Hayn-Leichsenring, G. U., Denzler, J., and Redies, C. (2014a). Evaluating the rule of thirds in photographs and paintings. *Art Percept.* 2, 163–182. doi: 10.1163/22134913-00002024

Amirshahi, S. A., Hayn-Leichsenring, G. U., Denzler, J., and Redies, C. (2014b). "Jenaesthetics subjective dataset: analyzing paintings by subjective scores," in *Workshop at the European Conference on Computer Vision* (Cham: Springer), 3–19.

Amirshahi, S. A., Koch, M., Denzler, J., and Redies, C. (2012). "PHOG analysis of self-similarity in esthetic images," in *Proceedings of SPIE (Human Vision and Electronic Imaging XVII)* (San Francisco, CA), 8291:82911J.

Aragón, J. L., Naumis, G. G., Bai, M., Torres, M., and Maini, P. K. (2008). Turbulent luminance in impassioned van Gogh paintings. *J. Math. Imag. Vis.* 30, 275–283. doi: 10.1007/s10851-007-0055-0

Arnheim, R. (1954). *Art and Visual Perception: A Psychology of the Creative Eye*. Berkeley, CA: University of California Press.

Arora, R. S., and Elgammal, A. (2012). "Towards automated classification of fine-art painting style: a comparative study," in *2012 21st International Conference on the Pattern Recognition (ICPR)* (Tsukuba: IEEE), 3541–3544.

Bar, M., and Neta, M. (2006). Humans prefer curved visual objects. *Psychol. Sci.* 17, 645–648. doi: 10.1111/j.1467-9280.2006.01759.x

Bar, Y., Levy, N., and Wolf, L. (2014). "Classification of artistic styles using binarized features derived from a deep neural network," in *Workshop at the European Conference on Computer Vision* (Cham: Springer), 71–84.

Bell, C. (1914). *Art*. London: Chatto & Windus.

Berezhnoy, I. E., Postma, E. O., and van den Herik, H. J. (2009). Automatic extraction of brushstroke orientation from paintings. *Mach. Vis. Applic.* 20, 1–9. doi: 10.1007/s00138-007-0098-7

Berezhnoy, I. E., Postma, E. O., and van den Herik, J. (2005). "Computerized visual analysis of paintings," in *International Conference on Association for History and Computing* (Amsterdam), 28–32.

Bergamo, A., and Torresani, L. (2012). "Meta-class features for large-scale object categorization on a budget," in *2012 IEEE Conference on the Computer Vision and Pattern Recognition (CVPR)* (Providence, RI: IEEE), 3085–3092.

Bergamo, A., Torresani, L., and Fitzgibbon, A. W. (2011). "Picodes: learning a compact code for novel-category recognition," in *Advances in Neural Information Processing Systems* (Granada), 2088–2096.

Berlyne, D. E. (1974). *Studies in the New Experimental Aesthetics: Steps Toward an Objective Psychology of Aesthetic Appreciation*. Oxford: Hemisphere.

Bertamini, M., Palumbo, L., Gheorghes, T. N., and Galatsidas, M. (2016). Do observers like curvature or do they dislike angularity? *Br. J. Psychol.* 107, 154–178. doi: 10.1111/bjop.12132

Bies, A. J., Blanc-Goldhammer, D. R., Boydston, C. R., Taylor, R. P., and Sereno, M. E. (2016a). Aesthetic responses to exact fractals driven by physical complexity. *Front. Hum. Neurosci.* 10:210. doi: 10.3389/fnhum.2016.00210

Bies, A. J., Boydston, C. R., Taylor, R. P., and Sereno, M. E. (2016b). Relationship between fractal dimension and spectral scaling decay rate in computer-generated fractals. *Symmetry* 8:66. doi: 10.3390/sym8070066

Birkhoff, G. D. (1933). *Aesthetic Measure*. Cambridge: Harvard University Press.

Bosch, A., Zisserman, A., and Munoz, X. (2007). "Representing shape with a spatial pyramid kernel," in *Proceedings of the 6th ACM International Conference on Image and Video Retrieval* (New York, NY: ACM), 401–408.

Brachmann, A., Barth, E., and Redies, C. (2017). Using CNN features to better understand what makes visual artworks special. *Front. Psychol.* 8:830. doi: 10.3389/fpsyg.2017.00830

Braun, J., Amirshahi, S. A., Denzler, J., and Redies, C. (2013). Statistical image properties of print advertisements, visual artworks and images of architecture. *Front. Psychol.* 4:808. doi: 10.3389/fpsyg.2013.00808

Bressan, M., Cifarelli, C., and Perronnin, F. (2008). "An analysis of the relationship between painters based on their work," in *2008 15th IEEE International Conference on Image Processing* (San Diego, CA: IEEE), 113–116.

Bruno, N., Gabriele, V., Tasso, T., and Bertamini, M. (2014). "Selfies" reveal systematic deviations from known principles of photographic composition. *Art Percept.* 2, 45–58. doi: 10.1163/22134913-00002027

Canclini, A., Cesana, M., Redondi, A., Tagliasacchi, M., Ascenso, J., and Cilla, R. (2013). "Evaluation of low-complexity visual feature detectors and descriptors," in *2013 18th International Conference on Digital Signal Processing (DSP)* (Fira: IEEE), 1–7.

Cela-Conde, C. J., Agnati, L., Huston, J. P., Mora, F., and Nadal, M. (2011). The neural foundations of aesthetic appreciation. *Progr. Neurobiol.* 94, 39–48. doi: 10.1016/j.pneurobio.2011.03.003

Cetinic, E., and Grgic, S. (2013). "Automated painter recognition based on image feature extraction," in *ELMAR, 2013 55th International Symposium* (Zadar: IEEE), 19–22.

Chatterjee, A., and Vartanian, O. (2014). Neuroaesthetics. *Trends Cogn. Sci.* 18, 370–375. doi: 10.1016/j.tics.2014.03.003

Condorovici, R., Florea, C., and Vertan, C. (2013). "Author identification for digitized paintings collections," in *2013 International Symposium on Signals, Circuits and Systems (ISSCS)* (Iasi: IEEE), 1–4.

Condorovici, R. G., Florea, C., and Vertan, C. (2015). Automatically classifying paintings with perceptual inspired descriptors. *J. Vis. Commun. Image Represent.* 26, 222–230. doi: 10.1016/j.jvcir.2014.11.016

Criminisi, A., Kemp, M., and Zisserman, A. (2002). *Bringing Pictorial Space to Life: Computer Techniques for the Analysis of Paintings*. Technical report, Microsoft Cooperation.

Criminisi, A., and Stork, D. G. (2004). "Did the great masters use optical projections while painting? Perspective comparison of paintings and photographs of Renaissance chandeliers," in *Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004, Vol. 4* (Cambridge, UK: IEEE), 645–648.

Čuljak, M., Mikuš, B., Jež, K., and Hadjić, S. (2011). "Classification of art paintings by genre," in *MIPRO, 2011 Proceedings of the 34th International Convention* (Opatija: IEEE), 1634–1639.

Cupchik, G. C. (1986). A decade after Berlyne: new directions in experimental aesthetics. *Poetics* 15, 345–369.

Cutzu, F., Hammoud, R., and Leykin, A. (2005). Distinguishing paintings from photographs. *Comput. Vis. Image Underst.* 100, 249–273. doi: 10.1016/j.cviu.2004.12.002

Danto, A. C. (1981). *The Transfiguration of the Commonplace: A Philosophy of Art*. Cambridge, MA: Harvard University Press.

Datta, R., Joshi, D., Li, J., and Wang, J. Z. (2006). "Studying aesthetics in photographic images using a computational approach," in *European Conference on Computer Vision* (Springer), 288–301.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. (2009). "Imagenet: a large-scale hierarchical image database," in *CVPR 2009. IEEE Conference on the Computer Vision and Pattern Recognition, 2009* (Miami, FL: IEEE), 248–255.

Denzler, J., Rodner, E., and Simon, M. (2016). "Convolutional neural networks as a computational model for the underlying processes of aesthetics perception," in *European Conference on Computer Vision* (Cham: Springer), 871–887.

Dhar, S., Ordonez, V., and Berg, T. L. (2011). "High level describable attributes for predicting aesthetics and interestingness," in *2011 IEEE Conference on the Computer Vision and Pattern Recognition (CVPR)* (Colorado Springs, CO: IEEE), 1657–1664.

Donahue, J., Jia, Y., Vinyals, O., Hoffman, J., Zhang, N., Tzeng, E., et al. (2014). "Decaf: a deep convolutional activation feature for generic visual recognition," in *International Conference on Machine Learning* (Beijing), 647–655.

Dong, Z., Shen, X., Li, H., and Tian, X. (2015). "Photo quality assessment with DCNN that understands image well," in *International Conference on Multimedia Modeling* (Springer), 524–535.

Elgammal, A., and Saleh, B. (2015). Quantifying creativity in art networks. *arXiv preprint arXiv:1506.00711*.

Fechner, G. T. (1876). *Vorschule der Aesthetik*, *Vol. 1*. Leipzig: Breitkopf & Härtel.

Fernandez, D., and Wilkins, A. J. (2008). Uncomfortable images in art and nature. *Perception* 37, 1098–1113. doi: 10.1068/p5814

Forsythe, A., Nadal, M., Sheehy, N., Cela-Conde, C. J., and Sawey, M. (2011). Predicting beauty: fractal dimension and visual complexity in art. *Br. J. Psychol.* 102, 49–70. doi: 10.1348/000712610X498958

Fukushima, K. (1980). Neocognitron: a self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biol. Cybern.* 36, 193–202.

Galanter, P. (2012). "Computational aesthetic evaluation: past and future," in *Computers and Creativity* (Berlin; Heidelberg: Springer), 255–293.

Gatys, L. A., Ecker, A. S., and Bethge, M. (2016). "Image style transfer using convolutional neural networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Las Vegas, NV).

Gershoni, S., and Hochstein, S. (2011). Measuring pictorial balance perception at first glance using Japanese calligraphy. *i-Perception* 2, 508–527. doi: 10.1068/i0472aap

Gómez-Puerto, G., Munar, E., and Nadal, M. (2015). Preference for curvature: a historical and conceptual framework. *Front. Hum. Neurosci.* 9:712. doi: 10.3389/fnhum.2015.00712

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., et al. (2014). "Generative adversarial nets," in *Advances in Neural Information Processing Systems* (Montréal, QC), 2672–2680.

Gopnik, B. (2014). "Aesthetic science and artistic knowledge," in *Aesthetic Science: Connecting Minds, Brains, and Experience*, eds A. Shimamura, and S. Palmer (Oxford: Oxford University Press), 129–159.

Graham, D., Schwarz, B., Chatterjee, A., and Leder, H. (2016). Preference for luminance histogram regularities in natural scenes. *Vis. Res.* 120, 11–21. doi: 10.1016/j.visres.2015.03.018

Graham, D. J., and Field, D. J. (2007). Statistical regularities of art images and natural scenes: spectra, sparseness and nonlinearities. *Spat. Vis.* 21, 149–164. doi: 10.1163/156856807782753877

Graham, D. J., and Field, D. J. (2008). Variations in intensity statistics for representational and abstract art, and for art from the Eastern and Western hemispheres. *Perception* 37, 1341–1352. doi: 10.1068/p5971

Graham, D. J., and Redies, C. (2010). Statistical regularities in art: Relations with visual coding and perception. *Vis. Res.* 50, 1503–1509. doi: 10.1016/j.visres.2010.05.002

Grammer, K., and Thornhill, R. (1994). Human (*Homo sapiens*) facial attractiveness and sexual selection: the role of symmetry and averageness. *J. Comparat. Psychol.* 108, 233–242.

Green, C. D. (1995). All that glitters: a review of psychological research on the aesthetics of the golden section. *Perception* 24, 937–968.

Greenfield, G. (2005). "On the origins of the term computational aesthetics," in *Proceedings of the First Eurographics Conference on Computational Aesthetics in Graphics, Visualization and Imaging* (Girona: Eurographics Association), 9–12.

Güçlütürk, Y., Jacobs, R. H., and Lier, R. v. (2016). Liking versus complexity: decomposing the inverted u-curve. *Front. Hum. Neurosci.* 10:112. doi: 10.3389/fnhum.2016.00112

Gunsel, B., Sariel, S., and Icoglu, O. (2005). "Content-based access to art paintings," in *IEEE International Conference on Image Processing 2005, Vol. 2* (Genova: IEEE).

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., and Witten, I. H. (2009). The weka data mining software: an update. *ACM SIGKDD Expl. Newsl.* 11, 10–18. doi: 10.1145/1656274.1656278

He, K., Zhang, X., Ren, S., and Sun, J. (2015). Deep residual learning for image recognition. *arXiv preprint arXiv:1512.03385*.

Hoenig, F. (2005). "Defining computational aesthetics," in *Proceedings of the First Eurographics Conference on Computational Aesthetics in Graphics, Visualization and Imaging* (Girona: Eurographics Association), 13–18.

Huang, G., Liu, Z., Weinberger, K. Q., and van der Maaten, L. (2016). Densely connected convolutional networks. *arXiv preprint arXiv:1608.06993*.

Hughes, J. M., Graham, D. J., and Rockmore, D. N. (2010). Quantification of artistic style through sparse coding analysis in the drawings of Pieter Bruegel the Elder. *Proc. Natl. Acad. Sci. U.S.A.* 107, 1279–1283. doi: 10.1073/pnas.0910530107

Ivanova, K., Stanchev, P., Velikova, E., Vanhoof, K., Depaire, B., Kannan, R., et al. (2012). "Features for art painting classification based on vector quantization of mpeg-7 descriptors," in *Data Engineering and Management* (Springer), 146–153.

Jacobsen, T. (2006). Bridging the arts and sciences: a framework for the psychology of aesthetics. *Leonardo* 39, 155–162. doi: 10.1162/leon.2006.39.2.155

Jacobsen, T., and Höfel, L. (2002). Aesthetic judgments of novel graphic patterns: analyses of individual judgments. *Percept. Motor Skills* 95, 755–766. doi: 10.2466/pms.2002.95.3.755

Jahanian, A., Vishwanathan, S., and Allebach, J. P. (2015). "Learning visual balance from large-scale datasets of aesthetically highly rated images," in *SPIE/IS&T Electronic Imaging*, *Vol. 9394* (San Francisco, CA: International Society for Optics and Photonics), 93940Y.

Jiang, S., Huang, Q., Ye, Q., and Gao, W. (2006). An effective method to detect and categorize digitized traditional chinese paintings. *Patt. Recogn. Lett.* 27, 734–746. doi: 10.1016/j.patrec.2005.10.017

Johnson, C. R., Hendriks, E., Berezhnoy, I. J., Brevdo, E., Hughes, S. M., Daubechies, I., et al. (2008). Image processing for artist identification. *IEEE Signal Proces. Magazine* 25, 37–48. doi: 10.1109/MSP.2008.923513

Jones-Smith, K., and Mathur, H. (2006). Fractal analysis: revisiting Pollock's drip paintings. *Nature* 444, E9–E10. doi: 10.1038/nature05398

Joye, Y. (2007). Fractal architecture could be good for you. *Nexus Netw. J.* 9, 311–320. doi: 10.1007/s00004-007-0045-y

Kao, Y., He, R., and Huang, K. (2016). Deep aesthetic quality assessment with semantic information. *arXiv preprint arXiv:1604.04970*.

Karayev, S., Trentacoste, M., Han, H., Agarwala, A., Darrell, T., Hertzmann, A., et al. (2013). Recognizing image style. *arXiv preprint arXiv:1311.3715*.

Ke, Y., Tang, X., and Jing, F. (2006). "The design of high-level features for photo quality assessment," in *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, *Vol. 1* (New York, NY: IEEE), 419–426.

Keren, D. (2002). "Painter identification using local features and naive bayes," in *Proceedings of the 16th International Conference on the Pattern Recognition, 2002, Vol. 2* (Quebec, QC: IEEE), 474–477.

Khan, F. S., van de Weijer, J., and Vanrell, M. (2010). "Who painted this painting," in *2010 CREATE Conference* (Gjøvik), 329–333.

Kim, D., Son, S.-W., and Jeong, H. (2014). Large-scale quantitative analysis of painting arts. *Sci. Reports* 4:7370.

Koch, M., Denzler, J., and Redies, C. (2010). 1/f2 characteristics and isotropy in the fourier power spectra of visual art, cartoons, comics, mangas, and different categories of photographs. *PLoS ONE* 5:e12268.

Kozbelt, A., and Kaufman, J. (2014). "Aesthetics assessment," in *The Cambridge Handbook of the Psychology of Aesthetics and the Arts* (Cambridge, UK: Cambridge University Press), 86–112.

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems* (Lake Tahoe, NV), 1097–1105.

Lecun, Y., and Bengio, Y. (1995). "Convolutional networks for images, speech, and time-series," in *The Handbook of Brain Theory and Neural Networks*, ed M. Arbib (Cambridge, MA: MIT Press).

Leder, H., Belke, B., Oeberst, A., and Augustin, D. (2004). A model of aesthetic appreciation and aesthetic judgments. *Br. J. Psychol.* 95, 489–508. doi: 10.1348/0007126042369811

Leder, H., Gerger, G., Brieber, D., and Schwarz, N. (2014). What makes an art expert? Emotion and evaluation in art appreciation. *Cogn. Emot.* 28, 1137–1147. doi: 10.1080/02699931.2013.870132

Leykin, A., and Cutzu, F. (2003). "Differences of edge properties in photographs and paintings," in *2003 International Conference on the Image Processing, 2003. ICIP 2003, Vol. 3* (Barcelona: IEEE).

Li, C., and Chen, T. (2009). Aesthetic visual quality assessment of paintings. *IEEE J. Select. Top. Signal Proces.* 3, 236–252. doi: 10.1109/JSTSP.2009.2015077

Li, J., and Wang, J. Z. (2004). Studying digital imagery of ancient paintings by mixtures of stochastic models. *IEEE Trans. Image Proces.* 13, 340–353. doi: 10.1109/TIP.2003.821349

Li, J., Yao, L., Hendriks, E., and Wang, J. Z. (2012). Rhythmic brushstrokes distinguish van Gogh from his contemporaries: findings via automated

brushstroke extraction. *IEEE Trans. Patt. Anal. Mach. Intell.* 34, 1159–1176. doi: 10.1109/TPAMI.2011.203

Little, A. C. (2014). Domain specificity in human symmetry preferences: symmetry is most pleasant when looking at human faces. *Symmetry* 6, 222–233. doi: 10.3390/sym6020222

Locher, P., Krupinski, E. A., Mello-Thoms, C., and Nodine, C. F. (2007). Visual interest in pictorial art during an aesthetic experience. *Spatial Vision* 21, 55–77. doi: 10.1163/156856807782753868

Locher, P., and Nodine, C. (1989). The perceptual value of symmetry. *Comput. Math. Applic.* 17, 475–484.

Locher, P. J., Stappers, P. J., and Overbeeke, K. (1999). An empirical evaluation of the visual rightness theory of pictorial composition. *Acta Psychol.* 103, 261–280.

Lombardi, T. E. (2005). *The Classification of Style in Fine-Art Painting*. Ph.D. thesis, Pace University, New York, NY.

Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.* 60, 91–110. doi: 10.1023/B:VISI.0000029664.99615.94

Lu, X., Lin, Z., Jin, H., Yang, J., and Wang, J. Z. (2015). Rating image aesthetics using deep learning. *IEEE Trans. Multimedia* 17, 2021–2034. doi: 10.1109/TMM.2015.2477040

Luo, Y., and Tang, X. (2008). "Photo and video quality evaluation: focusing on the subject," in *European Conference on Computer Vision* (Berlin; Heidelberg: Springer), 386–399.

Lyssenko, N., Redies, C., and Hayn-Leichsenring, G. U. (2016). Evaluating abstract art: Relation between term usage, subjective ratings, image properties and personality traits. *Front. Psychol.* 7:973. doi: 10.3389/fpsyg.2016.00973

Lyu, S., Rockmore, D., and Farid, H. (2004). A digital technique for art authentication. *Proc. Natl. Acad. Sci. U.S.A.* 101, 17006–17010. doi: 10.1073/pnas.0406398101

Mairal, J., Bach, F., and Ponce, J. (2014). Sparse modeling for image and vision processing. *Found. Trends Comput. Graph. Vis.* 8, 85–283. doi: 10.1561/0600000058

Mallon, B., Redies, C., and Hayn-Leichsenring, G. U. (2014). Beauty in abstract paintings: perceptual contrast and statistical properties. *Front. Hum. Neurosci.* 8:161. doi: 10.3389/fnhum.2014.00161

Malo, J., and Simoncelli, E. P. (2015). "Geometrical and statistical properties of vision models obtained via maximum differentiation," in *Human Vision and Electronic Imaging* (San Francisco, CA), 93940L.

Mandelbrot, B. B., and Pignoni, R. (1983). *The Fractal Geometry of Nature*. New York, NY: W.H. Freeman.

Marchesotti, L., Perronnin, F., Larlus, D., and Csurka, G. (2011). "Assessing the aesthetic quality of photographs using generic image descriptors," in *2011 International Conference on Computer Vision* (Barcelona: IEEE), 1784–1791.

Marković, S. (2012). Components of aesthetic experience: aesthetic fascination, aesthetic appraisal, and aesthetic emotion. *i-Perception* 3, 1–17. doi: 10.1068/i0450aap

Mather, G. (2014). Artistic adjustment of image spectral slope. *Art Percept.* 2, 11–22. doi: 10.1163/22134913-00002018

McManus, I., Cook, R., and Hunt, A. (2010). Beyond the golden section and normative aesthetics: Why do individuals differ so much in their aesthetic preferences for rectangles? *Psychol. Aesthet. Creat. Arts* 4, 113–126. doi: 10.1037/a0017316

McManus, I., Edmondson, D., and Rodger, J. (1985). Balance in pictures. *Br. J. Psychol.* 76, 311–324.

McManus, I., Stöver, K., and Kim, D. (2011a). Arnheim's Gestalt theory of visual balance: examining the compositional structure of art photographs and abstract images. *i-Perception* 2, 615–647. doi: 10.1068/i0445aap

McManus, I. C., Zhou, F. A., l'Anson, S., Waterfield, L., Stöver, K., and Cook, R. (2011b). The psychometrics of photographic cropping: the influence of colour, meaning, and expertise. *Perception* 40, 332–357. doi: 10.1068/p6700

Menzel, C., Hayn-Leichsenring, G. U., Langner, O., Wiese, H., and Redies, C. (2015). Fourier power spectrum characteristics of face photographs: attractiveness perception depends on low-level image properties. *PLoS ONE* 10:e0122801. doi: 10.1371/journal.pone.0122801

Montagner, C., Jesus, R., Correia, N., Vilarigues, M., Macedo, R., and Melo, M. J. (2016). Features combination for art authentication studies: brushstroke and materials analysis of Amadeo de Souza-Cardoso. *Mult. Tools Applic.* 75, 4039–4063. doi: 10.1007/s11042-015-3197-x

Munar, E., Gómez-Puerto, G., Call, J., and Nadal, M. (2015). Common visual preference for curved contours in humans and great apes. *PLoS ONE* 10:e0141106. doi: 10.1371/journal.pone.0141106

Mureika, J. R. (2005). Fractal dimensions in perceptual color space: a comparison study using Jackson Pollock's art. *Chaos Interdisc. J. Nonlinear Sci.* 15:043702. doi: 10.1063/1.2121947

Murray, N. (2012). *Predicting Saliency and Aesthetics in Images: A Bottom-up Perspective*. Ph.D. thesis, Departament de Ciencies de la Computacio, Autonomous University of Barcelona.

Murray, N., Marchesotti, L., and Perronnin, F. (2012). "Ava: a large-scale database for aesthetic visual analysis," in *2012 IEEE Conference on the Computer Vision and Pattern Recognition (CVPR)* (Providence, RI: IEEE), 2408–2415.

OED (2017). *Oxford English Dictionary Online*. Available online at: http://www.oed.com/viewdictionaryentry/Entry/293508 (Accessed May 24, 2017).

O'Hare, L., and Hibbard, P. B. (2011). Spatial frequency and visual discomfort. *Vis. Res.* 51, 1767–1777. doi: 10.1016/j.visres.2011.06.002

Oliva, A., and Torralba, A. (2006). Building the gist of a scene: the role of global image features in recognition. *Progr. Brain Res.* 155, 23–36. doi: 10.1016/S0079-6123(06)55002-2

Olshausen, B. A., and DeWeese, M. R. (2010). Applied mathematics: the statistics of style. *Nature* 463, 1027–1028. doi: 10.1038/4631027a

Osborne, H. (1986). Symmetry as an aesthetic factor. *Comput. Math. Applic.* 12, 77–82.

Palmer, S. E., and Schloss, K. B. (2010). An ecological valence theory of human color preference. *Proc. Natl. Acad. Sci. U.S.A.* 107, 8877–8882. doi: 10.1073/pnas.0906172107

Palmer, S. E., Schloss, K. B., and Griscom, W. S. (2016). Individual differences in perceptual preference. *Electr. Imag.* 2016, 1–6. doi: 10.2352/ISSN.2470-1173.2016.16.HVEI-113

Papaodysseus, C., Fragoulis, D. K., Panagopoulos, M., Panagopoulos, T., Rousopoulos, P., Exarhos, M., et al. (2006). Determination of the method of construction of 1650 BC wall paintings. *IEEE Trans. Patt. Anal. Mach. Intell.* 28, 1361–1371. doi: 10.1109/TPAMI.2006.183

Pinto, P. D., Linhares, J. M., Carvalhal, J. A., and Nascimento, S. M. (2006). Psychophysical estimation of the best illumination for appreciation of renaissance paintings. *Vis. Neurosci.* 23, 669–674. doi: 10.1017/S0952523806233340

Polatkan, G., Jafarpour, S., Brasoveanu, A., Hughes, S., and Daubechies, I. (2009). "Detection of forgery in paintings using supervised learning," in *2009 16th IEEE International Conference on Image Processing (ICIP)* (Cairo: IEEE), 2921–2924.

Rampone, G., O'Sullivan, N., and Bertamini, M. (2016). The role of visual eccentricity on preference for abstract symmetry. *PLoS ONE* 11:e0154428. doi: 10.1371/journal.pone.0154428

Reber, R., Schwarz, N., and Winkielman, P. (2004). Processing fluency and aesthetic pleasure: is beauty in the perceiver's processing experience? *Person. Soc. Psychol. Rev.* 8, 364–382. doi: 10.1207/s15327957pspr0804_3

Redies, C. (2007). A universal model of esthetic perception based on the sensory coding of natural stimuli. *Spat. Vis.* 21, 97–117. doi: 10.1163/156856807782753886

Redies, C. (2015). Combining universal beauty and cultural context in a unifying model of visual aesthetic experience. *Front. Hum. Neurosci.* 9:218. doi: 10.3389/fnhum.2015.00218

Redies, C., Amirshahi, S. A., Koch, M., and Denzler, J. (2012). "PHOG-derived aesthetic measures applied to color photographs of artworks, natural scenes and objects," in *European Conference on Computer Vision* (Berlin; Heidelberg: Springer), 522–531.

Redies, C., Brachmann, A., and Wagemans, J. (2017). High entropy of edge orientations characterizes visual artworks from diverse cultural backgrounds. *Vis. Res.* 133, 130–144. doi: 10.1016/j.visres.2017.02.004

Redies, C., Hänisch, J., Blickhan, M., and Denzler, J. (2007a). Artists portray human faces with the Fourier statistics of complex natural scenes. *Network Comput. Neural Syst.* 18, 235–248. doi: 10.1080/09548980701574496

Redies, C., Hasenstein, J., and Denzler, J. (2007b). Fractal-like image statistics in visual art: similarity to natural scenes. *Spatial Vis.* 21, 137–148. doi: 10.1163/156856807782753921

Renoult, J. P., Bovet, J., and Raymond, M. (2016). Beauty is in the efficient coding of the beholder. *R. Soc. Open Sci.* 3:160027. doi: 10.1098/rsos.160027

Rigau, J., Feixas, M., and Sbert, M. (2008). Informational aesthetics measures. *IEEE Comput. Graph. Appl.* 28, 24–34. doi: 10.1109/MCG.2008.34

Saleh, B., and Elgammal, A. (2015). Large-scale classification of fine-art paintings: learning the right metric on the right feature. *arXiv preprint arXiv:1505.00855*.

Schweinhart, A. M., and Essock, E. A. (2013). Structural content in paintings: artists overregularize oriented content of paintings relative to the typical natural scene bias. *Perception* 42, 1311–1332. doi: 10.1068/p7345

Shamir, L. (2015). What makes a Pollock Pollock: a machine vision approach. *Int. J. Arts Technol.* 8, 1–10. doi: 10.1504/IJART.2015.067389

Shamir, L., Macura, T., Orlov, N., Eckley, D. M., and Goldberg, I. G. (2010). Impressionism, expressionism, surrealism: automated recognition of painters and schools of art. *ACM Trans. Appl. Percept. (TAP)* 7:8. doi: 10.1145/1670671.1670672

Shamir, L., Orlov, N., Eckley, D. M., Macura, T., Johnston, J., and Goldberg, I. G. (2008). Wndchrm–an open source utility for biological image analysis. *Source Code Biol. Med.* 3:1. doi: 10.1186/1751-0473-3-13

Shamir, L., and Tarakhovsky, J. A. (2012). Computer analysis of art. *J. Comput. Cult. Herit.* 5:7. doi: 10.1145/2307723.2307726

Shen, J. (2009). Stochastic modeling Western paintings for effective classification. *Patt. Recogn.* 42, 293–301. doi: 10.1016/j.patcog.2008.04.016

Shimamura, A. P. (2014). "Toward a science of aesthetics: Issues and ideas," in *Aesthetic Science: Connecting Minds, Brains, and Experience*, eds A. Shimamura, and S. Palmer (Oxford: Oxford University Press), 3–28.

Siddiquie, B., Vitaladevuni, S. N., and Davis, L. S. (2009). "Combining multiple kernels for efficient image classification," in *2009 Workshop on the Applications of Computer Vision (WACV)* (Snowbird, UT: IEEE).

Silvia, P. J. (2005). Emotional responses to art: from collation and arousal to cognition and emotion. *Rev. Gen. Psychol.* 9, 342–357. doi: 10.1037/1089-2680.9.4.342

Silvia, P. J. (2014). "Human emotions and aesthetic experience: an overview of empirical aesthetics," in *Aesthetic Science: Connecting Minds, Brains, and Experience*, eds A. Shimamura, and S. Palmer (Oxford: Oxford University Press), 250–275.

Simoncelli, E. P., and Olshausen, B. A. (2001). Natural image statistics and neural representation. *Ann. Rev. Neurosci.* 24, 1193–1216. doi: 10.1146/annurev.neuro.24.1.1193

Simonyan, K., and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

Spehar, B., Walker, N., and Taylor, R. P. (2016). Taxonomy of individual variations in aesthetic responses to fractal patterns. *Front. Hum. Neurosci.* 10:350. doi: 10.3389/fnhum.2016.00350

Spratt, E. L., and Elgammal, A. (2014). "Computational beauty: Aesthetic judgment at the intersection of art and science," in *Workshop at the European Conference on Computer Vision* (Springer), 35–53.

Stork, D. G. (2009a). "Computer vision and computer graphics analysis of paintings and drawings: an introduction to the literature," in *International Conference on Computer Analysis of Images and Patterns* (Berlin; Heidelberg: Springer), 9–24.

Stork, D. G. (2009b). Learning-based authentication of jackson pollock's paintings. *SPIE Profes*. doi: 10.1117/2.1200905.1643

Stork, D. G., and Johnson, M. K. (2006). "Estimating the location of illuminants in realist master paintings computer image analysis addresses a debate in art history of the baroque," in *18th International Conference on Pattern Recognition (ICPR'06)*, *Vol. 1*(Hong Kong: IEEE), 255–258.

Tan, W. R., Chan, C. S., Aguirre, H. E., and Tanaka, K. (2016). "Ceci n'est pas une pipe: a deep convolutional network for fine-art paintings classification," in *2016 IEEE International Conference on the Image Processing (ICIP)* (Phoenix, AZ: IEEE), 3703–3707.

Taylor, R., Newell, B., Spehar, B., and Clifford, C. (2005). "Fractals: a resonance between art and nature," in *Mathematics and Culture II* (Berlin; Heidelberg: Springer), 53–63.

Taylor, R. P. (2002). Order in Pollock's chaos. *Sci. Am.* 287, 84–89. doi: 10.1038/scientificamerican1202-116

Taylor, R. P. (2006). Reduction of physiological stress using fractal art and architecture. *Leonardo* 39, 245–251. doi: 10.1162/leon.2006.39.3.245

Taylor, R. P., Micolich, A. P., and Jonas, D. (1999). Fractal analysis of Pollock's drip paintings. *Nature* 399, 422–422.

Taylor, R. P., and Spehar, B. (2016). "Fractal fluency: an intimate relationship between the brain and processing of fractal stimuli," in *The Fractal Geometry of the Brain* (New York, NY: Springer), 485–496.

Taylor, R. P., Spehar, B., Van Donkelaar, P., and Hagerhall, C. M. (2011). Perceptual and physiological responses to Jackson Pollock's fractals. *Front. Hum. Neurosci.* 5:50. doi: 10.3389/fnhum.2011.00060

Tong, H., Li, M., Zhang, H.-J., He, J., and Zhang, C. (2004). "Classification of digital photos taken by photographers or home users," in *Pacific-Rim Conference on Multimedia* (Springer), 198–205.

Torresani, L., Szummer, M., and Fitzgibbon, A. (2010). "Efficient object category recognition using classemes," in *European Conference on Computer Vision* (Berlin; Heidelberg: Springer), 776–789.

Wallraven, C., Fleming, R., Cunningham, D., Rigau, J., Feixas, M., and Sbert, M. (2009). Categorizing art: comparing humans and computers. *Comput. Graph.* 33, 484–495. doi: 10.1016/j.cag.2009.04.003

Wang, Y., and Takatsuka, M. (2012). "A framework towards quantified artistic influences analysis," in *2012 International Conference on the Digital Image Computing Techniques and Applications (DICTA)* (Fremantle, WA: IEEE), 1–8.

Westphal-Fitch, G., Oh, J., and Fitch, W. (2013). Studying aesthetics with the method of production: effects of context and local symmetry. *Psychol. Aesthet. Creat. Arts* 7:13. doi: 10.1037/a0031795

Widjaja, I., Leow, W. K., and Wu, F.-C. (2003). "Identifying painters from color profiles of skin patches in painting images," in *2003 International Conference on Image Processing, 2003. ICIP 2003, Vol. 1* (Barcelona: IEEE).

Yosinski, J., Clune, J., Nguyen, A., Fuchs, T., and Lipson, H. (2015). Understanding neural networks through deep visualization. *arXiv preprint arXiv:1506.06579*.

Zeki, S. (2013). Clive Bell's "Significant Form" and the neurobiology of aesthetics. *Front. Hum. Neurosci.* 7:730. doi: 10.3389/fnhum.2013.00730

Zhu, J.-Y., Park, T., Isola, P., and Efros, A. A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. *arXiv preprint arXiv:1703.10593*.

Zujovic, J., Gandy, L., Friedman, S., Pardo, B., and Pappas, T. N. (2009). "Classifying paintings by artistic genre: an analysis of features & classifiers," in *International Workshop on Multimedia Signal Processing, 2009. MMSP'09* (Rio De Janeiro: IEEE), 1–5.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Bioengineering and Biotechnology

ORIGINAL RESEARCH
published: 23 February 2018
doi: 10.3389/fbioe.2018.00012

Neuronal Mechanism for Compensation of Longitudinal Chromatic Aberration-Derived Algorithm

Yuval Barkan[1] *and Hedva Spitzer*[2]*

[1]*Biomedical Engineering Department, Faculty of Engineering, Tel Aviv University, Tel Aviv, Israel,* [2]*Electrical Engineering School, Faculty of Engineering, Tel-Aviv University, Tel-Aviv, Israel*

The human visual system faces many challenges, among them the need to overcome the imperfections of its optics, which degrade the retinal image. One of the most dominant limitations is longitudinal chromatic aberration (LCA), which causes short wavelengths (blue light) to be focused in front of the retina with consequent blurring of the retinal chromatic image. The perceived visual appearance, however, does not display such chromatic distortions. The intriguing question, therefore, is how the perceived visual appearance of a sharp and clear chromatic image is achieved despite the imperfections of the ocular optics. To address this issue, we propose a neural mechanism and computational model, based on the unique properties of the *S*-cone pathway. The model suggests that the visual system overcomes LCA through two known properties of the *S* channel: (1) omitting the contribution of the *S* channel from the high-spatial resolution pathway (utilizing only the *L* and *M* channels). (b) Having large and coextensive receptive fields that correspond to the small bistratified cells. Here, we use computational simulations of our model on real images to show how integrating these two basic principles can provide a significant compensation for LCA. Further support for the proposed neuronal mechanism is given by the ability of the model to predict an enigmatic visual phenomenon of large color shifts as part of the assimilation effect.

Keywords: aberration, chromatic adaptation, compensatory mechanisms, computer model, visual perception

OPEN ACCESS

Edited by:
Hagit Hel-Or,
University of Haifa, Israel

Reviewed by:
Inyoung Kim,
Virginia Tech, United States
Hauke Busch,
University of Lübeck, Germany

***Correspondence:**
Hedva Spitzer
hedva@eng.tau.ac.il

Specialty section:
This article was submitted to Bioinformatics and Computational Biology, a section of the journal Frontiers in Bioengineering and Biotechnology

Received: *07 October 2017*
Accepted: *23 January 2018*
Published: *23 February 2018*

Citation:
Barkan Y and Spitzer H (2018) Neuronal Mechanism for Compensation of Longitudinal Chromatic Aberration-Derived Algorithm. Front. Bioeng. Biotechnol. 6:12. doi: 10.3389/fbioe.2018.00012

INTRODUCTION

The human eye is affected by the imperfections of its optics, which degrade the quality of the retinal image and ultimately impose limits on vision. These imperfections have both spatial and chromatic implications. One of the most dominant chromatic implications is the phenomenon of longitudinal chromatic aberration (LCA). LCA is a significant and dominant attribute of the visual system and has been studied and measured extensively (e.g., Bedford and Wyszecki, 1957; Charman and Jennings, 1976).

Longitudinal chromatic aberration is induced by the dependence of the refractive power of the lens on wavelength. As can be seen in **Figure 1**, the ocular refractive power is higher for shorter wavelengths (Bedford and Wyszecki, 1957). The accommodation mechanism of human eyes can determine the focus for each wavelength, but it is impossible to bring all of the wavelengths to focus simultaneously (Wandell, 1995). The phenomenon of LCA has been measured extensively, both by psychophysically (Wald and Griffin, 1947; Ivanoff, 1953; Bedford and Wyszecki, 1957; Jenkins, 1963;

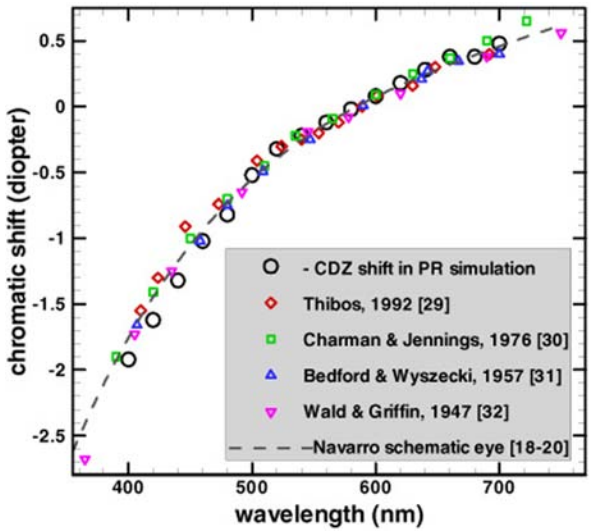


FIGURE 1 | Comparison of refractive power (chromatic shift) reported by several studies. Note that the chromatic shift is much larger for the short wavelengths (blue photoreceptor) than for the long wavelengths (red photoreceptor). All the data are adjusted vertically to have a zero value at the reference wavelength of 589 nm giving the longitudinal chromatic aberration a refractive power of about two diopters. This image has been taken with permission from The Optical Society (Chen et al., 2003).

Howarth and Bradley, 1986) and retinoscopy methods (Charman and Jennings, 1976; Rynders et al., 1998). These studies showed that LCA has a refractive power of about two diopters (*D*), across the visible spectrum (**Figure 1**).

An alternative method of representing the chromatic aberration is through the modulation transfer function (MTF), which describes the sensitivity as a function of the spatial frequency and the wavelength. Due to the LCA, the MTF of the *S*-cone (blue) channel has a lower frequency cutoff (by a factor of 3–5) than the MTF of the *M*/*L* cone channels (red–green) (Shevell, 2003).

An additional factor that limits the visual acuity of the *S*-pathway is the low density of the *S* photoreceptors at the retinal mosaic. It is plausible that this low density has evolved in the visual system, in order not to have more sensors than the optical MTF can utilize. The MTF thus would be limited by both the LCA and photoreceptor density which, as mentioned above, are not independent factors. Calkins (2001) showed that the *S*-cone density can be a consequence of efficient Nyquist sampling: ". . .the eye's optics together with what may be called 'typical' viewing conditions effectively limit any evolutionary pressure to pack S cones into the photoreceptor mosaic with a Nyquist rate greater than about 7–8 cycles deg^{-1}." If we approximate the *S* mosaic as triangular for ease of calculation, this sampling rate would correspond to an upper limit of foveal density in the human retina of 2,000–2,500 S cones mm^{-2}. Various anatomical measurements of the distribution of *S* cones in the human retina, both direct and indirect, converge to a similar estimate: *S* cones peak in density at about 2,000 cells mm^{-2}, just outside the center fovea, representing 5–10% of the cone population (Curcio et al., 1991).

The consequence of the LCA is that the retinal image will be focused only for the "green" wavelengths, and for the most part will be out of focus for the bluish wavelengths. The consequent image would be expected to have colored borders ("fringes")—similar to that seen with a cheap lens (Valberg, 2005). Although it is not possible to remove these chromatic defects from a lens, an efficient optical system should be designed to minimize the distortion caused by the LCA. For example, it is possible to correct chromatic aberration through a combination of two or more lenses, in such a way that the aberration of each lens compensates for the aberration of the other lens (achromatic lens). In the human visual system, this solution is impractical since we are continuously changing the focal distance.

A recent proposal suggests that Müller glial cells may play a role in reducing the chromatic aberration due to the fact that peripheral light at larger tilt angles will be rejected more readily (Labin and Ribak, 2010). Another suggestion is that the short-wavelength absorbing pigments of the ocular media may have a function in limiting the chromatic aberration (Walls, 1963; Nussbaum et al., 1981). However, spectral filtering in the ocular media has a relatively small effect on the MTF (Shevell, 2003) and none of these optical features (Walls, 1963; Labin and Ribak, 2010) is sufficient to explain the lack of perceived distortion at sharp achromatic edges.

It is therefore intriguing to understand how notwithstanding the imperfections of the ocular optics, including the LCA, the perceived visual appearance is still a sharp and clear image. Since the optical system of the eye cannot apparently account for the correction, it is reasonable to suppose that the neuronal system acts to reduce the distortion (Shevell, 2003; Valberg, 2005). It should be appreciated that a non-optical system, such as the neuronal mechanism, cannot fully compensate for the optical limitations, since some of the physical information is lost. (This is exhibited by the limited MTF.)

Several studies have indeed suggested that there must be neural compensation for the eye's aberrations. Although no specific mechanism has been described (Hay et al., 1963; Artal et al., 2004), a number of compensatory options have been suggested, most of which are related to the McCollough effect (ME) (Hay et al., 1963; Broerse et al., 1999; Grossberg et al., 2002). The ME is a long-term after-effect that can last from hours up to 3 months (Jones and Holding, 1975).

The rationale to associate the ME with the LCA phenomenon derives mainly from its long-lasting temporal property, and its relation to chromatic edges (McCollough, 1965). The proposed compensatory models are composed of oriented receptive fields (RFs) (multiplexed simple cells) consisting of both chromatic- and achromatic-separated subunits (Broerse et al., 1999; Grossberg et al., 2002). The elimination of the chromatic distortion is then explained by invoking a learning mechanism that inhibits the appearance of chromatic edges adjacent to achromatic edges.

These models have been supported by experiments that demonstrate that there is a long-term adaptation to chromatic aberration caused by a wedge prism. It has been demonstrated that dispersion of light passing through a wedge prism produces bluish and yellowish fringes on achromatic edges. These perceived fringes disappear when the prisms are worn for a long period of time (about 2 days) (Hay et al., 1963). This adaptation of the visual

system supports the existence of a long-term corrective neural compensation mechanism.

These models can be accounted for neuronal compensation only when the chromatic aberration refractive power is constant. However, the refractive power of the LCA constantly changes due to the pupil size (that is determined by the amount of light and the accommodation of the eye). The temporal scale of pupil size change is within the range of 200–500 ms, which is faster by orders of magnitude than the neuronal adaptation mechanisms described above (which can last hours to months). Consequently, there is necessity for an additional mechanism that compensates for chromatic aberration and is less dependent on a momentary magnitude of chromatic aberration.

This means that a neural mechanism that compensates for general LCA phenomenon still remains to be discovered. If such a neural mechanism exists, it is expected that not only will it have the ability to compensate for the LCA phenomenon but will also be able to predict the visual phenomena generated by the compensation neuronal mechanism.

In this paper, we propose a plausible computational model of the retina that can compensate for LCA. The model is based on well-known retinal color-coding RFs and does not require a learning process. The validity of the suggested model is supported by its ability to predict related visual phenomena.

MODEL

The model computes the perceived color in accordance with the response of retinal color-coding ganglion cells (Daw, 2012). This calculation involves two main stages. The first stage evaluates the response ganglion cells of type I (*L*/*M* and *M*/*L*, on center cells) and type II (*S*/*LM*, on coextensive cells). This stage includes the calculation of the RF response of each color-coding cell that also exhibits a remote adaptation mechanism. In addition, this stage also includes two separated pathways related to the luminance and chromatic knowledge of the two cell types. The second stage of the model proposes a novel transformation of the ganglion cell response into a perceived image by using an inverse function. The source code for the model simulation is available at https://github.com/yubarkan/LCAcompensation/.

Response of the Opponent RF

The retinal ganglion cells receive their input from the cones through several chemical and electrical processing layers (Shevell, 2003). The retinal ganglion cells then perform an adaptation of the first order. The adaptation of the first order is modeled here through adaptation of the cell inputs, rather than adaptation of the RF subregions (Spitzer and Semo, 2002; Spitzer and Barkan, 2005). We therefore define the adapted ganglion cell input signals as follows:

$$L_{\text{pr\_adapted}} = \frac{L_{\text{photo}-r}}{L_{\text{photo}-r} + \sigma_L\left(L_{\text{photo}-r} + L_{\text{remote}}\right)},$$

$$M_{\text{pr\_adapted}} = \frac{M_{\text{photo}-r}}{M_{\text{photo}-r} + \sigma_M\left(M_{\text{photo}-r} + M_{\text{remote}}\right)},$$

$$S_{\text{pr\_adapted}} = \frac{S_{\text{photo}-r}}{S_{\text{photo}-r} + \sigma_S\left(S_{\text{photo}-r} + S_{\text{remote}}\right)}, \quad (1)$$

where L_{adapted}, M_{adapted}, and S_{adapted} are the adapted inputs from the cones and $\sigma_{L,M,S}$ are remote and local adaptation signals and are defined as

$$\begin{aligned}\sigma_L &= a \cdot L_{\text{photo}-r} + b + c \cdot L_{\text{remote}},\\ \sigma_L &= a \cdot M_{\text{photo}-r} + b + c \cdot M_{\text{remote}},\\ \sigma_S &= a \cdot S_{\text{photo}-r} + b + c \cdot S_{\text{remote}},\end{aligned} \quad (2)$$

where the remote signals are defined as

$$\begin{aligned}L_{\text{remote}}(x,y) &= \iint_{\text{cen}-\text{area}} L_{\text{photo}-r}(x',y') \cdot f_{\text{remote}}(x-x',y-y') \cdot dx' \cdot dy',\\ M_{\text{remote}}(x,y) &= \iint_{\text{cen}-\text{area}} M_{\text{photo}-r}(x',y') \cdot f_{\text{remote}}(x-x',y-y') \cdot dx' \cdot dy',\\ S_{\text{remote}}(x,y) &= \iint_{\text{cen}-\text{area}} S_{\text{photo}-r}(x',y') \cdot f_{\text{remote}}(x-x',y-y') \cdot dx' \cdot dy'.\end{aligned} \quad (3)$$

The "remote" area is composed of an annulus-like shape around the entire RF region (Spitzer and Barkan, 2005). Its weight function (f_{remote}) is modeled as a decaying exponent at the remote area as follows:

$$f_{\text{remote}}(x,y) = \frac{1}{\pi \cdot \rho_{\text{remote}}} \exp\left(-\frac{x^2+y^2}{\rho_{\text{remote}}{}^2}\right); x,y \in \text{remote\_area}. \quad (4)$$

The spatial response profile of the two subregions of the retinal ganglion RF, "center" and "surround," is expressed by the known difference-of-Gaussians (DOG). It should be noted that the calculation of the DOG is performed on the adapted inputs.

The "center" signals of the two spectral regions, L_{cen}, M_{cen}, are defined as integrals of the adapted inputs (L_{adapted}, M_{adapted}; Eq. 1) over the center subregion, with a Gaussian decaying spatial weight function (f_c):

$$\begin{aligned}L_{\text{cen}}(x,y) &= \iint_{\text{cen}-\text{area}} L_{\text{pr\_adapted}}(x',y') \cdot f_c(x-x',y-y') \cdot dx' \cdot dy',\\ M_{\text{cen}}(x,y) &= \iint_{\text{cen}-\text{area}} M_{\text{pr\_adapted}}(x',y') \cdot f_c(x-x',y-y') \cdot dx' \cdot dy',\end{aligned} \quad (5)$$

while $L_{\text{cen}}(x,y)$ at each location represents the subregion response of the center area, which is centered at location $x, y, \ldots f_c$ and is defined as

$$f_c(x,y) = \frac{1}{\pi \cdot \rho_{\text{cen}}} \exp\left(-\frac{x^2+y^2}{\rho_{\text{cen}}{}^2}\right); x,y \in \text{center\_area}, \quad (6)$$

where ρ represents the radius of the center region of the RF. The "Surround" signals are defined in the same manner as follows (with a spatial weight function three times larger than that of the "center"):

$$\begin{aligned}L_{\text{sur}}(x,y) &= \iint_{\text{sur}-\text{area}} M_{\text{pr\_adapted}}(x',y') \cdot f_s(x-x',y-y') \cdot dx' \cdot dy',\\ M_{\text{sur}}(x,y) &= \iint_{\text{sur}-\text{area}} L_{\text{pr\_adapted}}(x',y') \cdot f_s(x-x',y-y') \cdot dx' \cdot dy',\end{aligned} \quad (7)$$

where f_s is defined as a decaying Gaussian over the surround region:

$$f_s(x,y) = \frac{1}{\pi \cdot \rho_{sur}} \exp\left(-\frac{x^2+y^2}{\rho_{sur}^2}\right); x,y \in \text{surround\_area}. \quad (8)$$

The total weight of f_c and f_s is 1.

The response of the cells is expressed by the subtraction of the center and surround-adapted responses as follows:

$$L^+M^-(x,y) = L_{cen}(x,y) - M_{sur}(x,y),$$
$$M^+L^-(x,y) = M_{cen}(x,y) - L_{sur}(x,y). \quad (9)$$

The *S*/*LM* retinal color-coding cell is known as the small bistratified ganglion cell. The RF of this cell is known in the literature to be coextensive (type II), i.e., it has mainly chromatic opponency rather than spatial opponency (Hubel and Wiesel, 1968; de Monasterio, 1978; Derrington et al., 1984). Accordingly, the response of the *S*-cone opponent is modeled here as a type-II RF. The *S*/*LM* signal was therefore modeled through integration of the chromatic difference (*S*/*LM*) over the whole RF of this cell type:

$$S^+LM^-(x,y) = \iint_{\text{blue}-\text{RF}-\text{area}} \left[S_{\text{adapted}}(x',y') - \frac{L_{\text{adapted}}(x',y') + M_{\text{adapted}}(x',y')}{2}\right] \cdot f_{s\_\text{center}}(x-x', y-y') \cdot dx' \cdot dy'. \quad (10)$$

The spatial weight function of the RF, f_{c_center}, is defined as in Eq. 7.

Transformation to Image

The purpose of this stage is to model how the visual system transforms the RF responses to a perceived image. We suggest that in order to eliminate the effect of the blurred *S*/*LM* channel, the visual system has to very precisely exclude this channel from the processing of the high-spatial resolution channel. This suggestion is in accordance with the consensus in the literature and with accumulated evidence indicating that the chromatic information that includes the *S*/*LM* information is processed through a unique pathway, i.e., the koniocellular pathway (Hendry and Reid, 2000). Additional support for our proposal is derived from the observation that the *L* and *M* data that code high-spatial resolution information are processed independently through the parvocellular pathway (Livingstone and Hubel, 1988; Van Essen and Gallant, 1994; Hendry and Reid, 2000; Sincich and Horton, 2005).

In order to perform a transformation from the opponent signals [$L+M-$, $M+L-$, and $S+(L+M)-$] to perceived triplet *LMS* values, we propose a functional minimization framework. We imply that the perceived values should satisfy the following equations:

$$L^+M^- = L_{\text{per}} - M_{\text{surround\_per}},$$
$$M^+L^- = M_{\text{per}} - L_{\text{surround\_per}}. \quad (11)$$

$L_{\text{surround\_per}}$ and $M_{\text{surround\_per}}$ are defined in Eq. 7, but here they are related to the perceived domain rather than adapted input signals. We define the following error function:

$$E(L_{\text{per}}, M_{\text{per}}) = \left[L_{\text{per}} - (L^+M^- + M_{\text{surround\_per}})\right]^2 + \left[M_{\text{per}} - (M^+L^- + L_{\text{surround\_per}})\right]^2. \quad (12)$$

This function is the square error between the estimation of L_{per}, M_{per}, and the satisfaction of Eq. 12. This error function can be minimized by various methods. For simplicity, we show the implication of the gradient descend method as follows (Snyman, 2005):

$$\frac{\partial L_{\text{per}}}{\partial t} = -\frac{\partial E(L_{\text{per}}, M_{\text{per}})}{\partial L_{\text{per}}},$$
$$\frac{\partial M_{\text{per}}}{\partial t} = -\frac{\partial E(L_{\text{per}}, M_{\text{per}})}{\partial M_{\text{per}}}. \quad (13)$$

Thus, we obtain the following iterative equations:

$$L^i{}_{\text{per}} = L^{i-1}{}_{\text{per}} + dt \cdot \left[2 \cdot \left(L^{i-1}{}_{\text{per}} - L^+M^- - M^{i-1}{}_{\text{surround\_per}}\right) + 2 \cdot fs(0,0) \cdot \left(M^{i-1}{}_{\text{per}} - M^+L^- - L^{i-1}{}_{\text{surround\_per}}\right)\right],$$
$$M^i{}_{\text{per}} = M^{i-1}{}_{\text{per}} + dt \cdot \left[2 \cdot \left(M^{i-1}{}_{\text{per}} - M^+L^- - L^{i-1}{}_{\text{surround\_per}}\right) + 2 \cdot fs(0,0) \cdot \left(L^{i-1}{}_{\text{per}} - L^+M^- - M^{i-1}{}_{\text{surround\_per}}\right)\right]. \quad (14)$$

This iteration process provides the perceived *L* and *M* values, independently of the *S*/*LM* channel (see the rationale above).

The perceived *S*-channel value (S_{per}) is calculated after evaluating the *L* and *M* perceived values (Eq. 14) by using the following equation:

$$S_{\text{per}} = S^+(L+M)^- + (L_{\text{per}} + M_{\text{per}})/2. \quad (15)$$

According to our model, the S_{per} contributes to the perceived color and not to the perceived luminance. Thus, the perceived brightness is expressed solely by the *L* and *M* values.

METHODS

In this section, we describe the different tools and parameters used in the model simulation. The same sets of parameters were used for all the simulated images that are presented in Section "Results."

Modeling Human Optics

In order to evaluate the ability of our model to compensate for chromatic aberration, it is necessary to simulate the results from human optics on test images. We have used the Image System Engineering Toolbox for Biology ISETBIO,[1] which provides a unique ability to simulate human optics in a real scene.

[1] https://github.com/isetbio/.

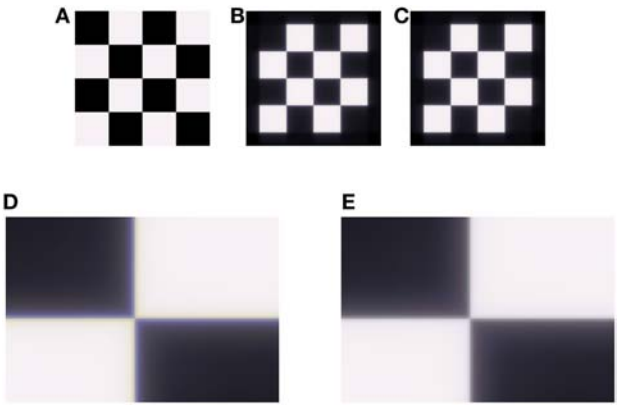


FIGURE 2 | Demonstration of the longitudinal chromatic aberration (LCA) model on an achromatic grid **(A)**. **(B)** Retinal image, simulated by ISETBIO toolbox (see Methods). **(C)** Model prediction for perceived image. **(D)** Enlarged section of the retinal image **(B)**, the LCA can be seen in vicinity of the edges as lines of a blue–yellow color. **(E)** Enlarged section of model prediction **(C)**, where there is a correction of the chromatic distortion.

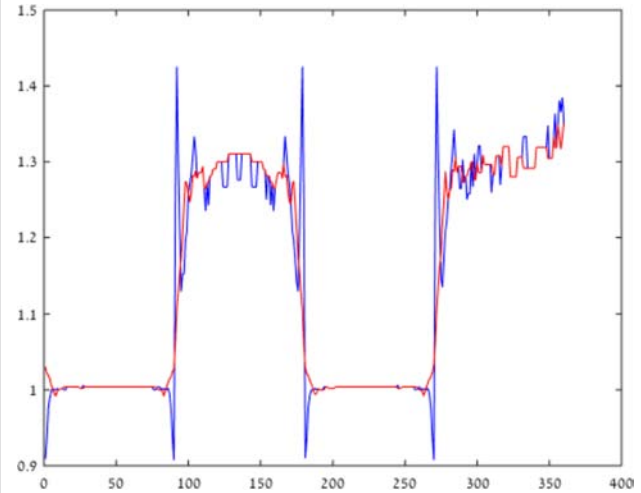


FIGURE 3 | Blue–yellow chromatic contrast. Blue–yellow chromatic contrast at a cross-section of the retinal image is presented, across a horizontal line in **Figure 2** (in blue), while the model correction for longitudinal chromatic aberration (LCA) is represented by the red line. It can be seen that the LCA, which is represented by the blue spikes, is reduced significantly by the model correction (red line), which eliminates the chromatic distortion.

For this purpose, we have used high-resolution, high-dynamic, multispectral image (HDRS) taken from the ISET High-Dynamic Range Multispectral Scene Database available by the Image Evaluation Tools.[2] ISETBIO also includes the WavefrontOptics code developed by David Brainard, Heidi Hofer, and Brian Wandell. Their code implements methods to model human eyes by taking adaptive optics data from wave-front sensors and calculating the optical blur as a function of the wavelength. The toolbox relies on data collected by Thibos et al. We have chosen an illumination of blackbody at 6,500 K and uses WavefrontOptics to simulate the retinal image produced by human optics. **Figure 2** is produced by this method.

Response of the Opponent RF

In the first stage of the model, the adapted signals are calculated (Eqs. 1–4). The remote area was simulated as an annulus with a diameter of 35 pixels. The adaptation parameters were chosen as follows: $a = 1$, $c = 1$, representing equal strength for the local and remote adaptations (Eq. 4). The parameter "b," which determines the strength of adaptation (Dahari and Spitzer, 1996; Spitzer and Barkan, 2005), was taken as $b = 3$.

The calculation of surround signals (Eq. 7) was calculated with fs (Eq. 8) having a decay constant (ρ) of 3 pixels. The response of the RFs was obtained by subtracting the center and surround-adapted responses (Eq. 9).

Transformation to Image (Inverse Function)

The purpose of this section is to perform a transformation from the RF responses to a perceived image. The transformation was performed using the Jacobi iterative method (Eq. 14). The iteration process was initiated ($i = 0$) by assuming achromatic stimuli. Specifically, all channels were initiated with the following values:

$$L^0_{\text{per}} = M^0_{\text{per}} = S^0_{\text{per}} = \frac{L_{\text{adapted}} + M_{\text{adapted}}}{2}.$$

[2]http://www.imageval.com/public/Products/ISET-SceneDatabase.html.

The iterative process converges to the predicted perceived image, while the color "fills-in" the stimulus.

RESULT

The ability of the model to reduce the effect of LCA was tested on both the artificial and natural images. Retinal images were simulated by using the ISETBIO toolbox, which takes into account the properties of the human optical system (see Methods). The LCA effect is very prominent when zooming into areas of luminance or chromatic edges (**Figure 2**).

Figure 2 demonstrates the model's performance on an artificial achromatic grid (**Figure 2A**) composed of equal energy squares. The image that is cast on the retina was calculated using ISETBIO (**Figure 2B**). It can be seen that this image (which simulates the eye's optics, including the LCA) has major chromatic distortions adjacent to the borders (**Figures 2B,D**). The distortion appears "yellowish" (lack of blue) on the bright side of the border and "bluish" on the darker side. **Figures 2C,E** present the effect of the model, which simulates the retinal response and its perceived image. **Figures 2B–E** show that the model succeeds in significantly reducing the chromatic-border distortion.

Figure 3 plots the chromatic contrast, defined as the ratio between the value of the blue and yellow channels [B/(R + G)], across the x-axis of **Figures 2B,C**. This chromatic contrast represents the chromatic deviation from neutral hue (achromatic region). An achromatic region is characterized by a contrast value of 1, while the higher and lower values represent deviations toward bluish and yellowish chroma, respectively.

The blue curve plots the chromatic contrast across the cast image (**Figure 2**). The fringes of the plot are indicated by the large negative and positive spikes next to the borders ($x = 90$). The results given by our model (red line) show a significant reduction of the spike magnitude, indicating a significant reduction of the

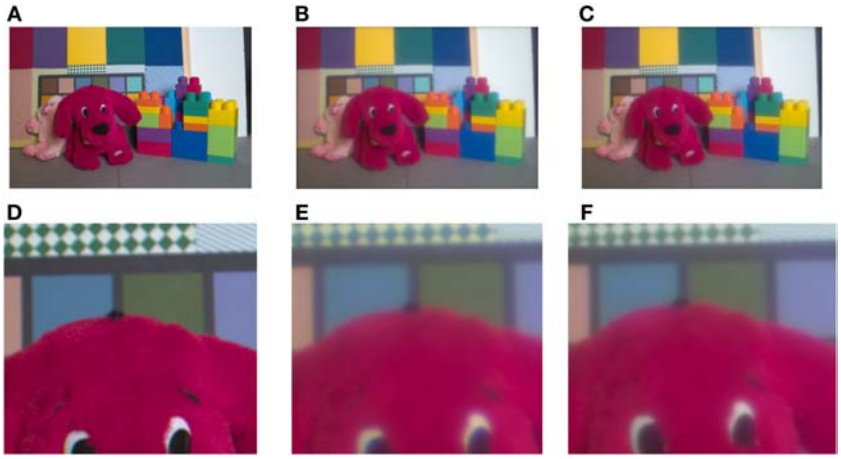


FIGURE 4 | Demonstration of the longitudinal chromatic aberration (LCA) model. Demonstration of the model performance on the toys' image **(A)** provided by Brian Wandell. **(B)** Retinal image, simulated by ISETBIO toolbox (see Methods). **(C)** Model prediction of the perceived image. **(D)** Enlarged section of retinal image (the LCA) can be seen in the vicinity of the edges as blue–yellow colored lines. **(D,E,F)** represent a magnified image of the puppy's eyes and the chromatic pattern zone in the background of the images **(A,B)**, and model prediction **(C)**. The correction can be observed only after enlargement **(F)**. The bluish color, a manifestation of the chromatic aberration, is prominent in **(E)** and the model's correction is seen clearly in **(F)**. The change in the bluish chroma is also clear in the background pattern in **(E)** and the greenish restoration in **(F)**.

chromatic fringes. The deviation from white is also significantly diminished. It should be noted that there is some constant hue generated mainly on the "black" squares, which is a side effect of the ISETBIO simulation, rather than an ideal achromatic appearance (contrast value of 1).

We also tested the model's ability to compensate for LCA on real images (**Figure 4A**), taken from the ISETBIO HDRS library. The optics of the eye was simulated using the ISETBIO (**Figure 4B**; see Methods). The results show that the model succeeds in correcting the chromatic distortions around borders (**Figure 4C**). The correction is prominent in the distorted puppy dog's eye color and the distorted green–white pattern behind the dog (**Figure 4D-F**). Although the model significantly reduces the distortion caused by LCA, it can also cause some minor chromatic artifacts.

The neuronal mechanism that we propose as capable of correcting for chromatic aberration is bound by the limitations of the spatial frequency of the *S*/*LM* channel (Eq. 10; see Model). In other words, a crucial part of the model suggests that the *S*/*LM* channel is processed through a spatial low-pass filter. If such a mechanism actually exists, we would predict that it would lead to visual phenomena that are prominent at stimuli with high frequencies of blue/yellow chromaticity. We would expect to see these phenomena as a blue–yellow assimilation effect, at high-spatial frequencies or among adjacent chromatic regions with sharp edges. These characteristics correspond closely to with a recent outstanding chromatic illusion, which is termed as "Chromatic induction from *S*-cone patterns" and described by Monnier and Shevell (2004) (**Figure 5**).

This illusion describes the perception of a chromatic specific narrow ring with color that differs completely, depending on the specific chromaticity of an adjacent ring (**Figure 5**). Psychophysical methods of analysis indicate that the chromatic shift is not directly dependent on the absolute blue channel intensity (*S*) of the blue component of the adjacent rings but rather on the relative amount of "blue" and "yellow" intensities (*S*/*LM*) in the adjacent rings (Shevell and Monnier, 2006).

We also tested our model on *S*-cone pattern stimuli, which have been reported by Monnier and Shevell (2004) to demonstrate prominent chromatic induction. The results (**Figure 5**) show that our model succeeds in predicting the trend of the perceived chromaticity shift toward the chromaticity of the adjacent ring (**Figure 5D**). The predicted chromatic shifts, between the two test chromaticities (the orange and pink rings) in terms of chromatic contrast $[S/(L+M)]$, are about 0.31. This shift agrees with the perceived colors as measured psychophysically by Shevell and Monnier.

DISCUSSION

This manuscript describes a neuronal mechanism and a computational model, based on retinal chromatic RFs and visual pathways, that compensate for LCA. The model can significantly reduce the chromatic distortion at both the artificial and natural images (**Figures 2** and **3**). The proposal is supported by the observation that an artifact of chromatic assimilation, which is a predicted consequence of the model, corresponds to a well-known chromatic assimilation phenomenon described previously (Shevell and Monnier, 2005).

The model is based on the specific spatial and chromatic structure of the blue–yellow channel $(S/L+M)$ RFs, which are spatially coextensive "type-II" small bistratified cell (SBC) (see Model; Hubel and Wiesel, 1968; de Monasterio, 1978; Derrington et al., 1984; Tailby et al., 2008; Crook et al., 2009; Martin and Lee, 2014) and correspond to the activities of the SBCs. These type-II RFs are incorporated into a retinal adaptation model (Spitzer and Barkan, 2005), and then the RF responses are subjected to an inverse function that mediates a transformation to perceived values. This transformation enables an evaluation of the model by

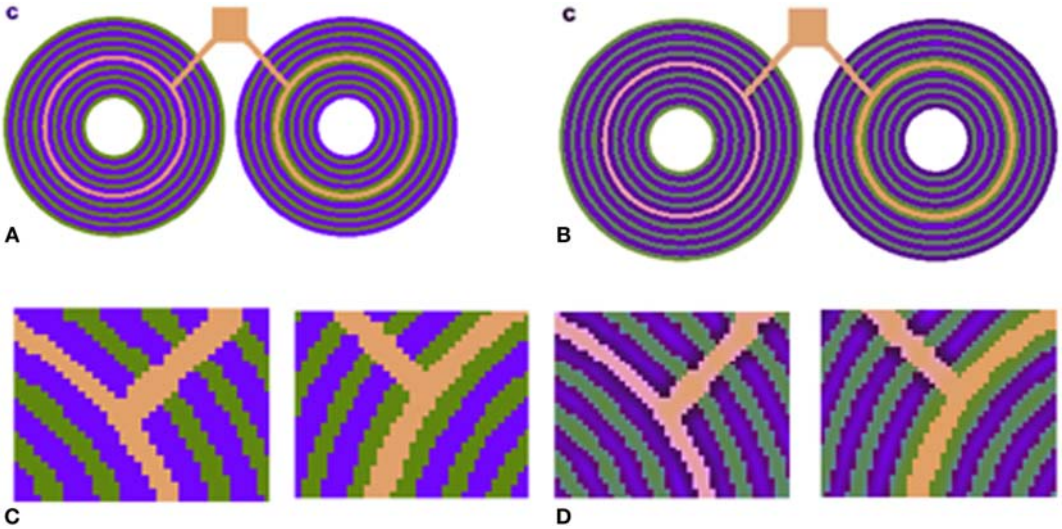


FIGURE 5 | **(A)** *S*-cone pattern reported by Shevell and Monnier (2005). The pink and orange rings are actually physically identical. **(B)** Our model prediction. **(C)** Zoom-in version of **(A)** shows that the central ring is identical. **(D)** Zoom-in version of **(B)** shows that the model successfully predicts the chromatic shift, and the left-hand side ring appears pinkish.

consideration of an image domain, rather than merely on the basis of the RF responses.

There has been some dispute in the literature regarding the spatial coextensive nature of the SBC. The coextensive nature of the SBC has been described by many electrophysiological researchers (Hubel and Wiesel, 1968; de Monasterio, 1978; Derrington et al., 1984). A recent experiment reported that the SBC RF may not be spatially coextensive (Field et al., 2007). However, these results have been criticized first because the data in Field et al. (2007) were collected in the far retinal periphery (30–75° eccentricity), where more recent and broad reports of the RF were recorded within the central 20°(Hubel and Wiesel, 1968; de Monasterio, 1978; Derrington et al., 1984). Crook et al. (2009) found that the *S*-ON and *LM*-OFF responses were spatially coextensive, or nearly so. Furthermore, this trend of results was supported by large previous papers including recent reports and a review (Tailby et al., 2008; Crook et al., 2009; Martin and Lee, 2014).

A logical conclusion may be that the development of visual system has been strongly influenced by the natural visual scenery. Most of the sun's spectral energy on earth is yellowish (550 nm) (Figure 1.2.1 in Wyszecki and Stiles, 1982), giving fewer chromatic edges in natural scenes than achromatic edges, and with a predominance of red–green chromatic edges over blue–yellow (Hansen and Gegenfurtner, 2009). The peak of the spectral luminance efficiency of the visual system (Wyszecki and Stiles, 1982) is similar to the peak of the sun's spectral energy with the ocular lens tuned for optimal focus at the same wavelength. The chromatic aberration occurs in the short wavelengths, where there is both less solar irradiance and fewer chromatic edges in natural images. It therefore appears that the ocular lens is designed to provide the optimal performance at the prominent natural wavelength (~550 nm) while allowing the aberration at shorter wavelengths, which are less significant both for spatial and luminance information.

Although the ocular lens is tuned to the most "important wavelengths," it still suffers from the consequences of the chromatic aberration. It is plausible that the neural system compensates for some of these optical imperfections (Wandell, 1995). We propose that the visual mechanism utilizes the absence of sharp blue–yellow edges to diminish the effect of chromatic distortions. In the model, this is replicated by the following mechanisms, whose existence is supported by psychophysics and neurophysiologic findings.

Luminance and high-spatial resolution chromatic information, under photopic light conditions, is obtained mainly from the *L* and *M* channels—which suffer less from LCA. This idea is supported by psychophysical evidence showing that the contribution of the *S* cone to luminance perception is negligible or null (Eisner and MacLeod, 1980; Wyszecki and Stiles, 1982). This knowledge has been also applied in the definition of the classical CIE color space where, for example, the $V(\lambda)s$ describing the spectral luminance efficiency (i.e., perceived brightness vs. wavelength) come mainly from greenish and red light (Wyszecki and Stiles, 1982). As a result, brightness is calculated by perceived *L* and *M* values with almost no input from the *S* channel (Eq. 14), while the calculation of the chromaticity takes the contribution of the *S* value into account as well as the contribution of the other chromatic channels (Eq. 15).

The opponent RF structure of the *S* channels (SBCs) is both spatially coextensive and chromatically complementary (Dacey, 1996; Rodieck, 1998; Eq. 10). Such an RF blurs the blue–yellow information, so that their chromatic mixture yields an achromatic color. In addition, the spatio-chromatic structure [of $S/(L+M)$ RF] yields a null response to achromatic edges, also in the presence of LCA affecting the *S* channel. In this way, the unique spatio-chromatic property minimizes the chromatic distortion (see Results; **Figure 2**).

In order to maintain the compensatory advantage at the retinal stage, which separates high-spatial frequency information from low-spatial frequency chromatic information, the system has to further process these two channels separately. There are

physiological findings, which show that the SBC RF (with B/Y chromatic structure) indeed feeds a distinct chromatic pathway, i.e., the koniocellular pathway (Hendry and Reid, 2000). The origin of the koniocellular pathway lies in the SBC in the retina, and the pathway is then relayed by the koniocellular layer in the LGN to the cytochrome-oxidase blobs in V1. Several studies have reported that information on color *per se* and information on form are separated (Livingstone and Hubel, 1988; Van Essen and Gallant, 1994; Sincich and Horton, 2005). The information on form is derived solely from the parvocellular pathway [which lacks the $S/(L+M)$ information]. The information on color, however, comes from both the koniocellular and parvocellular pathways. The parvocellular pathway sends inputs from layer 4$c\beta$ to the blobs in layer 2/3, area V1. The two separate pathways (color and form) do have different anatomical inputs in the V2 area. Here, the thin stripes that code the color information are fed both from the konio and parvo pathways, whereas the pale strips, which code the form information, are fed only by the parvo pathway. The "form" pathway is therefore not affected by the deficiencies of the $S/(L+M)$ pathway. Both pathways project to area V4 and additional higher visual areas.

Previous studies that proposed neuronal mechanisms to compensate for chromatic aberration (Hay et al., 1963; Broerse et al., 1999; Grossberg et al., 2002; Vladusich and Broerse, 2002) related these mechanisms to long-term after-effects, such as the ME—a long-term orientation-contingent color after-effect (McCollough, 1965). Vladusich and Broerse (2002) proposed a learning neuronal model that inhibits the fringes at luminance boundaries (caused by chromatic aberrations). Grossberg et al. (2002) proposed a learning mechanism whose primary function is to adaptively align the representations of the boundaries and surfaces, which are shifted due to the process of binocular fusion. Their mechanism was able to predict the ME. Since the ME has been previously suggested as the compensation mechanism for chromatic aberration, the model presented by Grossberg et al. (2002) was also regarded as a compensation model for LCA.

In our opinion, there are two main arguments against the idea that ME models can completely explain neuronal compensation to LCA. The first limitation of the above models (Broerse et al., 1999; Grossberg et al., 2002; Vladusich and Broerse, 2002) is that they assume that the magnitude of LCA effect depends solely on the magnitude of the luminance edge. However, the LCA effect also depends on additional optical factors, such as the pupil aperture (DeValois and DeValois, 1991), whose size changes dynamically in response to the level of ambient illumination and accommodation. Such learning mechanisms, therefore, would be expected to yield chromatic artifacts when the pupil aperture size changes and would therefore require continuous adaptation of the learning mechanism. The learning models described above may therefore be more applicable to transverse chromatic aberration (TCA), which does not depend on the pupil size. Thus, there could be two different and complementary mechanisms for the two types of aberrations, i.e., TCA and LCA.

An additional limitation of previous models (Broerse et al., 1999; Grossberg et al., 2002; Vladusich and Broerse, 2002) is their assumption that the LCA is triggered only by achromatic boundaries. In fact, chromatic aberration (and specifically the LCA) also occurs at iso-luminance chromatic boundaries, where there are no achromatic boundaries (**Figure 1**). Consequently, the above models fail to explain how the visual system processes chromatic fringes at non-achromatic borders.

The two types or mechanisms, the current proposed retinal model, and the above learning mechanisms can be synergetic in the visual system. The retinal mechanism performs an early-stage correction that eliminates most of the LCA effects, regardless of the degree of illumination and eye accommodation. The cortical learning mechanism (Watanabe et al., 1992; Broerse et al., 1999; Grossberg et al., 2002; Vladusich and Broerse, 2002; Grossberg, 2003) performs long-term adaptation that can adapt to specific ocular changes (such as lens defects that can be caused by aging or physical damage, etc.).

Although several studies have examined the improvement of visual acuity through optical correction of LCA (Campbell and Gubisch, 1967; Yoon and Williams, 2002; Artal et al., 2010), none found better than minor improvement (or none) of the contrast sensitivity. One may argue that these results suggest that LCA is not a real problem of the optical system, since correcting it does not create any significant improvement. However, in our opinion this would be an erroneous conclusion, since the whole visual pathway is already optimized to contend with the optical limitations. Therefore, correction of the optical limitations is not able to improve the situation further and it is necessary to invoke neuronal processing (including photoreceptor accommodation, RF structure and size, the different neuronal processing pathways, etc.).

Furthermore, LCA is expected to be manifested not only adjacently to achromatic edges but also in many other spatial and chromatic configurations. For example, one would also expect LCA at iso-luminance chromatic edges and non-oriented edges (such as textures or dots on a uniform background). In such configurations, the visual image is clear, despite the fact that the "leakage" of short-wavelength colors is still expected to influence the chromatic appearance, and the postulated models are unable to provide compensation.

The strength of a computational model can be enhanced by showing its ability to predict additional phenomena. Evidence for the competence of our model comes from its ability to predict the enigmatic visual phenomenon of the large chromatic shifts by *S*-cone pattern (Shevell and Monnier, 2005; **Figure 5**).

Shevell and Monnier (2006) and Cao and Shevell (2005) suggested that the large color shifts are mediated by a spatially antagonist $S+/S-$ cortical RF. The "*S*" term referred to the *S*-cone response normalized by the luminance. Cells with this type of response while not found in the retina have been identified in some neurons in V1 and V2 visual areas (Conway, 2001). Significantly, our model is based on retinal RFs (rather than cortical) (Hubel and Wiesel, 1968; de Monasterio, 1978; Derrington et al., 1984).

In addition, Shevell et al. also showed that the effect is more prominent with high-spatial frequency of the rings. We assume that this was the incentive to include spatially antagonist RFs in their qualitative model. We suggest, however, that an additional mechanism is recruited for low-frequency stimuli, i.e.,

simultaneous contrast mechanism (see Model, adaptation of the first order). Such a mechanism could originate from a retinal source (Spitzer and Barkan, 2005). This suggestion should be supported by additional experimental data, which should determine whether the effect originates from retinal vs. cortical mechanisms, as suggested previously (Cao and Shevell, 2005; Shevell and Monnier, 2006).

In summary, in this manuscript, we propose a model which explains how the visual system compensates for LCA. This compensatory mechanism can also explain additional visual phenomena, such as the large chromatic shifts by *S*-cone pattern, for which the underlying mechanism is still unknown. In addition, this mechanism can explain the necessity for two separate chromatic visual pathways, i.e., koniocellular and parvocellular pathways.

AUTHOR CONTRIBUTIONS

This is an original research done by YB under the supervision and parternship with HS.

REFERENCES

Artal, P., Chen, L., Fernandez, E. J., Singer, B., Manzanera, S., and Williams, D. R. (2004). Neural compensation for the eye's optical aberrations. *J. Vis.* 4, 281–287. doi:10.1167/4.4.4

Artal, P., Manzanera, S., Piers, P., and Weeber, H. (2010). Visual effect of the combined correction of spherical and longitudinal chromatic aberrations. *Opt. Exp.* 18, 1637–1648. doi:10.1364/OE.18.001637

Bedford, R. E., and Wyszecki, G. (1957). Axial chromatic aberration of the human eye. *J. Opt. Soc. Am.* 47, 564–565. doi:10.1364/JOSA.47.0564_1

Broerse, J., Vladusich, T., and O'Shea, R. P. (1999). Colour at edges and colour spreading in McCollough effects. *Vis. Res.* 39, 1305–1320. doi:10.1016/S0042-6989(98)00231-4

Calkins, D. J. (2001). Seeing with S cones. *Prog. Retin. Eye Res.* 20, 255–287. doi:10.1016/S1350-9462(00)00026-4

Campbell, F. W., and Gubisch, R. W. (1967). The effect of chromatic aberration on visual acuity. *J. Physiol.* 192, 345–358. doi:10.1113/jphysiol.1967.sp008304

Cao, D., and Shevell, S. K. (2005). Chromatic assimilation: spread light or neural mechanism? *Vis. Res.* 45, 1031–1045. doi:10.1016/j.visres.2004.10.016

Charman, W. N., and Jennings, J. A. (1976). Objective measurements of the longitudinal chromatic aberration of the human eye. *Vis. Res.* 16, 999–1005. doi:10.1016/0042-6989(76)90232-7

Chen, Y.-L., Tan, B., and Lewis, J. (2003). Simulation of eccentric photorefraction images. *Opt. Exp.* 11, 1628–1642. doi:10.1364/OE.11.001628

Conway, B. R. (2001). Spatial structure of cone inputs to color cells in alert macaque primary visual cortex (V-1). *J. Neurosci.* 21, 2768–2783. doi:10.1523/jneurosci.3577-09.2009

Crook, J. D., Davenport, C. M., Peterson, B. B., Packer, O. S., Detwiler, P. B., and Dacey, D. M. (2009). Parallel ON and OFF cone bipolar inputs establish spatially coextensive receptive field structure of blue-yellow ganglion cells in primate retina. *J. Neurosci.* 29, 8372–8387. doi:10.1523/JNEUROSCI.1218-09.2009

Curcio, C. A., Allen, K. A., Sloan, K. R., Lerea, C. L., Hurley, J. B., Klock, I. B., et al. (1991). Distribution and morphology of human cone photoreceptors stained with anti-blue opsin. *J. Comp. Neurol.* 312, 610–624. doi:10.1002/cne.903120411

Dacey, D. M. (1996). Circuitry for color coding in the primate retina. *Proc. Natl. Acad. Sci. U.S.A.* 93, 582–588. doi:10.1073/pnas.93.2.582

Dahari, R., and Spitzer, H. (1996). Spatiotemporal adaptation model for retinal ganglion cells. *J. Opt. Soc. Am. A. Opt. Img. Sci. Vis.* 13, 419–435. doi:10.1364/JOSAA.13.000419

Daw, N. (2012). *How Vision Works: The Physiological Mechanisms Behind What We See*, 1st Edn. Oxford Scholarship.

de Monasterio, F. M. (1978). Properties of concentrically organized X and Y ganglion cells of macaque retina. *J. Neurophysiol.* 41, 1394–1417. doi:10.1152/jn.1978.41.6.1435

Derrington, A. M., Krauskopf, J., and Lennie, P. (1984). Chromatic mechanisms in lateral geniculate nucleus of macaque. *J. Physiol.* 357, 241–265. doi:10.1113/jphysiol.1984.sp015499

DeValois, R. L., and DeValois, K. K. (1991). *Spatial Vision*. Oxford Psychology Series

Eisner, A., and MacLeod, D. I. (1980). Blue-sensitive cones do not contribute to luminance. *J. Opt. Soc. Am.* 70, 121–123. doi:10.1364/JOSA.70.000121

Field, G. D., Sher, A., Gauthier, J. L., Greschner, M., Shlens, J., Litke, A. M., et al. (2007). Spatial properties and functional organization of small bistratified ganglion cells in primate retina. *J. Neurosci.* 27, 13261–13272. doi:10.1523/JNEUROSCI.3437-07.2007

Grossberg, S. (2003). "Filling-in the forms: surface and boundaryinteractions in visual cortex," in *Filling-in: From Perceptual Completion to Cortical Reorganization*, ed. P. D. W. L. Pessoa (New York: Oxford University Press) 13–37.

Grossberg, S., Hwang, S., and Mingolla, E. (2002). Thalamocortical dynamics of the McCollough effect: boundary-surface alignment through perceptual learning. *Vis. Res.* 42, 1259–1286. doi:10.1016/S0042-6989(02)00055-X

Hansen, T., and Gegenfurtner, K. R. (2009). Independence of color and luminance edges in natural scenes. *Vis. Neurosci.* 26, 35–49. doi:10.1017/S0952523808080796

Hay, J. C., Pick, H. L. Jr., and Rosser, E. (1963). Adaptation to chromatic aberration by the human visual system. *Science* 141, 167–169. doi:10.1126/science.141.3576.167

Hendry, S. H., and Reid, R. C. (2000). The koniocellular pathway in primate vision. *Annu. Rev. Neurosci.* 23, 127–153. doi:10.1146/annurev.neuro.23.1.127

Howarth, P. A., and Bradley, A. (1986). The longitudinal chromatic aberration of the human eye, and its correction. *Vis. Res.* 26, 361–366. doi:10.1016/0042-6989(86)90034-9

Hubel, D. H., and Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *J. Physiol.* 195, 215–243. doi:10.1113/jphysiol.1968.sp008455

Ivanoff, A. (1953). *Les aberrations de l'oeil*. Paris: Éditions de la Revue d'optique théorique et instrumentale.

Jenkins, T. C. (1963). Aberrations of the eye and their effects on vision II. *Br. J. Physiol. Opt.* 20, 161–201.

Jones, P. D., and Holding, D. H. (1975). Extremely long-term persistence of the McCollough effect. *J. Exp. Psychol. Hum. Percept. Perform.* 1, 323–327. doi:10.1037/0096-1523.1.4.323

Labin, A. M., and Ribak, E. N. (2010). Retinal glial cells enhance human vision acuity. *Phys. Rev. Lett.* 104, 158102. doi:10.1103/PhysRevLett.104.158102

Livingstone, M., and Hubel, D. (1988). Segregation of form, color, movement, and depth: anatomy, physiology, and perception. *Science* 240, 740–749. doi:10.1126/science.3283936

Martin, P. R., and Lee, B. B. (2014). Distribution and specificity of S-cone ("blue cone") signals in subcortical visual pathways. *Vis. Neurosci.* 31, 177–187. doi:10.1017/S0952523813000631

McCollough, C. (1965). Color adaptation of edge-detectors in the human visual system. *Science* 149, 1115–1116. doi:10.1126/science.149.3688.1115

Monnier, P., and Shevell, S. K. (2004). Chromatic induction from S-cone patterns. *Vis. Res.* 44, 849–856. doi:10.1016/j.visres.2003.11.004

Nussbaum, J. J., Pruett, R. C., and Delori, F. C. (1981). Historic perspectives. Macular yellow pigment. The first 200 years. *Retina* 1, 296–310. doi:10.1097/00006982-198101040-00007

Rodieck, R. W. (1998). *The First Steps in Seeing*. Sunderland, MA: Sinauer Associates.

Rynders, M. C., Navarro, R., and Losada, M. A. (1998). Objective measurement of the off-axis longitudinal chromatic aberration in the human eye. *Vis. Res.* 38, 513–522. doi:10.1016/S0042-6989(97)00216-2

Shevell, S. K. (2003). *The Science of Color*, 2nd Edn. Amsterdam; London: Elsevier; Optical Society of America.

Shevell, S. K., and Monnier, P. (2005). Color shifts from S-cone patterned backgrounds: contrast sensitivity and spatial frequency selectivity. *Vis. Res.* 45, 1147–1154. doi:10.1016/j.visres.2004.11.013

Shevell, S. K., and Monnier, P. (2006). Color shifts induced by S-cone patterns are mediated by a neural representation driven by multiple cone types. *Vis. Neurosci.* 23, 567–571. doi:10.1017/S0952523806233303

Sincich, L. C., and Horton, J. C. (2005). The circuitry of V1 and V2: integration of color, form, and motion. *Annu. Rev. Neurosci.* 28, 303–326. doi:10.1146/annurev.neuro.28.061604.135731

Snyman, J. A. (2005). *Practical Mathematical Optimization: An Introduction to Basic Optimization Theory and Classical and New Gradient-Based Algorithms*. New York: Springer.

Spitzer, H., and Barkan, Y. (2005). Computational adaptation model and its predictions for color induction of first and second orders. *Vision Res.* 45, 3323–3342. doi:10.1016/j.visres.2005.08.002

Spitzer, H., and Semo, S. (2002). Color constancy: a biological model and its application for still andvideo images. *Pattern Recognit.* 35, 1645–1659. doi:10.1016/S0031-3203(01)00160-1

Tailby, C., Solomon, S. G., and Lennie, P. (2008). Functional asymmetries in visual pathways carrying S-cone signals in Macaque. *J. Neurosci.* 28, 4078–4087. doi:10.1523/JNEUROSCI.5338-07.2008

Valberg, A. (2005). *Light Vision Color*. Hoboken, NJ: John Wiley & Sons.

Van Essen, D. C., and Gallant, J. L. (1994). Neural mechanisms of form and motion processing in the primate visual system. *Neuron* 13, 1–10. doi:10.1016/0896-6273(94)90455-3

Vladusich, T., and Broerse, J. (2002). Color constancy and the functional significance of McCollough effects. *Neural Netw.* 15, 775–809. doi:10.1016/S0893-6080(02)00085-0

Wald, G., and Griffin, D. R. (1947). The change in refractive power of the human eye in dim and bright light. *J. Opt. Soc. Am.* 37, 321–336. doi:10.1364/JOSA.37.000321

Walls, G. L. (1963). *The Vertebrate Eye and Its Adaptive Radiation*. New York: Hafner Pub. Co.

Wandell, B. A. (1995). *Foundations of Vision*. Sunderland, MA: Sinauer Associates.

Watanabe, T., Zimmerman, G. L., and Cavanagh, P. (1992). Orientation-contingent color aftereffects mediated by subjective transparent structures. *Percept. Psychophys.* 52, 161–166. doi:10.3758/BF03206769

Wyszecki, G., and Stiles, W. S. (1982). *Colour Science: Concepts and Methods, Quantitative Data and Formulae*, 2nd Edn. New York; Chichester: Wiley.

Yoon, G. Y., and Williams, D. R. (2002). Visual performance after correcting the monochromatic and chromatic aberrations of the eye. *J. Opt. Soc. Am. A. Opt. Img. Sci. Vis.* 19. doi:10.1364/JOSAA.19.000266

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 02 March 2018
doi: 10.3389/fnins.2018.00123

Short and Long-Term Attentional Firing Rates Can Be Explained by ST-Neuron Dynamics

Oscar J. Avella Gonzalez[1,2]* *and John K. Tsotsos*[1,2]

[1] Department of Electrical Engineering and Computer Science, York University, Toronto, ON, Canada, [2] Laboratory for Active and Attentive Vision, Centre for Vision Research, York University, Toronto, ON, Canada

Attention modulates neural selectivity and optimizes the allocation of cortical resources during visual tasks. A large number of experimental studies in primates and humans provide ample evidence. As an underlying principle of visual attention, some theoretical models suggested the existence of a gain element that enhances contrast of the attended stimuli. In contrast, the Selective Tuning model of attention (ST) proposes an attentional mechanism based on suppression of irrelevant signals. In this paper, we present an updated characterization of the ST-neuron proposed by the Selective Tuning model, and suggest that the inclusion of adaptation currents (Ih) to ST-neurons may explain the temporal profiles of the firing rates recorded in single V4 cells during attentional tasks. Furthermore, using the model we show that the interaction between stimulus-selectivity of a neuron and attention shapes the profile of the firing rate, and is enough to explain its fast modulation and other discontinuities observed, when the neuron responds to a sudden switch of stimulus, or when one stimulus is added to another during a visual task.

Keywords: visual attention, single cell, ST-neuron, firing rate, neural selectivity

OPEN ACCESS

Edited by:
Xavier Otazu,
Universitat Autònoma de Barcelona, Spain

Reviewed by:
Keith Schneider,
University of Delaware, United States
Jihyun Yeonan-Kim,
San Jose State University, United States

***Correspondence:**
Oscar J. Avella Gonzalez
oscarjavella@gmail.com

Specialty section:
This article was submitted to Perception Science, a section of the journal Frontiers in Neuroscience

Received: *11 August 2017*
Accepted: *15 February 2018*
Published: *02 March 2018*

Citation:
Avella Gonzalez OJ and Tsotsos JK (2018) Short and Long-Term Attentional Firing Rates Can Be Explained by ST-Neuron Dynamics. Front. Neurosci. 12:123. doi: 10.3389/fnins.2018.00123

INTRODUCTION

Attention can be widely defined as "the selective prioritization of the neural representations that are most relevant to one's current behavioral goal" (Buschman and Kastner, 2015). Since James' pioneering work (James, 1891), research on attention has aimed to discover a precise and systematic description of how the brain is able to manage its limited resources for performing complex cognitive and behavioral tasks. Visual attention, as one component of attention, has received significant interest (Itti et al., 2005; Carrasco, 2011; Posner, 2011), leading to the proposal of detailed descriptions of aspects like bottom-up attention (Itti and Koch, 2001; Rutishauser et al., 2004; Itti, 2005) and top-down control (Corbetta and Shulman, 2002; Oliva et al., 2003; Buschman and Miller, 2007; Bressler et al., 2008), signal integration (Corbetta et al., 1991; Rao et al., 1997; Eagleman and Sejnowski, 2000), or focus of attention (Koch and Ullman, 1987; Desimone and Duncan, 1995; Tsotsos et al., 1995).

Mathematical models as a wide-spread strategy are used to make insightful predictions about neural communication, and brain dynamics in general (Hodgkin and Huxley, 1952; Destexhe et al., 1998; Kandel et al., 2000; Dayan and Abbott, 2001; Shriki et al., 2003; Izhikevich, 2004). Concerning visual attention, a number of relevant models have been proposed to study particular aspects concerned with the way single neurons and circuits process incoming information during visual tasks (Tsotsos, 1990; Niebur and Koch, 1994; Reynolds et al., 1999; Deco and Lee, 2002; Reynolds and Heeger, 2009); One of these aspects, treated by different studies and that currently

draws special interest, is the mechanism neurons use during attentional tasks to accurately encode, classify and prioritize dissimilar information using only their firing rates. For instance, in the biased competition model by Reynolds et al. (1999), stimuli compete for a cortical representation, and the average firing rate (response) of a neural population depends on the interaction between the selectivity of the cells for one particular type of stimulus or feature, and the modulation induced by attention. The feature similarity model of Martinez-Trujillo and Treue proposes that attention enhances neural selectivity (Martinez-Trujillo and Treue, 2004), thus causing neurons to increase their firing rate. The idea aligns well with the normalization model (Lee and Maunsell, 2009; Reynolds and Heeger, 2009) in which such enhancement relates to the contrast between the attended stimulus and the surrounding background perceived by a neural population. Other models also explore the relation between the detailed anatomy of the neurons and the response to the attentive signal. The Feedback model for example, acknowledges attention as a top-down process that operates via cortical feedback, and represents it using a gain factor that modulates the activity of impinging connections to a given neuron (Spratling and Johnson, 2004). It also takes into consideration physiological properties such as the roles of the basal (feedforward) and apical (feedback) connections, and how by adding those elements it is possible to resemble the response of pyramidal cells during attentional tasks (Spratling and Johnson, 2004). In the Selective Tuning model (ST) (Tsotsos, 1990, 2011; Rothenstein and Tsotsos, 2014), attention is also embodied as a top-down signal; but in contrast to other models, its selection mechanism fully relies on suppression of the irrelevant inputs to each neuron instead of the enhancement of their activity (Tsotsos, 1990, 2011), as supported by strong experimental evidence (Cutzu and Tsotsos, 2003; Loach et al., 2005; Hopf et al., 2006; Bartsch et al., 2017).

Adaptation mechanisms are well known for their facilitating role in detecting weak signals by means of stochastic resonance (Wiesenfeld and Moss, 1995) or through sub-threshold oscillations enhancement (Dorval and White, 2005). In a previous modeling study Rothenstein and Tsotsos (2014) found that by incorporating adaptation mechanisms, the overall performance of the ST neuron was improved during a simple attentional task. Thus, counterbalancing the rapid saturation of the firing rate due to the presentation of a highly affine stimulus, while resembling the shape of the firing profiles recorded in V4 visual cells (Kosai et al., 2014) (Figures 2, 3 therein). As a follow up of that study, in the present paper we perform a detailed characterization of the ST-neuron firing pattern with and without adaptation currents (Ih) (Pape, 1996). Next, and following the design by Reynolds et al., (Reynolds et al., 1999) we implement a simple circuit to explore various scenarios in which adaptation currents play a role in reshaping the firing profile of the neuron, either by fine-tuning it, or by increasing the sensitivity of the cell to the attentional signal.

The contribution of adaptation currents to the cell's dynamics is further highlighted, by simulating a set of experiments that strikingly uncovers the interplay between neural selectivity and attention as a twofold effect. It first creates a transitory and a stationary scenario in the firing response of the recorded cell; and second, induces the transition between the firing patterns evoked by two competitive stimuli in a task-dependent fashion. We also compare the results of our simulations against experimental findings, and show how the incorporation of Ih on the ST-model leads the response to closely resemble the transient and long-lasting effects observed in experimental data.

METHODS

Our model consists of four essential elements: the ST-neuron model, the circuit's design and connectivity, the neural selectivity, and the selection mechanism of attention.

The ST-Neuron

The Selective Tuning model of attention (ST) relies on the ST-neuron as its building block (Tsotsos, 1990, 2011; Tsotsos et al., 1995). The ST-neuron is responsible for the integration and propagation of signals across the visual hierarchy, and both implements attentional selection as well as displays modulations resulting from top-down attentional signals. As a rate-based model, the response is quantified by the temporal evolution of the firing rate (FR) according to Equation (1):

$$\frac{dFR}{dt} = \frac{1}{\tau} \cdot (-FR + S(P)) \tag{1}$$

In this expression, P is the synaptic input, $S(P) = \frac{MP^{\xi}}{\sigma^{\xi}+P^{\xi}}$ is the Naka-Rushton sigmoid function, whose value depends on the maximum firing rate M, the semi-saturation constant σ, i.e., the particular value of the input for which $S(\sigma) = \frac{1}{2}M$, and the constant factor ξ that determines the slope of S(P), i.e., how quickly it saturates. Aiming to resemble the time evolution of the firing rate FR, the response of the cells was restricted to the interval [0,1] by setting $M = 1$, and the semi-saturation constant $\sigma = \sigma_0$, with $\sigma_0 = 0.25 \cdot M$. The latter was chosen in order to prevent P from growing too fast and to avoid step-wise behavior of the activation function. The factor $\xi = 3$, is a heuristic parameter whose value for neurons in the visual cortex was previously reported by Wilson (1999). With this choice of values for all parameters we ensure that for $P = 1$, $S(P) = \frac{M}{0.25 \cdot M^{\xi}+1} \cong 0.98$; i.e. the reachable ceiling of the rate is not significantly attenuated irrespective of M (see **Figure 2A**). This represents a normalized and ideal scenario in which all impinging connections to a neuron are excitatory. Finally, τ represents the time constant of the activation and was set to $\tau = 10$ ms, thus satisfying the kinetics of gabaergic receptors such as GABAA, and matching the average duration of the post–inhibition refractory period (Whittington et al., 2000; van Aerde et al., 2009).

Similar to Rothenstein and Tsotsos (2014), we considered the effect of adaptation currents Ih on the ST-neuron, and incorporated them in the dynamic equation as additive factors that modulate the magnitude of the semi-saturation constant σ. The new σ(t) is then re-computed at every time-step using Equation (2) as follows:

$$\sigma(t) = \sigma_0 + f_{slow} \cdot H_{slow}(t) + f_{fast} \cdot H_{fast}(t) \tag{2}$$

where σ_0 is the original parameter. Adaptation currents consist of two different components H_{slow} and H_{fast}, each evolving within a particular time-scale, coupled to the value of the firing rate FR, and whose time course is scaled by the characteristic time constant τ_x with x being either *fast* or *slow*. In turn, f_{slow} and f_{fast} are the values of the amplitude for each contribution. The temporal evolution of the two components is given by the Equation (3):

$$\frac{dH(t)_{fast}}{dt} = \frac{1}{\tau_{fast}} \cdot \left(-H\left(t\right)_{fast} + FR\left(t\right)\right) \quad \text{and} \quad \frac{dH(t)_{slow}}{dt} = \frac{1}{\tau_{slow}} \cdot \left(-H\left(t\right)_{slow} + FR\left(t\right)\right) \tag{3}$$

Equations (1–3) are independently updated for each neuron at every time step ($\Delta t = 2$ ms) using a customized Runge-Kutta 4 algorithm implemented in MATLAB 2016a, (The MathWorks, Inc.). The original details of the implementation can be found in Wilson (1999).

Circuit Design and Connectivity

Following the original design by Reynolds et al. (1999), our circuit aims to represent a three tier structure, in which the response of the top-most unit quantifies the model's performance. The time course of this response was computed when the representations of two stimuli, each of which could be located either within or outside the cell's receptive field (RF), competed for representation (see **Figure 1C**). The bottom layer represented by two colored upwards arrows, contains the input representation. The Intermediate layer consists of two units, each accounting for the average response of individual populations (black ellipses) of ST-neurons, and are tuned to the stimulus directly below them. This level represents the activation of the populations at V1-V2 cortices. In turn, the neuron located at the top was defined as the main neuron (top circle). This unit represents a V4 cell, whose complex receptive field is able to process whole object representations.

Inputs at the bottom are represented by particular combinations of excitatory and inhibitory connection weights projected to the intermediate layer. Each intermediate population receives excitatory (red continuous arrows) and inhibitory connections (green dotted arrows) from the input, and project them to the top. The top unit receives both types of feed-forward inputs from the intermediate layer. **Figure 1B** shows a simplified version of the circuit in which a single stimulus is presented and processed. Connection weights were defined in the interval $[-1, 1]$, with the convention that w is inhibitory if $-1 \leq w < 0$, and excitatory if $0 \leq w \leq 1$. In consequence, any potential changes to the stimulus properties should be reflected as changes in the combination of connection weights representing it. During the time course of each simulation the set of excitatory and inhibitory connection weights from the intermediate layer onto the target (top) neuron remained fixed. Consistent with our assumptions, the representation of a given stimulus consisted of setting only the excitatory and inhibitory connection weights from the bottom to the intermediate layer. All other parameters were fixed within and across simulations, unless otherwise stated.

Neural Selectivity

Neural selectivity is the mechanism by which a neuron raises its firing rate when a stimulus has a certain feature matching its tuning curve. Thus, a preferred stimulus is one for which the neural selectivity is high. In order to incorporate selectivity into the circuit, and provided that neurons were connected through inhibitory and excitatory inputs with particular connection weights, we assumed for a preferred stimulus an excitatory (E) connection weight w_E belonging to the interval $0.75 < w_E \leq 1$, and consequently an inhibitory (I) weight $w_I = 1 - w_E$, belonging to $0 \leq w_I \leq 0.25$. In the case of a stimulus with low selectivity i.e., one for which the cell selectivity is low, the inhibitory weight approached $w_I = 1$ and the excitatory $w_E = 0$. For the sake of convenience, and bearing in mind that for the current normalized case the sum of weighted E and I inputs satisfies $\sum |w_I| \cdot I + |w_E| \cdot E = 1$, any stimuli with $0.7 \leq w_E \leq 0.75$ were considered as of neutral selectivity. Stimuli with $0.75 \leq w_E \leq 1$ were defined as preferred (or having high selectivity), and stimuli with $0.5 < w_E < 0.7$ were defined as non-preferred (or having low selectivity).

ST's Top-Down Attentional Signal

The attentional signal was implemented in consonance with the ST model, by creating a top-down branch-and-bound selection mechanism that picked the targets and suppressed the neural representation of the distractors, as described in Tsotsos (2011). The amplitude of the signal between belonged to the range [0, 1], and was computed like the absolute difference between the magnitude of the activation of the intermediate units, and the resulting factor was used to multiply the weights of the unit, associated to irrelevant input. This process has been fully described several times previously, most recently in Tsotsos (2011) and thus will not be repeated here.

RESULTS

Characterizing the ST Neuron Dynamics

In order to extend previous findings, we first characterized the time course of the neuron in relation to basic parameters, and then by modeling the response of the neuron after incorporating adaptation mechanisms, we evaluated their effect on the cell's firing dynamics during a set of simulated visual tasks.

In absence of adaptation mechanisms the activation of the ST-model neuron is determined by the two parameters σ and τ of the Naka-Rushton function (see equation 1. in section Methods). Although this function was first introduced in order to account for the adaptive saturation of photoreceptors to particular illumination conditions, its role in shaping the response of the ST-neuron was not previously addressed.

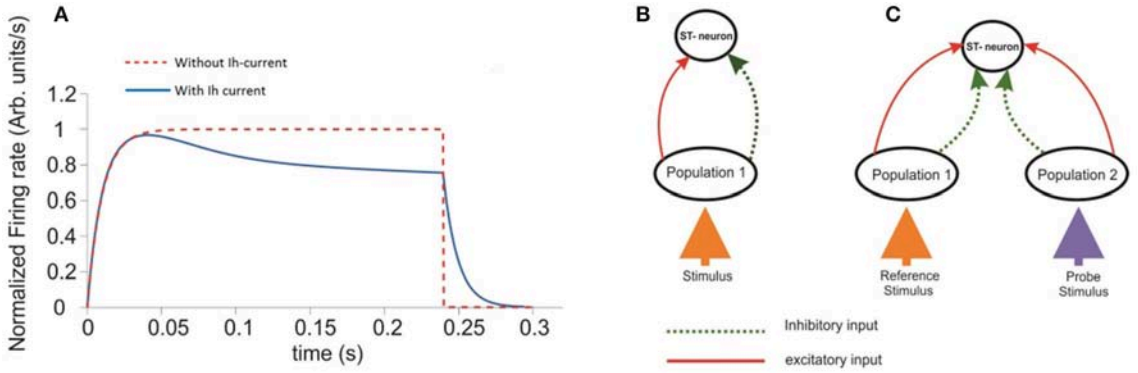


FIGURE 1 | Response of a single ST-Neuron to a fully preferred stimulus. **(A)** Activation of the neuron occurs after presenting the stimulus during a simulated attentional task with 300 ms total elapsed time. The red curve corresponds to the ST-neuron activation in the absence of the adaptation mechanism (Ih). The blue curve represents the same dynamics, when Ih currents are incorporated. **(B)** Schematic representation of the minimum circuit used to study the selectivity and attention aspects on the ST-neuron's response. The top unit represents a cell with a highly complex receptive field able to process abstract object representations. The unit at the bottom represents the average response of neurons selective for that stimulus. **(C)** An extended representation of the circuit used to model selection and attention. The diagram extends the circuit shown in **(B)**. Where each population (ellipses) has high selectivity for one of two incoming stimuli represented by the colored arrows.

As a two-step exercise we first fixed the value of τ and varied σ and then we flipped this, fixing σ while varying τ. In the first case, we assumed $M = 1.0$, and $\sigma = k \cdot M$, for $k = 0, 0.25, 0.5, 0.75$, and 1.0, obtaining the response curve shown in **Figure 2A**. Its shape followed a sigmoid pattern with amplitude of saturation (maxFR) proportional to the choice for σ, counterbalanced by P, and scaled by M (red curve in **Figure 1A**). Our simulations show that for every σ, the FR-profile saturated within the initial 50 ms. In the case of larger σ, any variation in k led to monotonic decrements of the saturation rate's magnitude (maxFR) (**Figures 2A,B**). The analytical relation was well described by the expression $maxFR = -0.54 \cdot \sigma^2 + 0.0076 \cdot \sigma$, with a resulting norm of residuals nr = 0.024696. This result suggests that in the limiting condition $\sigma \rightarrow 0$, the smaller the value of σ the closer maxFR is to M.

By fixing σ and varying τ within a biologically plausible range with $\tau = 0.0$, 5.0, 10.0, 15.0, and 20.0 ms rather than variations on maxFR, we observed significant effects on the timing required by the sub-saturation period (rising phase) to reach maxFR (see **Figures 2C,D**). In spite of the reasonable behavior of the model's output for $\tau \cong 10$–20 ms, we embraced experimental observations from previous studies (Jensen et al., 2005) choosing $\tau = 10$ ms, which on one hand accounts for an acceptable durations of the sub-saturation period of around 20 ms, and on the other coincides with the reported time constant of GABAergic synapses such as GABAA, aligning also with the idea that "..tonic inhibition in single neurons increases the firing threshold and reduces the membrane time constant …" (Hutt, 2012). In the case of τ shorter than 10 ms unrealistically fast saturation of the rate occurred, while for τ much larger than 20 ms, sub-saturation intervals were also extremely long. In general, the response of the model shows consistency with experimental findings (Kandel et al., 2000) deploying a relation between the duration of the time required for the firing rate to saturate, i.e., the sub-saturation period sSP and τ, given by the analytical expression $sSP = 130 \cdot \tau^2 + 6.6 \cdot \tau + 0.022$, with a norm of residuals $n = 0.00775$. Although the results for smaller τ's might reflect the action of other mechanisms, those do not necessarily represent the dynamics in the visual cortex (Cavelier et al., 2005).

A general result extracted from this simple analysis shows that far from interfering with one another, σ and τ control and modulate different parameters of the cell's activation, and their joint action reliably accounts for the efficacy of individual neurons to tune their firing to particular feature(s) of the synaptic representation of a certain stimulus.

Effects of the Adaptation Currents (Ih) on the Firing Rate of a Single Cell

An overall comparison between the FR-profile of the neuron without Ih and with Ih is depicted in **Figure 1A**. The stimulus onset occurred at $t = 0$ and the removal at $t = 250$ ms. Note the unaffected FR-profile's rising phase of the with-Ih scenario (blue trace) and the appreciable changes occurring during the post-saturation of the with-Ih case compared to the non-Ih case (red trace). As in Rothenstein and Tsotsos (2014) Ih currents are represented by the linear combination of a slow (H_s) and a fast (H_f) component, whose time courses are depicted in **Figure 2E** by the blue and purple traces respectively. The modulation imposed on the constant σ (yellow trace on top) shows a periodic signal that slowly raises from σ_0 to its maximum within ∼130 ms, and exponentially decays within a comparable interval (∼120 ms). As previously mentioned, the FR's rising phase remains unaffected and the overall effect is constrained to its post-saturation phase in a two step process (see **Figures 2E,F**): In the first, during a transitory interval (∼50 ms), the firing rate is driven by the activation of the Ih's fast component H_{fast}, leading the FR-profile to rapidly decay to ∼70–80% of its maximum (maxFR). In the second, and due to H_{fast} having reached its maximum, the slow activation of H_{slow} takes over the control and reduces the speed of the FR decay, leading to a pseudo-plateau in the FR-profile, in which, in absence of any further changes in the stimulus, the FR remains constant.

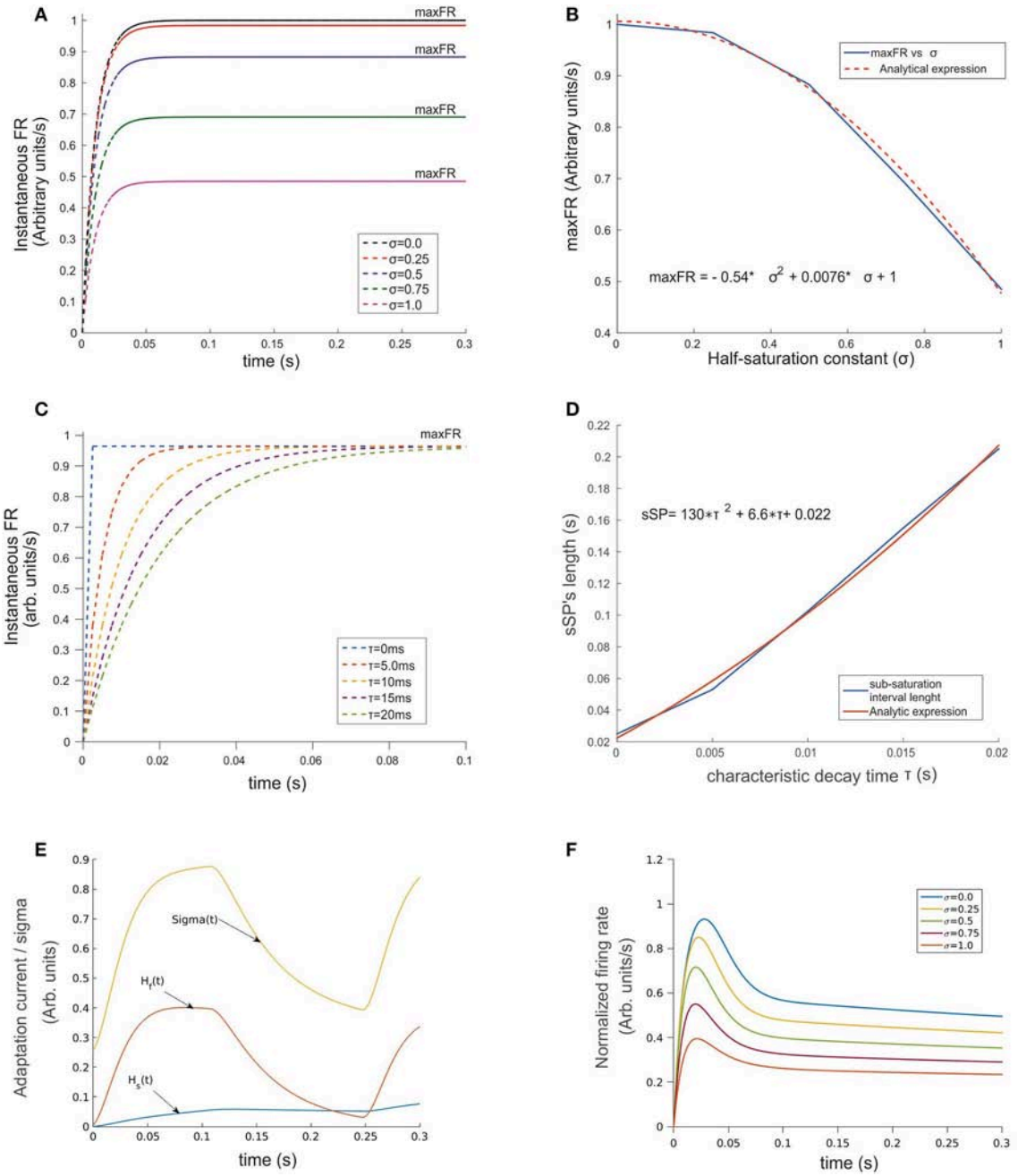


FIGURE 2 | Temporal evolution of the ST-neuron's firing rate. **(A)** For a constant input, the amplitude of the firing rate has a transitory pre-saturation period which is independent of the half saturation constant σ. However, after this point and depending on its magnitude, increasing σ led in a minor or major proportion the saturation rate of the cell to fall and reach smaller maxFRs. **(B)** Analytical expression of the relation between variables depicted in **(A)** firing rate and σ are related through a quadratic function for which small values of σ near 0 rapidly makes maxFR ≅ M. **(C)** A similar relation rules the effect of τ on the time required by the firing rate to saturate when σ was kept fixed. The simulation shows strong modulation before the 100 ms point of each simulation. In spite of maxFR remaining unchanged, the duration of the sub-saturation period increased proportionally to τ following the trend plotted in **(D).** A representation of the temporal pattern for the fast (H_f) and slow (H_S) components of the Ih-current is shown in **(E)**. The combined effect of the two components modulates the firing rate by adding temporal dependence to σ (see Equation 2 section Methods), whose dynamics is represented by the top trace in **(E)**. The response of the top cell in **(F)** shows the effect on the FR-profile when submitted to the action of the synaptic inputs and the activation of Ih. Here the values of σ are identical to **(A)**. Note in the latter the decaying post-saturation profile and the generation of bumps before reaching the stationary firing regime.

Response of ST-Neuron (With Ih) to Stimuli With Different Selectivity

To run this set of experiments we initially assumed attention not to be directed to the stimuli; thus the time course of the FR-profile only depended on the neuron's selectivity to a given stimulus. We simulated various (uniquely defined) types of inputs with selectivity being accounted for by the relative contribution of the inhibitory and excitatory connections.

In each experiment a given pair of stimuli was shown as input to the circuit of **Figure 1C** (for details see section Methods). To maintain consistency with psychophysical studies, we refer to the first stimulus as the reference, whose onset time occurred at $t = 0$ ms and its removal at $t = \mathrm{t}'$ with $\mathrm{t}' > 0$, coinciding with the onset of the second stimulus that remained active until the end of the simulation and was denoted by the probe. The time $t = \mathrm{t}'$ was designated as the switching time. In addition, the processing of each stimulus activated only one of the intermediate populations, and the probe stayed active until the end of the simulation, whose total duration of 300 ms was considered to be long enough to allow input-related information to propagate from the bottom to the top neuron (target).

Figure 3 shows the FR-profile's time course of the top neuron being initially driven by the reference, whose rising phase remained unaltered irrespective of how early t′ occurred, while being significantly affected on its post-saturation period in two ways. First, a latency appeared, caused by the decay of the initial FR and second, a sudden rebound appeared with maxFR depending on the probe alone. During the latency, and as an effect of switching inputs, the FR-profile became unstable leading to a transient drop and catch phase characterized by a discontinuous change of concavity and followed by a fast regain of firing. Once the FR surpassed maxFR due to the cell being engaged to the probe, the profile decays following the dynamics described in the previous section, with a pseudo-stationary state being ruled by the slow Ih's component. In every experiment a neutral reference i.e. excitatory synaptic weight $W_{E-ref} = 0.7$ (blue continuous trace) systematically preceded the probe, each of which had identical ($W_{E-p} = 0.7$), larger ($W_{E-p} = 0.75$, 0.80) or smaller selectivity ($W_{E-p} = 0.65$, 0.60, 0.55) than the reference. While the larger probes led to steeper jumps in the firing rate and bumps characterized by large maxFRs, stimuli with lower selectivity led to an even faster decay of the FR. The stationary response always equated the stationary response evoked by the probe in the absence of other inputs. Note that probes with identical selectivity to the reference did not align with the expected smooth profile evoked by the reference. An explanation to this is that the original tuning (i.e., the combination of weights) of the intermediate unit processing the probe was different from that of the reference and in consequence led to small bumps in the model (see purple traces in **Figure 3**) Measuring the plausibility of this effect needs further study and is left as an interesting open research point.

In general, the distortion in the reference's FR-profile was easier to recognize for probes presented briefly after the reference's onset i.e., t′ less than 200 ms. This result was consistent irrespective of the probe's selectivity (compare the shapes of the profiles in 3.A-3.C against those in 3.D-3.F). Note that in the case of a late t′, the transitory state did not interfere with the original time course of the FR-profile, but took place once the cell was close to the FR-profile's plateau, which could be interpreted as the replication of the original activity, but now due to the probe and with a different base rate.

Concerning the latency, our results show that for probes with less selectivity than the reference, the firing dropped and slowly recovered producing a smooth trough in the FR-profile, whose depth and width specifically depended on the relative difference of selectivity between both stimuli, being wider for less preferred probes, while in the case of probes with larger selectivity than the reference the width of the trough was negligible, and the FR-profile discontinuously lost and regained firing after switching stimuli. In general the particular shape and steepness of the bumps depended on the relative selectivity of the reference and probe, and once the transition occurred the rate slowly tended to stabilize around the stationary state evoked by the probe.

Adding the Probe to the Reference Modulates FR-Profile but Induces No Latencies

As a second scenario, instead of switching stimuli at t′, we modeled a condition in which the probe was added to the reference, while computing the time course of the top neuron's FR-profile (**Figure 4**). We ran the experiment for different probe selectivity and onset times t′ as follows: $W_{E-p} = 0.55$, 0. 60, 0.65, 0.70, 0.75, 0.80 (recalling that $W_{I-x} = 1\text{-}W_{E-x}$ with $x =$ ref or p), using a neutral reference (i.e., $W_{E-ref} = 0.7$) presented at $t = 0$. The FR-profiles in **Figure 4** show that in contrast to the previous case (see **Figure 3**), and in the absence of attention, adding the probe at $t = \mathrm{t}'$ produced no decaying latencies. Furthermore, probes with larger selectivity than the reference induced almost instantaneous rebounding bumps but in this case the amplitude of maxFR for the two stimuli never reached that of the reference alone, while less preferred probes led to a sudden drop followed by a less frequent but sustained and regular firing of the cell. Without exception for all probes, the value of maxFR was fixed across each of the diagram showing $\mathrm{t}' = 50$, 100, 150 ms. In contrast, for $\mathrm{t}' > 150$ ms, i.e., $\mathrm{t}' = 200$, 250, and 300 the amplitude of the maxFR for more preferred probes equated that of the reference alone, while for the less preferred it got closer to zero for late t′ followed by a smooth recovery with low but sustained firing.

In all cases the transient phases were followed by a recovery leading to a stationary rate. Since the sharp rebounding/dropping effect was a direct result of the presence of Ih and of the cell modulating its selectivity due to the probe being added, we hypothesize that as a result of trial and error such a change of concavity (inflection point in the first time derivative) may be utilized as a suitable selection cue to predict the stimulus' category. In particular, the computation of the instantaneous (not the average) derivative satisfies that requirement, and only demands local adaptation of the cell's firing.

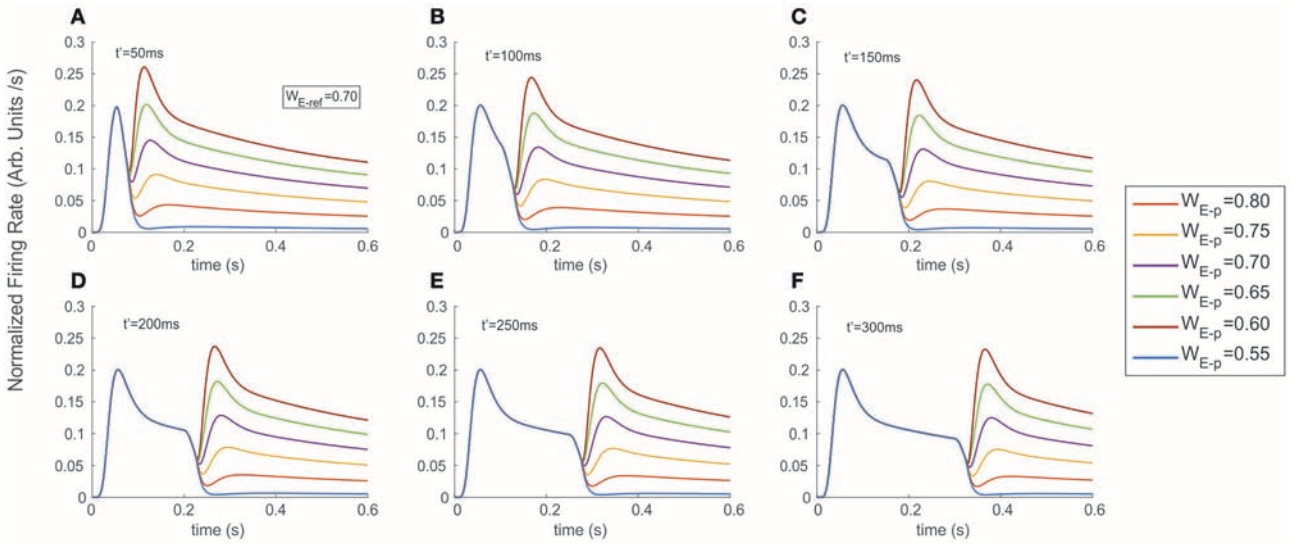


FIGURE 3 | Stimulus exchange leads to strong discontinuities and transients in the FR-profile. Experiments were run simulating a fixed interval of 600 ms, and exchanging the stimulus at **(A)** $t' = 50$ ms, **(B)** $t' = 100$ ms, **(C)** $t' = 150$ ms, **(D)** $t' = 200$ ms, **(E)** $t' = 250$ ms, and **(F)** $t' = 300$ ms. Colored traces indicate the probe's selectivity characterized by the excitatory weight WE-p (refer to labels in Methods for details). Switching from a neutral reference ($W_{E-ref} = 0.7$) to a probe with larger or smaller selectivity created unstable surges of firing, followed by a stationary state. Note that in the case of a late t', the transitory state did not interfere with the original time course of the FR-profile, but took place after the cell's recovery and near to the FR-profile's plateau.

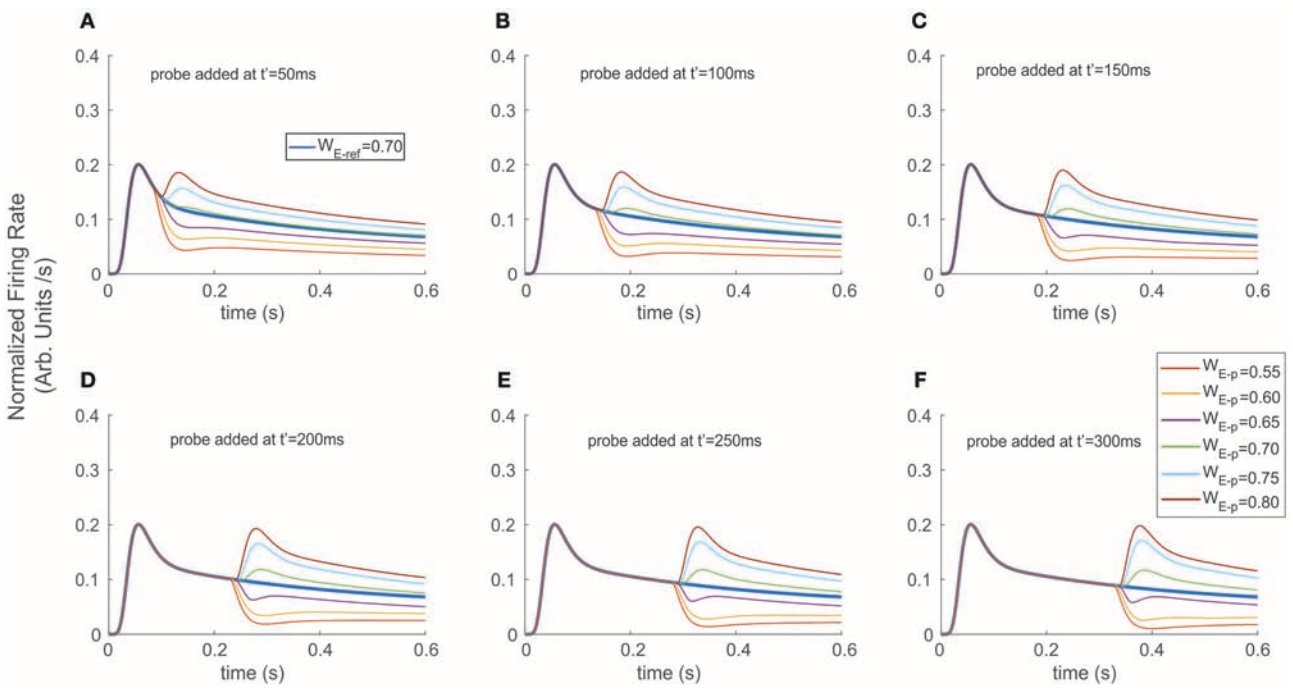


FIGURE 4 | Adding a probe to the reference destabilized and induced transients on the firing rate. The reference stimulus was presented at $t = 0$ ms ($W_{E-ref} = 0.7$), and different probes were added at **(A)** $t' = 50$ ms, **(B)** $t' = 100$ ms, **(C)** $t' = 150$ ms, **(D)** $t' = 200$ ms, **(E)** $t' = 250$ ms, and **(F)** $t' = 300$ ms. Similar to the exchange experiment, transient bumps/troughs indicated sharp variations in the FR-profile. However, the shape and amplitude ratios between the principal and secondary peaks depended on the probe's addition time t', for the case of probes with larger relative selectivity than the reference (see secondary bumps in **A–F**). In the case of probes with less selectivity, the transients exhibited variable concavities and lengths, thus led to cell responses with significantly reduced and more unstable firing rates (e.g., purple traces).

Comparing Selectivity Results in the Model With Experimental Findings

In a previous work on visual selection and color perception Fallah et al. (2007), measured the response of single neurons to a set of stimuli falling within its RF. Cells were located in the V4 extrastriate visual cortex in primates, and tuned to a particular hue. The animal was first exposed to a stimulus at $t = 0$, and at $t = t'$ a second with different hue structure was added. The recordings show a reshaping of the FR-profile in proportion to the relative match between the hue of the stimuli and the selectivity (selectivity) of the cell, producing FR patterns close to those shown in **Figure 4**, and depicted in **Figure 5A**. Ferrera et al. reported similar *in-vivo* dynamic while recording from cells in areas 7a, MT and V4 (Ferrera et al., 1994). Even though in both studies the outcome of the experiments clearly reflects correlations between the cell's response and feature-related information of the stimulus, the responsible mechanism was not characterized.

In order to explore the plausibility of the ST-cell dynamics with Ih in explaining those results, we implemented a high level simulation of Fallah's experiment using the circuit from **Figure 1C** The neutral reference ($W_{E-p} = 0.7$) was presented at $t = 0$ and a probe with larger or smaller selectivity was added at $t' = 300$ ms. As a first confirmation of the model's efficacy, we observed that when starting with a neutral reference, the addition of more preferred probe ($W_{E-p} = 0.80$) induced a sharp increase in the FR and a bump with similar characteristics to the effects described in the previous section for probes with selectivity larger than the reference (compare blue traces in **Figures 5A,B**). In turn, a less preferred probe ($W_{E-p} = 0.6$) led to a drop and stabilization of the FR-profile (see red traces in **Figures 5A,B**).

In spite of the qualitative similarities between simulations and experiment, once the second stimulus is added, the experiment shows a brief period of non-responsiveness prior to a sharp modulation of firing which is underestimated in the model, but not necessarily as its flaw.

Since the biological problem suggests that for a particular combination of inputs, the neuron activation remains close to the resting state, the cell may react either by raising its firing, whenever the threshold is reached (generating a silent period of non-sensitive change), or by getting hyperpolarized and in consequence reducing the firing, which does not demand a threshold crossing and in consequence, no insensitive periods are required. Thus, we believe this is an aspect that needs further analysis and to account for the result, experiments using a broader range of selectivities need to be considered in a future study, together with further computational exploration.

Effects of Attention on the FR-Profile

The most interesting aspect concerning the ST-characterization regards its behavior during attentional tasks. In this section we examine the extent to which attention could or not modulate the dynamics of the cell's selectivity.

As proposed by the Selective Tuning model (Tsotsos, 1990, 2011; Tsotsos et al., 1995), allocating/engaging attention in the model corresponds to the activation of the selection mechanism. Such mechanism was represented by a top down control signal responsible for suppressing information associated to irrelevant stimuli, while keeping unaffected the connections between the cells that processed information related to the attended stimulus in a task-dependent manner. We quantified the suppressive signal by computing the absolute difference between the weighted inputs impinging the top neuron, and used it to multiply the weight of the inputs from the unattended stimulus (see section Methods). This approach has proven to be fast and accurate at disambiguating stimuli, since rather than adding up the weighted contribution of all incoming signals, allows single neurons to efficiently filter them out and focus on the relevant ones. This idea is supported a key observation by Martinez-Trujillo et al., (Martinez-Trujillo and Treue, 2004; Khayat et al., 2010) according to which attention modulates the input to a given neuron instead of its direct response.

Using the circuit in **Figure 1C**, we studied the response of the top neuron when the reference and the probe were presented in isolation and simultaneously. In addition, to track possible variations in the stationary state, the attentional signal remained active until the end of the simulated period.

In agreement with real experiments, and regardless of the amount of selectivity associated to each, when two stimuli of different selectivity were exposed to the scrutiny of the top neuron, the average behavior of the FR-profile fell in between those evoked by each stimulus in isolation; see **Figures 6A,D**. However, in the case of stimuli being simultaneously presented, a late engagement of attention to one of them modulated the cell's FR and forced it to adjust it to the magnitude evoked by the attended stimulus regardless of its selectivity, consistent with the theory (Martinez-Trujillo and Treue, 2004). The behavior is shown in **Figures 6B,C**, where the neutral reference ($W_{E-ref} = 0.7$) and the probe with less selectivity ($W_{E-p} = 0.6$) were both located inside the classical receptive field of the top neuron and simultaneously presented at $t = 0$ ms. When attention was allocated at $t' = 50$, 100, 200, 400, and 600 ms, the FR rose or dropped accordingly to what stimulus was attended. Similar effects were obtained when the selectivity of the probe ($W_{E-p} = 0.8$) was larger than that of the reference, as shown in **Figures 6E,F**.

Irrespective of what stimuli was considered reference or probe, engaging attention to that of larger selectivity led the FR-profile to generate larger bumps (maxFR) than those observed for the attention away condition (dashed traces in **Figures 6C,E**); and FR with magnitude similar to the FR evoked by the largest stimulus in isolation. On the other hand, engaging attention to the stimulus with less selectivity produced FR-profiles characterized by troughs initiated at t'. In the case of **Figure 6B** the depth of the transient was more profound than in the case of the traces in **Figure 6F**, although in both cases the stationary response of the FR-profile coincided with that of the stimulus with less selectivity for the attention-away condition.

Comparing the Effect of Attention in the ST-Neuron With Experimental Recordings

Figures 7A,B correspond to the simulated conditions in which attention was either engaged to the reference with less selectivity (**Figure 7A**) or not allocated at all (**Figure 7B**). Interestingly, the resulting FR-profile in the first case shows a masking effect of attention that, in spite of a probe having larger selectivity than

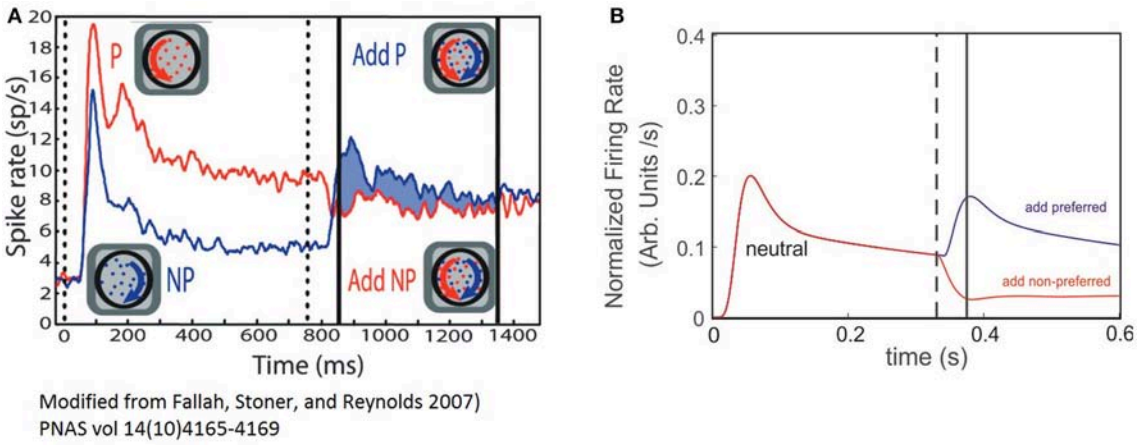


FIGURE 5 | The ST model cell FR-profile reproduce experimental observations of V4 neurons. **(A)** Experimental firing rate computed on the population's activity of V4 neurons of primates, adapted from Fallah et al. (2007). The vertical dotted line indicates stimulus appearance, and the continuous black lines the period over which modulation of the response was computed. The red trace indicates the population response for the "preferred" (P) stimulus alone followed by the addition of the non-preferred (NP); while the blue trace indicates the non-preferred alone, followed by the addition of the preferred. **(B)** Simulated experiment. Both traces represent the response of the neuron when a neutral stimulus was presented followed by the addition of the preferred stimulus (red trace), or the non-preferred one (blue trace). The dashed line indicates the time at which the probe addition occurred and the continuous line the time of the transient's peak.

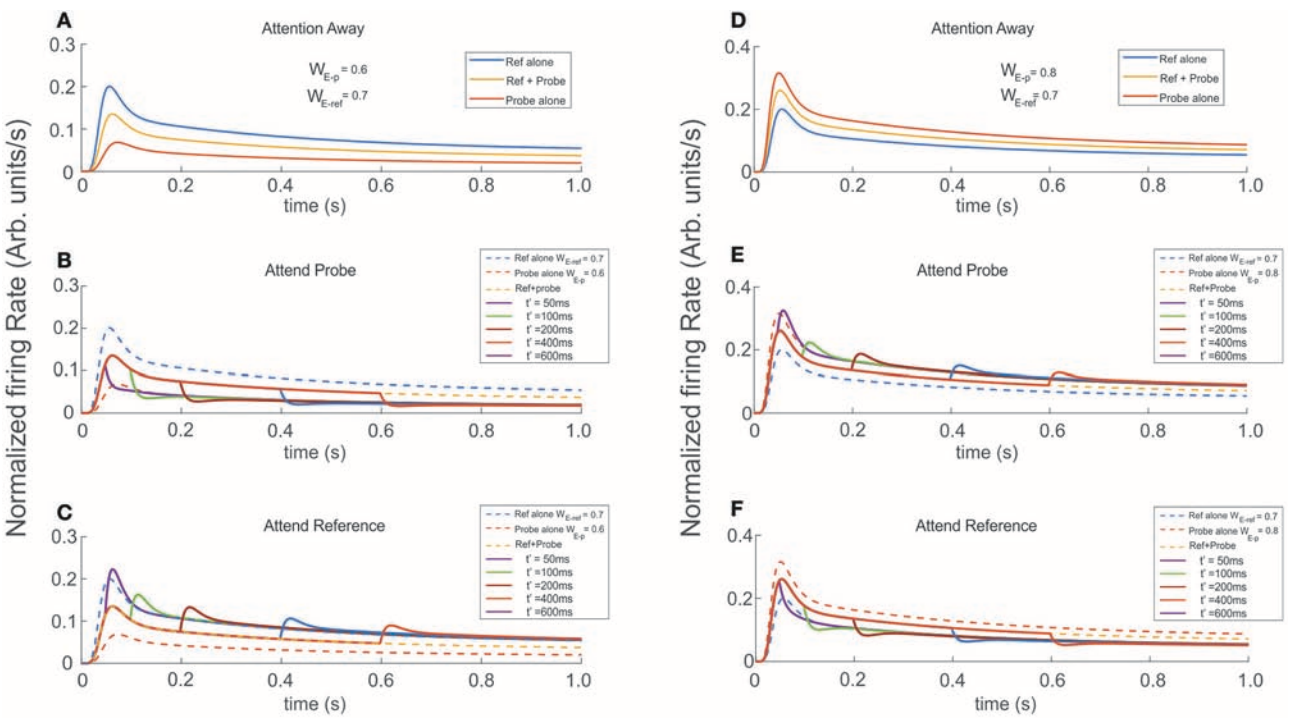


FIGURE 6 | Engaging attention modulates the transients, and modifies the amplitude of the stationary response. Reference and probe stimuli were simultaneously presented and attended as indicated for each trace (see labels). The stimuli were presented at t = 0 ms and attention was engaged at t′ = 50, 100, 200, 400, 600 ms after stimulus presentation. **(A,D)** Show the reference and probe stimuli presented in isolation (blue and red traces) and simultaneously (yellow traces). In the first scenario, the reference is stronger and in the second the probe is stronger (higher selectivity). As expected the cell's selectivity mechanism produced firing rates with well differentiated maxFR's, each proportional to the respective selectivity of the stimulus. In addition, simultaneous reference and probe presentation, led to FR-profiles with intermediate amplitudes. Experiments were run for attention oriented to the probe **(B,E)**, and attention oriented to the reference **(C,F)**. Besides the characteristic transients, directing attention to the probe shifted the tail of the fr-baselines to the profile produced by the probe alone. Attending the reference produced similar effect on the fr-baseline, shifting in this case the tails of the response toward the reference alone FR-profile. Those long rate responses were consistent and irrespective of the relative selectivity between the reference and the probe.

the reference, the FR gets modestly disrupted, remaining locked to the FR-profile of the reference. It contrasts the effect observed for the attention-away condition, in which the selectivity led the cell to rapidly increase the FR and adjust the FR-profile, matching that evoked by the probe alone, in this case with larger selectivity.

In an experimental study Luck et al. (1997) measured single cell responses of neurons located at V4 associated to the appearance of a particular target. Stimuli were defined as *effective* or *ineffective* on a selectivity basis. In their protocol a series of trials consisted in presenting sequentially/simultaneously pairs of simple stimuli characterized by color and orientation, which could be both inside the cell's receptive field, or one inside and the other outside it, and attention was deployed to one of the two regions. For further details please refer to Luck et al. (1997). By comparing our results with those experimental recordings (**Figures 7A,C** respectively), the simulation shows good agreement, not only in the shape, but also in the time course of the FR-profile. In contrast to the condition observed in those figures, **Figure 7B** shows that in the absence of attention (attention away condition) there is no masking at all of the scene, and any probe stimulus with larger selectivity than the reference will draw the largest part of the cell response when both stimuli are located inside the RF. As in the experiment, **Figure 7A** shows the response of the top neuron after presenting the reference and probe simultaneously at $t = 0$, and the attentional mechanism is deployed at $t = t'$. Both simulation (**Figure 7A**) and experiment **Figure 7C** are characterized by a small modulatory dent in the cell's FR-profile while attending a less selective reference. The match between model and experiment suggests that in effect from the model's perspective, Ih makes the neuron highly sensitive to the effects of attention on selectivity (recall that in the absence of Ih the cell reached saturation, and the FR couldn't be modulated, see red trace in **Figure 1A**), but also from the biological perspective, the model suggests that attention and selection compete for resources when stimuli with low selectivity are attended. However, as it will be discussed later, the results in **Figure 8** show that collaborative enhancement is also possible.

Attention Competes Against or Reinforces Neural-Selectivity

In our final experimental design, we ran simulations in which the reference was presented at $t = 0$ and the probe was presented and attended at t' = 50, 100, 200, 400, and 600 ms. Probes had either larger or smaller selectivity than the reference. In the attention away condition, a probe with less selectivity than the reference produced a decaying FR-profile characterized by shallow troughs and durations of the transient close to 150 ms, followed by a slow recovery of the FR in the direction of the stationary state (**Figure 8D**). In the same condition, probes with larger selectivity than the reference created rebounding firing rates with increasing amplitude, especially for late stimulus onset t'.

Running the same set of experiments while attention was allocated to the probe at t' simultaneously with the probe's presentation, shows that attention has an ambiguous effect depending on whether the transient or the stationary dynamics of the cell's response were analyzed. As reference, **Figures 8A,B** show the FR-profile of the ST-model neuron in the attention away condition. All traces show that consistent with previous studies (Martinez-Trujillo and Treue, 2004), and based on its selectivity, the cell has larger maxFR for a more preferred stimulus and vice versa, while when the pair is active, the response always falls in between the FR-profile of the other two.

The effect of the selection mechanism of attention seemed to have a transitory component characterized by reinforcement of the cell's selectivity, while in the long term its behavior turned competitive. Although the affirmation may look contradictory, a careful check of **Figures 8B,C** shows that although the depth of the trough is larger for the attend-to-probe scenario, suggesting a steeper reduction of the FR (inhibition's reinforcement), the cell's response to the same onsets of the probe (indicated by traces of the same color in both figures) also corresponds to shorter widths (duration) of the trough in the attend-to-probe condition. In turn, when the FR was restored, the FR-profile matched that of the probe alone, in contrast to the attention-away condition (**Figure 8B**), in which the stationary state matched the FR-profile of the pair.

Interestingly, when a probe with larger selectivity than the reference was presented, it resulted in the opposite response of the neuron. A comparision between individual colored traces in **Figures 8E,F** shows that due to its large selectivity, a bump in the FR-profile occurred almost after the probe's onset in the attention away scenario, and that its magnitude increased by increasing the delay t' between the onset of the reference and the probe, in a non-linear fashion (see bumps in **Figure 8E**). In the stationary state the solely effect of selectivity led the cell's FR-profile to match the response evoked by the pair.

In contrast, when the attentional mechanism was turned on while presenting the probe, the reduction in firing was represented by a deep and short trough characterizing the transitory response, exhibiting a duration of around 20 ms, similar for all t', and depth with magnitude near to 20% of the maximum FR, except for $t' = 50$ ms, (close to 30%).

This period that we called "latency," preceded a bump in the FR-profile whose peak FR, was similar for most t', and in general larger than the maxFR of the cell obtained when the pair was active, as shown in **Figure 8F**. Consistent with the case of the troughs, the peak of the bump for $t' = 50$ ms was also slightly larger than for any other t', suggesting that a short delay between the probe's onset and the activation of the attentional mechanism eases the processing of the stimulus of interest. Regarding the stationary response, we found the engagement of the FR to the response obtained when the probe was presented alone, in contrast to the attention away scenario, in which the FR was engaged to the FR-profile of the pair (see **Figures 8E,F**). It is important to note that in all simulations we implemented the selection mechanism of attention proposed by the ST model, which is based on inhibition of non-relevant inputs. In an earlier work by Busse et al. (2008), shifting attention from a cue located outside or inside of an MT cell's receptive to a probe in the opposite region was preceded by a drop in the firing rate of the cell. Authors claimed that the "short-latency decrease of responses" was caused by an interruption of endogenous attention, due to focusing

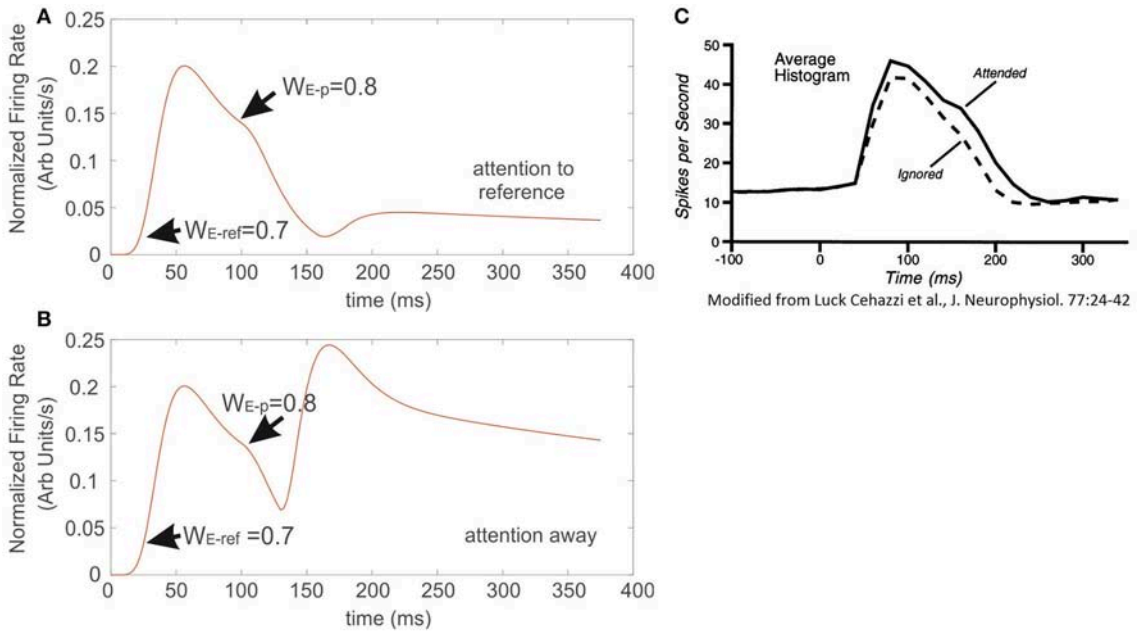


FIGURE 7 | When attention competes against the effect of neural selectivity, it occludes the latter. Sequential presentation of stimuli with different selectivity inside the receptive field of the model neuron. **(A)** The reference stimulus is presented and attended at t=0, once the stimulus switches to another with larger selectivity, but outside the still attended spot, the incoming stimulus barely altered the instantaneous firing of the cell, and only produced a negligible bump on the FR-profile at the exchange time. Condition **(B)** represents an identical setup to the one described in **(A)**, but in this case attention was not engaged to any of the stimuli. The absence of attention forced the firing of the cell to get immediately locked to the incoming input. The traces in **(C)** were adapted from Luck et al. (1997). They show that adding a probe with high selectivity to the receptive field of the top cell, while attending the reference also in the same receptive field, barely affects the ongoing response evoked by the reference.

on a stimulus that delayed the expected response toward the target.

By restricting our analysis to the case in which attention switches from the outside to the stimulus in the inside (red trace in **Figure 9A**), similar to the Busse et al. experiment, our findings show a two-step process: first a drastic drop in the FR, and second, the steep recovery of firing that precedes a bump. It validates our observation that when a cell is initially active due to a cue with certain selectivity, attention leads the single cell's response to a brief interruption in the FR, represented by short and deep troughs in the FR-profile, regardless of the selectivity of a second stimulus; and to recover the FR following a time course whose shape (**Figure 9A**) is closely resembled by the model, as depicted by the red traces in **Figures 9B,C**. In our simulations the circuit in **Figure 1C** was initially exposed to the effect of a neutral reference ($W_{E-p} = 0.7$) and at $t = t'$ a probe with more/less selectivity was added to the cell's receptive field and attended. The model predicts a deeper trough for the preferred probe ($W_{E-p} = 0.8$) (**Figure 9B**) than for a non-preferred probe ($W_{E-p} = 0.6$) (**Figure 9C**), and both latencies having similar duration. However, additional experiments are required for a solid validation of this point. The study also suggests that the intention of switching attention generates a similar effect (black trace in **Figure 9A**), but because that there is no optimal way to simulate the intention of switching attention in the model, we represented that condition by leaving the reference stay during the whole simulation (see black traces in **Figures 9B,C**).

DISCUSSION

Attention is responsible for modulating the amount of input received by a neuron from the stimulus in its RF. In order to quantify the nature and magnitude of this modulatory effect, earlier studies (Pestilli et al., 2007) have reported significant correlation between attention and the dynamics of the threshold and contrast sensitivity processed single neurons, supporting some of their claims on the results of computational studies like the biased competition (Reynolds et al., 1999) and the multiplicative response gain model, that endow attention with an enhancment role of single neuron's activity (McAdams and Maunsell, 1999; Williford and Maunsell, 2006). In a theoretical study, Ladenbauer et al. (2014) presented a description of the effects of adaptation mechanisms, on the single cell's firing rate, highlighting a major influence on the gain of firing and threshold modulation, that agrees with the idea that external inhibitory synaptic inputs are relevant modulators of the input-output curve of single neurons.

A second intriguing element concerns the eventual generation of transients (bumps and troughs) in the firing rate of single cells (Martinez-Trujillo and Treue, 2004; Fallah et al., 2007; Busse et al., 2008), when a rapid stimulus switch takes place during attentional tasks, and that this particular response is due to suppression of irrelevant stimuli as previously posed by Lennert and Martinez-Trujillo (2011). In an earlier paper, Tsotsos (1990) first predicted such behavior, suggesting that inhibition of

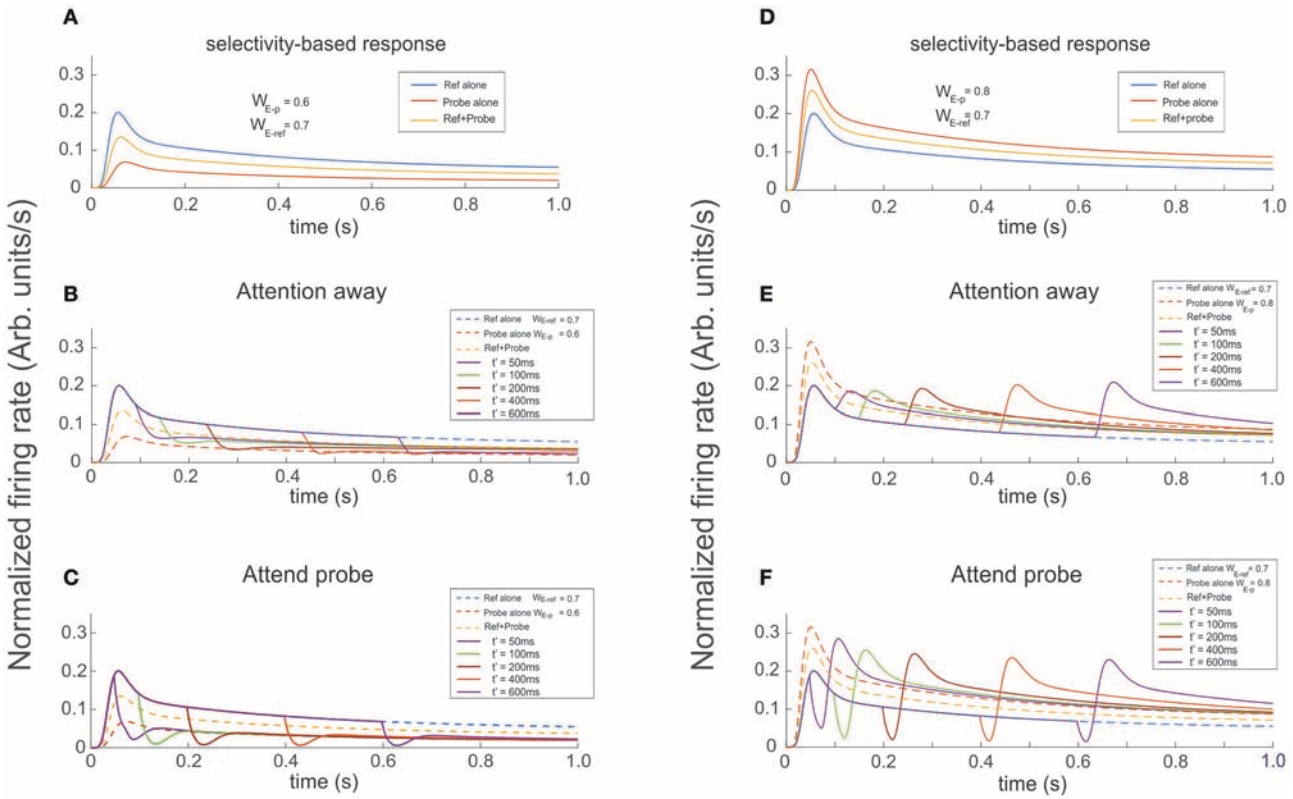


FIGURE 8 | Presenting and attending a probe modulates the cell's selectivity effect. The reference was presented at $t = 0$ and the probe presented and attended as indicated for each color trace at $t' = 50, 100, 200, 400, 600$ ms (see labels). **(A,D)** Show the firing rate for the stimuli presented in isolation (reference -blue, probe -red), and simultaneously (yellow trace). The three curves are also shown as dashed lines in **(B,C,E,F)** when the probe was added to the reference (both located inside the cell's receptive field) it has the effect to increase or reduce the cell's firing according to the cell's selectivity for the probe. In the attention away scenarios **(B,E)** the sole effect of selectivity, characterized by transients and baseline shifts, was observed. When Attention was engaged to the probe at t', as shown in **(C,F)** it induced the occurrence of large transients with sharp changes of concavity, whose magnitude significantly depended on the respective cell's selectivity to the probe, relative to the response in the attention away scenarios. In addition, the magnitude of maxFR in the rebounding conditions were in average 30% larger, with slower decay times and tails shifting toward the FR-profile of the probe alone for more preferred probes, and toward the curve of the reference alone for probes with less selectivity, while in the attention away scenario the stationary response converged toward the profile evoked by both stimuli simultaneously presented.

distractors allows the target neuron to restore its firing rate to the level evoked by the attended stimulus in isolation.

In this study we presented a revisited version of the ST neuron model, and characterized the effect on the firing rate of incorporating adaptation currents (Ih) into its dynamic equation, quantifying the neuron's response when submitted to various simulated experiments. We also strengthen the results of Rothenstein and Tsotsos (2014) describing the capabilities of the ST-neuron in reproducing experimental FR-profiles observed in simple attentional tasks, by separating the effects related to the cell's selectivity when Ih currents were active, from those related to attention. To our knowledge, this is the first time that adaptation current mechanisms are combined with an inhibition based model of the top-down attentive signal, to study the response of neurons in the visual cortex during attentive states.

With regard to the ST-neuron characterization, we found that in the absence of further mechanisms, the time course of the firing rate was driven by the balance between the constant σ of the Naka-Rushton term and the characteristic decay time of the inhibitory inputs. In turn, the modulation provided by Ih (depicted in **Figures 1A**, **2F**) determined the existence of two regions in the FR-profile: the first quantifying the variability of the initial FR activation, and the second the post-saturation effect. Using a similar circuit to the originally proposed by Reynolds, we simulated the activation of V4 neurons, showing that selectivity creates a strong differentiation between patterns of response (FR-profiles), each possessing a unique maxFR (peak FR) and a stationary rate, correlated to the relevance of the input for the neuron. As an important aspect, the obtained FR-profile could be linked to different features of the stimulus or even to the whole stimulus (as in the case of V4 neurons) being represented not only by variations in the contrast or firing threshold.

The biological plausibility of the ST-neuron proved to be successful at reproducing different experimental scenarios, by only modulating the relation between inputs weights representing each stimulus. Our simulation of Fallah et al. experiment (Fallah et al., 2007), highlights the modulatory effect of Ih to reshape the FR, when responding to stimuli

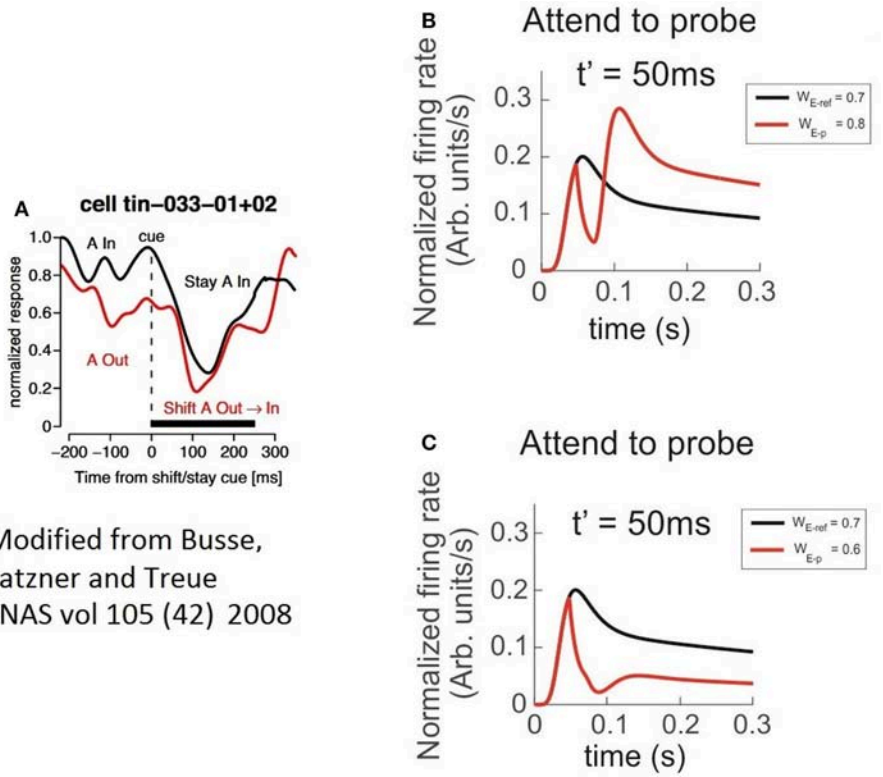


FIGURE 9 | The characteristics of latencies preceding the attentional rebound depends on the relative selectivity between reference and probe. In the figure, a shift of attention occurs when a given cue (represented by a vertical dashed line) indicates the subject to attend outside of the receptive field of the measured neuron. **(A)** The red trace shows the neuron's response when the cue indicates the shift of the focus of attention from the outside to the inside, while the black trace reflects the response when the cue instructs attention to remain in the outside of the receptive field. Adapted from Busse et al. (2008). **(B)** Depicts the response of the top neuron (in the circuit of **Figure 1C**) to an initially non-attended neutral reference with onset at $t = 0$, while a preferred probe ($W_{E-ref} = 0.7$) is presented and attention is engaged at $t = t'$. In this case, the FR shows a short and sharp transient trough followed by a rebounding bump that engages the stationary FR-profile of the probe alone. **(C)** Represents a similar condition to **(B)** when using a less preferred probe than the reference. In spite of a significant reduction in the peak during the rebound, the stationary cell's response remains engaged to the FR-profile of the probe alone. The black traces in **(A,B)**, denote no shift of attention, cases in which the probe was absent and the reference remained in place until the end of the simulated period.

with significant differences in the selectivity in the absence of attention. Although the model predicts changes in the transitory state of the FR, further experiments are required to verify the prediction.

The significance of the Ih dynamics proved its relevance also in more complex scenarios that included activation of the attentional signal. As described in Results, we showed that by incorporating the selection mechanism of attention proposed by Tsotsos (1990), the FR-profile resembled the response of real V4 neurons, and that by using Reynold's design (**Figure 1C**), as seen in **Figure 7**. A no enhancement is necessary to account for the time course of the firing rate when stimuli with different levels of neural selectivity are presented in isolation or simultaneously. Furthermore, our simulations show that by including the activation of the attentional mechanism, the FR was able to differentially represent possible conditions for the onsets of attention, or its shift in a non-redundant way, for different experimental designs, regardless of how similar can be the stimuli. In this scenario we show the interplay between selectivity and attention (**Figures 6A,B**) is crucial to define the dynamics of the FR when two stimuli suddenly switch with each other, affecting both the transitory and the stationary phases of the FR-profile. We predict the existence of a dual role played by attention, in which it can enhance or compete against selectivity during the transitory stage, and the opposite during the stationary stage, depending on how preferred each stimulus is for the neuron. The plausibility of our results is strongly backed up by the significant resemblance obtained by simulating the Luck et al. (1997), and Busse et al., experiments (Busse et al., 2008), in which the change of selectivity in the first (**Figure 7C**) together with the deployment of attention, and the shift of the focus of attention in the second (**Figure 9**), are well accounted by the significant changes in both phases of the FR-profile. Overall, the behavior of the ST-model reflects the context-based competitive or enhancing effect of the cross-talk between attention and selectivity.

Our results coincided with the claim posed by the ST-model (Tsotsos, 1990; Tsotsos and Rothenstein, 2011) that suppressing irrelevant activity in the surround of the attentional focus forces the cell to adapt its firing and match the rate evoked by the

attended stimulus in isolation, in the sense that when attended, the FR-profile of the neuron in all simulations depended on its selectivity to that stimulus regardless of stimulus context. It made the response produced by all stimuli within the receptive field to be larger in the unattended scenario than when one of them was attended, due to the presence of distractors with high selectivity in the surrounding.

Since a significant amount of the information was encoded by the transient (latency), we hypothesize that this period of average duration in the range 20–30 ms, during which the firing rate suddenly drops and raises, could be required for the cells to re-accommodate to the confluent and ongoing bottom-up effect of selectivity and the top-down signal of attention; however, future work will require experiments in single cells and populations to test the functioning principles of the latency periods, so as to characterize their time courses. Secondly, based on our hypotheses it will be necessary to also check if the interplay between attention and selectivity is enough to fully disambiguate stimuli with complex combinations of features within a single visual scene.

AUTHOR CONTRIBUTIONS

This research work was carried out in collaboration between all authors. OA and JT defined the research theme. OA and JT designed methods and simulations, OA analyzed the data, OA and JT interpreted the results and wrote the paper. OA and JT discussed analyses, interpretation, and data presentation. All authors have contributed to, seen and approved the manuscript.

FUNDING

This research was performed in the frame of the STAR project funded by the Air Force Office of Scientific Research; Grant no. FA9550-14-1-0393.

ACKNOWLEDGMENTS

We want to express our gratefulness to Professor Julio-Cesar Martinez-Trujillo and his research group at Western University in London Canada, for valuable discussions.

REFERENCES

Bartsch, M. V., Loewe, K., Merkel, C., Heinze, H. J., Schoenfeld, M. A., Tsotsos, J. K., et al. (2017). Attention to color sharpens neural population tuning via feedback processing in the human visual cortex hierarchy. *J. Neurosci.* 37, 10346–10357. doi: 10.1523/JNEUROSCI.0666-17.2017

Bressler, S. L., Tang, W., Sylvester, C. M., Shulman, G. L., and Corbetta, M. (2008). Top-down control of human visual cortex by frontal and parietal cortex in anticipatory visual spatial attention. *J. Neurosci.* 28, 10056–10061. doi: 10.1523/JNEUROSCI.1776-08.2008

Buschman, T. J., and Kastner, S. (2015). From behavior to neural dynamics: an integrated theory of attention. *Neuron* 88, 127–144. doi: 10.1016/j.neuron.2015.09.017

Buschman, T. J., and Miller, E. K. (2007). Top-down versus bottom-up control of attention in the prefrontal and posterior parietal cortices. *Science* 315, 1860–1862. doi: 10.1126/science.1138071

Busse, L., Katzner, S., and Treue, S. (2008). Temporal dynamics of neuronal modulation during exogenous and endogenous shifts of visual attention in macaque area MT. *Proc. Natl. Acad. Sci. U.S.A.* 105, 16380–16385. doi: 10.1073/pnas.0707369105

Carrasco, M. (2011). Visual attention: the past 25 years. *Vision Res.* 51, 1484–1525. doi: 10.1016/j.visres.2011.04.012

Cavelier, P., Hamann, M., Rossi, D., Mobbs, P., and Attwell, D. (2005). Tonic excitation and inhibition of neurons: ambient transmitter sources and computational consequences. *Prog. Biophys. Mol. Biol.* 87, 3–16. doi: 10.1016/j.pbiomolbio.2004.06.001

Corbetta, M., Miezin, F. M., Dobmeyer, S., Shulman, G. L., and Petersen, S. E. (1991). Selective and divided attention during visual discriminations of shape, color, and speed: functional anatomy by positron emission tomography. *J. Neurosci.* 11, 2383–2402.

Corbetta, M., and Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nat. Rev. Neurosci.* 3, 201–215. doi: 10.1038/nrn755

Cutzu, F., and Tsotsos, J. K. (2003). The selective tuning model of attention: psychophysical evidence for a suppressive annulus around an attended item. *Vision Res.* 43, 205–219. doi: 10.1016/S0042-6989(02)00491-1

Dayan, P., and Abbott, L. F. (2001). *Theoretical Neuroscience*. Cambridge, MA: MIT Press.

Deco, G., and Lee, T. S. (2002). A unified model of spatial and object attention based on inter-cortical biased competition. *Neurocomputing* 44, 775–781. doi: 10.1016/S0925-2312(02)00471-X

Desimone, R., and Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annu. Rev. Neurosci.* 18, 193–222. doi: 10.1146/annurev.ne.18.030195.001205

Destexhe, A., Mainen, Z. F., and Sejnowski, T. J. (1998). Kinetic models of synaptic transmission. *Methods Neuronal Modeling* 2, 1–25.

Dorval, A. D., and White, J. A. (2005). Channel noise is essential for perithreshold oscillations in entorhinal stellate neurons. *J. Neurosci.* 25, 10025–10028. doi: 10.1523/JNEUROSCI.3557-05.2005

Eagleman, D. M., and Sejnowski, T. J. (2000). Motion integration and postdiction in visual awareness. *Science* 287, 2036–2038. doi: 10.1126/science.287.5460.2036

Fallah, M., Stoner, G. R., and Reynolds, J. H. (2007). Stimulus-specific competitive selection in macaque extrastriate visual area V4. *Proc. Natl. Acad. Sci. U.S.A.* 104, 4165–4169. doi: 10.1073/pnas.0611722104

Ferrera, V. P., Rudolph, K. K., and Maunsell, J. H. (1994). Responses of neurons in the parietal and temporal visual pathways during a motion task. *J. Neurosci.* 14, 6171–6186.

Hodgkin, A. L., and Huxley, A. F. (1952). A quantitative description of membrane current and its application to conduction and excitation in nerve. *J. Physiol.* 117, 500–544. doi: 10.1113/jphysiol.1952.sp004764

Hopf, J. M., Boehler, C. N., Luck, S. J., Tsotsos, J. K., Heinze, H. J., and Schoenfeld, M. A. (2006). Direct neurophysiological evidence for spatial suppression surrounding the focus of attention in vision. *Proc. Natl. Acad. Sci. U.S.A.* 103, 1053–1058. doi: 10.1073/pnas.0507746103

Hutt, A. (2012). The population firing rate in the presence of GABAergic tonic inhibition in single neurons and application to general anaesthesia. *Cogn. Neurodyn.* 6, 227–237. doi: 10.1007/s11571-011-9182-9

Itti, L. (2005). Models of bottom-up attention and saliency. *Neurobiol. Attent.* 582, 576–582. doi: 10.1016/B978-012375731-9/50098-7

Itti, L., and Koch, C. (2001). Computational modelling of visual attention. *Nat. Rev. Neurosci.* 2, 194–203. doi: 10.1038/35058500

Itti, L., Rees, G., and Tsotsos, J. K. (2005). *Neurobiology of Attention*. Waltham, MA: Elsevier/Academic Press.

Izhikevich, E. M. (2004). Which model to use for cortical spiking neurons? *IEEE Trans. Neural Netw.* 15, 1063–1070. doi: 10.1109/TNN.2004.832719

James, W. (1891). *The Principles of Psychology*, Vol. 2, London, UK: Macmillan.

Jensen, O., Goel, P., Kopell, N., Pohja, M., Hari, R., and Ermentrout, B. (2005). On the human sensorimotor-cortex beta rhythm: sources and modeling. *Neuroimage* 26, 347–355. doi: 10.1016/j.neuroimage.2005.02.008

Kandel, E. R., Schwartz, J. H., Jessell, T. M., Siegelbaum, S. A., and Hudspeth, A. (2000). *Principles of Neural Science*. New York, NY: McGraw-hill.

Khayat, P. S., Niebergall, R., and Martinez-Trujillo, J. C. (2010). Attention differentially modulates similar neuronal responses evoked by varying contrast and direction stimuli in area MT. *J. Neurosci.* 30, 2188–2197. doi: 10.1523/JNEUROSCI.5314-09.2010

Koch, C., and Ullman, S. (1987). "Shifts in selective visual attention: towards the underlying neural circuitry," in *Matters of Intelligence*, ed L. M. Vaina (Dordrecht: Springer), 115–141.

Kosai, Y., El-Shamayleh, Y., Fyall, A. M., and Pasupathy, A. (2014). The role of visual area V4 in the discrimination of partially occluded shapes. *J. Neurosci.* 34, 8570–8584. doi: 10.1523/JNEUROSCI.1375-14.2014

Ladenbauer, J., Augustin, M., and Obermayer, K. (2014). How adaptation currents change threshold, gain, and variability of neuronal spiking. *J. Neurophysiol.* 111, 939–953. doi: 10.1152/jn.00586.2013

Lee, J., and Maunsell, J. H. (2009). A normalization model of attentional modulation of single unit responses. *PLoS ONE* 4:e4651. doi: 10.1371/journal.pone.0004651

Lennert, T., and Martinez-Trujillo, J. (2011). Strength of response suppression to distracter stimuli determines attentional-filtering performance in primate prefrontal neurons. *Neuron* 70, 141–152. doi: 10.1016/j.neuron.2011.02.041

Loach, D. P., Tombu, M., and Tsotsos, J. K. (2005). Interactions between spatial and temporal attention: an attentional blink study. *J. Vis.* 5:109. doi: 10.1167/5.8.109

Luck, S. J., Chelazzi, L., Hillyard, S. A., and Desimone, R. (1997). Neural mechanisms of spatial selective attention in areas V1, V2, and V4 of macaque visual cortex. *J. Neurophysiol.* 77, 24–42. doi: 10.1152/jn.1997.77.1.24

Martinez-Trujillo, J. C., and Treue, S. (2004). Feature-based attention increases the selectivity of population responses in primate visual cortex. *Curr. Biol.* 14, 744–751. doi: 10.1016/j.cub.2004.04.028

McAdams, C. J., and Maunsell, J. H. (1999). Effects of attention on orientation-tuning functions of single neurons in macaque cortical area V4. *J. Neurosci.* 19, 431–441.

Niebur, E., and Koch, C. (1994). A model for the neuronal implementation of selective visual attention based on temporal correlation among neurons. *J. Comput. Neurosci.* 1, 141–158. doi: 10.1007/BF00962722

Oliva, A., Torralba, A., Castelhano, M. S., and Henderson, J. M. (2003). "Top-down control of visual attention in object detection," in *Image Processing, 2003. ICIP 2003. Proceedings. 2003 International Conference on: IEEE* (Barcelona), I253–I256.

Pape, H. C. (1996). Queer current and pacemaker: the hyperpolarization-activated cation current in neurons. *Annu. Rev. Physiol.* 58, 299–327. doi: 10.1146/annurev.ph.58.030196.001503

Pestilli, F., Viera, G., and Carrasco, M. (2007). How do attention and adaptation affect contrast sensitivity? *J. Vis.* 7:9. doi: 10.1167/7.7.9

Posner, M. I. (2011). *Cognitive Neuroscience of Attention*. New York, NY: Guilford Press.

Rao, S. C., Rainer, G., and Miller, E. K. (1997). Integration of what and where in the primate prefrontal cortex. *Science* 276, 821–824. doi: 10.1126/science.276.5313.821

Reynolds, J. H., Chelazzi, L., and Desimone, R. (1999). Competitive mechanisms subserve attention in macaque areas V2 and V4. *J. Neurosci.* 19, 1736–1753.

Reynolds, J. H., and Heeger, D. J. (2009). The normalization model of attention. *Neuron* 61, 168–185. doi: 10.1016/j.neuron.2009.01.002

Rothenstein, A. L., and Tsotsos, J. K. (2014). Attentional modulation and selection–an integrated approach. *PLoS ONE* 9:e99681. doi: 10.1371/journal.pone.0099681

Rutishauser, U., Walther, D., Koch, C., and Perona, P. (2004). "Is bottom-up attention useful for object recognition?," in *Computer Vision and Pattern Recognition, 2004. CVPR 2004, Proceedings of the 2004 IEEE Computer Society Conference on: IEEE* (Washington, DC), II37–II44.

Shriki, O., Sompolinsky, H., and Hansel, D. (2003). Rate models for conductance based cortical neuronal networks. *Neural Comput.* 15, 1809–1841. doi: 10.1162/08997660360675053

Spratling, M. W., and Johnson, M. H. (2004). A feedback model of visual attention. *J. Cogn. Neurosci.* 16, 219–237. doi: 10.1162/089892904322984526

Tsotsos, J. K. (1990). Analyzing vision at the complexity level. *Behav. Brain Sci.* 13, 423–445. doi: 10.1017/S0140525X00079577

Tsotsos, J. K. (2011). *A Computational Perspective on Visual Attention*. Cambridge: MIT Press.

Tsotsos, J. K., Culhane, S. M., Wai, W. Y. K., Lai, Y., Davis, N., and Nuflo, F. (1995). Modeling visual attention via selective tuning. *Artif. Intell.* 78, 507–545. doi: 10.1016/0004-3702(95)00025-9

Tsotsos, J., and Rothenstein, A. (2011). Computational models of visual attention. *Scholarpedia* 6:6201. doi: 10.4249/scholarpedia.6201

van Aerde, K. I., Mann, E. O., Canto, C. B., Heistek, T. S., Linkenkaer-Hansen, K., Mulder, A. B., et al. (2009). Flexible spike timing of layer 5 neurons during dynamic beta oscillation shifts in rat prefrontal cortex. *J. Physiol.* 587(Pt 21), 5177–5196. doi: 10.1113/jphysiol.2009.178384

Whittington, M. A., Traub, R. D., Kopell, N., Ermentrout, B., and Buhl, E. H. (2000). Inhibition-based rhythms: experimental and mathematical observations on network dynamics. *Int. J. Psychophysiol.* 38, 315–336. doi: 10.1016/S0167-8760(00)00173-2

Wiesenfeld, K., and Moss, F. (1995). Stochastic resonance and the benefits of noise: from ice ages to crayfish and SQUIDs. *Nature* 373, 33–36. doi: 10.1038/373033a0

Williford, T., and Maunsell, J. H. (2006). Effects of spatial attention on contrast response functions in macaque area V4. *J. Neurophysiol.* 96, 40–54. doi: 10.1152/jn.01207.2005

Wilson, H. R. (1999). *Spikes, Decisions, and Actions: the Dynamical Foundations of Neurosciences*, Don Mills, ON: Oxford University Press, Inc.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 28 March 2018
doi: 10.3389/fpsyg.2018.00417

A Neurodynamic Model of Feature-Based Spatial Selection

*Mateja Marić and Dražen Domijan**

Department of Psychology, Faculty of Humanities and Social Sciences, University of Rijeka, Rijeka, Croatia

Huang and Pashler (2007) suggested that feature-based attention creates a special form of spatial representation, which is termed a Boolean map. It partitions the visual scene into two distinct and complementary regions: selected and not selected. Here, we developed a model of a recurrent competitive network that is capable of state-dependent computation. It selects multiple winning locations based on a joint top-down cue. We augmented a model of the WTA circuit that is based on linear-threshold units with two computational elements: dendritic non-linearity that acts on the excitatory units and activity-dependent modulation of synaptic transmission between excitatory and inhibitory units. Computer simulations showed that the proposed model could create a Boolean map in response to a featured cue and elaborate it using the logical operations of intersection and union. In addition, it was shown that in the absence of top-down guidance, the model is sensitive to bottom-up cues such as saliency and abrupt visual onset.

Keywords: boolean map, feature-based attention, lateral inhibition, neural network, winner-take-all

OPEN ACCESS

Edited by:
Hedva Spitzer,
Tel Aviv University, Israel

Reviewed by:
Xavier Otazu,
Universitat Autònoma de Barcelona, Spain
Marius Usher,
Tel Aviv University, Israel
David Golomb,
Ben-Gurion University of the Negev, Israel

***Correspondence:**
Dražen Domijan
mmaric2@ffos.hr;
ddomijan@ffri.hr

Specialty section:
This article was submitted to Perception Science, a section of the journal Frontiers in Psychology

Received: *17 May 2017*
Accepted: *13 March 2018*
Published: *28 March 2018*

Citation:
Marić M and Domijan D (2018) A Neurodynamic Model of Feature-Based Spatial Selection. Front. Psychol. 9:417. doi: 10.3389/fpsyg.2018.00417

INTRODUCTION

In the literature on visual attention, significant progress has been made in characterizing the principles of selection. Visual attention can be allocated flexibly to a circumscribed region of space, the whole object or feature dimensions such as color and orientation (Nobre and Kastner, 2014). Indeed, early work suggested that a restricted circular region of space is a representational format of attentional selection. Posner (1980) proposed that attention operates like a spotlight that highlights a single circular region of space with a fixed radius. All locations that fall inside the spotlight are selected, and everything outside is left out. An extension of this proposal, which is called the zoom-lens model, suggests that the spotlight of attention can change its radius depending on the spatial resolution that one wants to achieve (Eriksen and St. James, 1986). If high resolution is required, the spotlight can be narrowed to capture details in the selected region, whereas the radius of the spotlight can be widened when a lower resolution is sufficient.

Other studies point to an object as a unit of selection. Duncan (1984) showed that it is easier to report two attributes if they appear on the same object, relative to the scenario in which each attribute appears on a different object. This finding implies that the object is selected as a whole and has been replicated many times using different stimuli and behavioral paradigms (Scholl, 2001). This effect cannot be explained by spatial attention because objects were spatially superimposed, that is, they shared the same locations. More recently, it was shown that attention can also be allocated to a visual feature such as color or direction of motion independent of spatial location (Saenz et al., 2002, 2003). Single-unit recordings have shown that feature-based attention is accompanied by the global location-independent modulation of neural response in a range of areas in the visual cortex. Attentional modulation was described as a

multiplicative gain change that increases responses of neurons that are selective to attended feature values and decreases responses of neurons that are tuned to unattended feature values (Treue and Martinez-Trujillo, 1999; Martinez-Trujillo and Treue, 2004).

Object-based attention, however, is not necessarily detached from spatial representation. There is behavioral and neurophysiological evidence that object-based attention involves selection of all spatial locations that are occupied by the same object. Specifically, it was suggested that attention selects a grouped array of locations (O'Grady and Müller, 2000). In other words, attention spreads from one spatial location along the shape of the object and highlights all locations that belong to the object (Richard et al., 2008; Vatterott and Vecera, 2015). Neurophysiological studies showed that object-based selection is indeed achieved by the spreading of the enhanced firing rate along the shape of the object (Roelfsema, 2006; Roelfsema and de Lange, 2016).

In a similar way, feature-based attention might involve the selection of all locations that are occupied by the same feature value, as shown by Huang and Pashler (2007). They proposed that attention is limited because it may access only one feature value (e.g., red) per dimension (e.g., color) at any given moment. However, the accessed feature value is bound to space in parallel, without capacity limits. Feature-based attention is allocated in space via the formation of a binary or Boolean map. When a conscious decision is made to attend to a specific feature value, the Boolean map indicates all spatial locations that are occupied by the chosen feature value because they are labeled by a positive value (e.g., 1), while all other locations are labeled with zero. In each selection process, selected locations need not be contiguous in space, but they must share the same feature value. After a Boolean map is formed, it is possible to operate on its output by applying the set operations of intersection and union. Recent work suggests that a spatial representation, such as a Boolean map, might mediate perceptual grouping by similarity (Huang, 2015; Yu and Franconeri, 2015). Moreover, the idea has been recently applied successfully in the computer vision literature on developing algorithms for saliency detection (Zhang and Sclaroff, 2016; Qi et al., 2017).

Figure 1 illustrates a Boolean map that is formed in response to three different stimulus configurations and sequential application of two top-down feature cues. **Figure 1A** shows a simple stimulus that consists of red and green squares. An observer might attempt to isolate only red or only green items. To do so, a top-down cue should be supplied to the feature map that encodes the desired feature value. For example, when attention is directed to the red color, the top-down cue highlights all locations that are occupied by red squares. The Boolean map picks up on this feature cue and forms a spatial representation in which cued locations are labeled with 1 (white) and non-cued locations are labeled with 0 (black). In terms of a neural network, these labels correspond to the active (excited) and inactive (inhibited) states of the corresponding nodes in the network (Boolean Map – 1). Later, the observer might wish to switch to green color (Boolean Map – 2). Again, in a response to a new feature cue, the Boolean map now shows all locations that are occupied by green squares.

Figure 1B shows a typical stimulus that is used in visual search experiments. It consists of red and green horizontal and vertical bars. The task is to find a red horizontal bar. This is an example of a conjunction search task in which two feature dimensions should be combined to find the target object. According to Huang and Pashler (2007), the conjunction task is solved in two steps. In the first step, a Boolean map is formed by top-down cueing of red items, irrespective of their orientations. In the second step, only horizontal items are cued. However, since red items have already been selected, the second Boolean map will correspond to the intersection of red and horizontal items. There is only one item that satisfies these selection criteria: the target. In this way, visual search is substantially faster compared to the strategy of sequentially visiting each item by moving the attentional spotlight across the visual field. It is also possible to reverse the order of the applied feature cues. In the first step, horizontal items might be cued, and the intersection is formed by highlighting red items in the second step. Importantly, there is behavioral evidence that observers indeed implement such a *subset selection* strategy in conjunction search tasks (Egeth et al., 1984; Kaptein et al., 1995). Moreover, Huang and Pashler (2012) showed that the same strategy is used in the perception of spatial structure in a stimulus that is composed of multiple items that differ in several dimensions.

Figure 1C illustrates an example of the union of two Boolean maps. As in the previous example, the observer starts by cueing red items and creating a Boolean map that consists of a representation of their locations. In the second step, the observer wishes to combine red with horizontal items. Therefore, in the second step, one should cue horizontal items but simultaneously maintain locations of the remaining items in memory. The resulting new Boolean map now represents the locations of all red and all horizontal items that were found in the image. Computing with Boolean maps might not be restricted to only two steps, as **Figure 1** suggests. It is possible to incorporate more feature dimensions, such as motion, texture, or size, that can also be engaged in creating Boolean maps that are more complicated.

Feature-based spatial selection, as illustrated by the Boolean map, provides a strong constraint on the computational models of visual attention because it requires simultaneous selection of arbitrarily many locations based on an arbitrary criterion that is set by the observer. Computational models of attention often rely on a winner-take-all (WTA) network to select a single, most salient location from the input image (Itti and Koch, 2000, 2001). The WTA network consists of an array of excitatory nodes that are connected reciprocally with inhibitory interneurons. This anatomical arrangement creates lateral inhibition among excitatory nodes that lead to the selection of a single node that receives maximal input and the suppression of all other nodes, which receive non-maximal input. However, when faced with the input where multiple (potentially many) nodes share the same maximal input level, the typical WTA network tends to suppress all winning nodes due to a strong mutual inhibition among them instead of selecting them together. For example, Usher and Cohen (1999) showed that, under the conditions of strong recurrent excitation and weak lateral inhibition, the WTA network reaches a steady state with multiple active winners.

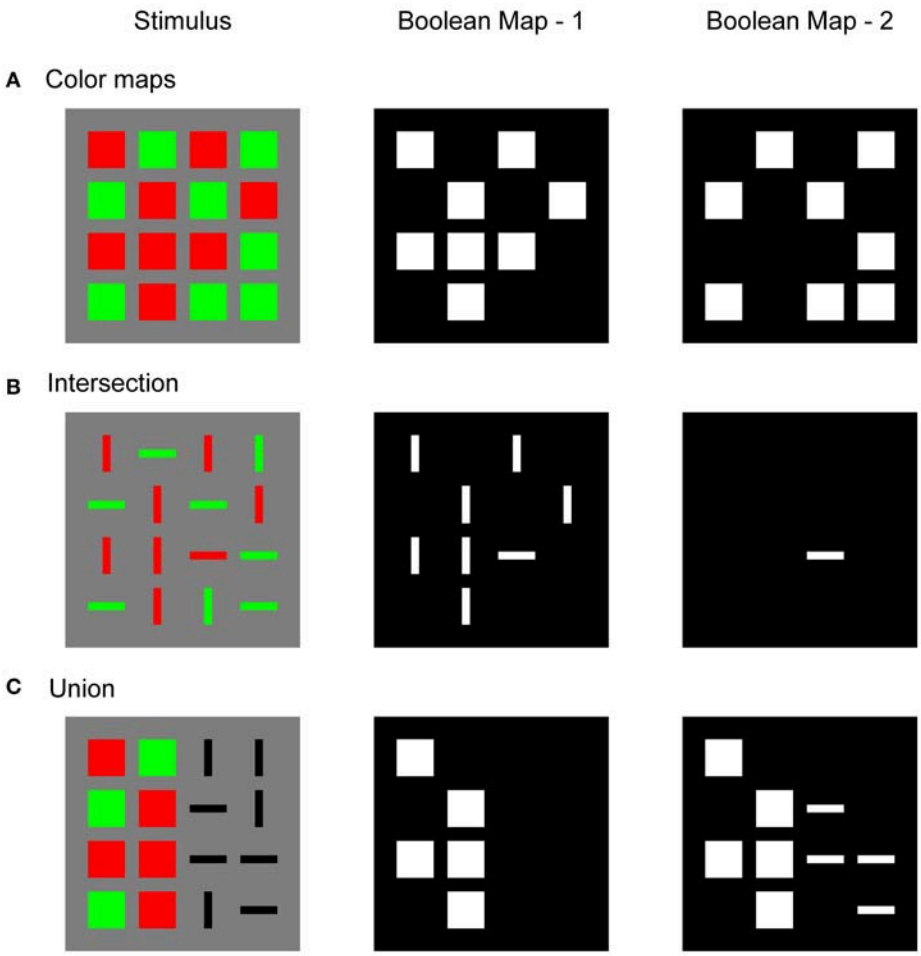


FIGURE 1 | Illustration of the Boolean map that was created in response to the input image (Stimulus) after the first feature cue was applied to the spatial representation (Boolean Map – 1) and after the second feature cue was applied (Boolean Map – 2). **(A)** Boolean maps that were created by two-color cues: red in the first step and green in the second step; **(B)** intersection of two Boolean maps, where red is cued in the first step and horizontal orientation in the second step; **(C)** union of two Boolean maps, where red is cued in the first step and horizontal orientation in the second step.

Importantly, activation of the winning nodes decreases linearly toward zero as their quantity increases. In other words, this network design suffers from the capacity limitation. This is a useful property in modeling short-term memory and frontal lobe function (Haarmann and Usher, 2001) but it is inadequate for understanding how the Boolean map might arise in a large retinotopic map, as exemplified by **Figure 1**.

Another problem is that the dynamics of the WTA network are not sensitive to transient changes in the input amplitude. Due to strong self-excitation and the resulting persistent activity, the WTA network settles into one of its memory states (fixed points). Importantly, each memory state is independent of later inputs. If self-excitation is weakened, the network will become sensitive to input. However, at the same time, it will lose its ability to form a memory state and will behave like a feedforward network (Rutishauser and Douglas, 2009). One way to solve this problem is to apply an external reset signal to the network before a new input is processed (Grossberg, 1980; Kaski and Kohonen, 1994; Itti and Koch, 2000, 2001). However, this is not sufficient in the context of feature-based attention. An intersection or union operation between two Boolean maps requires that the currently active memory state (formed after the first feature cue) be updated by taking into account new input (the second feature cue). Therefore, the dynamics of the WTA network should allow uninterrupted transition between memory states that are governed by external inputs. In other words, the WTA network should be capable of state-dependent computation (Rutishauser and Douglas, 2009).

To summarize, a WTA network that is capable of computing with Boolean maps should simultaneously satisfy two computational constraints:

1. It should be able to select together all locations that share a common feature value. This should be achieved without degrading the representation of the winners.

2. It should exhibit state-dependent computation, in which new inputs are combined with the current memory state to produce a new resultant state (e.g., intersection or union).

Here, we have developed a new WTA network that satisfies these constraints and provides the neural implementation of the Boolean map theory of attention (Huang and Pashler, 2007).

MODEL DESCRIPTION

The aim of the current work is to provide an explanation of how a Boolean map may be formed in a recurrent competitive network that can implement feature-based winner-take-all (F-WTA) selection. To this end, we have extended the previously proposed network model based on the linear-threshold units (Hahnloser, 1998; Hahnloser et al., 2003; Rutishauser and Douglas, 2009). Concretely, the model circuit is presented in **Figure 2**. It consists of a single inhibitory unit, which is reciprocally connected to a group of excitatory units. In addition to these basic elements, we introduce two processing components into the WTA circuit to expand its computational power. The first is a dendritic non-linearity, which prevents excessive excitation that arises from self-recurrent and nearest-neighbor collaterals. We modeled the dendritic tree as a separate electrical compartment with its own non-linear output that is supplied to the node's body (Häusser and Mel, 2003; London and Häusser, 2005; Branco and Häusser, 2010; Mel, 2016). The second is modulation of synaptic transmission by retrograde inhibitory signaling (Tao and Poo, 2001; Alger, 2002; Zilberter et al., 2005; Regehr et al., 2009). This is a form of presynaptic inhibition, where postsynaptic cells release a neurotransmitter that binds to the receptors that are located on the presynaptic terminals. Retrograde signaling creates a feedback loop that dynamically regulates the amount of transmitter that is released from the presynaptic terminals. Here, we have hypothesized that such interactions occur in recurrent pathways from the excitatory nodes to the inhibitory interneuron and back from the interneuron to the excitatory nodes. In the excitatory-to-inhibitory pathway, retrograde signaling enables the inhibitory interneuron to compute the maximum instead of the sum of its inputs. Computation of the maximum arises from the limitation that the activity of the inhibitory interneuron cannot grow beyond the maximal input that it receives from the excitatory nodes. Furthermore, retrograde signaling in the inhibitory-to-excitatory pathway enables the excitatory nodes that receive maximal input to protect themselves from the common inhibition. In this way, the network can select all excitatory nodes with maximal input, irrespective of their quantity or arrangement in visual space.

At first sight, it might appear strange to propose that an excitatory unit can inhibit its input by releasing a neurotransmitter that binds to the presynaptic terminal. However, several signaling molecules have been identified to support such interactions, including endogenous cannabinoids (Alger, 2002). Moreover, Zilberter (2000) found that glutamate is released from dendrites of pyramidal neurons in the rat neocortex and suppresses the inhibition that impinges on them. In addition, similar action has been found for GABA (Zilberter et al., 1999), which suggests that conventional neurotransmitters can engage in retrograde signaling.

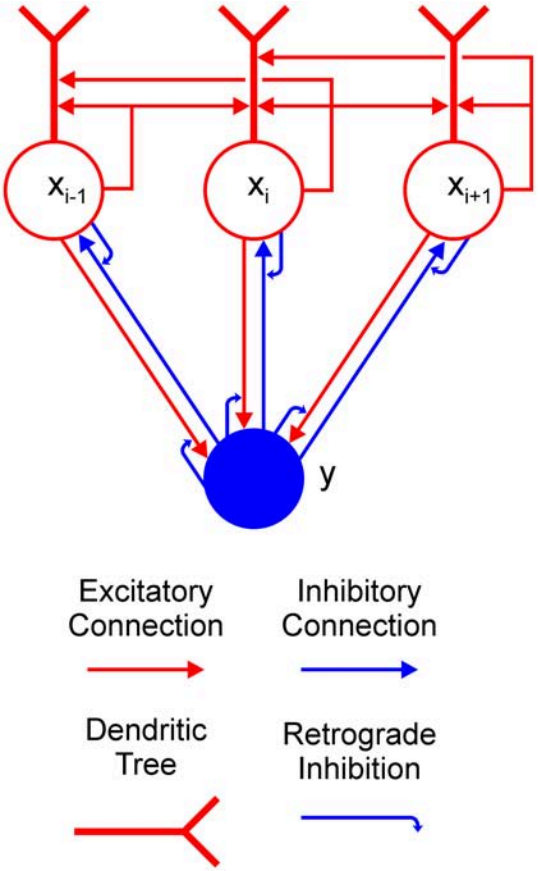


FIGURE 2 | Feature-based winner-take-all (F-WTA) circuit. Connections between excitatory (red circles) and inhibitory (blue disk) units are modulated by retrograde inhibition (curved blue arrows). Self-excitation and nearest-neighbor excitation are mediated by the dendrites of the excitatory units. The same motif is repeated for all excitatory nodes in the recurrent map.

To situate the proposed F-WTA circuit in a larger neural architecture that describes the cortical computations that underlie top-down attentional control, we have adopted the model that was proposed by Hamker (2004). He showed how attentional selection of a target arises from the recurrent interactions within a distributed network that consists of model cortical area V4, the inferotemporal cortex (IT), the posterior parietal cortex (PPC), and the frontal eye fields (FEF). **Figure 3** illustrates part of these interactions that are involved in feature-based attentional guidance. Top-down signals that provide feature cues originate in the IT, which contains a spatially invariant representation of relevant visual features. The IT sends feature-specific feedback projections to the V4, where topographically organized feature maps for each feature value are located. For simplicity, we consider only maps for two colors (red and green), and two orientations (vertical and horizontal). We do not explicitly model IT and V4 dynamics. Rather, they serve here as a tentative explanation of how input to the F-WTA network

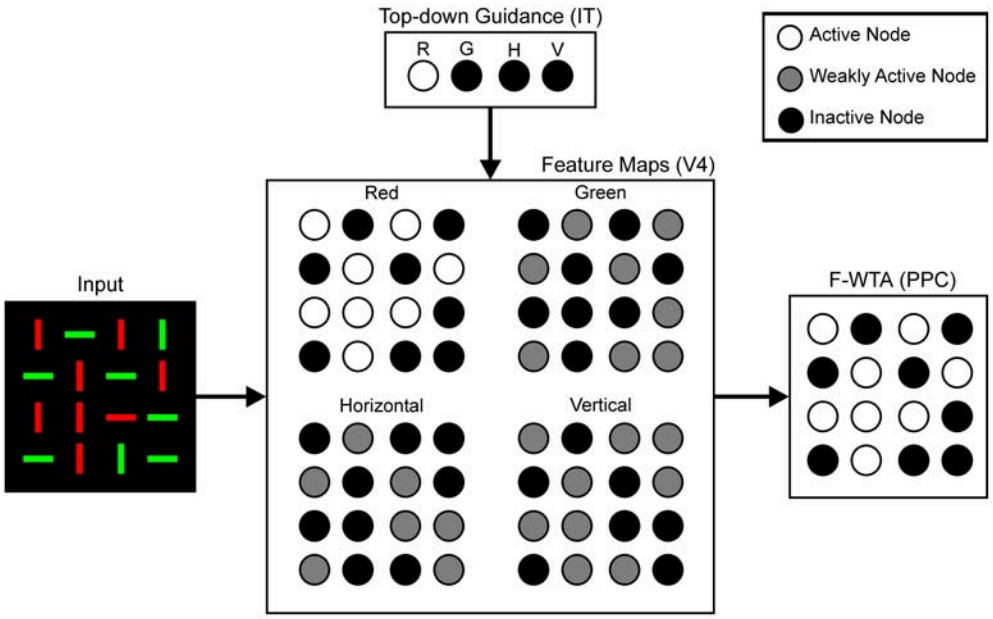


FIGURE 3 | Neural architecture for the top-down guidance of attention by feature cues, following Hamker (2004). Input is processed in retinotopically organized feature maps for colors and orientations. These maps also receive top-down signals, which provide feature guidance. In this example, input image is taken from **Figure 1B** and red color is cued by the top-down signals. Therefore, the activity of the nodes in the Red map is enhanced (white discs) relative to the activity in all other feature maps (gray discs) because latter receives only feedforward signals. Black discs represent inactive nodes. In the feature maps, this indicates the absence of a feature at a given locations. The F-WTA network sums output of all feature maps. Its activity represents all locations occupied by the cued feature. Parentheses contain reference to cortical areas thought to be involved in proposed computations. R, red; G, green; H, horizontal; V, vertical.

arises within the ventral visual pathway. Also, we omitted the contribution of the FEF and its spatial reentry signals to the V4 activity.

We hypothesize that the feature-based WTA network resides in the PPC, where it receives summed input over all feature maps from the V4. Top-down guidance is implemented by a temporary increase in activity in one of the V4 feature maps. For example, when the decision is made to attend to the red color, the IT representation of red color sends feedback signals to the Red Map in the V4. Top-down signals to the feature map are modeled as a multiplicative gain of neural activity, which is consistent with neurophysiological findings (Treue and Martinez-Trujillo, 1999; Martinez-Trujillo and Treue, 2004; Maunsell and Treue, 2006).

The following neural network equations represent the quantitative description of the model. Each unit is defined by its instantaneous firing rate (Dayan and Abbott, 2000). The time evolution of the activity of excitatory node x at position i in the recurrent map is given by the following differential equation:

$$\tau_x \frac{dx_i}{dt} + x_i = \left[I_i(t) + \alpha f\left(x_i + x_{i+1} + x_{i-1}\right) - \beta_1 g\left(y - x_i - T_y\right) \right]^+. \tag{1}$$

The time evolution of the activity of inhibitory interneuron y is given by

$$\tau_y \frac{dy}{dt} + y = \left[\beta_2 \sum_i g\left(x_i - y - T_x\right) \right]^+. \tag{2}$$

Parameters τ_x and τ_y are integration time constants for excitatory and inhibitory nodes, respectively. We assume that inequality $\tau_x > \tau_y$ holds, which accords with the observation in electrophysiological measurements that inhibitory cells exhibit faster dynamics than excitatory cells (McCormick et al., 1985). The second term on the left-hand side of Equations (1) and (2) describes the passive decay that drives the unit's activity to the resting state in the absence of external input. Firing rate activation function $[u]^+$ is a non-saturating rectification nonlinearity, which is defined by

$$[u]^+ = \max(u, 0). \tag{3}$$

Following Hamker (2004), we assume that feedforward input I_i at time t to the excitatory node x_i in the F-WTA network is given by the sum over activity in all V4 feature maps $I_i^{(m)}$,

$$I_i(t) = \sum_m I_i^{(m)} G^{(m)}(t). \tag{4}$$

In Equation (4), m denotes available feature maps with $m \in \{red, green\}$ in the simulation that is reported in section Simulation of the Formation of a Single Boolean Map and $m \in \{red, green, horizontal, vertical\}$ in the simulation that is reported in section Simulation of the Intersection and Union of Two Boolean Maps. Parameter G^m refers to the feature-specific, global multiplicative gain that all units $I_i^{(m)}$ within the same feature map m receive via top-down projections. As shown in **Figure 2**,

these projections arrive from the feature representation in the IT. Multiplicative gating is generally consistent with previous models that describe the effect of feature-based attention on the responses of neurons in the early visual cortex (Boynton, 2005, 2009). Equation (4) ensures that the F-WTA network is not particularly sensitive to any feature value. Rather, it signals the behavioral relevance of locations in a spatial map. Here, the relevance can be set according to differences in the bottom-up input $I_i^{(m)}$ that arise from competitive interactions in the early visual cortex. Alternatively, relevance can be signaled by the top-down feature cues G^m that change the gain of all locations that are occupied by the same feature value.

Dendritic output *f(u)* is described by the sigmoid response function

$$f(u) = \frac{S_d}{1 + e^{-\lambda(u - T_d)}} \tag{5}$$

where λ and T_d control the shape of the sigmoid function and S_d is its upper asymptotic value. We set λ to a high value to achieve a steep rise of the dendritic activity immediately after its input crosses the dendritic threshold, which is denoted as T_d. Such strong non-linearity is justified by experimental data, which show all-or-none behavior in real dendrites (Wei et al., 2001; Polsky et al., 2004). In Equation (1), parameter α controls the strength of the impact that the dendritic compartment exerts on the soma.

Self-recurrent x_i and nearest-neighbor collaterals x_{i-1} and x_{i+1} arrive on the dendrite of the excitatory node, which is consistent with the anatomical observation that most recurrent excitatory connections are made on the dendrites of the excitatory cells (Spruston, 2008). Nodes at the edge of the network receive excitation only from a single available neighbor. That is, node x_1 receives excitation only from x_2, and x_N receives excitation only from x_{N-1}. Nearest-neighbor excitatory interactions enable feature cues to spread activity enhancement automatically to all connected locations that contain a given feature value. This is not essential for the simulation of Boolean maps but we included it in our model because recurrent connections among nearby neurons are prominent feature of the synaptic organization of the cortex (Douglas and Martin, 2004). Also, we wanted to show that the proposed model is capable of simulating object-based attention (Roelfsema, 2006; Roelfsema and de Lange, 2016). Moreover, Wannig et al. (2011) found direct evidence for activity spreading among neurons that encode the same feature value in the primary visual cortex.

The output of the presynaptic interactions *g(u)* is defined by the rectification non-linearity of the form

$$g(u) = [u]^+ = \max(u, 0). \tag{6}$$

In Equation (1), the term $-\ g(y - x_i - T_y)$ describes the output of the presynaptic terminal that delivers inhibition from interneuron y to excitatory node x_i (**Figure 4A**). However, we did not explicitly model the dynamics of retrograde signaling. We assumed that the release of the retrograde transmitter occurs simultaneously with the activation of the postsynaptic node and that it is proportional to its firing rate. Therefore, it is represented by the term $-\ x_i$.

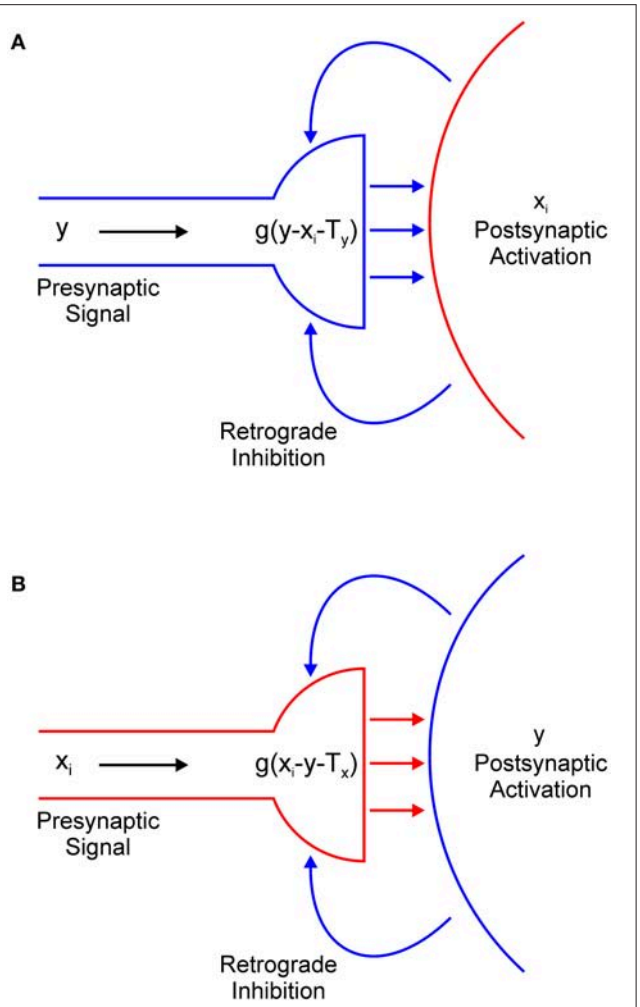


FIGURE 4 | Retrograde inhibitory signaling (blue curved arrows) from excitatory node x_i to the presynaptic terminal of inhibitory interneuron y **(A)** and from the inhibitory interneuron to the presynaptic terminal of the excitatory node **(B)**. Both terminals compute half-wave rectification $g(u)$ of their input. Terminals release respective inhibitory **(A)** or excitatory **(B)** neurotransmitter (straight horizontal arrows) only when they receive net positive input.

Function $g(u)$ ensures that the presynaptic terminal will release the inhibitory transmitter only when the electrical signal from node y exceeds the inhibitory retrograde signal $-x_i$ and the threshold for presynaptic activation, which is denoted as T_y. In other words, node x_i will be inhibited only if $y > x_i + T_x$. If this is not the case, node x_i will effectively isolate itself from the inhibitory influence of node y. This is always the case for the winning node because $x(t) > y(t)$ for $t > 0$. Moreover, this result extends to all other nodes whose input magnitude is sufficiently close to the maximal input. The strength of the inhibition is determined by parameter β_1. In a similar vein, in Equation (2), the term $-g(x_i - y - T_x)$ describes the action of the retrograde signal that is released from inhibitory interneuron y on the presynaptic terminal that delivers excitation from node x_i (**Figure 4B**). Here, parameter T_x describes the threshold for the

activation of the presynaptic terminal of the excitatory node and β_2 determines the strength of the excitation.

We have proposed a model of a one-dimensional network, although it attempts to simulate phenomena that occur in 2-D, as illustrated by **Figure 1**. We have chosen to work with the 1-D version of the network simply because we want to focus on the analysis of its temporal dynamics and its ability to combine information over time. Without loss of generality, the computer simulations that are reported in section Computer Simulations should be considered as a cross-section of a 2-D network.

For simplicity, the thresholds that control the activation of the excitatory and inhibitory nodes are all set to zero and are omitted from the model description. Parameters were set as follows: $\tau_x = 5$; $\tau_y = 2$; $\alpha = 1$; $\beta_1 = 1$; $\beta_2 = 10$; $S_d = 1$; $\lambda = 100$; $T_d = 0.1$; $T_x = 0.1$; and $T_y = 0.1$. Parameters were chosen in a way to simultaneously achieve intersection and union. Systematic variations on the parameters α, β_1 and β_2 showed that intersection is observed when $1 \leq (\alpha, \beta_1) \leq 5$. In contrast, union is observed when $0.8 \leq (\alpha, \beta_1) \leq 1$. Parameter β_2 can be set to any value above the default without changing the results.

MODEL EXTENSIONS

The network that is defined by Equations (1) and (2) is chosen in a way that achieves the desired behavior with the minimal number of computational elements. This simplicity heuristic is important for understanding model properties without adding extra neuroscientific complexity (Ashby and Hélie, 2011). However, at the same time, this approach sacrifices anatomical and biophysical plausibility of the proposed model. In this section, we present several extensions and generalizations of the basic model that bring it closer to satisfying the neurobiological constraints.

Inhibitory Pool

The model has just one inhibitory interneuron for computational convenience, which is not realistic. It is known that excitatory neurons outnumber inhibitory neurons by a factor of four in the cortex (Braitenberg and Schüz, 1991). However, it is possible to design an F-WTA network with a pool of inhibitory interneurons and the appropriate ratio between excitatory and inhibitory nodes that achieves the same behavior as the original model. An extended F-WTA network is presented in **Figure 5A**. Here, each inhibitory interneuron receives input from a subset of the excitatory nodes. We depicted each excitatory subset as a vertical arrangement of four nodes that do not overlap in their projections to the inhibitory pool. Therefore, each excitatory node projects to just one inhibitory node. Naturally, this does not need to be the case. It is possible that each excitatory node projects to more than one node without compromising the network output. Importantly, all inhibitory interneurons are mutually connected. In addition, each inhibitory interneuron projects its output to all excitatory nodes (denoted by thick blue arrow). As in the original model, we assume that all inhibitory and excitatory nodes are endowed with the capability of retrograde signaling on their synaptic contacts.

Within the pool of inhibitory nodes, retrograde signaling enables computation of the MAX function, as in the original model. To see this, consider the inhibitory node that receives maximal input. Due to the retrograde signaling, it will reach a steady state that corresponds to the computation of the MAX function over input from its excitatory subset. Moreover, it will not receive inhibition from the other members of the pool. All other inhibitory nodes, which receive less excitatory support, will be silenced because their retrograde signaling is not sufficiently strong to prevent lateral inhibition from the winning node. However, if there are multiple inhibitory nodes with the same level of activity, they will remain active together. Finally, the winning nodes send inhibition to all excitatory subsets. Since excitatory nodes also engage in retrograde signaling, the nodes that receive maximal input will block inhibition and remain active. Therefore, the network output will look much like the original model because the MAX computation on the inhibitory nodes makes irrelevant the number of them that are active simultaneously.

Localized Inhibition

An important shortcoming of the previous model is that it assumes that inhibitory projections extend across the whole network of excitatory units. This is clearly not the case in real neural networks, where the spatial spread of inhibition is limited. To account for this property, we have constructed a more elaborate version of the basic model, which is shown in **Figure 5B**. It contains a new pool z_j of excitatory nodes with long-range projections. The z_j nodes receive input from the subset of the x_i nodes. Additionally, each z_j node sends its projection to at least one y_j node from the pool of inhibitory nodes. The number of z nodes must equal the number of inhibitory nodes y_j so that they can be indexed by the same subscript j. Again, we assume that the z_j nodes are equipped with the ability of retrograde signaling on their synapses. Therefore, they also compute the MAX function over all their inputs, including feedforward input from the corresponding subset of x_i nodes and recurrent input from other z_j nodes. In this design, the maximum level of activity that is sensed by the x_i nodes in one part of the network is easily propagated via z_j nodes to all other parts of the network. Furthermore, z_j nodes transfer this activity to inhibitory nodes. Therefore, each inhibitory node will eventually receive the maximal level of activity and apply it to the subset of x_i nodes to which it is connected. In this design, it is not necessary for inhibitory nodes to interact with one another. The excitatory nodes x_i that receive maximal input will block inhibition by their retrograde signaling and remain active in the same manner as described in the previous section. In this way, the proposed circuit achieves the same result as the original model.

Output Functions

The model employs threshold-linear output functions for the soma and the logistic sigmoid function for dendrites. This is inconsistent with the observation that somatic output also saturates and is also often modeled by the sigmoid function.

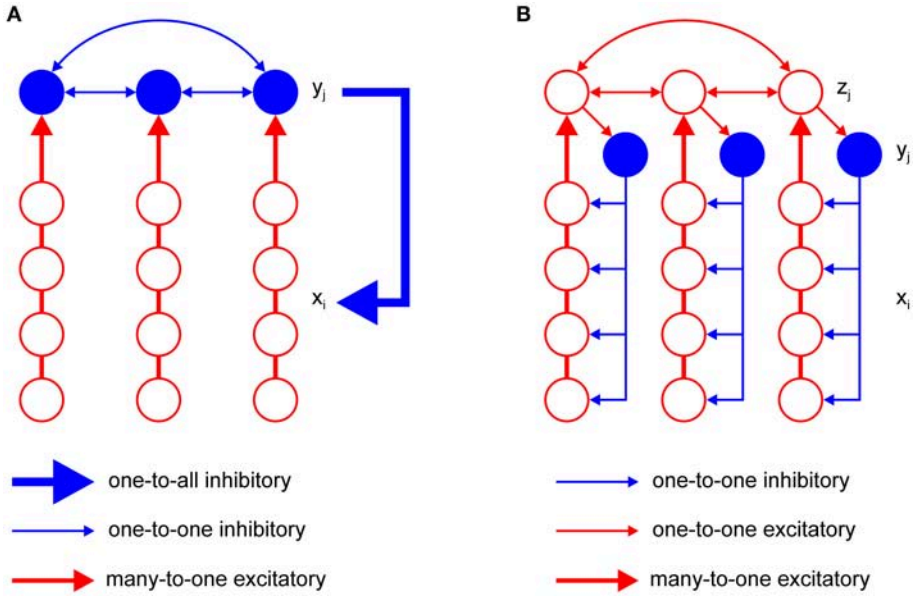


FIGURE 5 | Two variations of the F-WTA circuit design that are computationally equivalent to the basic circuit that is shown in **Figure 2**. **(A)** Circuit with a set of inhibitory nodes, which are denoted as y_j. Each y_j receives input from a subset of excitatory nodes. Inhibitory nodes compete with one another and the winning node encodes the maximum of its input. It delivers inhibition to all excitatory nodes in the same way as single inhibitory node y in the basic circuit. **(B)** Circuit with an additional set of excitatory nodes z_j with long-range horizontal projections. These nodes propagate the locally computed maximum level of activity to all parts of the network. Therefore, the whole set of z_j converges to a global maximum. Furthermore, they contact inhibitory nodes y_j that deliver inhibition to a subset of excitatory nodes x_i.

However, in normal circumstances, neurons operate in a linear mode that is far from their saturation level (Rutishauser and Douglas, 2009). To provide a more systematic approach to the output functions that are used in the model, we introduce a piecewise-linear approximation to the sigmoid function $s_q(u)$ of the form

$$s_q(u) = \begin{cases} 0 & if \quad u \leq 0 \\ u & if \quad 0 < u < S_q \\ S_q & if \quad u \geq S_q \end{cases} \tag{7}$$

where S_q denotes the upper saturation point, which can be set differently for different computational units $q \in \{c, d, p\}$, which correspond to the somatic, dendritic, and presynaptic terminal outputs, respectively. With the output function $s_q(u)$ applied to all computational elements of a single node, the model equations, namely, Equations (1) and (2), can be restated as

$$\tau_x \frac{dx_i}{dt} + x_i = s_c \left[I_i(t) + \alpha s_d (x_i + x_{i+1} + x_{i-1} - T_d) - \beta_1 s_p (y - x_i - T_y) \right] \tag{8}$$

and

$$\tau_y \frac{dy}{dt} + y = s_c \left[\beta_2 \sum_i s_p (x_i - y - T_x) \right]. \tag{9}$$

An important constraint of the model that is defined by Equations (8) and (9) is that saturation point for the dendritic output S_d should be chosen to be smaller than S_c, which is the saturation point of the somatic output. In this way, feedforward input I_i can be combined with the dendritic output without causing saturation at the output of the node. In contrast, if dendrites are allowed to saturate at the same activity level as the node, the dendritic output will overshadow the feedforward input. Consequently, the network will lose its sensitivity to the input changes. This is undesirable with respect to the requirements that are imposed by the sequential formation of the multiple Boolean maps. Therefore, the choice between the linear or the sigmoid output function for the node is not important if the dendritic output is restricted to a smaller interval relative to the output of the node itself.

LINEAR STABILITY ANALYSIS

Fixed Points

Fixed point is found iteratively starting from the set of nodes receiving maximal input, x_M. We assume that the winning nodes and inhibitory interneuron are activated above their thresholds, so we set $[u]^+ = u$. Next, we observe that the winning nodes do not receive inhibition from the interneuron y since $x_M(t) > y(t)$ for $t > 0$. This holds because the activity of the inhibitory node is bounded above by $x_M + T_x > y$ where T_x is a positive constant. Then, retrograde signaling ensures that $g(y - x_M - T_y) = 0$ for all times t. Consequently, nodes receiving maximal input are driven solely by excitatory terms. Since the recurrent excitation is bounded above by its asymptotic value S_d, dendritic output

function $f(u)$ in Equation (1) is replaced with S_d. This yields the following approximation to the steady state of the winning nodes:

$$x_M \approx I_M + \alpha S_d. \tag{10}$$

After the x_M, inhibitory interneuron y also reaches its steady state because its activity is driven primarily by the input from x_M. As the activity of y grows, terms $g(x_i - y - T_x)$ in Equation (2) vanish for all nodes that do not receive maximal input x_i where $i \notin M$. In contrast, the presynaptic terminals of x_M are above the threshold for their activation just before y reaches equilibrium, that is, $g(x_M - y - T_x) > 0$. Therefore, the output function of the presynaptic terminal $g(u)$ can be replaced by u. Then, Equation (2) is solved as

$$y = \frac{\beta_2 k\,(x_M - T_x)}{\beta_2 k + 1}, \tag{11}$$

where k is the number of x_M. When β_2 is chosen to be sufficiently large, and/or there are many nodes with maximal input x_M, then

$$y \to x_M - T_x. \tag{12}$$

Continuity of the function defined by Equation (2) implies that y cannot grow above $x_M - T_x$, that is, $y(t) > x_M(t) - T_x$ cannot hold at any time t unless $y(t_0) = x_M(t_0) - T_x$ at some earlier time $t_0 < t$. However, equality $y(t_0) = x_M(t_0) - T_x$ implies that $dy/dt = 0$ at time t_0 because $g(x_M(t_0) - y(t_0) - T_x) = 0$. In other words, node y loses all its excitatory drive when it reaches $x_M - T_x$. This is true irrespective of the number k of x_M. Thus, node y computes the maximum over its input.

The x_M nodes, together with the inhibitory node, create a quenching threshold (QT) for the network, which is defined by

$$QT = y - T_y = x_M - T_x - T_y. \tag{13}$$

Grossberg (1973) introduced the concept of the quenching threshold to describe the property of contrast enhancement in recurrent competitive networks. Nodes whose activity is above QT are enhanced and stored in the memory state, while all nodes whose activity is below QT are suppressed and removed from the memory representation. In the same manner, the remaining excitatory nodes converge to one of two states, depending on whether they exceed QT or not:

$$x_{i \notin M} \approx \begin{cases} I_i + \alpha S_d & \text{if } x_i \geq QT \\ 0 & \text{if } x_i < QT. \end{cases} \tag{14}$$

QT and its relationship with the activity of the winning and non-winning nodes and inhibitory interneuron is illustrated in **Figure 6**. According to Equations (10), (11), and (14), the fixed-point linearly combines input and recurrent excitation. As maximal input increases or decreases, the fixed point will move up or down and track these changes. Moreover, the input may cease, and the winning nodes will settle into the activity level that is provided by the recurrent excitation alone, which is expressed as αS_d. In other words, the network remembers who the last winner was. The same is true in the case where the winner is determined by transient cues that are applied sequentially on a sustained input. This is a protocol that is used in the computer simulations that are reported in section Computer Simulations.

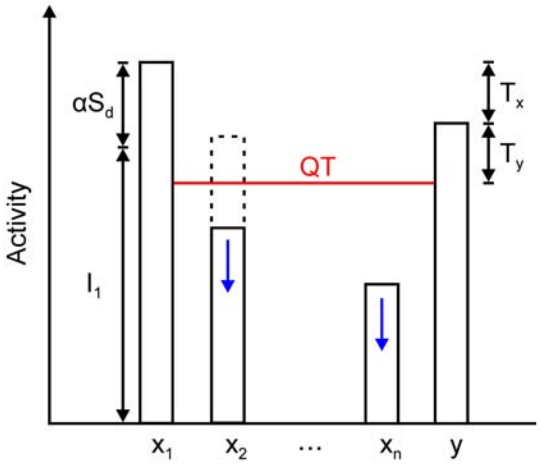


FIGURE 6 | Relationship among the steady state of the winning node x_1, inhibitory node y, and all other excitatory nodes in the network, $x_2 \ldots x_n$. The activity of the winning node is given by the sum of its feedforward input I_1 and the output of its dendrite mediating self- and nearest neighbor excitation, which is expressed as αS_d. Inhibitory node y approximately converges to $x_1 - T_x$. It sets the quenching threshold (QT) that separates excitatory nodes into two sets. Nodes $x_2 \ldots x_n$ are spared from inhibition if their activity is above the QT (dashed line); otherwise, they are silenced to zero (solid line). QT equals $y - T_y$ (or $x_1 - T_x - T_y$) because the activity of the inhibitory node must exceed the threshold on its presynaptic terminals that contact the excitatory nodes.

Linearization Near Fixed Points

To simplify the stability analysis, we consider an F-WTA network with two excitatory nodes and one inhibitory node: [x_1, x_2, y]. This system has three fixed points: x_1 is the only winner, x_2 is the only winner, and both excitatory nodes are winners. To which fixed point the network will converge depends on the relationship between inputs I_1 and I_2.

Local stability of the fixed point is estimated from the eigenvalues of the Jacobian matrix, which is the matrix of partial derivatives of the system of equations. If the real parts of all eigenvalues of the Jacobian are negative, the fixed point will be asymptotically stable (Rutishauser and Douglas, 2009). However, before we can compute the Jacobian matrix, we note that a linear-threshold function is continuous, but not differentiable. To sidestep this problem, we follow the approach that was described by Rutishauser et al. (2011) of inserting dummy terms that correspond to the derivate. That is, we need three separate dummy terms: c_i and p_{xi}, which correspond to the somatic and presynaptic output functions of excitatory node i, and a set of p_{yi} dummy terms that describe the presynaptic output function of inhibitory node y. The dummy terms are defined as.

$$c_i = p_{xi} = p_{yi} = \frac{d}{du}[u_i(t)]^+ = \begin{cases} 0 & \text{if } u_i(t) \leq 0 \\ 1 & \text{if } u_i(t) > 0. \end{cases} \tag{15}$$

Based on the above definition of the dummy terms, we have constructed the Jacobian matrix of

the system that consists of Equations (1) and (2):

$$J = \begin{bmatrix} \tau_x^{-1}\left(c_1\left(\alpha D_1 f + \beta_1 p_{y1}\right) - 1\right) & \tau_x^{-1} c_1 \alpha D_2 f & \tau_x^{-1} c_1 \beta_1 p_{y1} \\ \tau_x^{-1} c_2 \alpha D_1 f & \tau_x^{-1}\left(c_2\left(\alpha D_2 f + \beta_1 p_{y2}\right) - 1\right) & \tau_x^{-1} c_2 \beta_1 p_{y2} \\ \tau_y^{-1} \beta_2 p_{x1} & \tau_y^{-1} \beta_2 p_{x2} & -\tau_y^{-1}\left(\beta_2\left(p_{x1} + p_{x2}\right) - 1\right) \end{bmatrix} \quad (16)$$

where $D_1 f$ and $D_2 f$ denote the partial derivatives of the sigmoid function with respect to x_1 and x_2. Now, we examine the Jacobian matrix at the three fixed points that are mentioned above. If x_1 is the only winner, then $c_1 = 1$. However, $D_{x1} f \approx 0$ because the recurrent excitation of the winning node approaches its asymptotic value, which is S_d. In addition, $p_{y1} = 0$ because the winning node blocks inhibition from node *y,* as discussed above. Node x_2 is inhibited below its somatic threshold, that is, $c_2 = 0$. Presynaptic signaling by inhibitory node *y* blocks excitation from x_1 and x_2 is inactive, so $p_{x1} = p_{x2} = 0$. Consequently, the Jacobian matrix at the fixed point reduces to a diagonal matrix of the form

$$J_{W1} = J_{W2} = J_{W12} = \begin{bmatrix} -\tau_x^{-1} & 0 & 0 \\ 0 & -\tau_x^{-1} & 0 \\ 0 & 0 & -\tau_y^{-1} \end{bmatrix}. \quad (17)$$

All eigenvalues of the J_{W1} are negative, and the fixed point is asymptotically stable. In the case when x_2 is the sole winner, the same arguments are applied to set the dummy terms, thereby leading to the same diagonal matrix J_{W2} as shown in Equation (17). Moreover, if both excitatory nodes are winners, then $c_1 = c_2 = 1$, $D_{x1} f = D_{x2} f \approx 0$ and $p_{x1} = p_{x2} = 0$. Again, the Jacobian matrix J_{W12} is diagonal. Thus, all three fixed points are asymptotically stable.

The same analysis can be generalized to a network of arbitrary size and arbitrarily many fixed points. Retrograde signaling and dendritic saturation will ensure that the Jacobian matrix of any size will be diagonal and that the network dynamics will be independent of the network parameters, namely, α, β_1, and β_2. Local stability analysis suggests that the system behaves much like a feedforward network that is driven by the input. However, an important difference is that the F-WTA network has memory states like the recurrent network (Usher and Cohen, 1999; Rutishauser and Douglas, 2009).

COMPUTER SIMULATIONS

We performed a set of computer simulations to illustrate the model behavior. We employed a vector of 200 excitatory units and one inhibitory unit. Differential Equations (1) and (2) were solved numerically using MATLAB's *ode15s* solver. The simulations were run for 250 time steps. In subsequent figures, we followed the convention that activity of the node at position *i* as a function of time is depicted by a shade of gray, with white representing the maximal value and black representing zero.

Simulation of the Formation of a Single Boolean Map

First, we demonstrate how a Boolean map arises in the F-WTA network in response to the presentation of the color cue, as illustrated by **Figure 1A**. In **Figure 7A**, we recreate a similar stimulus condition in the 1-D map. The input consists of red and green items of equal sizes, which are intermixed in space on a black background. Input magnitude I was set to 1 in both maps and to 0.2 in the empty space around items to represent spontaneous activity in the absence of visual stimulation. Initially, the top-down or attentional gain is set to $G^m = 1$ in both feature maps$m \in \{red, green\}$. At $t = 50$, the red color is attended, which is reflected in the input to the network by increasing the gain for all nodes in the Red map ($G^{red} = 2$) and simultaneously reducing the gain in the Green map by the same factor ($G^{green} = 1/G^{red} = 1/2$). Top-down gain is also applied to the empty space between items, which is consistent with the finding that feature-based attention spreads across the whole visual field (Saenz et al., 2002, 2003; Serences and Boynton, 2007). The duration of the top-down cue is 50 simulated time steps. For simplicity, top-down signals are suddenly switched on and off without exponential decay. At $t = 150$, the green color is cued in the same way.

At the beginning of the simulation, before the top-down signals are applied, the F-WTA network simply selects all presented items together, irrespective of their color. Next, when the red color is cued by applying top-down signals to the corresponding feature map, the network responds to the new input by selectively increasing and sustaining the activity of nodes that encode locations of red items in the input and suppressing locations that encode green items. That is, the network creates a Boolean map by highlighting the spatial pattern that is associated with the red color. Furthermore, due to a self-excitation, the network maintains locations of the cued feature value in working memory after the top-down signals cease to influence the feature map. When the observer decides to switch attention to another feature value, the network can select the locations of the new feature value and suppress the locations that are associated with the previously cued value without requiring an external reset. Namely, the network is sensitive to input changes even though it also exhibits activity persistence.

Importantly, the activity level at selected locations is invariant with respect to the number of active nodes. At the beginning of the simulation, the number of active nodes was four times larger than after the cue was delivered. However, the active nodes remained at the same activity level as they were at the beginning of the simulation. This is a consequence of retrograde inhibitory signaling in recurrent pathways. It prevents unbounded growth of inhibition due to the dynamic regulation of its strength. To illustrate this point further, we run another simulation with items that are almost double in size (**Figure 7B**). Even though the total size of the cued items is increased, the activity of the cued nodes converges to the same level as before. In this simulation, we also checked that the network successfully operates even if we remove gain reduction from the non-attended feature map.

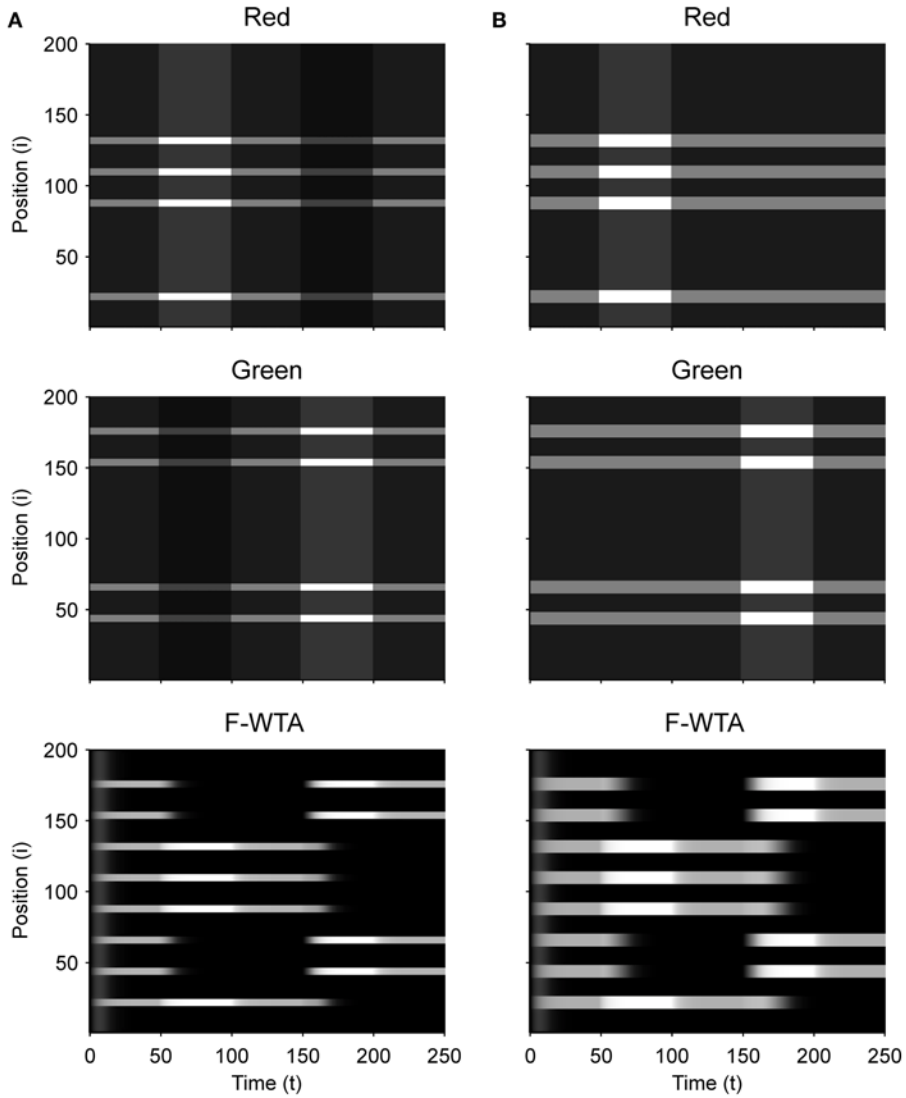


FIGURE 7 | (A) Simulation of the Boolean map formation in the F-WTA network in response to the sequential presentation of two color cues (red appeared between 50th and 100th and green appeared between 150th and 200th simulated time unit). **(B)** The same simulation with larger items and without gain modulation applied on the unattended feature map.

Next, we determined the minimal feature gain that must be applied on the input to produce the desired behavior. When the gain modulation is applied simultaneously on attended feature map G^A and on unattended feature map G^{NA} (where $G^{NA} = 1/G^A$), we found that $G^A \geq 1.7$ is sufficient for creating a Boolean map and switching to another one. In contrast, when the gain modulation is not applied on the unattended feature map, as shown in **Figure 7B**, the feature gain in the attended map should be set to $G^A \geq 2$ to achieve the same behavior.

Figure 8 illustrates that the F-WTA network can support space- and object-based attention alongside feature-based attention. When the spatial cue is applied to a single location in one of the feature maps, the network responds by selecting only this location. Neighboring nodes are not selected even though they are reciprocally connected to the cued node. The reason is that they receive weaker input relative to the cued node. Furthermore, recurrent excitation that arrives from the cued node is bound by the dendritic non-linearity. Thus, it is not sufficiently strong to keep them active. Interestingly, when the spatial cue is removed, the network activity starts to propagate from the cued node toward the boundary of the whole item. In this case, the network selects not just the cued location, but all locations that are connected to it. Therefore, the F-WTA network exhibits object-based selection, which is consistent with neurophysiological studies that show spreading of enhanced activity along the shape of the object (Roelfsema, 2006). This property arises because the removal of the cue equalizes the input magnitude along the object, which allows activity enhancement to propagate via local lateral connections.

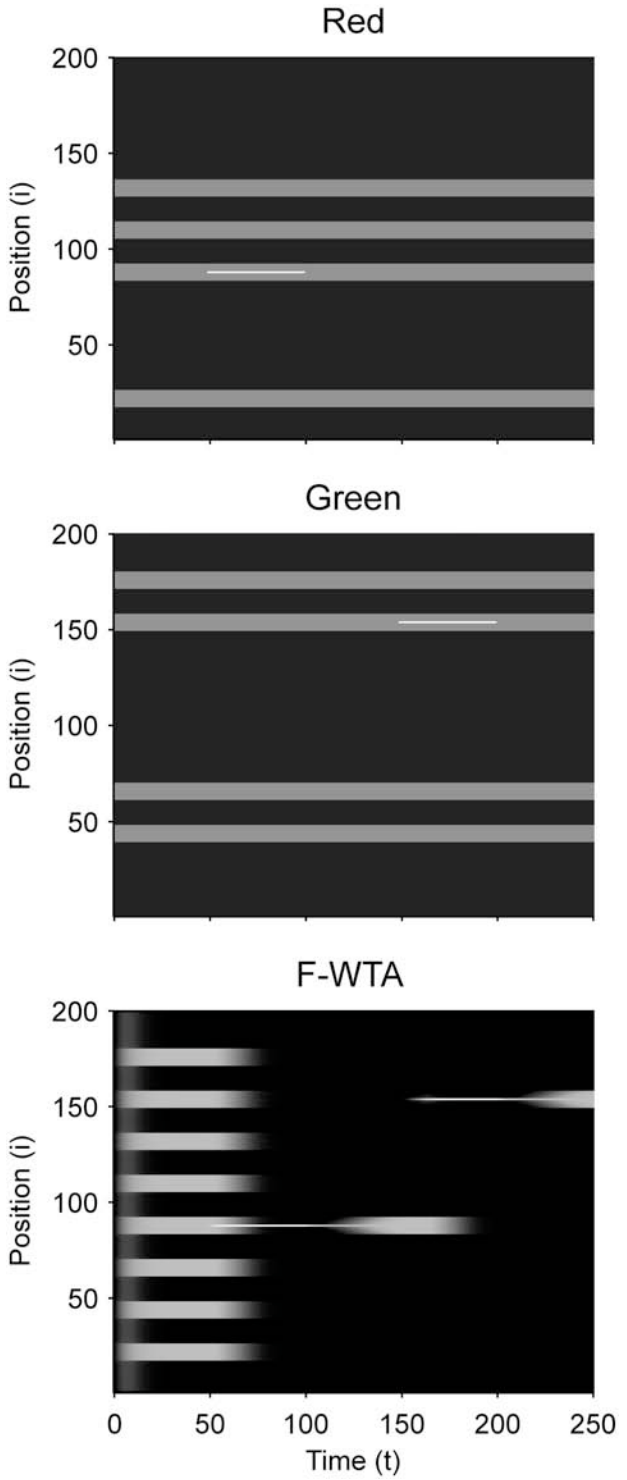


FIGURE 8 | Simulation of space- and object-based attention in the F-WTA network.

In addition, this simulation shows that spatial attention can be easily oriented toward a new location in a single jump without the need for attentional pointers that move attention across the map (Hahnloser et al., 1999).

Simulation of the Intersection and Union of Two Boolean Maps

Figure 9 illustrates that the model can sequentially combine two Boolean maps when the network is cued by top-down signals from two separate feature dimensions. In this simulation, we have employed a visual input that consists of red and green horizontal and red and green vertical bars, like those that are illustrated in **Figure 1B**. First, the F-WTA network is cued to select red bars, irrespective of their orientation. In the second step, it is cued to select horizontal bars, irrespective of their color. However, green vertical bars are already suppressed and the top-down signal that is supplied to them is not sufficient to override the inhibition that arises from red vertical bars. The net result is the selection of a subset of red horizontal bars. In other words, the network activity converges to an intersection between a set of red bars and a set of horizontal bars, thereby resulting in the selection of red horizontal bars.

Next, we examined how the network achieves the union of two Boolean maps (**Figure 10**). Here, we assumed that the input consists of two non-overlapping components: colored squares that activate color maps but do not activate orientation maps, and achromatic horizontal and vertical bars that activate orientation maps but do not activate color maps, as shown in **Figure 1C**. Red-colored items occupy locations between 1 and 100 and oriented bars occupy locations between 101 and 200. This closely resembles the stimulus that is used by Huang and Pashler (2007) to demonstrate the union of color and texture. Taken together, the data show that the union of two Boolean maps is possible only when two top-down cues overlap in time or when the second cue closely follows the withdrawal of the first cue. In **Figure 10**, the cue for the red map is applied in the interval [50, 100] and the cue for the horizontal map is applied in the interval [110, 160]. In this case, the F-WTA network converges to the union of red and horizontal items. However, when top-down cues do not overlap, as shown in **Figure 11**, the second cue overrides the network activity that remains from the first cue. We suggest that this property partly explains why the union is difficult to achieve, as observed by Huang and Pashler (2007).

In addition, we examine the boundary conditions on the choice of the feature gain parameter. We parametrically vary the feature gain in steps of 0.1 starting from $G = 2$ and moving below and above to determine when the ability to form the intersection or union breaks down. When the gain modulation is applied simultaneously on attended (G^A) and unattended (G^{NA}) feature maps, we find that G^A should be chosen from the interval [1.5, 2.1] to achieve the intersection between two maps. When $G^A < 1.5$, the network fails to segregate cued from non-cued locations in the first step. In contrast, when $G^A > 2.1$, the network successfully segregates cued from non-cued locations in the first step. However, the gain is too high, so all horizontal items are selected together in the second step. That is, the representation of red horizontal items is merged with the representation of green horizontal items. When $G^{NA} = 1$ throughout the simulation, G^A should be chosen from the interval [1.8, 2.0] to achieve intersection.

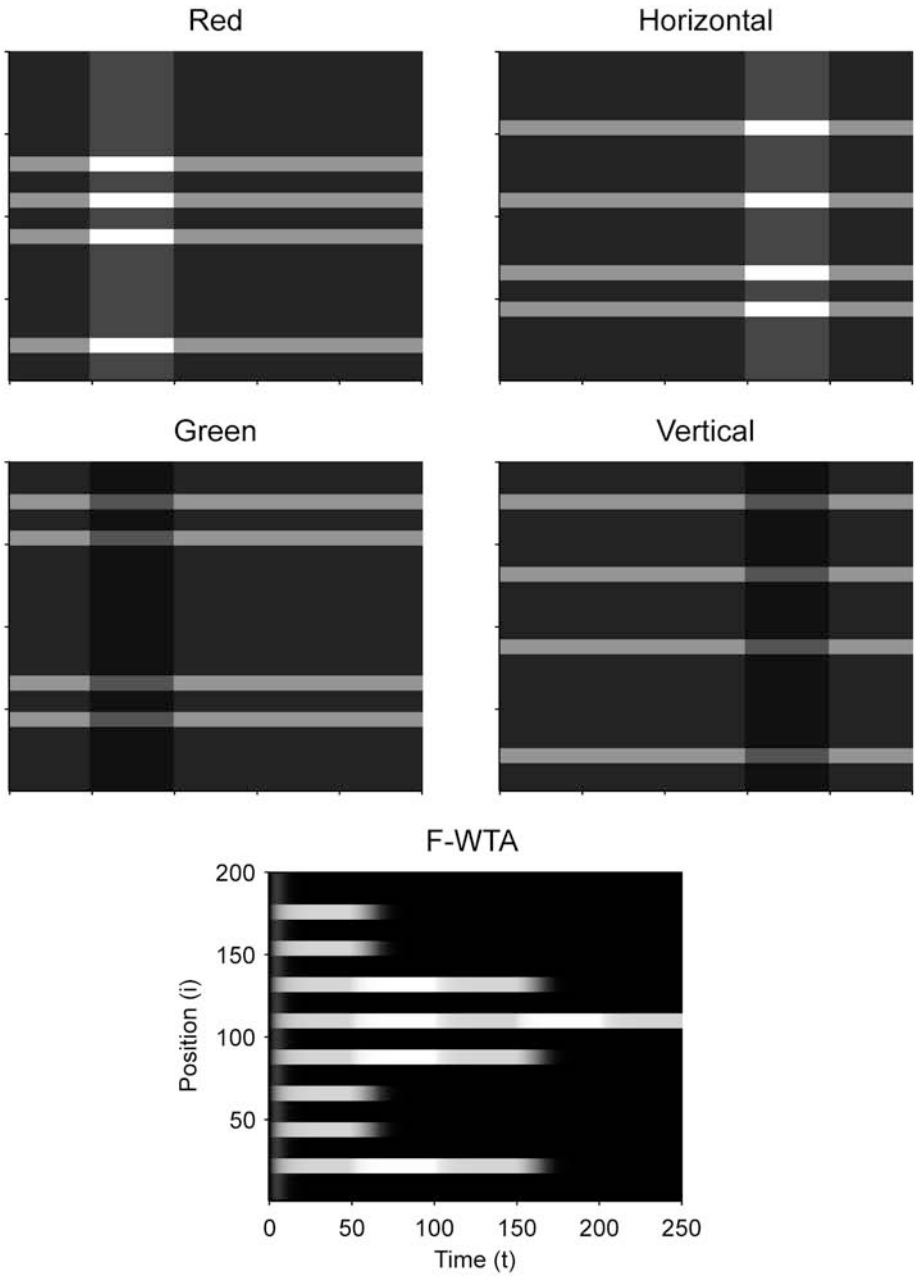


FIGURE 9 | Simulation of the intersection of red and horizontal items.

With respect to the union of two maps, the feature gain G^A should be chosen from the interval [1.4, 2.0] when $G^{NA} = 1/G^A$ and from the interval [1.6, 2.0] when $G^{NA} = 1$. When G^A is chosen below the suggested intervals, feature gain is too weak, and the second cue will not be able to raise the activity level of the nodes that represent horizontal items above the quenching threshold. Therefore, the network ends up with the Boolean map of red items that is formed in the first step. When G^A is chosen above the suggested interval, the network switches between the representation of the red items in the first step to the representation of the horizontal items in the second step. In this case, the feature cue is too high, and the activity of the nodes that represent horizontal items simply overrides the activity of the nodes that represent the red items. These constraints are derived from the situation in which the two top-down cues overlap in time. As shown above, temporal lag of the second cue relative to the first cue also destroys the ability of the network to form the union of two Boolean maps.

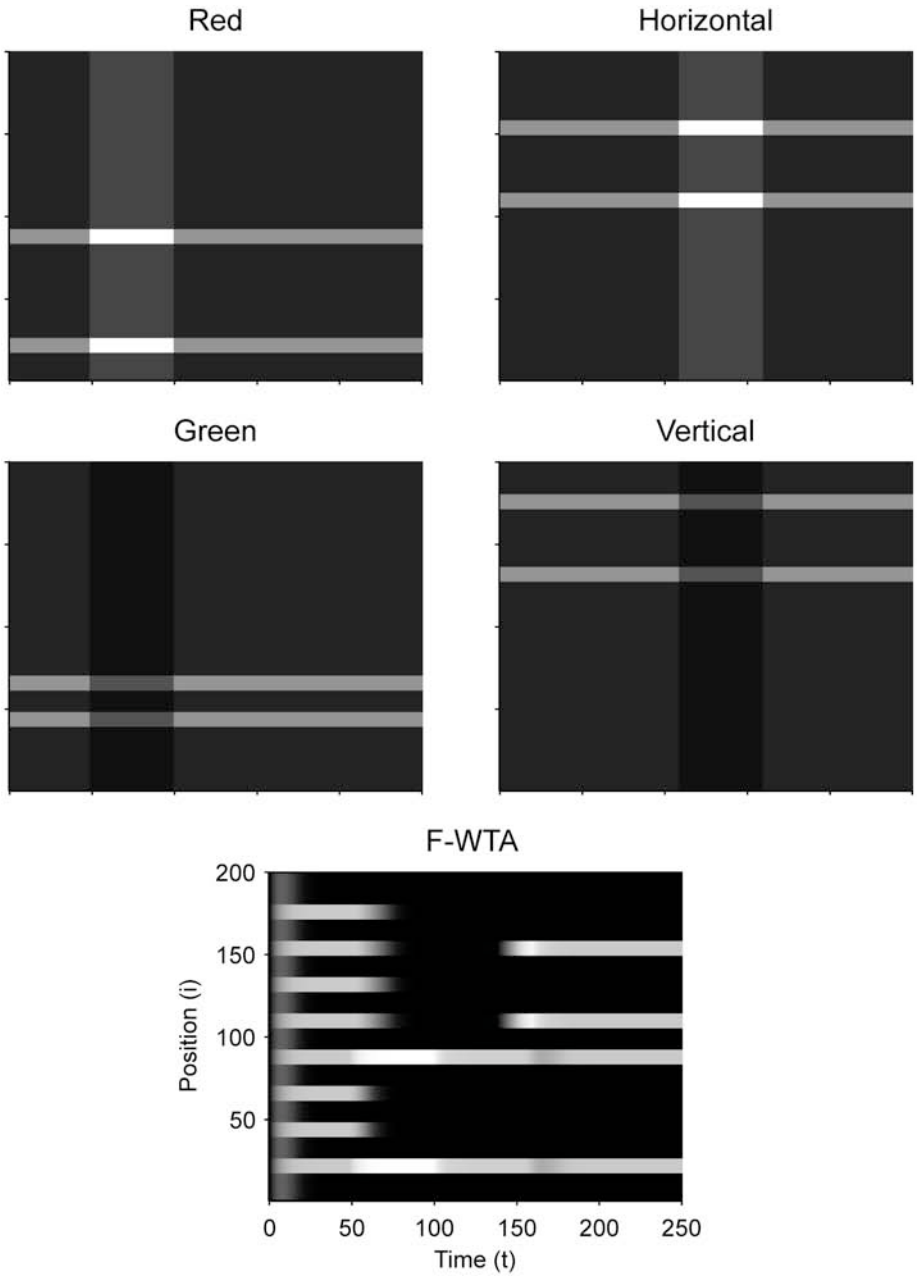


FIGURE 10 | Simulation of the union of red and horizontal items.

Simulation of Bottom-Up Spatial Selection

Finally, we have shown that when there is no top-down guidance, the network selects the most-salient locations based on the bottom-up salience that is computed within feature maps (**Figure 12**). We did not explicitly model competition among maps, but it is reasonable to assume that in a scene with many multi-featured objects, their input magnitudes (i.e., saliencies) will be different. Therefore, we arbitrarily assigned different input magnitudes to different items. As shown in **Figure 12A**, the F-WTA network selects the most salient object if the difference in input magnitude between the two most active nodes is sufficiently large. However, when this difference is small, as shown in **Figure 12B**, the F-WTA model chooses two most salient items together. Furthermore, in both examples, the network activity retains the input amplitude of the winning item (or items), thereby illustrating the ability to compute the function maximum (Yu et al., 2002).

The precision of saliency detection depends on the threshold for the activation of synaptic receptors on the inhibitory interneuron. In all reported simulations, it was set to $T_y = 0.1$.

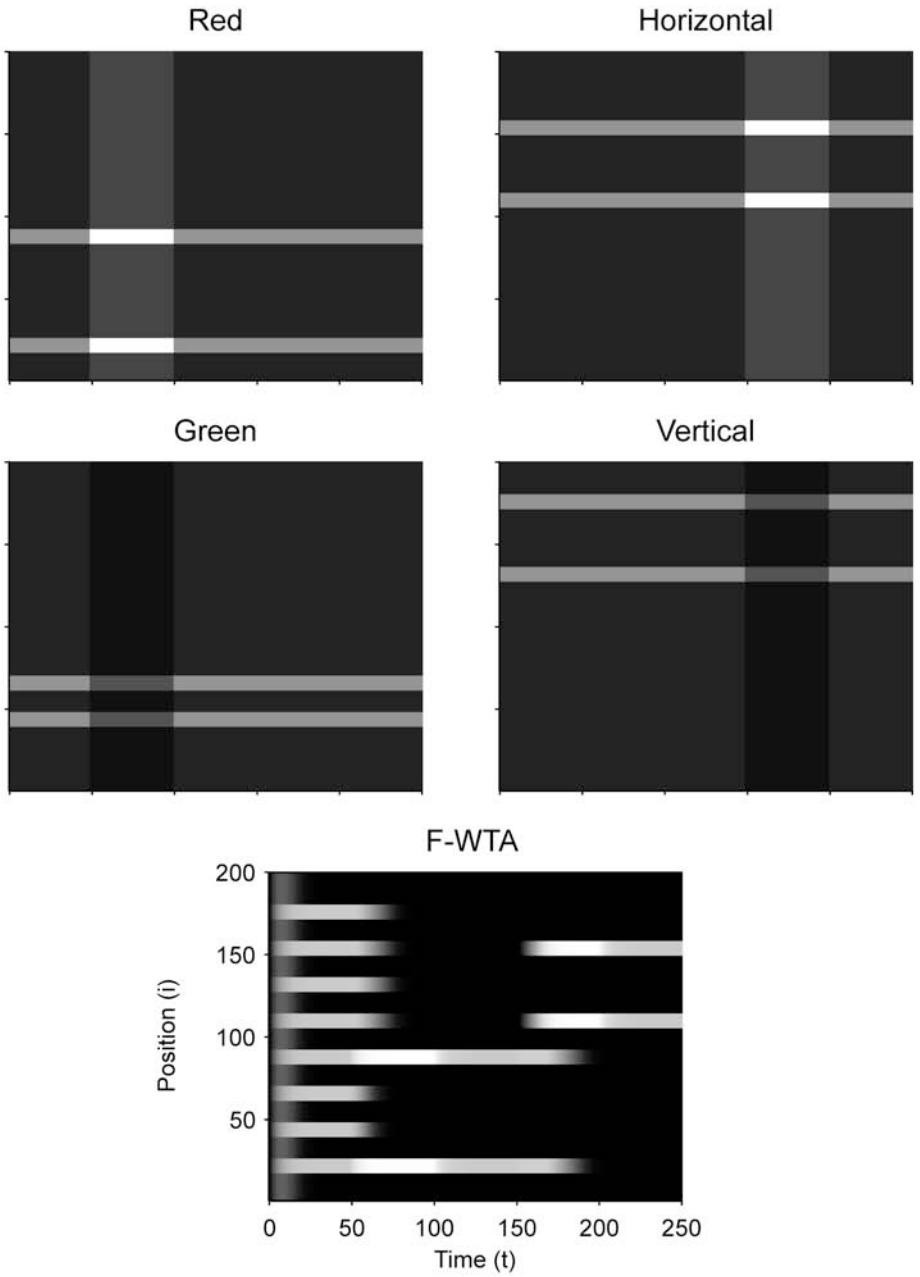


FIGURE 11 | Breakdown of union of red and horizontal items when the delivery of top-down cues for red and horizontal items is separated by a large temporal gap.

If smaller values were chosen, the network would improve in terms of precision and be able to separate the two objects that are presented in **Figure 12B**. However, this comes at the price of losing the ability to form a union of two Boolean maps. Therefore, there is a trade-off between the precision of saliency detection and the ability to form Boolean maps.

An important aspect of stimulus-driven attentional control is attentional capture by peripheral cues. Behavioral studies have shown that the abrupt onset of a new object in a visual scene can automatically capture attention even if it is irrelevant for the current goal (Theeuwes, 2010). **Figure 13** illustrates the sensitivity of the F-WTA network to abrupt visual onset. To simulate this effect, we have made the additional assumption that the network receives input not only from a sustained channel that is comprised of feature maps in V4 but also from a transient channel that responds vigorously only to changes in input (Kulikowski and Tolhurst, 1973; Legge, 1978). Thus, when the abrupt onset is accompanied by a strong transient signal that exceeds the activity level of the currently attended item, the F-WTA network temporarily switch activity toward the location of

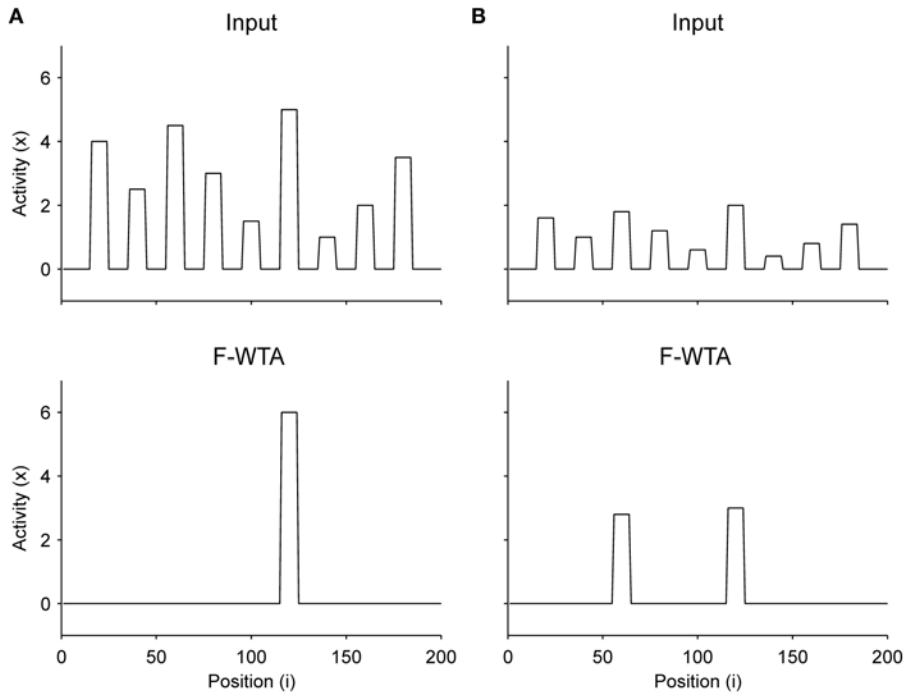


FIGURE 12 | Selection of the most salient item in the absence of top-down guidance. **(A)** When the most salient item is sufficiently distinctive from other items, the F-WTA network selects it. **(B)** When the saliency of all items is relatively low, the F-WTA network may select more than one item because it has a limit on the precision by which it separates inputs of different magnitudes.

the onset (**Figure 13A**). Here, the input at the locations that are occupied by the winning item in the center of the map was set to $I_W = 2$. Input to all other items was set to $I_i = 1$. Finally, the transient input that appears on the sides of the map was set to $I_T = 4$. It is sufficient to set $I_T \geq I_W + 0.8$ to achieve sensitivity to abrupt onsets. Moreover, the same relation holds even if we choose a larger value for I_W.

Next, when abrupt onset produces only weak transient signals ($I_T = 2$) that do not satisfy the inequality that is stated above ($I_W = 2$), the activity in the F-WTA network resists abrupt onset and stays on the previously attended item (**Figure 13B**). This observation is consistent with behavioral findings that abrupt onset can be ignored (Theeuwes, 2010), perhaps by attenuating the response of the transient channel. Another possibility is that the top-down gain for the attended location can be increased so that it exceeds the activity of the transient channel. In this case, intense focus on the current object prevents attentional capture, which is consistent with the psychological concept of the attentional window (Belopolsky and Theeuwes, 2010).

DISCUSSION

We have proposed a new model of the WTA network that can simultaneously select multiple spatial locations based on a shared feature value. We named the model the feature-based WTA (F-WTA) network because the unit of selection is not a point in space or object, but rather an abstract feature value that is set by the top-down signals. We have demonstrated how the F-WTA network implements the central proposal of the Boolean theory of visual attention that there exists a spatial map that divides the visual space into two mutually exclusive sets. One set represents all locations that are occupied by the chosen feature value. The other set contains all other locations, which are not of interest. The Boolean map controls spatial selection and access to the consciousness (Huang and Pashler, 2007). Moreover, we have shown that the network successfully integrates information across space and time to form the intersection or union of two maps that are defined by different feature cues. Previous models of the WTA network are not capable of such integration because they require that the current winner be externally inhibited to allow attentional focus to move from one location to another (Kaski and Kohonen, 1994; Itti and Koch, 2000, 2001). Another possibility to move activity across locations in the network is to introduce dynamic thresholds that simulate habituation or fatigue in individual neurons. In this case, current winner loses its competitive advantage due to the raise of its threshold. This allows non-winners to gain access to working memory (Horn and Usher, 1990). However, both approaches are not suitable for forming the intersection or union of a set of previous winners and a set of later winners.

Another important property of the F-WTA network that sets it apart from previous models of WTA behavior is the ability to select and store arbitrarily many locations in the memory. This is achieved by inhibitory retrograde signaling, which effectively isolates winning nodes from mutual inhibition. First, the amount

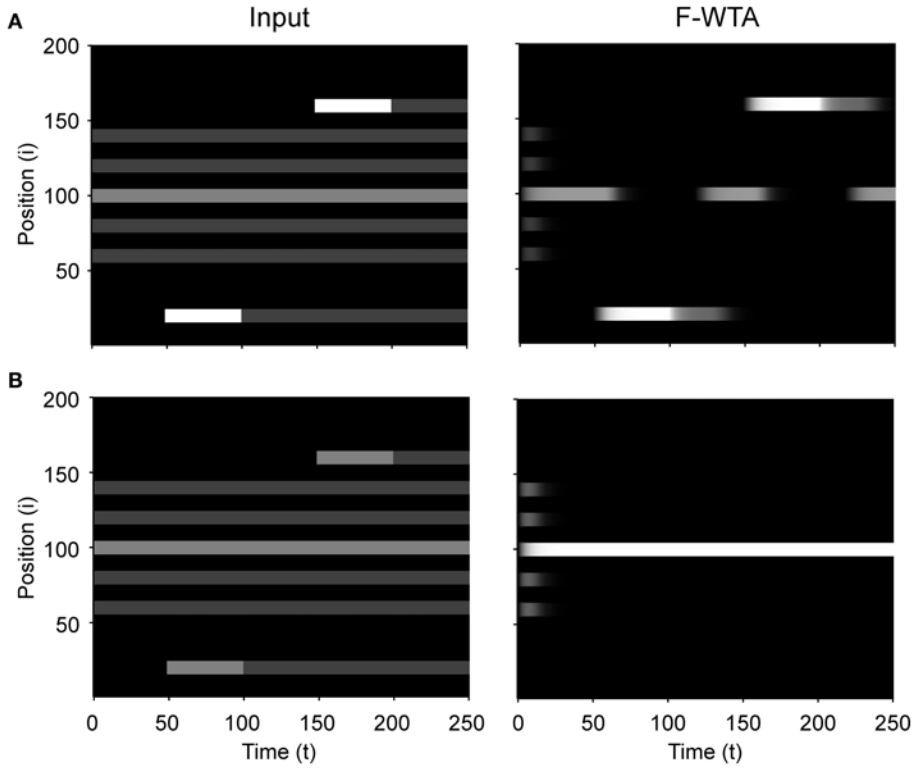


FIGURE 13 | Sensitivity to abrupt visual onsets. **(A)** When the transient signal that is produced by the abrupt onset of a new object is sufficiently strong, it temporarily draws attention to itself. **(B)** When the transient signal is weak, attention resists abrupt onset and stays on the item that was selected at the beginning of the simulation.

of inhibition in the network is significantly reduced because the inhibitory interneuron computes the maximum instead of the sum of the recurrent input that it receives from the excitatory nodes. Second, the winning excitatory nodes release their retrograde signals and block inhibition from the interneuron. Consequently, arbitrarily many winners can participate in representing the selected locations without degrading their activation. In other words, there is no capacity limit on the number of objects that can be simultaneously selected. This is consistent with recent behavioral findings that suggest that our ability to select multiple objects is not fixed. Rather, spatial attention should be considered a fundamentally continuous resource without a strict capacity limit (Davis et al., 2000, 2001; Alvarez and Franconeri, 2007; Liverence and Franconeri, 2015; Scimeca and Franconeri, 2015).

In addition, the network is sensitive to the sudden appearance of a new object in the scene, which suggests that it can also be guided by bottom-up feature cues (Theeuwes, 2013). We hypothesize that the network receives strong input from the transient channel. Such input overrides the network's current memory state, thereby making it sensitive to abrupt onsets. Moreover, the transient channel can be activated by any type of change in the spatiotemporal energy of the input, and not just by the sudden appearance (or disappearance) of objects. For example, it will be activated by a sudden change in the direction of motion (Farid, 2002). When the network simultaneously receives transient input from different locations, they all will be selected together. In this way, the network achieves temporal grouping of synchronous transient input. That is, the network can discover spatial structures that are defined purely by temporal cues (Lee and Blake, 1999; Rideaux et al., 2016).

Biophysical Considerations

As noted above, the model of the F-WTA network rests upon three key computational elements: the dendrite as an independent computational unit, retrograde signaling on synaptic contacts, and computing the maximum over inputs. Here, we review supporting neuroscientific evidence that suggests that all three biophysical mechanisms are plausible candidates for computation in real neural networks.

There is a growing body of evidence that the excitatory pyramidal cell should not be viewed as a single electrical compartment. Rather, it consists of multiple independent synaptic integration zones arranged in a two-layer hierarchy (Häusser and Mel, 2003; London and Häusser, 2005; Branco and Häusser, 2010; Mel, 2016). Using a detailed biophysical model of the pyramidal neuron, Poirazi et al. (2003) showed that its output is well approximated by a two-layer neural network. In the first layer of the network, dendrites independently integrate their synaptic input and produce sigmoidal output. In the second layer, the dendritic output is summed at the soma to produce the neuron's firing rate. Importantly, the somatic and dendritic

output functions need not be the same (Jadi et al., 2014). For example, Behabadi and Mel (2014) showed that the soma of the model neuron generates nearly linear output, while the dendritic output is sigmoid. In our model, the dendrite conveys recurrent excitation to the node. Due to the dendritic non-linearity, there is no risk of unbounded activity growth in the node. Furthermore, the dendritic output is summed with the external input at the soma of the node. By using a linear output function at the soma, we have ensured that the F-WTA network remains sensitive to input fluctuations.

Synaptic transmission can be dynamically regulated in an activity-dependent manner, as shown by the existence of depolarization-induced suppression of inhibition (DSI) (Pitler and Alger, 1992) and depolarization-induced suppression of excitation (DSE) (Kreitzer and Regehr, 2001). DSI (DSE) refers to the reduction in inhibitory (excitatory) post-synaptic potentials following depolarization of the postsynaptic cell. These processes have been observed in various brain regions, including the cerebellum, hippocampus, and neocortex. A retrograde messenger that is released from postsynaptic cell due to its depolarization mediates DSI and DSE. After release, the retrograde messenger binds to the receptors at the presynaptic axon terminals and suppresses the release of the transmitter. Based on these properties, Regehr et al. (2009) suggested that a possible physiological function of DSI and DSE is to provide negative feedback that reduces the impact of the synaptic input on the ongoing neural activity.

The model behavior rests upon the assumption that the inhibitory interneuron computes the maximum instead of the sum of its inputs. There is some direct physiological evidence that real cortical neurons indeed compute the MAX function. For example, Sato (1989) examined responses of neurons in the primate inferior temporal cortex to the presentation of one or two bars in their receptive field. He concluded that the responses to two bars that were presented simultaneously were well described by the maximum of the responses to each separately. In a similar vein, Gawne and Martin (2002) recorded the activity of neurons in primate V4 and found that their firing rate in response to the combination of stimuli is best described by the maximum function over the firing rates that are evoked by each stimulus alone. Furthermore, Lampl et al. (2004) directly measured membrane potentials in the complex cells of the cat primary visual cortex and found evidence for the MAX-like behavior in response to the pair of optimal bars.

Indirectly, the importance of the MAX-like operation in cortical information processing can be appreciated by considering the many computational models of visual functions that have employed it in simulating rich and complex datasets. For example, Riesenhuber and Poggio (1999) employed hierarchical computation of the MAX function in a model of invariant object recognition. Spratling (2010, 2011) used it in simulating a large range of classical and non-classical receptive field properties of V1 neurons. Moreover, Tsui et al. (2010) used MAX-like input integration to explain diverse properties of MT neurons and Hamker (2004) used it in his model of top-down guidance of spatial attention. Furthermore, Kouh and Poggio (2008) developed a canonical cortical circuit that is capable of many non-linear operations, including computation of the MAX function. Here, we have shown that a single inhibitory node that is endowed with retrograde signaling can compute the maximum.

Based on the proposed model, we have derived two testable predictions. The cortical network that is involved in spatial selection will contain inhibitory interneurons that can compute the MAX function. Moreover, both the excitatory and inhibitory neurons in this network will be endowed with the anatomical structures that support retrograde signaling (presynaptic receptors and postsynaptic transmitter release sites).

Comparison With Other WTA Network Models

Several models of biophysical mechanisms have been proposed for implementing WTA behavior in a neural network, including linear-threshold units (Hahnloser, 1998; Rutishauser and Douglas, 2009), non-linear shunting units (Grossberg, 1973; Fukai and Tanaka, 1997), and oscillatory units (Wang, 1999; Borisyuk and Kazanovich, 2004).

A simple model of a competitive network that is based on linear-threshold units has been extensively studied. Stability analysis revealed that this network requires fine-tuning of the connectivity to achieve stable dynamics that can perform cognitively relevant computations, such as choice behavior (Hahnloser, 1998; Hahnloser et al., 2003; Rutishauser et al., 2015). Recently, Binas et al. (2014) showed that a biophysically plausible learning mechanism could tune the network connections in a way that keeps the network dynamics in the stable regime. Here, we have shown how dendritic and synaptic non-linearities ensure that the network dynamics near fixed points depends only on the time constants of the nodes and not on the parameters that control recurrent excitation and lateral inhibition. Therefore, a precise balance between excitation and inhibition is not necessary for achieving a stable memory state. Moreover, the network is sensitive to the input and can iteratively combine the current memory state with new input to form the intersection or union of them.

An important problem for WTA networks that are based on the linear-threshold or sigmoid output functions is that they lack a mechanism for controlling inhibition between the winning nodes. Therefore, they have limited capacity to represent multiple winners. Usher and Cohen (1999) showed that their activation decreases up to the point of complete inactivation as the number of winning nodes increases. This is due to the increased amount of mutual inhibition. The problem cannot be solved simply by reducing the strength of the lateral inhibition because it is not known in advance how many locations will be cued. On the other hand, feature-based spatial selection requires that the network be able to adjust automatically the amount of inhibition to accommodate the selection of a very small or very large number of winners.

Grossberg (1973) proposed a recurrent competitive map model that was based on shunting non-linear interaction between the synaptic input and the membrane potential. The output of the model depends on the exact form of the signal function that is used to convert membrane potential into the firing rate.

When the signal function is chosen to grow faster than linear, the network exhibits WTA behavior. By contrast, when the signal function is sigmoid, the network can select multiple winners if they have similar activity levels. The most important property of this model is the existence of the quenching threshold. All nodes whose activity is above QT are enhanced and all nodes whose activity is below QT are suppressed. This behavior is similar to the operation of the F-WTA network that was proposed here. However, an important difference is that in the shunting model, QT is fixed and dependent on the parameters of the network. In contrast, the feature-based WTA network exhibits dynamic QT that depends on the input to the network and not on its parameters. In this way, the F-WTA network rescales its sensitivity to the input fluctuations.

More recently, a version of the recurrent competitive map was applied in modeling object-based attention (Fazl et al., 2009). It was shown that sustained network activity in the model PPC encompasses the whole object as an attentional shroud around it. Such spatial representation of a single object supports view-invariant object recognition within a larger neural architecture, namely, ARTSCAN. In an extension of the model, Foley et al. (2012) proposed two separate competitive networks that account for distinct properties of object- and space-based attention. A network with strong inhibition is limited to the selection of a single object. The other network utilizes weaker inhibition to support multifocal spatial selection. To increase the capacity of this network to represent multiple objects, Foley et al. (2012) suggested that the amount of lateral inhibition could be controlled externally. As the number of objects that should be selected together increases, the lateral inhibition should become weaker to counteract the effect of the larger number of nodes that participate in the competition. In contrast, the F-WTA network does not require such external adjustments of the strength of the lateral inhibition to accommodate the selection of arbitrarily many objects of arbitrary size. Moreover, in the F-WTA network, object-based and multifocal spatial attention coexist within the same circuit. Whether the network exhibits object-based spatial selection depends on the type of cue that is presented to the network and not on its parameters.

Wang (1999) proposed a model of object-based attention that relies on the phase synchronization and desynchronization among oscillatory units. At each location of the recurrent map, there is a pair of excitatory and inhibitory units with distinct temporal dynamics that creates a relaxation oscillator. Excitatory units are also mutually connected with their nearest neighbors and with a global inhibitor. The network is initialized with random phase differences between oscillators at different network locations. The activity of the global inhibitor further enforces phase separation among excitatory units. However, local excitatory interactions among nearest neighbors oppose global inhibition and result in phase synchronization that spreads among nodes that encode the same object. The net result of these interactions is temporal segmentation and selection of one active object representation at a time in a multi-object input image. Importantly, the network can switch its activity from one object representation to another. However, this transition is generated internally by the oscillator dynamics. It is not possible to drive the object selection by external cues such as top-down gain control or bottom-up cues such as abrupt onsets. Moreover, it is not possible to enforce simultaneous selection of more than one object by a joint feature value because the global inhibitor will desynchronize all nodes that encode non-connected items. Therefore, it is not clear how synchronous oscillations could support feature-based attentional selection. Taken together, it is still an open issue whether they are relevant for perception and cognition (Ray and Maunsell, 2015).

Limitations

The proposed model of spatial selection successfully simulates the formation of the Boolean map and its elaboration by the set operations of intersection and union but does not fully implement all aspects of the theory that was proposed by Huang and Pashler (2007). Precisely, it does not explain why attention is limited to only one feature value per dimension or how the observer sequentially chooses one feature value after another or combines feature dimensions into intersections or unions of Boolean maps. It is likely that this severe limitation arises from some form of the WTA network. However, this constraint requires a more elaborate model of the interactions among the spatially invariant representation of the feature values in the IT cortex and the interactions between the IT and the prefrontal cortex, where decisions and plans are made.

In all simulations that are reported here, we kept items segregated in space. This was not the case in the stimuli that were used by Huang and Pashler (2007). They employed a matrix of colored squares that were connected to one another. This is because activity spreading can occur among adjacent nodes even if they encode different feature values. Activity spreading is observed after top-down signals stop favoring one feature value over the other. In this case, all feature maps contribute equally to the input of the F-WTA network and the network is no longer able to discriminate between selected and unselected feature values. One way to solve this issue is to assume that the top-down signals are constantly present during the whole trial. In this way, the activity magnitude on the cued locations is kept above that on the non-cued locations. Therefore, non-cued locations are treated as background noise and suppressed, despite their proximity to the cued locations. Another possibility is to impose boundary signals that act upon recurrent collaterals of the nodes in the F-WTA network in a way that is similar to how activity spreading is stopped in the network models of brightness perception (Grossberg and Todorović, 1988), visual segmentation (Domijan, 2004), and figure-ground organization (Domijan and Šetić, 2008).

Finally, input to the network does not follow the distance-dependent activity profile that is usually observed in the visual cortex. However, this is not a critical issue for the model's performance because the precision of selection depends on the thresholds for presynaptic terminal activation, namely, T_x, and T_y. If they are set to very small values, the network will tend to select the centers of the objects when the input pattern is convolved with a Gaussian filter. In contrast, if they are set to larger values, the network will be able to select extended parts of

the objects and possibly even the whole objects. In the same way, the model achieves resistance to the input noise. As thresholds are set to larger values, the network can tolerate a larger amount of noise. However, this comes at a cost of less-precise selection, as demonstrated by the simulation that is shown in **Figure 12**.

CONCLUSIONS

We have demonstrated how the feature-based WTA network achieves spatial selection of all locations that are occupied by the same feature value without suffering from capacity limitations. The network responds to the top-down cue by storing in memory spatial pattern that corresponds to the cued feature value, while non-cued feature values are suppressed. In this way, we have shown how the Boolean map is formed. In addition, we have shown that it is possible to create more complex spatial representations that involve the intersection or the union of two or more Boolean maps. In this way, the F-WTA network goes beyond the capabilities of previous models of the competitive neural network, which cannot integrate information across space and time. Our work suggests that dendritic non-linearity and retrograde signaling are biophysically plausible mechanisms that are essential for model success.

AUTHOR CONTRIBUTIONS

DD designed the study and write the manuscript. MM performed computer simulations and write the manuscript.

FUNDING

This research was supported by the Croatian Science Foundation Research Grant HRZZ-IP-11-2013-4139 and the University of Rijeka Grant 13.04.1.3.11.

REFERENCES

Alger, B. E. (2002). Retrograde signaling in the regulation of synaptic transmission: focus on endocannabinoids. *Prog. Neurobiol.* 68, 247–286. doi: 10.1016/S0301-0082(02)00080-1

Alvarez, G. A., and Franconeri, S. L. (2007). How many objects can you track? evidence for a resource-limited attentive tracking mechanism. *J. Vision* 7, 14.1–14.10. doi: 10.1167/7.13.14

Ashby, F. G., and Hélie, S. (2011). A tutorial on computational cognitive neuroscience: modeling the neurodynamics of cognition. *J. Math. Psychol.* 55, 273–289. doi: 10.1016/j.jmp.2011.04.003

Behabadi, B. F., and Mel, B. W. (2014). Mechanisms underlying subunit independence in pyramidal neuron dendrites. *Proc. Natl. Acad. Sci. U.S.A.* 111, 498–503. doi: 10.1073/pnas.1217645111

Belopolsky, A. V., and Theeuwes, J. (2010). No capture outside the attentional window. *Vision Res.* 50, 2543–2550. doi: 10.1016/j.visres.2010.08.023

Binas, J., Rutishauser, U., Indiveri, G., and Pfeiffer, M. (2014). Learning and stabilization of winner-take-all dynamics through interacting excitatory and inhibitory plasticity. *Front. Comput. Neurosci.* 8:68. doi: 10.3389/fncom.2014.00068

Borisyuk, R. M., and Kazanovich, Y. B. (2004). Oscillatory model of attention-guided object selection and novelty detection. *Neural Netw.* 17, 899–915. doi: 10.1016/j.neunet.2004.03.005

Boynton, G. M. (2005). Attention and visual perception. *Curr. Opin. Neurobiol.* 15, 465–469. doi: 10.1016/j.conb.2005.06.009

Boynton, G. M. (2009). A framework for describing the effects of attention on visual responses. *Vision Res.* 49, 1129–1143. doi: 10.1016/j.visres.2008.11.001

Braitenberg, V., and Schüz, A. (1991). *Anatomy of the Cortex. Statistics and Geometry,* Vol. 18. Berlin: Springer-Verlag.

Branco, T., and Häusser, M. (2010). The single dendritic branch as a fundamental functional unit in the nervous system. *Curr. Opin. Neurobiol.* 20, 494–502. doi: 10.1016/j.conb.2010.07.009

Davis, G., Driver, J., Pavani, F., and Shepherd, A. (2000). Reappraising the apparent costs of attending to two separate visual objects. *Vision Res.* 40, 1323–1332. doi: 10.1016/S0042-6989(99)00189-3

Davis, G., Welch, V. L., Holmes, A., and Shepherd, A. (2001). Can attention select only a fixed number of objects at a time? *Perception* 30, 1227–1248. doi: 10.1068/p3133

Dayan, P., and Abbott, L. F. (2000). *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems*. Cambridge, MA: MIT Press.

Domijan, D. (2004). Recurrent network with large representational capacity. *Neural Comput.* 16, 1917–1942. doi: 10.1162/0899766041336422

Domijan, D., and Šetić, M. (2008). A feedback model of figure-ground assignment. *J. Vis.* 8, 1–27. doi: 10.1167/8.7.10

Douglas, R., and Martin, K. (2004). Neuronal circuits of the neocortex. *Annu. Rev. Neurosci.* 27, 419–451. doi: 10.1146/annurev.neuro.27.070203.144152

Duncan, J. (1984). Selective attention and the organization of visual information. *J. Exp. Psychol.* 113, 501–517. doi: 10.1037/0096-3445.113.4.501

Egeth, H. E., Virzi, R. A., and Garbart, H. (1984). Searching for conjunctively defined targets. *J. Exp. Psychol.* 10, 32–39. doi: 10.1037/0096-1523.10.1.32

Eriksen, C. W., and St. James, J. D. (1986). Visual attention within and around the field of focal attention: a zoom lens model. *Percept. Psychophys.* 40, 225–240. doi: 10.3758/BF03211502

Farid, H. (2002). Temporal synchrony in perceptual grouping: a critique. *Trends Cogn. Sci.* 6, 284–288. doi: 10.1016/S1364-6613(02)01927-7

Fazl, A., Grossberg, S., and Mingolla, E. (2009). View-invariant object category learning, recognition, and search: how spatial and object attention are coordinated using surface-based attentional shrouds. *Cogn. Psychol.* 58, 1–48. doi: 10.1016/j.cogpsych.2008.05.001

Foley, N. C., Grossberg, S., and Mingolla, E. (2012). Neural dynamics of object-based multifocal visual spatial attention and priming: object cueing, useful-field-of-view, and crowding. *Cogn. Psychol.* 65, 77–117. doi: 10.1016/j.cogpsych.2012.02.001

Fukai, T., and Tanaka, S. (1997). A simple neural network exhibiting selective activation of neuronal ensembles: from winner-take-all to winners-share-all. *Neural Comput.* 9, 77–97. doi: 10.1162/neco.1997.9.1.77

Gawne, T. J., and Martin, J. M. (2002). Responses of primate visual cortical V4 neurons to simultaneously presented stimuli. *J. Neurophysiol.* 88, 1128–1135. doi: 10.1152/jn.2002.88.3.1128

Grossberg, S. (1973). Contour enhancement, short-term memory, and constancies in reverberating neural networks. *Stud. Appl. Math.* 52, 217–257. doi: 10.1002/sapm1973523213

Grossberg, S. (1980). How does a brain build a cognitive code? *Psychol. Rev.* 87, 1–51. doi: 10.1037/0033-295X.87.1.1

Grossberg, S., and Todorović, D. (1988). Neural dynamics of 1-D and 2-D brightness perception: a unified model of classical and recent phenomena. *Percept. Psychophys.* 43, 241–277. doi: 10.3758/BF03207869

Haarmann, H., and Usher, M. (2001). Maintenance of semantic information in capacity limited item short-term memory. *Psychon. Bull. Rev.* 8, 568–578. doi: 10.3758/BF03196193

Hahnloser, R. H., Seung, H. S., and Slotine, J.-J. (2003). Permitted and forbidden sets in symmetric threshold-linear networks. *Neural Comput.* 15, 621–638. doi: 10.1162/089976603321192103

Hahnloser, R. L. (1998). On the piecewise analysis of networks of linear threshold neurons. *Neural Netw.* 11, 691–697. doi: 10.1016/S0893-6080(98)00012-4

Hahnloser, R., Douglas, R. J., Mahowald, M., and Hepp, K. (1999). Feedback interactions between neuronal pointers and maps for attentional processing. *Nat. Neurosci.* 2, 746–752. doi: 10.1038/11219

Hamker, F. H. (2004). A dynamic model of how feature cues guide spatial attention. *Vision Res.* 44, 501–521. doi: 10.1016/j.visres.2003.09.033

Häusser, M., and Mel, B. W. (2003). Dendrites: bug or feature? *Curr. Opin. Neurobiol.* 13, 372–383. doi: 10.1016/S0959-4388(03)00075-8

Horn, D., and Usher, M. (1990). Excitatory–inhibitory networks with dynamical thresholds. *Int. J. Neural Syst.* 1, 249–257. doi: 10.1142/S0129065790000151

Huang, L. (2015). Grouping by similarity is mediated by feature selection: evidence from the failure of cue combination. *Psychon. Bull. Rev.* 22, 1364–1369. doi: 10.3758/s13423-015-0801-z

Huang, L., and Pashler, H. (2007). A boolean map theory of visual attention. *Psychol. Rev.* 114, 599–631. doi: 10.1037/0033-295X.114.3.599

Huang, L., and Pashler, H. (2012). Distinguishing different strategies of across-dimension attentional selection. *J. Exp. Psychol.* 38, 453–464. doi: 10.1037/a0026365

Itti, L., and Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Res.* 40, 1489–1506. doi: 10.1016/S0042-6989(99)00163-7

Itti, L., and Koch, C. (2001). Computational modelling of visual attention. *Nat. Rev. Neurosci.* 2, 194–203. doi: 10.1038/35058500

Jadi, M., Behabadi, B. F., Poleg-Polsky, A., Schiller, J., and Mel, B. W. (2014). An augmented two-layer model captures nonlinear analog spatial integration effects in pyramidal neuron dendrites. *Proc. IEEE Inst. Electr. Electron. Eng.* 102, 782–798. doi: 10.1109/JPROC.2014.2312671

Kaptein, N. A., Theeuwes, J., and van der Heijden, A. H. C. (1995). Search for a conjunctively defined target can be selectively limited to a color-defined subset of elements. *J. Exp. Psychol.* 21, 1053–1069. doi: 10.1037/0096-1523.21.5.1053

Kaski, S., and Kohonen, T. (1994). Winner-take-all networks for physiological models of competitive learning. *Neural Netw.* 7, 973–984. doi: 10.1016/S0893-6080(05)80154-6

Kouh, M., and Poggio, T. (2008). A canonical neural circuit for cortical nonlinear operations. *Neural Comput.* 20, 1427–1451. doi: 10.1162/neco.2008.02-07-466

Kreitzer, A. C., and Regehr, W. G. (2001). Retrograde inhibition of presynaptic calcium influx by endogenous cannabinoids at excitatory synapses onto Purkinje cells. *Neuron* 29, 717–727. doi: 10.1016/S0896-6273(01)00246-X

Kulikowski, J. J., and Tolhurst, D. J. (1973). Psychophysical evidence for sustained and transient detectors in human vision. *J. Physiol.* 232, 149–162. doi: 10.1113/jphysiol.1973.sp010261

Lampl, I., Ferster, D., Poggio, T., and Riesenhuber, M. (2004). Intracellular measurements of spatial integration and the MAX operation in complex cells of the cat primary visual cortex. *J. Neurophysiol.* 92, 2704–2713. doi: 10.1152/jn.00060.2004

Lee, S. H., and Blake, R. (1999). Visual form created solely from temporal structure. *Science* 284, 1165–1168. doi: 10.1126/science.284.5417.1165

Legge, G. E. (1978). Sustained and transient mechanisms in human vision: temporal and spatial properties. *Vision Res.* 18, 69–81. doi: 10.1016/0042-6989(78)90079-2

Liverence, B. M., and Franconeri, S. L. (2015). "Resource limitations in visual cognition," in *Emerging Trends in the Social and Behavioral Sciences*, eds R. Scott and S. Kosslyn (Hoboken, NJ: John Wiley and Sons), 1–13.

London, M., and Häusser, M. (2005). Dendritic computation. *Annu. Rev. Neurosci.* 28, 503–532. doi: 10.1146/annurev.neuro.28.061604.135703

Martinez-Trujillo, J. C., and Treue, S. (2004). Feature-based attention increases the selectivity of population responses in primate visual cortex. *Curr. Biol.* 14, 744–751. doi: 10.1016/j.cub.2004.04.028

Maunsell, J. H. R., and Treue, S. (2006). Feature-based attention in visual cortex. *Trends Neurosci.* 29, 317–322. doi: 10.1016/j.tins.2006.04.001

McCormick, D. A., Connors, B. W., Lighthall, J. W., and Prince, D. A. (1985). Comparative electrophysiology of pyramidal and sparsely spiny stellate neurons of the neocortex. *J. Neurophysiol.* 54, 782–806. doi: 10.1152/jn.1985.54.4.782

Mel, B. W. (2016). "Towards a simplified model of an active dendritic tree," in *Dendrites*, *3rd Edn*, eds G. Stuart, N. Spruston, and M. Häusser (Oxford: Oxford University Press), 465–486.

Nobre, A. C., and Kastner, S. (2014). *The Oxford Handbook of Attention*. Oxford: Oxford University Press.

O'Grady, R. B., and Müller, H. J. (2000). Object-based selection operates on a grouped array of locations. *Percept. Psychophys.* 62, 1655–1667. doi: 10.3758/BF03212163

Pitler, T. A., and Alger, B. E. (1992). Postsynaptic spike firing reduces synaptic GABAA responses in hippocampal pyramidal cells. *J. Neurosci.* 12, 4122–4132.

Poirazi, P., Brannon, T. M., and Mel, B. W. (2003). Pyramidal neuron as two-layer neural network. *Neuron* 37, 989–999. doi: 10.1016/S0896-6273(03)00149-1

Polsky, A., Mel, B. W., and Schiller, J. (2004). Computational subunits in thin dendrites of pyramidal cells. *Nat. Neurosci.* 7, 621–627. doi: 10.1038/nn1253

Posner, M. I. (1980). Orienting of attention. *Q. J. Exp. Psychol.* 32, 3–25. doi: 10.1080/00335558008248231

Qi, W., Han, J., Zhang, Y., and Bai, L. (2017). Saliency detection via Boolean and foreground in a dynamic Bayesian framework. *Vis. Comput.* 33, 209–220. doi: 10.1007/s00371-015-1176-x

Ray, S., and Maunsell, J. H. R. (2015). Do gamma oscillations play a role in cerebral cortex? *Trends Cogn. Sci.* 19, 78–85. doi: 10.1016/j.tics.2014.12.002

Regehr, W. G., Carey, M. R., and Best, A. R. (2009). Activity-dependent regulation of synapses by retrograde messengers. *Neuron* 63, 154–170. doi: 10.1016/j.neuron.2009.06.021

Richard, A. M., Lee, H., and Vecera, S. P. (2008). Attentional spreading in object-based attention. *J. Exp. Psychol. Hum. Percept. Perform.* 34, 842–853. doi: 10.1037/0096-1523.34.4.842

Rideaux, R., Badcock, D. R., Johnston, A., and Edwards, M. (2016). Temporal synchrony is an effective cue for grouping and segmentation in the absence of form cues. *J. Vis.* 16:23. doi: 10.1167/16.11.23

Riesenhuber, M., and Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nat. Neurosci.* 2, 1019–1025. doi: 10.1038/14819

Roelfsema, P. R. (2006). Cortical algorithms for perceptual grouping. *Annu. Rev. Neurosci.* 29, 203–227. doi: 10.1146/annurev.neuro.29.051605.112939

Roelfsema, P. R., and de Lange, F. P. (2016). Early visual cortex as a multiscale cognitive blackboard. *Ann. Rev. Vision Sci.* 2, 131–151. doi: 10.1146/annurev-vision-111815-114443

Rutishauser, U., and Douglas, R. J. (2009). State-dependent computation using coupled recurrent networks. *Neural Comput.* 21, 478–509. doi: 10.1162/neco.2008.03-08-734

Rutishauser, U., Douglas, R. J., and Slotine, J.-J. (2011). Collective stability of networks of winner-take-all circuits. *Neural Comput.* 23, 735–773. doi: 10.1162/NECO_a_00091

Rutishauser, U., Slotine, J.-J., and Douglas, R. J. (2015). Computation in dynamically bounded asymmetric systems. *PLoS Comput. Biol.* 11:e1004039. doi: 10.1371/journal.pcbi.1004039

Saenz, M., Buracas, G. T., and Boynton, G. M. (2002). Global effects of feature-based attention in human visual cortex. *Nat. Neurosci.* 5, 631–632. doi: 10.1038/nn876

Saenz, M., Buracas, G. T., and Boynton, G. M. (2003). Global feature-based attention for motion and color. *Vision Res.* 43, 629–637. doi: 10.1016/S0042-6989(02)00595-3

Sato, T. (1989). Interactions of visual stimuli in the receptive fields of inferior temporal neurons in awake macaques. *Exp. Brain Res.* 77, 23–30. doi: 10.1007/BF00250563

Scholl, B. J. (2001). Objects and attention: the state of the art. *Cognition* 80, 1–46. doi: 10.1016/S0010-0277(00)00152-9

Scimeca, J. M., and Franconeri, S. L. (2015). Selecting and tracking multiple objects. *Wiley Interdiscip. Rev. Cogn. Sci.* 6, 109–118. doi: 10.1002/wcs.1328

Serences, J. T., and Boynton, G. M. (2007). Feature-based attentional modulations in the absence of direct visual stimulation. *Neuron* 55, 301–312. doi: 10.1016/j.neuron.2007.06.015

Spratling, M. W. (2010). Predictive coding as a model of response properties in cortical area V1. *J. Neurosci.* 30, 3531–3543. doi: 10.1523/JNEUROSCI.4911-09.2010

Spratling, M. W. (2011). A single functional model accounts for the distinct properties of suppression in cortical area V1. *Vision. Res.* 51, 563–576. doi: 10.1016/j.visres.2011.01.017

Spruston, N. (2008). Pyramidal neurons: Dendritic structure and synaptic integration. *Nat. Rev. Neurosci.* 9, 206–221. doi: 10.1038/nrn2286

Tao, H. W., and Poo, M. (2001). Retrograde signaling at central synapses. *Proc. Natl. Acad. Sci. U.S.A.* 98, 11009–11015. doi: 10.1073/pnas.191351698

Theeuwes, J. (2010). Top-down and bottom-up control of visual selection. *Acta Psychol.* 135, 77–99. doi: 10.1016/j.actpsy.2010.02.006

Theeuwes, J. (2013). Feature-based attention: it is all bottom-up priming. *Philos. Trans. R. Soc. B* 368:20130055. doi: 10.1098/rstb.2013.0055

Treue, S., and Martinez-Trujillo, J. C. (1999). Feature-based attention influences motion processing gain in macaque visual cortex. *Nature*, 399, 575–579. doi: 10.1038/21176

Tsui, J. M. G., Hunter, J. N., Born, R. T., and Pack, C. C. (2010). The role of V1 surround suppression in MT motion integration. *J. Neurophysiol.* 103, 3123–3138. doi: 10.1152/jn.00654.2009

Usher, M., and Cohen, J. D. (1999). "Short term memory and selection processes in a frontal-lobe model," in *Connectionist Models in Cognitive Neuroscience*, eds D. Heinke, G. W. Humphreys, and A. Olson (London: Springer-Verlag), 78–91.

Vatterott, D. B., and Vecera, S. P. (2015). The attentional window configures to object and surface boundaries. *Vis. Cogn.* 23, 561–576. doi: 10.1080/13506285.2015.1054454

Wang, D. L. (1999). Object selection based on oscillatory correlation. *Neural Netw.* 12, 579–592. doi: 10.1016/S0893-6080(99)00028-3

Wannig, A., Stanisor, L., and Roelfsema, P. R. (2011). Automatic spread of attentional response modulation along Gestalt criteria in primary visual cortex. *Nat. Neurosci.* 18, 1243–1244. doi: 10.1038/nn.2910

Wei, D. S., Mei, Y. A., Bagal, A., Kao, J. P., Thompson, S. M., and Tang, C. M. (2001). Compartmentalized and binary behavior of terminal dendrites in hippocampal pyramidal neurons. *Science* 293, 2272–2275. doi: 10.1126/science.1061198

Yu, A. J., Giese, M. A., and Poggio, T. A. (2002). Biophysically plausible implementations of the maximum operation. *Neural Comput.* 14, 2857–2881. doi: 10.1162/089976602760805313

Yu, D., and Franconeri, S. L. (2015). Similarity grouping as feature-based selection. *Vis. Cogn.* 23, 843–847. doi: 10.1080/13506285.2015.1093234

Zhang, J., and Sclaroff, S. (2016). Exploiting surroundedness for saliency detection: a Boolean map approach. *IEEE Trans. Pattern Anal. Mach. Intell.* 38, 889–902. doi: 10.1109/TPAMI.2015.2473844

Zilberter, Y. (2000). Dendritic release of glutamate suppresses synaptic inhibition of pyramidal neurons in rat neocortex. *J. Physiol.* 528(Pt 3), 489–496. doi: 10.1111/j.1469-7793.2000.00489.x

Zilberter, Y., Harkany, T., and Holmgren, C. D. (2005). Dendritic release of retrograde messengers controls synaptic transmission in local neocortical networks. *Neuroscientist* 11, 334–344. doi: 10.1177/1073858405275827

Zilberter, Y., Kaiser, K. M., and Sakmann, B. (1999). Dendritic GABA release depresses excitatory transmission between layer 2/3 pyramidal and bitufted neurons in rat neocortex. *Neuron* 24, 979–988. doi: 10.1016/S0896-6273(00)81044-2

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer, MU, and handling editor declared their shared affiliation.

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 07 June 2018
doi: 10.3389/fnins.2018.00391

Globally Normal Bistable Motion Perception of Anisometropic Amblyopes May Profit From an Unusual Coding Mechanism

Jiachen Liu[1], *Yifeng Zhou*[1,2] *and Tzvetomir Tzvetanov*[1,3]*

[1] *Hefei National Laboratory for Physical Sciences at Microscale, School of Life Science, University of Science and Technology of China, Hefei, China,* [2] *State Key Laboratory of Brain and Cognitive Science, Institute of Biophysics, Chinese Academy of Science, Beijing, China,* [3] *Anhui Province Key Laboratory of Affective Computing and Advanced Intelligent Machine and School of Computer and Information, Hefei University of Technology, Hefei, China*

Anisometropic amblyopia is a neurodevelopmental disorder of the visual system. There is evidence that the neural deficits spread across visual areas, from the primary cortex up to higher brain areas, including motion coding structures such as MT. Here, we used bistable plaid motion to investigate changes in the underlying mechanisms of motion integration and segmentation and, thus, help us to unravel in more detail deficits in the amblyopic visual motion system. Our results showed that (1) amblyopes globally exhibited normal bistable perception in all viewing conditions compared to the control group and (2) decreased contrast led to a stronger increase in percept switches and decreased percept durations in the control group, while the amblyopic group exhibited no such changes. There were few differences in outcomes dependent upon the use of the weak eye, the strong eye, or both eyes for viewing the stimuli, but this was a general effect present across all subjects, not specific to the amblyopic group. To understand the role of noise and adaptation in such cases of bistable perception, we analyzed predictions from a model and found that contrast does indeed affect percept switches and durations as observed in the control group, in line with the hypothesis that lower stimulus contrast enhances internal noise effects. The combination of experimental and computational results presented here suggests a different motion coding mechanism in the amblyopic visual system, with relatively little effect of stimulus contrast on amblyopes' bistable motion perception.

Keywords: plaid motion, anisometropic amblyopia, motion coding mechanism, bistable percept, model prediction

OPEN ACCESS

Edited by:
Hedva Spitzer,
Tel Aviv University, Israel

Reviewed by:
Richard J. A. Van Wezel,
University of Twente, Netherlands
Benjamin Thompson,
University of Waterloo, Canada

***Correspondence:**
Tzvetomir Tzvetanov
tzvetan@hfut.edu.cn

Specialty section:
This article was submitted to Perception Science, a section of the journal Frontiers in Neuroscience

Received: *09 December 2017*
Accepted: *22 May 2018*
Published: *07 June 2018*

Citation:
Liu J, Zhou Y and Tzvetanov T (2018) Globally Normal Bistable Motion Perception of Anisometropic Amblyopes May Profit From an Unusual Coding Mechanism. Front. Neurosci. 12:391. doi: 10.3389/fnins.2018.00391

INTRODUCTION

Amblyopia is a neurodevelopmental disorder of the visual system. The condition is caused by an imbalance in visual input during cortex development, mostly in infancy (Wong, 2012; Hess and Thompson, 2015). Anisometropic amblyopia is typically due to the presence of a chronic blur. These conditions result in a weakening or suppression of the input from the amblyopic eye, and, thus, this input is processed abnormally within the visual cortex (Hubel and Wiesel, 1965, 1970; Kiorpes and McKee, 1999; Hess and Thompson, 2015). Such an abnormal processing causes amblyopes to see differently from neurotypical subjects in visual perception tasks; for example, amblyopes may exhibit a reduction in contrast sensitivity, stereo-acuity (3D, depth perception), or visual acuity (Bradley and Freeman, 1981; Levi et al., 2011).

In contrast, suprathreshold contrast perception seems equivalent between both eyes of amblyopes (Hess and Bradley, 1980), while prolonged observations of static gratings by amblyopes make them report illusory static or dynamic patterns in the stimulus (Sireteanu et al., 2008; Thiel and Iftime, 2016).

In addition to the above basic visual features, other spatial and temporal processing are also affected by amblyopia in early visual cortices (Barnes et al., 2001; Bonhomme et al., 2006; Hess et al., 2010; Li et al., 2011). Increasing evidence has demonstrated that amblyopia is also associated with abnormal function of the MT/MST areas, which are highly motion-sensitive and related to local and global motion integration (Britten et al., 1992; Born and Bradley, 2005; Majaj et al., 2007). There is strong neurophysiological evidence to suggest that motion integration and segregation processing involve area MT (Newsome and Parés, 1988; Salzman et al., 1990). In addition, psychophysical studies have shown abnormal global motion perception in amblyopia, even after adjusting for the deficits in contrast sensitivity. These results strongly suggest that the motion-sensitive areas MT/MST are affected by this disorder (Ellemberg et al., 2002; Constantinescu et al., 2005; Simmers et al., 2006; Aaen-Stockdale et al., 2007; Thompson et al., 2008; Ho and Giaschi, 2009; El-Shamayleh et al., 2010), and a recent neuroimaging study found evidence of abnormal cortical processing of pattern motion in amblyopia (Thompson et al., 2012).

In psychophysical research, plaid motion is a particular stimulus used to investigate the underlying neural mechanisms of motion integration and segregation (Adelson and Movshon, 1982). Plaid stimuli are typically constructed from two drifting gratings within a circular aperture. The drifting directions of both gratings are different. When the two gratings have similar temporal and spatial properties, the stimulus will produce an initial percept of a single patterned surface drifting in a "global" direction, which is a unique combination of both component directions. With prolonged observation of the pattern, a perceptual switching phenomenon occurs; the plaid motion can be seen either as "coherent motion" (a single object moving rigidly) or as "transparent motion" (two independent gratings sliding over each other), dubbed bistable motion perception. Because of the advances in the theoretical understanding of bistable perception, we considered that plaid motion would be a particularly useful probe for investigating the mechanisms of motion segmentation and integration and help us to unravel in more detail the deficits in the amblyopic visual motion system.

The various observations of bistable perception have inspired models of multistability, which mainly focus on bistable rivalry (Lago-Fernández and Deco, 2002; Laing and Chow, 2002; Moreno-Bote et al., 2007). In such models, the random alternation of percepts is influenced by the competition between two neuronal populations via reciprocal inhibition, noise levels in the neural inputs and some sort of adaptation, e.g., spike frequency adaptation and/or synaptic depression. Such models are extendable to tristable percepts, of which plaid motion perception is argued to be an example (Huguet et al., 2014). In all of these models, the exact number of percept switches together with the durations of the two major types of percepts are very sensitive to internal variables, especially internal noise. Thus, any changes in internal variables differentially affect all measurable variables.

This manuscript first describes results of three experiments performed to compare the bistable motion perception in anisometropic amblyopes (AMB) and neurotypical observers (NTE). Experiment 1 was mainly performed as an exploratory study to search for plausible differences between AMB and NTE in plaid motion perception. This experiment led to the hypothesis of differential effects associated with stimulus strength between AMB and NTE that was tested in Experiment 2. Experiment 3 was a control test of the main finding of contrast effects. In the last part, with the help of simulations, we analyzed one model predictions (Moreno-Bote et al., 2007) in order to compare to the experimental results, and thus to propose putative changes in the mechanisms of motion coding in the amblyopic visual system.

METHODS

Observers

A total of 32 observers participated in the experiments, including 17 normal-sighted subjects (five women and 12 men; including two authors; age range 20–42) and 15 anisometropic amblyopes (one woman and 14 men; age range: 23–27). A portion of the observers in these two groups participated in experiments 1, 2, and 3. The exact number of subjects within a given experiment is stated in the corresponding section. All amblyopes had anisometropic amblyopia; amblyope #10 had bilateral amblyopia. For that person, the eye with the best visual acuity (strong eye) was treated as the fellow eye in all the analysis. Detailed ophthalmologic characteristics of these observers, including amblyopia type and optical correction, were obtained during normal university medical examinations at the department of ophthalmology in the hospital of USTC. The amblyopic group was defined according to the Preferred Practice Protocol (PPP) of The American Academy of Ophthalmology (Wallace et al., 2018), with anisometropic type was defined as the difference of dioptre sphere above 1.5 and/or the difference of dioptre of cylinder over 1.0 who can not fuse image in retina well binocularly. Nonamblyopes had normal or corrected-to-normal eyesight, while amblyopes wore their best refractive corrections. All observers provided informed consent and received a fee of 60 CNY/hour for participating in the experiments. The experiments were approved by the ethics committee of the School of Life Science of USTC and followed the tenets of the Declaration of Helsinki for experiments with human subjects. **Table 1** presents the eyes characteristics of the amblyopes.

Apparatus

Stimuli were presented on an ASUS VG248 monitor with a 1,920 × 1,080-pixel resolution at a frame rate of 120 Hz. Observers were comfortably seated 100 cm in front of the screen in a dark room, with their chin and forehead resting on a chinrest. When the eye signal was available, binocular or monocular eye movements (randomly) were monitored and recorded for a portion of the observers (13 amblyopes/10 normal observers) with an Eyelink 1,000 eye recording setup and sampled at 500 Hz to confirm correct eye fixation at the stimulus location.

TABLE 1 | Ophthalmic details of the observers with amblyopia.

Obs	Age/sex	Type	Refraction	SA	VA (MAR)
Amb1	25/M	RE anis	+6.00 DS/+1.00 DCx25	100	10.00
		LE	Ø		1.000
Amb2	27/M	RE anis	+4.00 DS/+1.00 DCx85	50	10.00
		LE	Ø		1.00
Amb3	26/M	RE anis	+2.50 DS/+1.00 DCx170	160	3.16
		LE	−0.75 DS		0.63
Amb4	23/F	RE anis	−2.00 DCx110	100	2.00
		LE	−6.00 DS		1.00
Amb5	26/M	RE anis	+1.50 DS/+1.50 DCx60	400	6.31
		LE	−1.00 DS/−0.50 DCx160		1.00
Amb6	23/M	RE anis	+1.00 DCx105	400	3.98
		LE	−4.00 DS/−1.25 DCx30		0.79
Amb7	25/M	RE	−0.500 DS	25	1.00
		LE anis	+2.500 DS/0.500 DCx90		1.58
Amb8	25/M	RE	+3.00 DS/+1.00 DCx85	400	0.79
		LE anis	+5.50 DS/0.75 DCx95		3.16
Amb9	23/M	RE	−1.250 DS/−1.00 DCx160	63	0.79
		LE anis	1.00 DCx80		1.58
Amb10	23/M	RE anis	−5.50 DS/−2.00 DCx10	32	3.98
		LE anis	−5.250 DS/-5.00 DCx175		6.31
Amb11	25/M	RE	−1.50 DS/−0.50 DCx30	400	0.79
		LE anis	+4.50 DS/0.50 DCx35		5.01
Amb12	25/M	RE	−2.75 DS/−0.50 DCx10	400	1.00
		LE anis	+1.00 DS/+1.00 DCx95		3.98
Amb13	24/M	RE	−2.75 DS/−1.00 DCx20	50	1.00
		LE anis	+3.75 DS/0.75 DCx115		2.00
Amb14	26/M	RE anis	−2.50 DS	100	3.16
		LE	+2.00 DS/+0.50 DCx96		0.63
Amb15	26/M	RE anis	+1.00 DS/+1.50 DCx95	32	3.98
		LE	Ø		0.79

Obs, observer; Amb, anisometropic amblyope; M, male; F, female; RE, right eye; LE, left eye; anis, anisometropic; DS, dioptre sphere; DC, dioptre of cylinder; Ø, plano; SA, stereo acuity; VA, visual acuity; MAR, minimum angle of resolution.

Stimuli

The stimulus comprised two rectangular-wave gratings presented through a circular aperture 7.7° in diameter on a middle-gray background of RGB 126. Gratings moved at 3°/s (defined in the direction normal to their orientation) in directions 90° apart (angle α hereafter), with a spatial frequency of 3 c/d and duty cycle of 50%. The mean direction of motion of both gratings was either vertical upward or horizontal leftward, thus making the coherent pattern perceived as moving upwards or leftwards, respectively. Grating contrast was defined in RGB units, and two contrasts of 30% (high) and 5% (low) values were possible, with both gratings having the same contrast. A pink fixation point was added in the middle of the circular aperture to help subjects locate the stimulus center and minimize optokinetic nystagmus (Huguet et al., 2014), and subjects were instructed to fixate this point throughout the stimulus presentation.

Experimental Procedure

Subjects were first familiarized with the stimuli and procedure. They had to report the time of percept change with two keyboard keys, with each key indicating that they perceived either coherent motion or transparent motion. They were instructed to passively report the percepts, without trying to influence them. Each observer was exposed to both global coherent directions (upward and leftward) to avoid motion direction adaptation, one (Experiment 1) or two (Experiment 2) contrast levels (for Experiment 1, 30% contrast; for Experiment 2, 30 and 5% contrast), and three eye conditions (binocular, left, right eye monocular), corresponding to a total of 6 or 12 different stimulus configurations. Presentation time was 120 s for each stimulus, and observers were tested on each configuration one time. The order of presentation was random. Because the first percept is known to always be coherent in normal-sighted observers (Hupé and Rubin, 2003), and amblyopes are able to demonstrate possible grating misperceptions/illusions (Hess et al., 1978; Hess and Bradley, 1980; Thompson et al., 2008; Thiel and Iftime, 2016), each observer was debriefed at the end of each 120-s trial about their first percept (coherent or not) and overall visibility of the pattern. All participants reported that they could clearly see the stimuli, a single moving plaid stimulus and two grating surfaces sliding over each other, in all conditions, even at the lowest contrast used in this study. No amblyopes reported differences between AE and fellow eye perception of the moving gratings, out of the switch rate/duration differences. The dominant eye of each subject was assessed with the hole-in-card experiment. Stereo acuity was assessed with the Titmus Stereopsis Test. Visual acuity was measured using a standard wall-mounted Tumbling E chart, from a distance of 5 metres, and defined as the score associated with a correct judgment rate of 75% at the minimum angle of resolution.

Model Simulation and Numerical Procedures

We implemented the tristable model of motion coherence/transparency proposed by Huguet et al. (2014). This model is a firing rate-based tristable model that includes three pools of neuronal populations that encode three different percepts: coherence (C), transparent with the leftward moving grating on top (T_L), and transparent with the rightward moving grating on top (T_R). The equations describing the dynamics of the three populations are:

$$\begin{aligned}
\tau\frac{dr_c}{dt} &= -r_c + S(-\beta_1 r_{T_R} - \beta_1 r_{T_L} - a_c + I_c + n_c)\\
\tau\frac{dr_{T_R}}{dt} &= -r_{T_R} + S(-\beta_1 r_c - \beta_2 r_{T_L} - a_{T_R} + I_{T_R} + n_{T_R})\\
\tau\frac{dr_{T_L}}{dt} &= -r_{T_L} + S(-\beta_1 r_c - \beta_2 r_{T_R} - a_{T_L} + I_{T_L} + n_{T_L}) \quad (1)
\end{aligned}$$

with a_i, I_i, and n_i representing adaptation, external input, and noise for each population, respectively. The time constant τ was $\tau = 10$ ms. β_1 is the cross-inhibition strength between population C and T (including T_R and T_L), while β_2 is the inhibition strength

between T_R and T_L. The intensity of external input changes is represented with I_C and $I_T = I_{T_R} = I_{T_L}$.

The function S is a sigmoidal transducer of input-output function:

$$S(x) = \frac{1}{1 + \theta^{-(x-\theta)/k}} \quad (2)$$

with threshold $\theta = 0.2$ and $k = 0.1$.

The adaptation of firing activity was done through the terms a_C, a_{T_R}, a_{T_L} and all followed the same time evolution:

$$\tau \frac{da_i}{dt} = -a_i + \gamma r_i \quad (3)$$

with $\tau = 2{,}500$ ms, and a maximum strength of $\gamma = 0.25$ for all populations.

Noise input is modeled with an Ornstein-Uhlenbeck process as:

$$\frac{dn_i}{dt} = -\frac{n_i}{\tau_s} + \sigma \sqrt{\frac{2}{\tau_s}} \times \xi(t) \quad (4)$$

with $\tau_S = 200$ ms, $\sigma = 0.08$, and $\xi(t)$ is a white-noise process whose mean value is zero with a standard deviation of one and no temporal correlations.

In this model (Huguet et al., 2014), we adjusted the cross-inhibition strength values β_1 and β_2, external input value I_C and I_T (I_{T_R} and I_{T_L} were set equal), noise strength value σ, and adaptation strength value γ to reproduce our behavioral results with other parameters remaining unchanged. The time window of simulations was set to 120 s, corresponding to the length of one block of measure in the psychophysical experiment, and repeated simulations were performed to obtain the mean and variability of the variables analyzed in the experiments.

Since we focused on the bistable condition, we report only transparent and coherent states by considering T_R and T_L as the transparent percept. A coherent percept was defined when r_C was simultaneously higher than r_{T_R} and r_{T_L} and otherwise defined as transparent. For each 120 s of simulations, we computed the number of switches and durations of coherent and transparent states.

Data Analysis

For each 120-s trial, the number of percept changes was computed from the first report of a transparent percept to the end of the trial, as in work by Hupé and Rubin (2003). The dominance durations were measured between successive presses of the two keys. The duration of the last interrupted percept was not computed. The first percept was coherent in all trials (as reported in the debriefing), but in some conditions, a few subjects did not first press the "coherent" percept key, due to their knowledge of this appearance. Dominance durations were log10-transformed (Moreno-Bote et al., 2010).

Each dependent variable was analyzed with within-between analysis of variance, while all statistical levels used Geisser-Greenhouse epsilon-hat-adjusted values where appropriate. In the first analysis, the dependent variable was the number of key-presses for each condition, which allowed for the comparison of the frequencies of perception switches in different conditions and observers (amblyopes/normal observers). This analysis included the data from all subjects. In the second analysis, the dependent variable was the mean duration of the percept, with an additional within-subject factor in the ANOVA corresponding to coherent and transparent conditions. In this analysis, observers who were unable to see perceptual switches in at least one condition were not included due to lack of the corresponding variable. This phenomenon only appeared in 3 out of 15 anisometropic amblyopes (2 in Experiment 1 and 2 in Experiment 2) and 1 out of 17 NTE subjects (in Experiment 1), and it was mostly present for horizontal motion directions. We also calculated the mean value and standard deviation for each condition across all normal subjects and found that 1 of the 11 subjects in Experiment 2 had percept durations that deviated above 2 *SD* from the between-subjects mean of the condition in 8 out of 24 conditions. In contrast, the other subjects had such deviations in a maximum of 2 conditions. For this reason, we also removed this subject data in the analysis of percept durations.

RESULTS

Experiment 1

In the first experimental test, we measured the performance of each subject in three eye conditions (binocular, monocular with strong eye, and monocular with weak eye) with only a strong contrast of the gratings (30%) and global moving directions upwards and leftwards. We focused on the number of perceptual switches and mean duration of each percept type. Twenty subjects participated in this experiment; 10 of them were anisometropic amblyopes (AMB), and the remaining were neurotypical subjects (including two authors) that had no known visual deficits (NTE). During the experiment, all amblyopic subjects reported that they did not feel any difference between the fellow eye or binocular condition when using the amblyopic eye to watch the stimulus.

Frequency of Perceptual Switches

Figure 1 illustrates the number of key-presses in each viewing condition for the two groups. There was a significant difference between the two moving directions [$F_{(1,\ 18)} = 15.865$, $p = 0.001$] showing that, globally, the number of perceptual switches for the vertical motion directions were higher than for the horizontal directions. Eye viewing conditions also showed significant differences in perceptual switches [$F_{(1.987,\ 35.758)} = 5.836$, $p = 0.006$], with the *post-hoc* Bonferroni test revealing a difference between the binocular and weak eye conditions [$F_{(1,\ 18)} = 10.860$, $p = 0.004$]. Statistical analysis showed that there was no difference between the two groups of subjects [$F_{(1,\ 18)} = 1.061$, $p = 0.317$], nor a significant interaction between the observer groups and the other factors (see **Table 2** for full ANOVA results).

Duration of the Two Percept Types in Different Conditions

Figure 2 summarizes the results of the duration of the percepts. Statistical analysis showed that there was no difference between the two groups of subjects [$F_{(1,\ 15)} = 0.559$, $p = 0.466$],

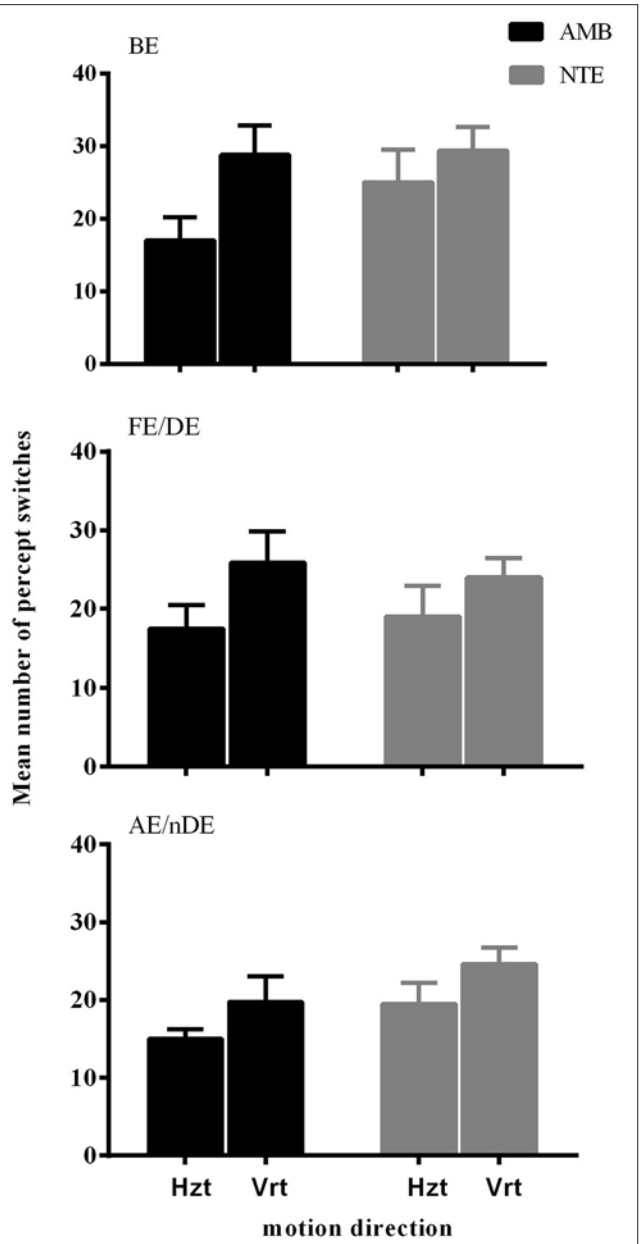


FIGURE 1 | The mean number of percept switches in Experiment 1, split between the factors motion direction (Hzt: horizontal, Vrt: vertical), eye viewing (BE-binocular, AE/nDE-nondominant eye, FE/DE- fellow/dominant eye), and Group (AMB/NTE). Error bars indicate between-subject SEM.

indicating that the mean perceived duration of each percept type was similar in normal and amblyopic people. A significant difference was found in the durations of each percept type [$F_{(1, 15)} = 10.925$, $p = 0.005$], with duration of coherent percept being longer than the duration of the transparent percept, independent of the subject group (see **Figure 2A**). We also found significant differences in motion direction [$F_{(1, 15)} = 22.272$, $p < 0.001$] with the mean of log10-transformed duration of horizontal direction being longer than that of the vertical direction (mean of horizontal = 0.673, mean of vertical = 0.570) and a significant interaction between direction and group [$F_{(1, 15)} = 10.062$, $p = 0.006$; see **Figure 2B**]. This last interaction was due to the much longer percept duration for the horizontal motion directions than for the vertical ones in AMB, while NTE exhibited similar values for both directions. There was also an interaction between eye condition and direction [$F_{(1.927, 28.906)} = 3.927$, $p = 0.031$; **Figure 2C**]. For the horizontal direction, the means of the log10-transformed durations for each eye condition were similar but were distinct when the global motion direction was vertical. This difference may indicate that there are different strategies to address different motion directions. Additionally, with the change in the direction, the weak eye showed a relatively stable log10-transformed duration. *Post-hoc* Bonferroni-adjusted comparisons showed a difference between the weak eye and binocular condition in its interaction with direction [$F_{(1, 15)} = 8.787$, $p = 0.01$]. No significant differences were found in other factors (see **Table 3** for complete ANOVA results).

TABLE 2 | ANOVA results on Presses Number of Experiment 1.

Variables	df	*F*	Sig.	Partial Eta Squared
Eye	1.987, 35.758	5.836	0.006	0.245
Eye * Group	1.987, 35.758	1.988	0.152	0.099
Dir	1, 18	15.865	0.001	0.468
Dir * Group	1, 18	1.146	0.298	0.060
Eye * Dir	1.736, 31.241	3.141	0.064	0.149
Eye * Dir * Group	1.736, 31.241	1.319	0.279	0.068
Group	1, 18	1.601	0.317	0.056

Experiment 2

From the above Experiment 1 results, we observed that there were few differences between amblyopes and non-amblyopes in their perception of a bistable plaid motion stimulus. This outcome was unexpected because, based on previous reports of stronger noise in the motion amblyopic system (Simmers et al., 2006) and possibly a very different visual motion coding system in amblyopes (Thompson et al., 2012), we expected that motion rivalry, due to its keen sensitivity to internal noise and inhibition strength (Huguet et al., 2014), would result in strong systematic differences between the two observer types. Given the non-significant differences, we realized that our experimental design might have missed the effects because of the relatively high contrast of the gratings. Thus, if the activation of the motion system was too high such that the signal-to-noise (SNR) ratio was relatively large, then any internal noise differences might have gone unnoticed. Therefore, we performed a second experiment that was identical to the first in all aspects except that one more factor was added, the contrast of the stimuli, with two levels, high (30%) and low (5%) contrast. By decreasing the contrast, we expected that the SNR would also decrease, and differences between the groups would be observed, with a prediction that there would be a main effect of lower contrast in which the low-contrast condition would be associated with more perceptual

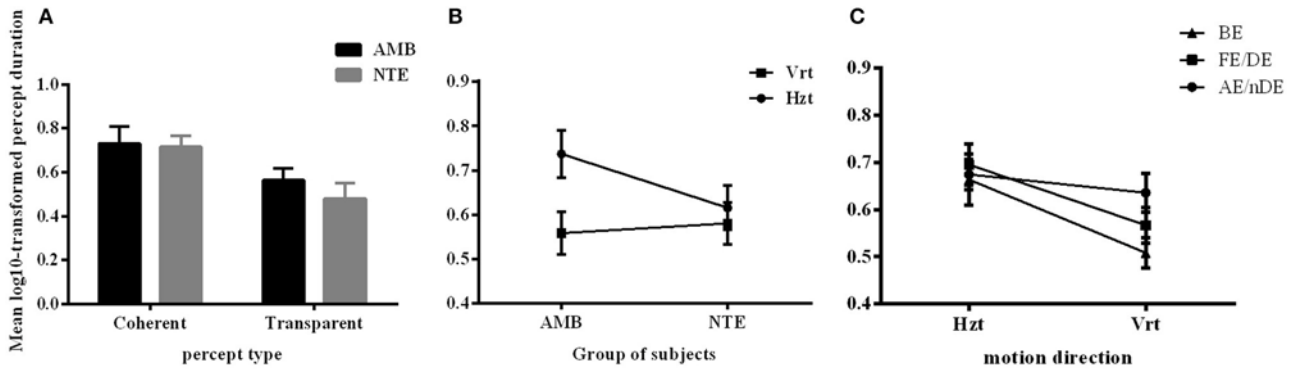


FIGURE 2 | The mean log10-transformed percept durations showing **(A)** a main effect of percept type, **(B)** main effect of motion direction, and **(C)** an interaction between eye condition and direction. Mean of the log10-transformed percept durations expressed in seconds. C, coherent; T, transparent; NTE, Neurotypical/Normal; AMB, Amblyopes; Vrt, vertical; Hzt, horizontal; BE, binocular condition; AE/nDE, weak eye of subjects; FE/DE, fellow/dominant eye of subjects. Error bars indicate between-subject SEM.

TABLE 3 | ANOVA results on Mean of log-10 Durations of Experiment 1.

Variables	df	*F*	Sig.	Partial Eta Squared
Per	1, 15	10.925	0.005	0.412
Per * Group	1, 15	0.297	0.594	0.019
Eye	1.869, 28.038	2.723	0.086	0.154
Eye * Group	1.869, 28.038	0.836	0.437	0.053
Dir	1, 15	22.272	0.000	0.598
Dir * Group	1, 15	10.062	0.006	0.401
Per * Eye	1.772, 26.586	0.722	0.479	0.046
Per * Eye * Group	1.772, 26.586	0.371	0.669	0.024
Per * Dir	1, 15	1.164	0.298	0.072
Per * Dir * Group	1, 15	0.033	0.858	0.002
Eye * Dir	1.927, 28.906	3.927	0.032	0.207
Eye * Dir * Group	1.927, 28.906	2.100	0.142	0.123
Per * Eye * Dir	1.740, 26.101	1.175	0.319	0.073
Per * Eye * Dir * Group	1.740, 26.101	0.468	0.605	0.030
Group	1, 15	0.559	0.466	0.036

switches in amblyopes when compared to the high-contrast condition.

Twenty-one subjects participated in this experiment, with 10 anisometropic amblyopes (AMB; 5 of them also participated in Experiment 1), and the remaining were neurotypical subjects (NTE; 4 of them participated in Experiment 1).

Frequency of Perceptual Switches

Here, we still used the number of key-presses to represent the frequency of perceptual switches. Analysis included data from all 21 subjects (10 AMB and 11 NTE). **Figure 3** shows the main significant effects and interaction of how the press number increased with lower contrast and that the frequency of percept switches was globally lower in the weak eye condition than in the other conditions. There was no significant difference in the performance of normal and amblyopic subjects [$F_{(1, 19)} = 0.287$, $p = 0.598$]. However, there was a significant difference in contrast [$F_{(1, 19)} = 5.575$, $p = 0.029$], direction [$F_{(1, 19)} = 5.697$, $p = 0.028$], and eye condition [$F_{(1.904, 36.171)} = 4.446$, $p = 0.020$]. The number of presses increased with the decrease in contrast, potentially due to an increase in internal noise or, equivalently, a decrease in the signal-to-noise ratio. Upon examination of the effects of the global direction of motion, both groups had higher percept switches when stimuli were moving upward (as in Experiment 1). *Post-hoc* comparisons (Bonferroni-corrected) for eye conditions showed a difference between the binocular and weak eye conditions [$F_{(1, 19)} = 6.426$, $p = 0.02$] and a difference between the weak and strong eye conditions [$F_{(1, 19)} = 5.472$, $p = 0.03$].

An interaction between contrast and eye condition was also found in this case [$F_{(1.904, 30.537)} = 5.492$, $p = 0.013$]. However, no other interactions were significant (see **Table 4** for complete ANOVA results).

Duration of Two Percept Types in Different Conditions

Here, we analyzed the duration of both percept types (i.e., coherent and transparent) for different contrast, eye, and moving direction conditions and whether there were differences between neurotypical subjects and anisometropic amblyopes; 2/10 AMB were not included because of at least one condition with no percept switch, and 1/11 NTE was excluded as an outlier (see section Methods).

Figure 4A illustrates the durations of both direction and eye conditions for subject groups and stimulus contrast conditions. Statistical analysis showed that there were no differences between the two groups of subjects [$F_{(1, 16)} = 0.298$, $p = 0.593$], indicating that globally, percept durations were similar in normal and amblyopic people. Significant differences were found across contrast conditions [$F_{(1, 16)} = 5.173$, $p = 0.037$] and percept type [$F_{(1, 16)} = 19.241$, $p = 0.0005$; **Figures 4B,C**]. Lower contrasts globally decreased percept duration, paralleling the increase in number of switches. The duration in the coherent percept was always longer than that in the transparent percept regardless of subject group (**Figure 4C**).

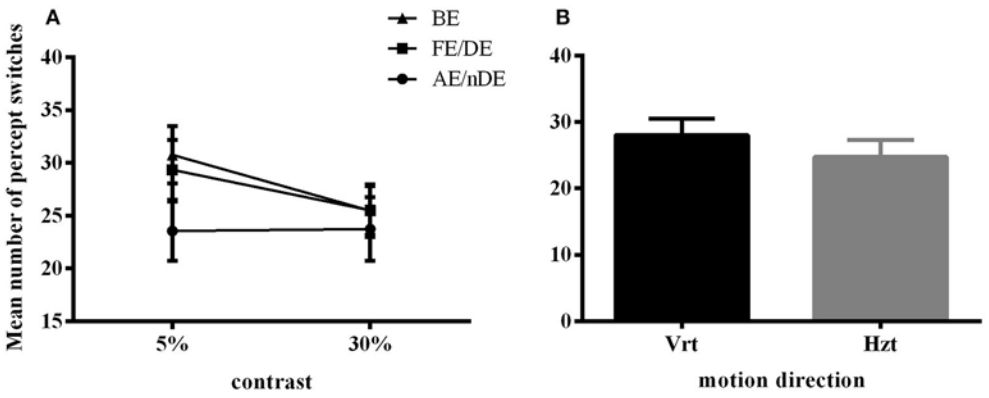


FIGURE 3 | Main significant effects and interaction in Experiment 2 for number of percept switches. **(A)** Interaction plot for contrast and eye conditions. DE (dominant eye) and FE (fellow eye) corresponded to the strong eye for normal and amblyope observers, respectively; nDE (non-dominant eye) and AE (amblyopic eye) corresponded to the weak eye for normal and amblyope observers, respectively. BE was the binocular condition. **(B)** Main effect of global motion direction. Mean number of switches was higher in the vertical condition (Vrt) than in the horizontal condition (Hzt). Error bars indicate between-subjects SEM.

TABLE 4 | ANOVA results on Presses Number of Experiment 2.

Variables	df	*F*	Sig.	Partial Eta Squared
Crt	1, 19	5.575	0.029	0.227
Crt * Group	1, 19	2.725	0.115	0.125
Dir	1, 19	5.697	0.028	0.231
Dir * Group	1, 19	2.954	0.102	0.135
Eye	1.904, 36.171	4.446	0.020	0.190
Eye * Group	1.904, 36.171	1.591	0.218	0.077
Crt * Dir	1, 19	1.911	0.183	0.091
Crt * Dir * Group	1, 19	0.019	0.893	0.001
Crt * Eye	1.607, 30.537	5.492	0.013	0.224
Crt * Eye * Group	1.607, 30.537	1.042	0.351	0.052
Dir * Eye	1.867, 35.469	1.550	0.227	0.075
Dir * Eye * Group	1.867, 35.469	0.046	0.946	0.002
Crt * Dir * Eye	1.828, 34.728	1.737	0.193	0.084
Crt * Dir * Eye * Group	1.828, 34.728	1.441	0.250	0.070
Group	1, 19	0.287	0.598	0.015

ANOVA also showed significant interactions between subject groups and contrast condition [group vs. contrast, $F_{(1, 16)} = 9.326$, $p = 0.008$; **Figure 4B**]. In NTE, percept duration decreased with a decrease in contrast, while amblyopes had no clear variation. This effect suggested that amblyopes seem to have a different motion processing mechanism from NTE. Another interaction showed a significant effect of the contrast and eye condition [$F_{(1.973, 31.575)} = 4.420$, $p = 0.021$; **Figure 4D**]. The performance in the binocular condition and stronger eye condition was similar across contrast conditions, while results differed according to contrast when the observer was using the weak eye to do the task. In this latter viewing condition, duration was slightly decreased when contrast increased, and the duration was always longer than the duration in the other two eye conditions. Thus, this interaction was mainly caused by the weak eye. *Post-hoc* Bonferroni-corrected comparisons for interaction between contrast and eye conditions showed that the dominant/fellow eye had a strong tendency for resulting in a different outcome than the binocular viewing condition [$F_{(1, 16)} = 4.463$, $p = 0.051$], while the weak eye had a different outcome than the binocular condition [$F_{(1, 16)} = 7.624$, $p = 0.014$]. No other effects were significant (**Table 5**).

Experiment 3: Control of Contrast Effects

We performed a control experiment to cross-check the effect of contrast in a different manner. We measured 5 AMB and 6 NTE (all participated in Experiment 1 or Experiment 2) in only the vertical condition to avoid a low number of switches with 6 levels of contrast (0.03, 0.05, 0.1, 0.15, 0.35, 0.5) with the hypothesis that the AMB should exhibit no variation with contrast, while the NTE should show an increase in the number of switches with a lower contrast. The results showed a clear interaction between the linear slopes of the number of switches versus contrast in AMB and NTE [group vs. contrast: $F_{(1, 8)} = 11.9$, $p = 0.009$], with the slope from AMB not different from zero (b = 4.2, CI = [−11.74, 20.22], $R^2 = 0.12$, $p = 0.502$) and a significantly negative slope from NTE ($b = -18.27$, CI = [−26.76, 9.78], $R^2 = 0.90$, $p = 0.0039$; see **Figure 5**). These results were also present when analyzing overall mean percept duration vs. contrast [Group vs. Contrast: $F_{(1, 8)} = 9.31$, $p = 0.016$; **Figure 5**]. The results were nearly identical when regressing in log-contrast space (number of switches vs. log-contrast, interaction group vs. contrast: $F_{(1, 8)} = 11.898$, $p = 0.009$; percept duration vs. log-contrast, interaction group vs. contrast: $F_{(1, 8)} = 9.037$, $p = 0.017$].

In summary, as expected, we found that contrast affected percept switches and percept durations by increasing the number of switches and decreasing the durations of the percepts with lower contrasts of gratings. In line with our expectation, this effect was mainly observed in NTE, and AMB showed no clear changes in percept duration with changes in contrast. Thus, based on our original hypothesis of decreased SNR with lower stimulus contrast, AMB seemed to show weak changes in plaid motion perception when contrast of the stimulus varied.

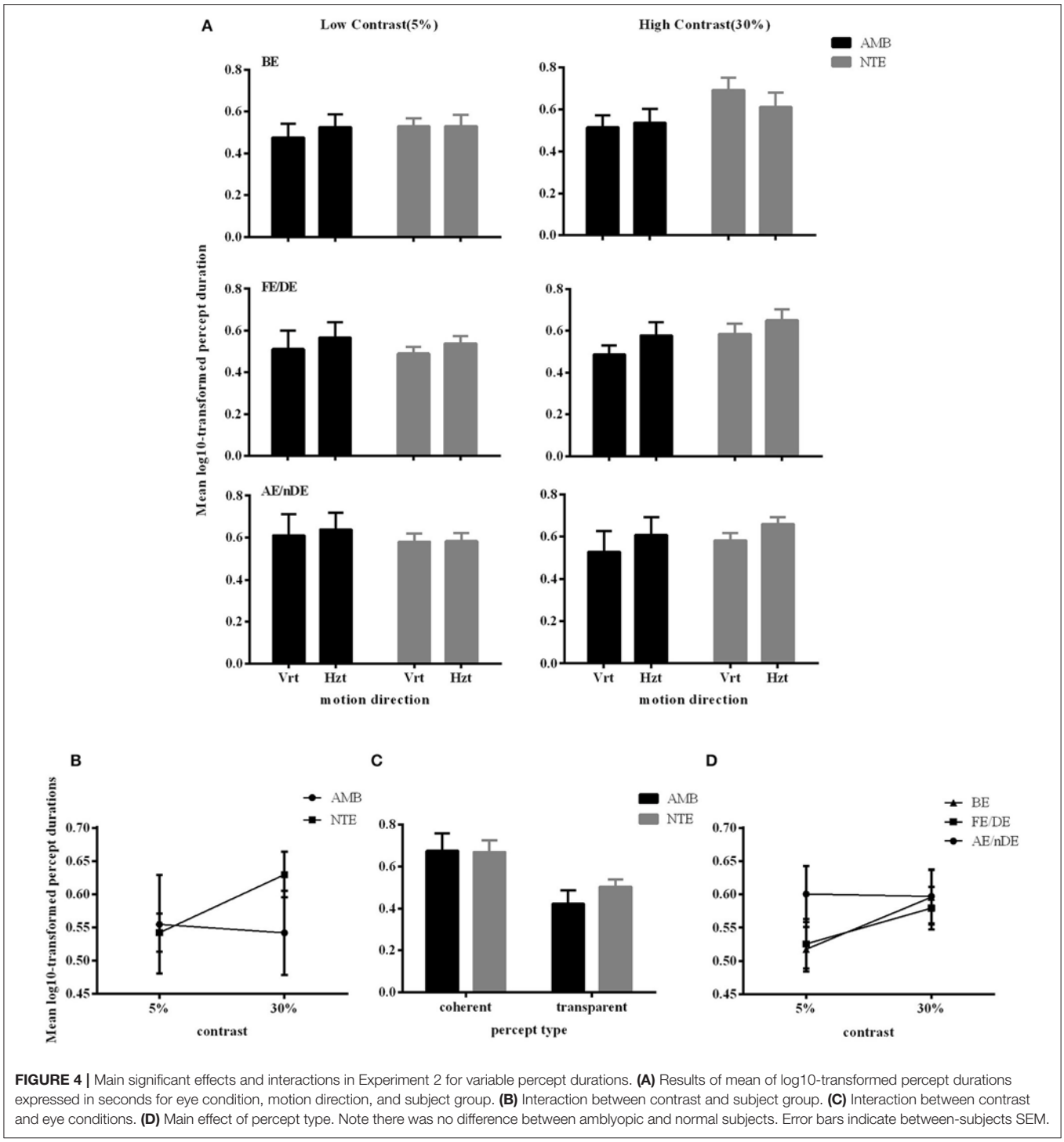


FIGURE 4 | Main significant effects and interactions in Experiment 2 for variable percept durations. **(A)** Results of mean of log10-transformed percept durations expressed in seconds for eye condition, motion direction, and subject group. **(B)** Interaction between contrast and subject group. **(C)** Interaction between contrast and eye conditions. **(D)** Main effect of percept type. Note there was no difference between amblyopic and normal subjects. Error bars indicate between-subjects SEM.

Correlation Between Bistability and VA or SA

We tested the correlation of the classic visual deficits as measured with the visual acuity (VA) and stereo acuity (SA) tests with the strength of bistability as measured through the number of switches. **Table 6** shows that there were no significant correlations for all monocular conditions in the amblyopic group in both Experiments 1 and 2.

Model Predictions of Bistable Motion Perception and Consequences for the Amblyopic Visual Motion System

We used the tristable model defined by Huguet et al. (2014) to identify the plausible internal mechanisms underlying the results of Experiment 2. Because these authors argued and presented evidence that moving plaid stimuli consist of not two but three

TABLE 5 | ANOVA results on Mean of log-10 Durations of Experiment 2.

Variables	df	*F*	Sig.	Partial Eta Squared
Crt	1, 16	5.173	0.037	0.244
Crt * Group	1, 16	9.326	0.008	0.368
Dir	1, 16	2.785	0.115	0.148
Dir * Group	1, 16	0.623	0.442	0.037
Eye	1.742, 27.868	3.155	0.064	0.165
Eye * Group	1.742, 27.868	1.524	0.236	0.087
Per	1, 16	19.241	0.000	0.546
Per * Group	1, 16	0.813	0.381	0.048
Crt * Dir	1,16	0.147	0.706	0.009
Crt * Dir * Group	1,16	0.074	0.789	0.005
Crt * Eye	1.973, 31.575	4.420	0.021	0.216
Crt * Eye * Group	1.973, 31.575	0.039	0.961	0.002
Dir * Eye	1.711, 27.375	2.178	0.139	0.120
Dir * Eye * Group	1.711, 27.375	0.545	0.559	0.033
Crt * Dir * Eye	1.783, 28.527	2.714	0.089	0.145
Crt * Dir * Eye * Group	1.783, 28.527	0.288	0.727	0.018
Crt * Per	1, 16	1.430	0.249	0.082
Crt * Per * Group	1, 16	0.682	0.421	0.041
Dir * Per	1, 16	0.360	0.557	0.022
Dir * Per * Group	1, 16	0.332	0.572	0.020
Crt * Dir * Per	1, 16	0.119	0.735	0.007
Crt * Dir * Per * Group	1, 16	0.400	0.536	0.024
Eye * Per	1.872, 29.944	2.213	0.130	0.121
Eye * Per * Group	1.872, 29.944	0.026	0.968	0.002
Crt * Eye * Per	1.799, 28.777	0.482	0.603	0.029
Crt * Eye * Per * Group	1.872, 29.944	1.325	0.279	0.076
Dir * Eye * Per	1.885, 30.154	0.256	0.763	0.016
Dir * Eye * Per * Group	1.885, 30.154	0.009	0.988	0.001
Crt * Dir * Eye * Per	1.607, 25.713	1.012	0.362	0.060
Crt * Dir * Eye * Per * Group	1.607, 25.713	0.925	0.390	0.055
Group	1, 16	0.298	0.593	0.018

different percepts, i.e., the transparent condition with two clearly perceived sliding gratings can have two states with different depth orderings, and that there are perceptual switches across the three states, we considered this model as more relevant to our experiments even though the experimental task was only a simple dual report of either transparent or coherent motion. Their model incorporates three populations of neurons that code three possible percepts: coherence (C), transparent with the leftward (counterclockwise) moving grating on top (T_L), and transparent with the rightward (clockwise) moving grating on top (T_R); in the use of the model here, we considered the transparent state (T) only when the C state was not active. A schematic of the model is presented in **Figure 6**, and it contains 6 parameters (β_1, β_2, γ, σ, I_C, $I_T = I_{TL} = I_{TR}$). The model is used in a range of parameters providing winner-takes-all behavior where only one of the three populations can be active at a given time, thus representing the active percept. Competitive inhibition between the three neuronal populations, together with spike-frequency adaptation and internal noise, provide the substrate for perceptual switches between the percepts.

As described in Huguet et al. (2014), the model parameters play essential roles in determining the mean number of percept switches and their duration. We parametrically varied the parameters in order to understand their effects on the two main measures. **Figure 6** presents representative simulation results for model parameters of $\beta_1 = 0.9$, $\beta_2 = 0.7$, $\sigma = 0.06$, $\gamma = 0.2$, $I_C = 1$, and $I_T = I_{TL} = I_{TR} = 0.9I_C$, when varying one of the last four parameters. An increase in internal noise σ strongly increases the number of percept switches and concurrently decreases the durations of the two percepts of C and T states (**Figure 7A**). An increase in the adaptation strength γ also increases the number of perceptual switches but differentially affects the C and T states (**Figure 7B**), with the C state duration showing a stronger relation (decrease) to an increase in adaptation than the T state, making C durations longer than the T duration at low γ and the reverse pattern observed with stronger γ. When the input strength is varied (with relative input T-to-C as constant; **Figure 7C**), the number of percept switches rapidly decreases at low inputs, corresponding to rapid increases in the signal-to-noise ratio. However, the number of percept switches is also observed to exhibit a minimum after which it begins to increase again. From multiple simulations, we found that this minimum was strongly dependent on the relative input strengths (I_T/I_C) as well as on the inhibitory strengths (β_1, β_2; results not shown). The durations of the two types of percepts, C and T, concurrently changed with a strong change in the number of switches. The percepts also showed a change in their relative durations with low input strengths showing T states longer than C ones and a reversal at higher input values. Finally, a change in the relative strength between C and T inputs demonstrated a typical bell-shaped curve for the number of switches (Brascamp et al., 2015), with the maximum value near input equality, together with their concurrent C and T state duration changes (**Figure 7D**). These last effects mimicked the expected effects of relative input strengths onto the two variables as observed in previous reports (Moreno-Bote et al., 2010; Brascamp et al., 2015).

Similar observations were obtained for other inhibitory strengths (β_1, β_2) but with the absolute values of noise, input, adaptation, and relative input strengths correspondingly changed.

The above simulations show two important effects. First, the number of perceptual switches and percept durations are very sensitive to the internal noise and adaptation strength (**Figures 7A,B**). This observation supports the original hypothesis that plaid gratings would show differences between the two groups of subjects that putatively have different noise levels in their motion visual system (Mansouri and Hess, 2006). In contrast to this prediction, Experiment 1 did not show any differences between AMB and NTE. Second, a striking effect was present in the simulation for the absolute input strengths I_C and I_T that represent the inputs of the C and T states. At very low input levels, the internal noise of the system is much stronger than the input strengths and thus makes the system oscillate much faster between the two states. This effect is in line

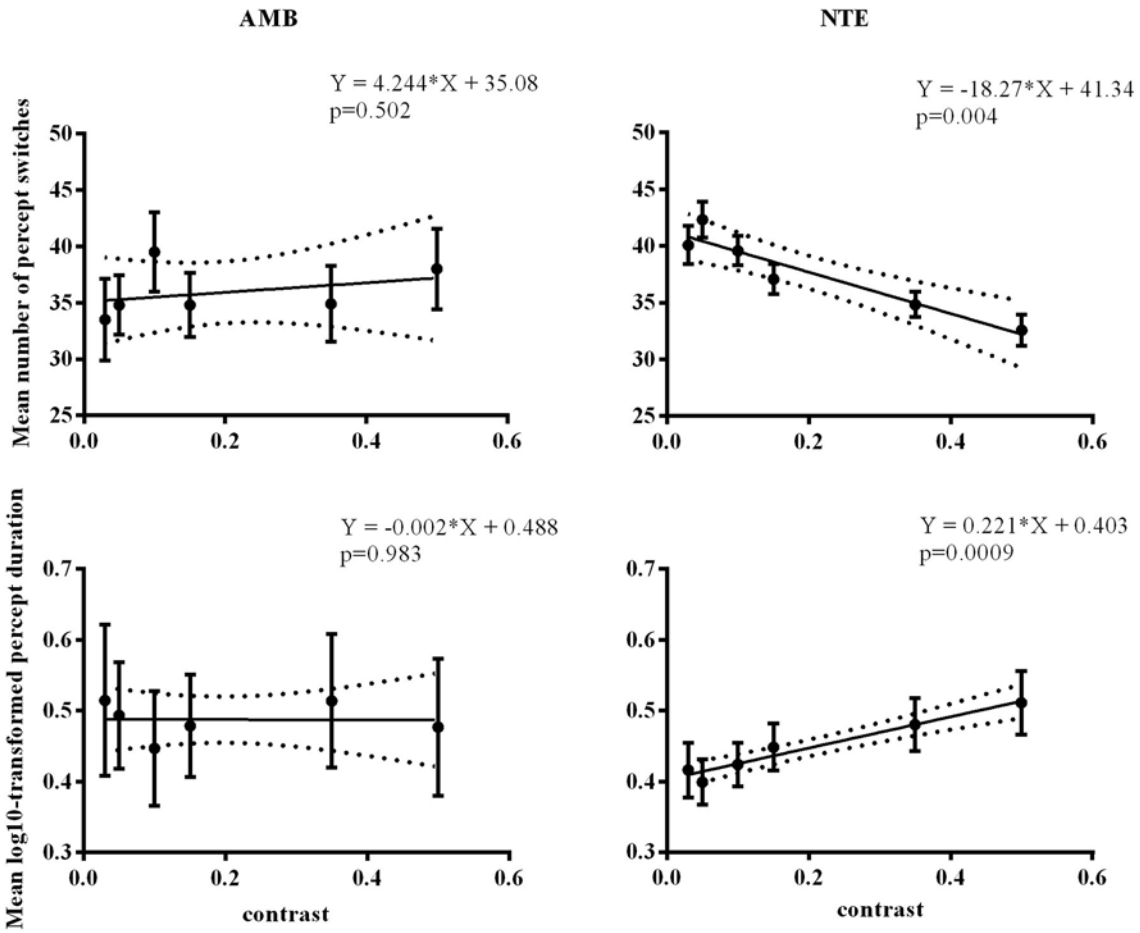


FIGURE 5 | Linear regression across different contrast conditions in AMB and NTE. Top showing percept switches; bottom showing percept duration. Left column graphics for AMB; right column for NTE. Solid line was the best-fit line, while the dashed line indicates the 95% confidence band of the best-fit line. Error bar indicates the SEM.

TABLE 6 | Correlation between bistablity and SA or VA.

	Number of percept switches vs.	Pearson Correlation	Sig.	*N*
Experiment 1	Log of SA	−0.272	0.447	10
	Log of VA of AE	−0.192	0.596	10
	Log of VA of FE	0.217	0.546	10
Experiment 2	Log of SA	−0.034	0.927	10
	Log of VA of AE	−0.383	0.274	10
	Log of VA of FE	−0.372	0.290	10

with our hypothesis that lower grating contrasts would increase the number of switches and percept durations, which led us to perform Experiment 2 with the idea that AMB should exhibit an increase in the number of switches and also show a decrease in the durations of the percepts. However, the results differed from our expectation, with NTE showing the predicted effect, but AMB showing no changes with lower grating contrasts.

DISCUSSION

We investigated putative differences in the visual motion system between anisometropic amblyopes and neurotypical observers through the use of bistable plaid motion perception. First, our group of amblyopes globally exhibited normal bistable perception in any viewing condition (binocular, monocular with amblyopic or fellow eye) when compared to the control group. Second, we hypothesized that lower contrast of the plaid stimulus should emphasize the internal noise differences between the two groups and thus lead to a stronger increase in percept switches and decrease in percept durations. The results confirmed this hypothesis only in the control group, while the amblyopic group exhibited no changes. These latter results are at odds with the idea of stronger noise in the amblyopic motion system, and plausible explanations of these discrepancies are discussed below.

Bistable perception of plaid square gratings was found to be normal in anisometropic amblyopes when compared to that in the neurotypical controls. These results are in agreement with

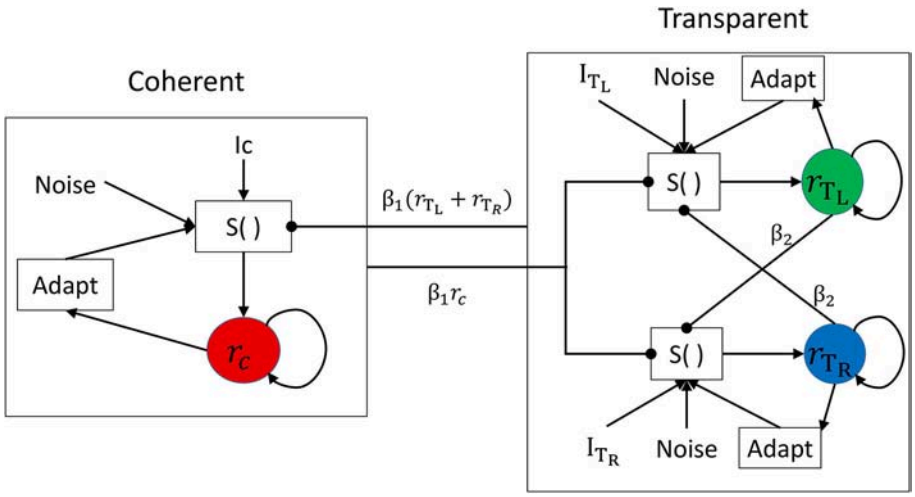


FIGURE 6 | Network architecture for the neuronal competition model with direct mutual inhibition. The activity of each population is associated with a different percept: coherent (C), transparent right (T_R), or transparent left (T_L). Each population receives an excitatory deterministic input of strength I and independent noise *n*. Spike-frequency adaptation is present in each population. The function S() represents the sigmoidal transducer.

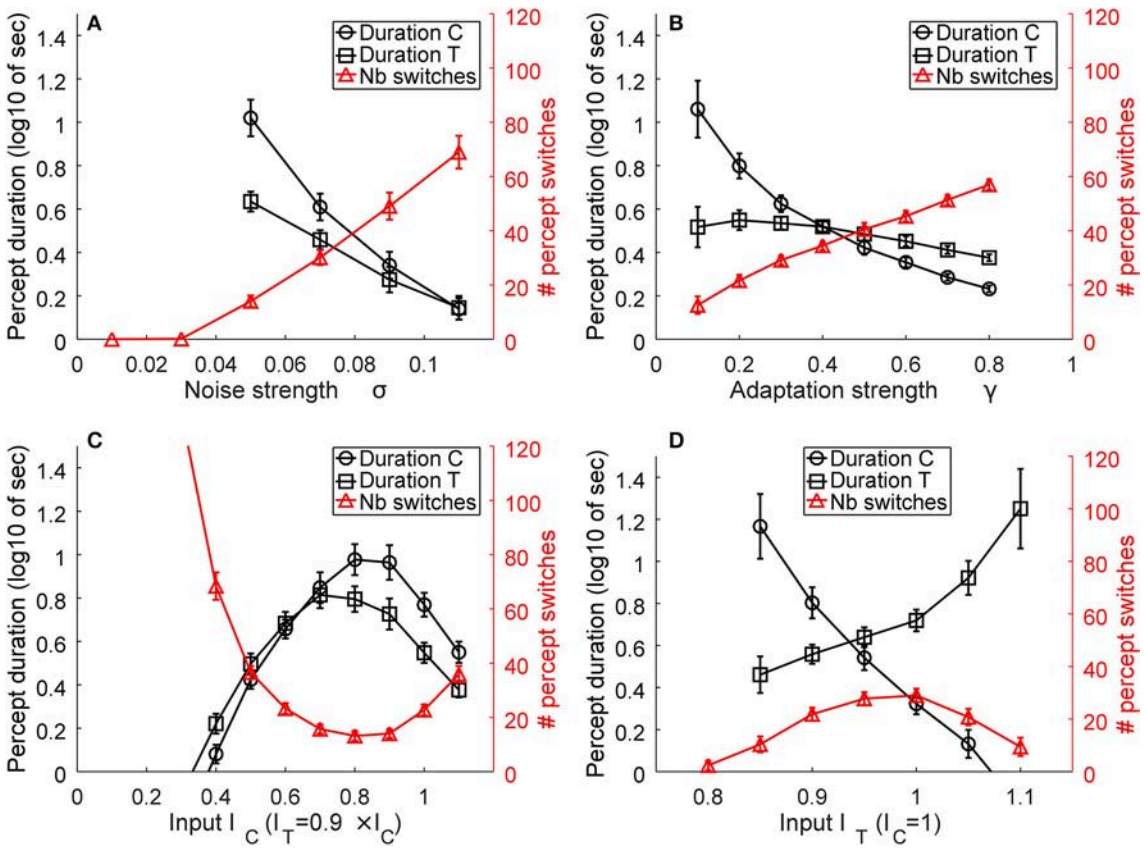


FIGURE 7 | Representative model results. The effects of noise **(A)**, adaptation **(B)**, input strength, as the absolute value of I_C and $I_T = I_{TL} = I_{TR}$ **(C)**, and relative transparent-to-coherent input **(D)** on the mean number of percept switches (red curves and right y-axis) and percept durations (black curves and left y-axis, expressed in log10 of seconds). Error bars indicate standard deviation of $n = 30$ simulations for each datum.

previous reports of normal perception of bistable sine-grating plaids in such group of subjects (Thompson et al., 2008, 2012; Hamm et al., 2014), even when first-order contrast deficits are taken into account (Tang et al., 2012). In our study, these earlier reports are confirmed through analysis of perceptual bistability applied on square gratings.

While bistability of the percepts was similarly seen and stochastic across eye-viewing conditions and groups of subjects, our methods and results unveiled a new and unexpected effect of contrast on plaid motion perception in amblyopes. Based on reports of possibly stronger internal noise in the amblyopic visual motion system (Simmers et al., 2003; Mansouri and Hess, 2006; Hamm et al., 2014) and theoretical insights into perceptual bistability and neural noise (Brascamp et al., 2006; Moreno-Bote et al., 2007; Shpiro et al., 2009; Huguet et al., 2014), lower contrasts of the stimulus were argued to decrease the duration of each percept in amblyopes when compared to that in the control group. This effect was found, but it was reversed between groups, with the control group showing decreased percept stability (decrease in percept durations), while the amblyopes did not exhibit such an effect.

This result is interesting in at least two aspects. First, contrast sensitivity, the reciprocal of contrast threshold that is used to describe subjects' ability to visually detect a target, is known to be strongly affected in amblyopic eyes (Woodruff, 1991). Earlier research has shown that contrast sensitivity is highly decreased in the amblyopic eye, especially at high spatial frequencies, but the sensitivity of the fellow eye is also affected when compared with the eyes in normal subjects (Bradley and Freeman, 1981). Interestingly, amblyopes do not exhibit clear deficits in contrast perception at suprathreshold stimulus contrasts, indicating that there is no clear contrast coding abnormality for the suprathreshold contrast range in amblyopes (Hess and Bradley, 1980; Loshin and Levi, 1983). On the contrary, suprathreshold static grating perception is affected but in a very different manner. Amblyopes staring at images of classic square gratings perceive perceptual distortions of the stimulus that could be of static or dynamic nature (Hess et al., 1978; Sireteanu et al., 2008; Thiel and Iftime, 2016). Thus, the two facts that (1) our group of amblyopes perceived the 120-s moving plaids normally, with classic perceptual bistability and no reports of differences in perception between the weak and fellow eyes, and (2) amblyopes did not show an effect of contrast on the global bistability of the percept hint to a motion coding system in their visual pathway that uses dynamic visual input in a different way from neurotypical subjects. The results of neurotypical subjects experimentally confirmed the inversed "Levelt IV rule" at low contrasts (Brascamp et al., 2015), but the overall pattern of results led us to consider in further detail the models of plaid motion perception and a plausible explanation of the effects observed in amblyopes.

In analyzing and applying a model (Huguet et al., 2014), we found that input intensity indeed affected percept switches and durations as hypothesized. These effects also suggested that, for amblyopes, contrast of the stimulus is decoupled from or very weakly related to the "input" variable of the model. This suggests that there may be different motion coding system in the amblyopic visual system from that in the neurotypical one, with the perceptual switches observed in the former visual motion system related to different mechanisms.

From a neurophysiological perspective, motion coding and decoding of plaid stimuli might not be performed at a single stage, but instead, multiple areas may be involved (Thompson et al., 2012; Villeneuve et al., 2012). Thus, the segregation of motion (transparency) or the assimilation of motion (coherency) may be coded in a distributed manner across the early cortices. The differences between our amblyopic and control groups in contrast effects might stem from the fact that, in the amblyopic system, motion coherency and transparency coding could be more widely distributed than in neurotypical subjects, as suggested by a recent study (Thompson et al., 2012). From a different and more detailed perspective, the major motion area MT is known to contain cells that can selectively respond to the pattern or components of moving plaid gratings (Rust et al., 2006) and, furthermore, has some depth coding structure (Born and Bradley, 2005) that should help to create depth ordering of different motion surfaces. Although MT cells in the macaque monkey seem to have dominance over fellow eye inputs, the distribution of cells sensitive to pattern and the components of plaid gratings were found equal (El-Shamayleh et al., 2010), thus showing global similar plaid motion coding. Therefore, we might assume that the equivalent percepts of coherence and transparency are decoded through a simple rule: to decode only one neuronal population—component or pattern cells. Because MT cells receive major input from V1 cells, the contrast dependence of all MT cells should be similar. The observation in control subjects of stronger perceptual changes at lower contrast supports the idea that pattern and component cells should be similarly activated by contrast strength. On the other hand, the lack of contrast effects in amblyopes seems to indicate that pattern and component cells have different input relations to the contrast of the stimulus. This difference provides an interesting possibility and its exact nature is far from the scope of the current study.

Importantly, the model used here is more qualitative in nature, helping to grasp essential structural differences and changes in the multistable perception of plaid motion stimuli but not providing a realistic implementation of motion coding. Recent studies reported that, closely related to our work, tristable motion perception could be explained by a more detailed motion-tuned neuronal population (Meso et al., 2016; Medathati et al., 2017) that more closely resembles MT physiology. Further investigations and theoretical modeling also incorporating depth coding should help to unravel the plausible changes in the amblyopic motion system.

A systematic and interesting difference we found was the global direction effect. Both amblyopes and normal subjects had more percept switches when global motion direction was upward, i.e., vertical, than when it was horizontal. We did not find systematic effects between the two groups across the first two experiments. Differences between cardinal axes have already been reported in previous studies of visual motion perception in ambiguous conditions (Castet et al., 1999; Hupé and Rubin, 2004). The exact nature of the asymmetry in bistability between

vertical and horizontal global motions may lie in the eye movement differences between these two cardinal directions. The global effect present across all observers might stem from clear differences in eye movement dynamics of horizontal and vertical eye movement (fixational, reflexive, or voluntary pursuit eye movements) (Baloh et al., 1988; Sparks, 2002). This explanation partly supports a separate control of vertical and horizontal pursuit, which may contribute to the direction difference that is systematically reported. Furthermore, eye movement may influence the percept through retinal motion. Van Dam et al. demonstrated that the retinal image shift, caused by saccade, can change the bistable percept (van Dam and van Ee, 2005, 2006). For clarification of the exact mechanism of such a direction effect and determination of whether amblyopes with clear changes or deficits in eye movements exhibit an effect on perception of plaid motion, further studies are still needed with proper measures and controls for eye movements in neurotypical and amblyopic groups.

In summary, by using bistable plaid motion as a probe of the visual motion system, we found a systematic and clear effect of stimulus contrast on perceptual bistability in neurotypical subjects that was not present in anisometropic amblyopes. The former effect is explained by classic models of multistability and thus hints toward a generally different motion coding and decoding system in the amblyopes.

ETHICS STATEMENT

This study was carried out in accordance with the recommendations of Ethical review of biomedical research involving human beings, Committee on biomedical ethics of university of science and technology of China with written informed consent from all subjects. All subjects gave written informed consent in accordance with the Declaration of Helsinki. The protocol was approved by the Committee on biomedical ethics of university of science and technology of China.

AUTHOR CONTRIBUTIONS

JL carried out the experiments, and wrote the manuscript with the support of TT and YZ. Both JL and TT contributed to the design and implementation of research, to the analysis of results. YZ supervised the project. All the authors reviewed the manuscript and conceptualized this study.

ACKNOWLEDGMENTS

This study was supported by the National Natural Science Foundation of China (NSFC 31230032 and 91749102 to YZ) and the Fundamental Research Funds for the Central Universities (TT).

REFERENCES

Aaen-Stockdale, C., Ledgeway, T., and Hess, R. F. (2007). Second-order optic flow deficits in amblyopia. *Invest. Ophthalmol. Vis. Sci.* 48, 5532–5538. doi: 10.1167/iovs.07-0447

Adelson, E. H., Movshon, J. A. (1982). Phenomenal coherence of moving visual patterns. *Nature* 300, 523–525. doi: 10.1038/300523a0

Baloh, R., Yee, R., Honrubia, V., and Jacobson, K. (1988). A comparison of the dynamics of horizontal and vertical smooth pursuit in normal human subjects. *Aviat. Space Environ. Med*. 59, 121–124.

Barnes, G., Hess, R., Dumoulin, S., Achtman, R., and Pike, G. (2001). The cortical deficit in humans with strabismic amblyopia. *J. Physiol*. 533. 281–297. doi: 10.1111/j.1469-7793.2001.0281b.x

Bonhomme, G. R., Liu, G. T., Miki, A., Francis, E., Dobre, M.-C., Haselgrove, J. C. et al. (2006). Decreased cortical activation in response to a motion stimulus in anisometropic amblyopic eyes using functional magnetic resonance imaging. *J. AAPOS* 10, 540–546. doi: 10.1016/j.jaapos.2006.07.008

Born, R. T., and Bradley, D. C. (2005). Structure and function of visual area MT. *Annu. Rev. Neurosci.* 28, 157–189. doi: 10.1146/annurev.neuro.26.041002.131052

Bradley, A., and Freeman, R. D. (1981). Contrast Sensitivity in anisometropic amblyopia. *Invest. Ophthalmol. Vis. Sci*. 21, 467–76.

Brascamp, J. W., Klink, P., and Levelt, W. J. (2015). The 'laws' of binocular rivalry: 50 years of Levelt's propositions. *Vision Res.* 109, 20–37. doi: 10.1016/j.visres.2015.02.019

Brascamp, J. W., van Ee, R., Noest, A. J., Jacobs, R. H., and van den Berg A. V., (2006). The time course of binocular rivalry reveals a fundamental role of noise. *J. Vis*. 6, 1244–1256. doi: 10.1167/6.11.8

Britten, K. H., Shadlen, M. N., Newsome, W. T., and Movshon, J. A. (1992). The analysis of visual motion: a comparison of neuronal and psychophysical performance. *J. Neurosci.* 12, 4745–4765. doi: 10.1523/JNEUROSCI.12-12-04745.1992

Castet, E., Charton, V., and Dufour, A. (1999). The extrinsic/intrinsic classification of two-dimensional motion signals with barber-pole stimuli. *Vision Res.* 39, 915–32. doi: 10.1016/S0042-6989(98)00146-1

Constantinescu, T., Schmidt, L., Watson, R., and Hess, R. F. (2005). A residual deficit for global motion processing after acuity recovery in deprivation amblyopia. *Invest. Ophthalmol. Vis. Sci*. 46, 3008–3012. doi: 10.1167/iovs.05-0242

Ellemberg, D., Lewis, T. L., Maurer, D., Brar, S., and Brent, H. P. (2002). Better perception of global motion after monocular than after binocular deprivation. *Vision Res.* 42, 169–179. doi: 10.1016/S0042-6989(01)00278-4

El-Shamayleh, Y., Kiorpes, L., Kohn, A., and Movshon, J. A. (2010). Visual motion processing by neurons in area MT of macaque monkeys with experimental amblyopia. *J. Neurosci.* 30, 12198–12209. doi: 10.1523/JNEUROSCI.3055-10.2010

Hamm, L. M., Black, J., Dai, S., and Thompson, B. (2014). Global processing in amblyopia: a review. *Front. Psychol.* 5:583. doi: 10.3389/fpsyg.2014.00583

Hess, R., Campbell, F., and Greenhalgh, T. (1978). On the nature of the neural abnormality in human amblyopia; neural aberrations and neural sensitivity loss. *Pflügers Arch*. 377, 201–207. doi: 10.1007/BF00584273

Hess, R. F., and Bradley, A. (1980). Contrast perception above threshold is only minimally impaired in human amblyopia. *Nature* 287, 463–464. doi: 10.1038/287463a0

Hess, R. F., and Thompson, B. (2015). Amblyopia and the binocular approach to its therapy. *Vision Res.* 114, 4–16. doi: 10.1016/j.visres.2015.02.009

Hess, R. F., Thompson, B., Gole, G. A., and Mullen, K. T. (2010). The amblyopic deficit and its relationship to geniculo-cortical processing streams. *J. Neurophysiol.* 104, 475–483. doi: 10.1152/jn.01060.2009

Ho, C. S., and Giaschi, D. E. (2009). Low-and high-level motion perception deficits in anisometropic and strabismic amblyopia: evidence from fMRI. *Vision Res.* 49, 2891–2901. doi: 10.1016/j.visres.2009.07.012

Hubel, D. H., and Wiesel, T. N. (1965). Receptive fields and functional architecture in two nonstriate visual areas (18 and 19) of the cat. *J. Neurophysiol.* 28, 229–289. doi: 10.1152/jn.1965.28.2.229

Hubel, D. H., and Wiesel, T. N. (1970). The period of susceptibility to the physiological effects of unilateral eye closure in kittens. *J. Physiol*. 206, 419–436. doi: 10.1113/jphysiol.1970.sp009022

Huguet, G., Rinzel, J., and Hupé, J.-M. (2014). Noise and adaptation in multistable perception: noise drives when to switch, adaptation determines percept choice. *J. Vis.* 14, 1–24. doi: 10.1167/14.3.19

Hupé, J. M., Rubin, N. (2003). The dynamics of bi-stable alternation in ambiguous motion displays: a fresh look at plaids. *Vision Res*. 43, 531–548. doi: 10.1016/S0042-6989(02)00593-X

Hupé, J. M., and Rubin, N. (2004). The oblique plaid effect. *Vision Res.* 44, 489–500. doi: 10.1016/j.visres.2003.07.013

Kiorpes, L., and McKee, S. P. (1999). Neural mechanisms underlying amblyopia. *Curr. Opin. Neurobiol.* 9, 480–486. doi: 10.1016/S0959-4388(99)80072-5

Lago-Fernández, L. F., and Deco, G. (2002). A model of binocular rivalry based on competition in IT. *Neurocomputing* 44 503–507. doi: 10.1016/S0925-2312(02)00408-3

Laing, C. R., and Chow, C. C. (2002). A spiking neuron model for binocular rivalry. *J. Comput. Neurosci.* 12, 39–53. doi: 10.1023/A:1014942129705

Levi, D. M., McKee, S. P., and Movshon, J. A. (2011). Visual deficits in anisometropia. *Vision Res.* 51, 48–57. doi: 10.1016/j.visres.2010.09.029

Li, X., Mullen, K. T., Thompson, B., and Hess, R. F. (2011). Effective connectivity anomalies in human amblyopia. *Neuroimage* 54, 505–516. doi: 10.1016/j.neuroimage.2010.07.053

Loshin, D. S., and Levi, D. M. (1983). Suprathreshold contrast perception in functional amblyopia. *Doc. Ophthalmol.* 55, 213–236. doi: 10.1007/BF00140810

Majaj, N. J., Carandini, M., and Movshon, J. A. (2007). Motion integration by neurons in macaque MT is local, not global. *J. Neurosci.* 27, 366–370. doi: 10.1523/JNEUROSCI.3183-06.2007

Mansouri, B., and Hess, R. F. (2006). The global processing deficit in amblyopia involves noise segregation. *Vision Res.* 46, 4104–4117. doi: 10.1016/j.visres.2006.07.017

Medathati, V. N. K., Rankin, J., Meso, A. I., Kornprobst, P., and Masson, G. S. (2017). Recurrent network dynamics reconciles visual motion segmentation and integration. *Sci. Rep.* 7:11270. doi: 10.1038/s41598-017-11373-z

Meso, A. I., Rankin, J., Faugeras, O., Kornprobst, P., and Masson, G. S. (2016). The relative contribution of noise and adaptation to competition during tri-stable motion perception. *J. Vis*. 16, 1–24. doi: 10.1167/16.15.6

Moreno-Bote, R., Rinzel, J., and Rubin, N. (2007). Noise-induced alternations in an attractor network model of perceptual bistability. *J. Neurophysiol.* 98, 1125–1139. doi: 10.1152/jn.00116.2007

Moreno-Bote, R., Shpiro, A., Rinzel, J., and Rubin, N. (2010). Alternation rate in perceptual bistability is maximal at and symmetric around equi-dominance. *J. Vision* 10, 1–18. doi: 10.1167/10.11.1

Newsome, W. T., and Paré, E. B. (1988). A selective impairment of motion perception following lesions of the middle temporal visual area (MT). *J. Neurosci*. 8, 2201–2211. doi: 10.1523/JNEUROSCI.08-06-02201.1988

Rust, N. C., Mante, V., Simoncelli, E. P., and Movshon, J. A. (2006). How MT cells analyze the motion of visual patterns. *Nat. Neurosci.* 9, 1421–1431. doi: 10.1038/nn1786

Salzman, C. D., Britten, K. H., and Newsome, W. T. (1990). Cortical microstimulation influences perceptual judgements of motion direction. *Nature* 346, 174–177. doi: 10.1038/346174a0

Shpiro, A., Moreno-Bote, R., Rubin, N., and Rinzel, J. (2009). Balance between noise and adaptation in competition models of perceptual bistability. *J. Comput. Neurosci.* 27, 37–54. doi: 10.1007/s10827-008-0125-3

Simmers, A. J., Ledgeway, T., Hess, R. F., and McGraw, P. V. (2003). Deficits to global motion processing in human amblyopia. *Vision Res.* 43, 729–738. doi: 10.1016/S0042-6989(02)00684-3

Simmers, A., Ledgeway, T., Mansouri, B., Hutchinson, C., and Hess, R. (2006). The extent of the dorsal extra-striate deficit in amblyopia. *Vision Res.* 46, 2571–2580. doi: 10.1016/j.visres.2006.01.009

Sireteanu, R., Baumer, C. C., and Iftime, A. (2008). Temporal instability in amblyopic vision: relationship to a displacement map of visual space. *Invest. Ophthalmol. Vis. Sci.* 49, 3940–3954. doi: 10.1167/iovs.07-0351

Sparks, D. L. (2002). The brainstem control of saccadic eye movements. *Nat. Rev. Neurosci.* 3, 952–964. doi: 10.1038/nrn986

Thiel, A., and Iftime, A. (2016). Temporal instabilities in amblyopic perception: a quantitative approach. *Perception* 45, 443–465. doi: 10.1177/0301006615625796

Thompson, B., Aaen-Stockdale, C. R., Mansouri, B., and Hess, R. F. (2008). Plaid perception is only subtly impaired in strabismic amblyopia. *Vision Res.* 48, 1307–1314. doi: 10.1016/j.visres.2008.02.020

Thompson, B., Villeneuve, M., Casanova, C., and Hess, R. (2012). Abnormal cortical processing of pattern motion in amblyopia: evidence from fMRI. *Neuroimage* 60, 1307–1315. doi: 10.1016/j.neuroimage.2012.01.078

Tang, Y., Chen, L., Liu, Z., Liu, C., and Zhou, Y. (2012). Low-level processing deficits underlying poor contrast sensitivity for moving plaids in anisometropic amblyopia. *Vis. Neurosci.* 29, 315–323. doi: 10.1017/S095252381200034X

van Dam, L. C., and van Ee, R. (2005). The role of (micro)saccades and blinks in perceptual bi-stability from slant rivalry. *Vision Res.* 45, 2417–2435. doi: 10.1016/j.visres.2005.03.013

van Dam, L. C., and van Ee, R. (2006). Retinal image shifts, but not eye movements per se, cause alternations in awareness during binocular rivalry. *J. Vis.* 6, 1172–1179. doi: 10.1167/6.11.3

Villeneuve, M. Y., Thompson, B., Hess, R. F., and Casanova, C. (2012). Pattern-motion selective responses in MT, MST and the pulvinar of humans. *Eur. J. Neurosci.* 36, 2849–2858. doi: 10.1111/j.1460-9568.2012.08205.x

Wallace, D. K., Repka, M. X., Lee, K. A., Melia, M., Christiansen, S. P., Morse, C. L., et al. (2018). Amblyopia preferred practice pattern®. *Ophthalmology* 125, P105–P142. doi: 10.1016/j.ophtha.2017.10.008

Wong, A. M. (2012). New concepts concerning the neural mechanisms of amblyopia and their clinical implications. *Can. J. Ophthalmol.* 47, 399–409. doi: 10.1016/j.jcjo.2012.05.002

Woodruff, M. (1991). Amblyopia: basic and clinical aspects. *Optom. Vis. Sci.* 68, 365–396. doi: 10.1097/00006324-199105000-00017

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Computational Neuroscience

METHODS
published: 12 June 2018
doi: 10.3389/fncom.2018.00042

A Retinotopic Spiking Neural Network System for Accurate Recognition of Moving Objects Using NeuCube and Dynamic Vision Sensors

Lukas Paulun[1,2], *Anne Wendt*[1]* *and Nikola Kasabov*[1]

[1] *Knowledge Engineering and Discovery Research Institute, Auckland University of Technology, Auckland, New Zealand,*
[2] *Mathematical Institute, Albert Ludwigs University of Freiburg, Freiburg im Breisgau, Germany*

This paper introduces a new system for dynamic visual recognition that combines bio-inspired hardware with a brain-like spiking neural network. The system is designed to take data from a dynamic vision sensor (DVS) that simulates the functioning of the human retina by producing an address event output (spike trains) based on the movement of objects. The system then convolutes the spike trains and feeds them into a brain-like spiking neural network, called NeuCube, which is organized in a three-dimensional manner, representing the organization of the primary visual cortex. Spatio-temporal patterns of the data are learned during a deep unsupervised learning stage, using spike-timing-dependent plasticity. In a second stage, supervised learning is performed to train the network for classification tasks. The convolution algorithm and the mapping into the network mimic the function of retinal ganglion cells and the retinotopic organization of the visual cortex. The NeuCube architecture can be used to visualize the deep connectivity inside the network before, during, and after training and thereby allows for a better understanding of the learning processes. The method was tested on the benchmark MNIST-DVS dataset and achieved a classification accuracy of 92.90%. The paper discusses advantages and limitations of the new method and concludes that it is worth exploring further on different datasets, aiming for advances in dynamic computer vision and multimodal systems that integrate visual, aural, tactile, and other kinds of information in a biologically plausible way.

Keywords: Spiking neural networks (SNN), NeuCube, dynamic vision sensor (DVS), MNIST-DVS, retinotopy, deep learning in SNN

OPEN ACCESS

Edited by:
Xavier Otazu,
Universidad Autónoma de Barcelona, Spain

Reviewed by:
Timothée Masquelier,
Centre National de la Recherche Scientifique (CNRS), France
Pablo Martinez-Cañada,
Universidad de Granada, Spain

***Correspondence:**
Anne Wendt
anne.wendt@aut.ac.nz

Received: *11 September 2017*
Accepted: *24 May 2018*
Published: *12 June 2018*

Citation:
Paulun L, Wendt A and Kasabov N (2018) A Retinotopic Spiking Neural Network System for Accurate Recognition of Moving Objects Using NeuCube and Dynamic Vision Sensors. Front. Comput. Neurosci. 12:42. doi: 10.3389/fncom.2018.00042

INTRODUCTION

During the past years, the quest for accurate image recognition systems has been one of the driving forces behind major advances in the field of artificial neural networks such as the development of convolutional neural networks (Lecun et al., 1998). Today, algorithms for image recognition are well advanced and can be found in many applications such as search engines, security systems, industrial robots, medical devices, and virtual reality. Besides the many areas of application, another reason for the fast progress in image recognition might be the vast knowledge about the human visual system. The eye is arguably the best studied human sensory organ and the visual cortex has

been the main object of interest in a large number of neuroscientific studies. Findings from vision science have inspired the development of new hardware as well as novel algorithms and computational tools. High-definition and high-speed cameras have long surpassed the capacities of the human eye in terms of spatial and temporal resolution. On the software side though, it still proves to be a difficult task to extend the scope of present achievements in static image recognition to dynamic visual recognition of moving objects or a moving scene.

The benefit of accurate and fast dynamic visual recognition is apparent: each of the above-mentioned applications of image recognition constitutes a potential application area for dynamic visual recognition systems. Any kind of robot that must navigate within a three-dimensional environment or perform tasks on moving objects would benefit from an accurate and fast dynamic visual system. The popular topic of self-driving cars is only one example. Other potential implementations include security systems, automated traffic prediction and tolls, monitoring of manufacturing processes, navigational tools in air and ship traffic, or diagnostic assistants for inspections or surgery. Since the human visual system's adaptability and efficiency are still highly superior to computer systems when it comes to tasks of dynamic vision, it is natural to let biology serve as an inspiration for the development of new computational models.

Previous works have used a combination of bio-inspired visual sensors and spiking neural networks for the recognition of human postures (Perez-Carrasco et al., 2010), the extraction of car trajectories on a freeway (Bichler et al., 2012), or the control of robotic movements (Jimenez-Fernandez et al., 2009; Perez-Peña et al., 2013). We consider these very promising approaches, though the mentioned works lack benchmarking results that make them comparable.

This paper introduces a new system for dynamic visual recognition that combines a silicon retina device with a brain-like spiking neural network (SNN). As we introduce the different parts of our proposed system, we include findings from vision science that inspired us or that might provide promising approaches for future improvements. We present the setup and the results of a benchmarking experiment carried out on the MNIST-DVS dataset and show that our system achieves a classification accuracy of 92.90% on this dataset. The SNN architecture NeuCube is very flexible in terms of its connectivity and learning algorithms and allows for the visualization of the learning processes inside the SNN. After discussing the advantages and limitations of the system, we conclude by suggesting further exploration of the system's performance with modified algorithms and different datasets.

THE PROPOSED SYSTEM ARCHITECTURE

The Dynamic Vision Sensor

The Dynamic Vision Sensor (DVS) was developed at the Institute for Neuroinformatics in Zürich as a fast and storage efficient silicon retina system (Delbruck, 2008). Unlike conventional frame-based video cameras that capture multiple frames per second and store a large number of pixels for each of these frames, the DVS only captures changes in the brightness of single pixels caused by movement of the scene or an object (Lichtsteiner et al., 2008). This is called an Address Event Representation (AER) since the output of the sensor consists of a time series of events together with their location (address), representing the temporal contrast of a specific pixel at a specific time. By responding to temporal contrast on the pixel-level rather than taking a continuous series of snapshots of the whole scene, the DVS mimics the functioning of the human retina much better than conventional video cameras (Purves, 2012).

Together with its focus on movements within a scene there is another reason to choose the DVS over a conventional video camera for a dynamic vision system based on a spiking neural network: the address event output of the DVS comes in the form of a series of spike trains, each spike train corresponding to one pixel of the sensor. Every single spike in the train of one specific pixel represents a change in brightness in that pixel at a specific time. However, there are two difficulties with taking the raw DVS output as spike trains and directly feeding them into a spiking neural network: firstly, the sensor can achieve a very high temporal resolution of 1 μs and a spike train for a single pixel will initially consist of many time steps, e.g., 2,000,000 time steps for a 2 s video, and a relatively small number of spikes. Feeding such a spike train into a spiking neural network would result in very low overall spiking activity and probably unsatisfying performance. Secondly, although the sensor's spatial resolution of $128 \times 128 = 16{,}384$ pixels is low compared to conventional video cameras, it is desirable to reduce computational cost by integrating the signals of multiple pixels into single input neurons for the SNN rather than creating 16,384 input neurons.

For this purpose, we propose an algorithm for the compression of time and the convolution and pooling of the DVS pixels into a total of 128 spike trains consisting of roughly 100 time steps for each second of video data that can then be fed into 128 input neurons of an SNN.

Proposed Encoding Algorithm of DVS Data as Input Data for the SNN System

The algorithm we propose is inspired by the structure and organization of retinal ganglion cells. These cells receive information from photoreceptors on the retina and transmit them to the brain (Purves, 2012). There are different types of retinal ganglion cells, but we focus on two global properties shared by the majority of all ganglion cells: first, the distribution of retinal ganglion cells across the retina, which is used to determine which photoreceptors converge into one retinal ganglion cell and, thus, how many DVS pixels converge into one input neuron for our SNN. Second, the mechanism by which retinal ganglion cells fire and, thus, the algorithm that generates the input spike trains for the SNN.

Pooling of DVS Output Into 128 Input Neurons of the SNN System

Despite large differences across individuals, there are roughly 100 million photoreceptor cells on the retina and around 1 million retinal ganglion cells providing information transmission to the brain (Curcio et al., 1990). Thus, on average, one ganglion cell integrates information from roughly 100 photoreceptor

cells. However, the number of photoreceptors converging into one ganglion cell depends highly on the retinal location of the photoreceptors. Ganglion cells connecting to the *fovea centralis*, the small central spot of the retina specialized in sharp and detailed vision, receive information from only a single photoreceptor cell, implying that information from these photoreceptors is transmitted directly to the brain without any pooling (Purves, 2012). The receptive fields of ganglion cells increase with distance from the fovea and ganglion cells connecting to peripheral parts of the retina integrate the signals of many photoreceptors at once (Croner and Kaplan, 1995).

The way our encoding algorithm pools information from multiple DVS pixels into single spike trains adapts this property of detailed information transmission from central parts of the retina and averaging over larger numbers of photoreceptors in the periphery. Overall, the algorithm generates 128 spike trains that will serve as input for the SNN. Each spike train represents one retinal ganglion cell with its own receptive field on the 128 × 128-pixel output of the DVS (**Figure 1**).

In our algorithm, the central 8 × 8 pixels of the DVS output represent the fovea (**Figure 1A**), and for each of these central 64 pixels, there is a single ganglion cell only considering the output of that single pixel. Furthermore, there are four groups of 16 ganglion cells each, with receptive fields that increase from the center to the periphery. The first group consists of the central 16 × 16 pixels, divided into 16 squares that integrate an area of four by four pixels each (**Figure 1B**). The next group consists of the central 32 × 32 pixels, again divided into 16 squares, this time with an area of 8 × 8 pixels each (**Figure 1C**). The same happens for the central 64 × 64 pixels (**Figure 1D**) and the total of 128 × 128 pixels (**Figure 1E**), resulting in 16 squares per group, of size 16 × 16 and 32 × 32, respectively. In this pooling mechanism, an average of 170.5 pixels converge into one ganglion cell. The size of the receptive fields can easily be adapted to higher or non-square video resolutions.

Having set the distribution of the ganglion cells across the DVS output, the next step is to determine how the information of the DVS pixels is encoded into spike trains for the ganglion cells.

Firing Mechanism

The Dynamic Vision Sensor provides a very high temporal resolution of up to 1 μs. Preserving is detailed temporal information is desirable from a computational point of view, but as described below we reduce this resolution to 10 ms to maintain biological plausibility. While some spike encoding algorithms like Poisson models focus merely on the spike count within a given time interval and disregard the exact spike timing, it has been shown that the spike timing of mammalian retinal ganglion cells conveys several times more information than the spike count (Berry et al., 1997; van Rullen and Thorpe, 2001; Uzzell and Chichilnisky, 2004). Furthermore, retinal ganglion cells fire very briefly as a response to specific stimuli rather than emitting a high frequency of background firing. Spikes emitted by retinal ganglion cells of rabbits and salamanders, presented with random flicker, covered less than 5% of the total stimulus time (Berry et al., 1997). The maximum firing rate of retinal ganglion cells varies between different animal species and depends on the type of visual stimuli. Transient peak rates of up to 250 Hz have been observed in retinal ganglion cells of mice (Krieger et al., 2017), but for the sustained firing of human retinal ganglion cells, an upper bound of 100 Hz can be reasonably assumed (Nelson, 1995).

As described in section The Dynamic Vision Sensor, the DVS output consists of a series of events, including their timing in microseconds and their location in pixel coordinates. In fact, each event also includes a polarity of +1 or −1, depending on whether the event indicates a pixel becoming brighter or darker. Our encoding algorithm ignores the event polarity, but it might be worthwhile for future experiments to consider a translation of positive and negative events into positive and negative spikes.

Our spike encoding algorithm is illustrated in **Figure 2**. In the first step, the algorithm takes the time series of the DVS and groups it into windows of 10,000 μs or 10 ms. The new time series consists of 10 ms steps, and for every ganglion cell, it must be decided at which of these steps the cell will fire. Since each time step represents 10 ms of video data, the maximum firing rate of the ganglion cells cannot exceed 100 Hz. The encoding for the central 64 pixels that represent the fovea is straightforward: if there is at least one event for a pixel at time step t_i, the ganglion cell that corresponds to that pixel will fire at t_i. There are no parameters to tune for these central 64 pixels and the spike trains of the ganglion cells that correspond to these pixels are completely determined by the DVS output. For the 64 ganglion cells that integrate the events of multiple DVS pixels, the situation is slightly different. For each of these cells, the algorithm counts how many events occurred in each time window within the receptive field of that ganglion cell. If the number of events from pixels within the receptive field of cell C_j at time step t_i exceeds a certain threshold, C_j will fire at t_i.

Theoretically, this threshold can be set for each ganglion cell individually, but since the 16 cells of each group have receptive fields of the same size, our algorithm assigns the same threshold to all 16 cells of a group, resulting in a total of 4 thresholds that can be tuned. Clearly, the value of the thresholds will determine the average spike rate of the final spike trains, with higher thresholds leading to fewer spikes, and it is possible to imitate biological evidence about spike rates under certain stimuli. We discuss the tuning of the thresholds in more detail in section Model Design and Implementation.

Inspired by the structure and organization of retinal ganglion cells, our algorithm pools 128 × 128 DVS pixels into 128 ganglion cells that will serve as input neurons for the SNN. The algorithm compresses the microsecond resolution of the DVS output into time steps of 10 ms, but it preserves the timing of the DVS events instead of generating a Poisson process with random spike timing. The next section describes the structure of a brain-like SNN architecture called NeuCube, and our imitation of the retinotopic mapping of retinal ganglion cells into the visual cortex.

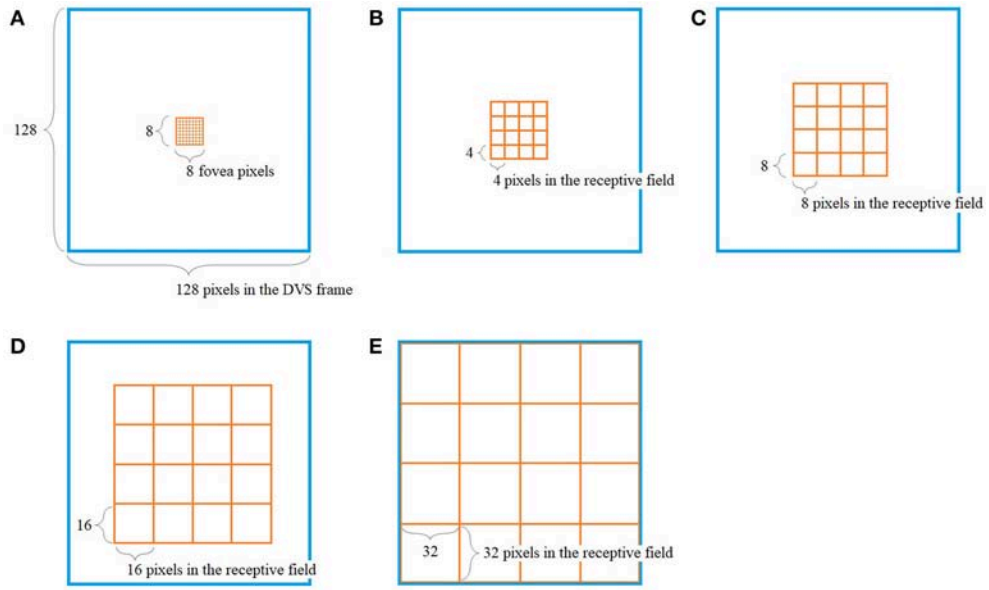


FIGURE 1 | Pooling of 128 × 128 DVS pixels into 128 ganglion cells. 64 foveal ganglion cells that correspond to the central 64 DVS pixels **(A)** and four groups of 16 ganglion cells each with increasing size of receptive fields toward the periphery **(B–E)**. The image seen by the DVS camera is marked with a blue frame and the receptive fields are marked with orange frames.

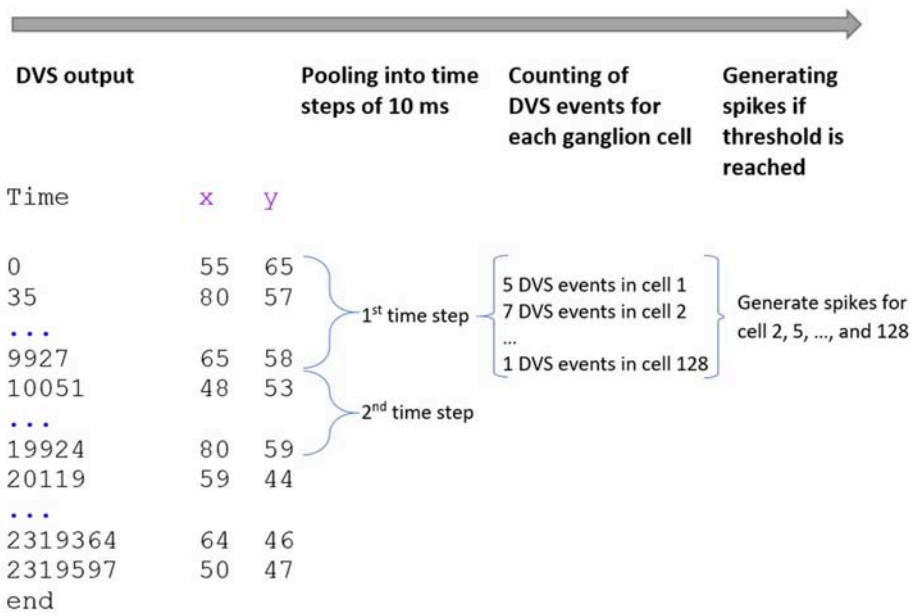


FIGURE 2 | Encoding of spike trains from DVS output. The DVS time series is grouped into windows of 10 ms. For each time step, the DVS events within the receptive fields of all 128 ganglion cells are counted. If the number of DVS events within the receptive field of one ganglion cell exceeds a certain threshold, the cell fires at that time step.

The Brain-Like SNN Neucube and the Proposed Retinotopic Mapping

The NeuCube SNN architecture incorporates several different principles of SNN and combines them into a single model for mapping, learning, and understanding of spatio-temporal data (Kasabov, 2014). Signals are processed along successive stages as shown in **Figure 3**. Before going into detail about the learning algorithms used by NeuCube, we want to focus on the three-dimensional structure of NeuCube and the bio-inspired way we mapped the 128 input neurons into this structure. Our system

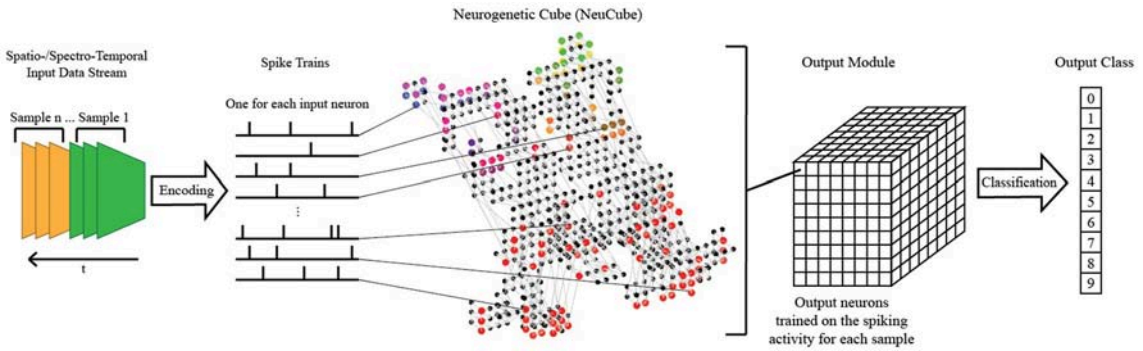


FIGURE 3 | A schematic diagram of a general NeuCube architecture, consisting of: input encoding module; NeuCube module; output function module. Our system only makes use of the NeuCube module and the output function module.

uses a NeuCube initialized with 732 neurons, using the MNI coordinates of neurons from the primary visual cortex (V1, Brodman area 17), taken from the Atlas of the Human Brain (downloaded together with the xjView toolbox: http://www.alivelearn.net/xjview). The number of neurons is only bounded by computational limitations; it is possible to add further neurons from the secondary or tertiary visual cortex or to represent the whole brain. Initial connections between the neurons are based on the "small-world" paradigm, where random connections are formed within a pre-defined maximum distance of each neuron, 80% of the time as excitatory and 20% of the time as inhibitory connections. The mapping of the 128 input neurons into the 732 neurons of NeuCube mimics two important characteristics of the human visual cortex: cortical magnification and retinotopic mapping (**Figure 4**).

Cortical magnification describes the overrepresentation of foveal signals inside the primary visual cortex. Although the fovea has a diameter of only 1.2 mm (Purves, 2012), its signals are processed by almost 50% of all neurons in V1 (Krantz, 2012; Born et al., 2015). Therefore, we chose exactly 64 of our 128 input neurons to correspond to the central 64 DVS pixels with a one-to-one relationship. This way, 50% of input neurons automatically correspond to the central pixels of the DVS, just like 50% of the primary visual cortex correspond to the central photoreceptors on the retina.

The second characteristic of the primary visual cortex that we adopted in our mapping is the preservation of spatial relationships between photoreceptors on the retina and their neural representation in the primary visual cortex, the so-called retinotopy (Rosa, 2002). Signals from the top left of our visual field are mapped to the bottom right of V1 and vice versa. What humans see is flipped upside down and mirrored, but objects that appear next to each other in the visual field will still be represented next to each other in V1. Both the foveal as well as the peripheral ganglion cells follow this principle, although foveal signals are mapped into the posterior part and peripheral signals into the anterior part of V1 (Purves, 2012). **Figure 5** shows how the principle of retinotopy is applied to the mapping of the 128 input neurons to the 732 neurons of NeuCube.

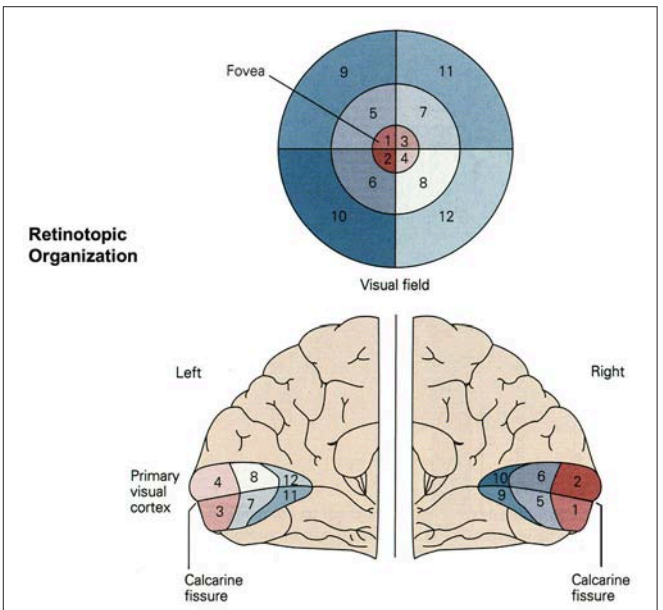


FIGURE 4 | Retinotopic organization of the primary visual cortex. Up to 50% of the primary visual cortex processes foveal signals (cortical magnification). Signals from the top left of the visual field are mapped to the bottom right of the visual cortex (retinotopy). Source: Jaygandhi786 (2015).

Unsupervised and Supervised Learning of Dynamic Visual Patterns in the Neucube Architecture

Learning in the NeuCube is performed in two stages: in the first step, unsupervised learning is performed to modify the initial connection weights. In our system we use pair-based multiplicative spike-timing-dependent plasticity (STDP, van Rossum et al., 2000), but in principle, the NeuCube architecture allows for a flexible implementation of different learning algorithms. The SNN will learn to activate the same groups of spiking neurons when similar input stimuli are presented and to change existing connections that preserve the spatio-temporal patterns of the input data (Kasabov and

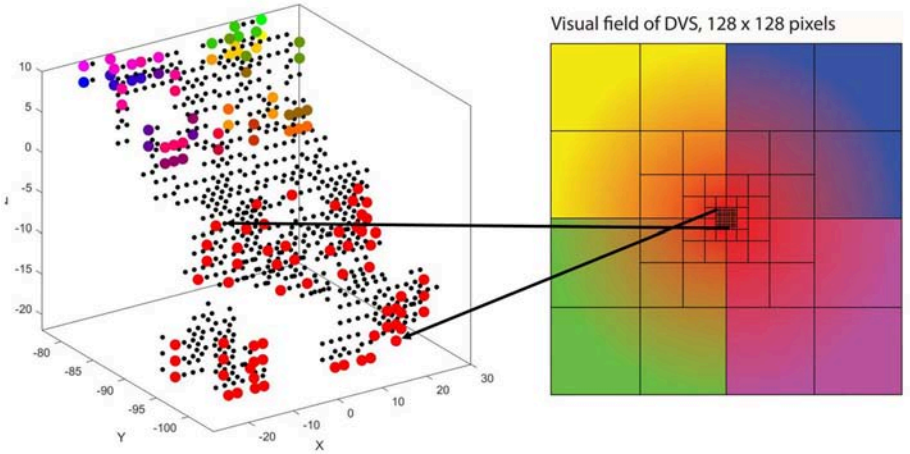


FIGURE 5 | Retinotopic mapping of 128 input neurons to NeuCube, initialized with 732 neurons, using the MNI coordinates of neurons from the primary visual cortex, taken from the Atlas of the Human Brain.

Capecci, 2015). Previous works have shown that STDP is well suited to train neurons to respond to discriminative visual features (Masquelier and Thorpe, 2007). The neurons become selective to successive coincidences of particular patterns and learn to detect them robustly even in the presence of noise (Masquelier et al., 2009). Our approach using the NeuCube differs from these works mainly in the structure of the network, which is not based on layers, but rather a three-dimensional network shaped like the primary visual cortex. However, our results are similar to those works in that certain neurons and connections can be identified that seem to play a major role in discriminating between the different classes. NeuCube allows for a visualization of the learning process and we discuss how the visualization can be used for a better understanding of the data and the neural processes after presenting our experimental results.

In the second step, supervised learning is applied to the spiking neurons in the output classification module, where the same spike trains used for the unsupervised training are now propagated again through the trained SNN and output neurons are generated and trained to classify the spiking activity of the SNN into pre-defined classes (Kasabov and Capecci, 2015). Again, the NeuCube architecture allows for the application of different algorithms for the evolving classifier. The output function we used is called the dynamic evolving SNN algorithm (deSNN, Kasabov et al., 2013), which makes use of rank-order learning (Thorpe and Gautrais, 1999). This kind of evolving classifier is computationally inexpensive and puts emphasis on the order in which input spikes arrive, making it suitable for on-line learning and early prediction of temporal events (Kasabov, 2014). Similar to previous works on image recognition based on reward-modulated STDP (Mozafari et al., 2017), the deSNN algorithm uses a "highest" layer of neurons to discriminate between classes. While Mozafari et al. (2017) used an existing layer of output neurons, the deSNN algorithm creates and trains one new output neuron per sample by connecting it to all 732 neurons in the network and propagating the signal through the network once more. The connection weights that are learned in this process are then classified using a K-nearest neighbor (KNN) algorithm and the labels that are known for all the samples. Here our method differs from the aforementioned (Mozafari et al., 2017) in that we do not apply "anti-STDP" for misclassified samples before applying KNN. This means that the results of the deSNN's decisions are not fed back into the network since we create a new output neuron for each sample.

For a more detailed description of the NeuCube architecture see Kasabov (2014).

Summary of the Proposed Methodology

The methodology we propose for dynamic visual recognition consists of the following steps:

(1) Event-based video recording with DVS.
(2) Pooling and encoding of DVS output into spike trains for the input neurons of the SNN.
(3) Training NeuCube on the spike data using unsupervised learning, e.g., STDP.
(4) Training of an output classifier in a supervised mode.
(5) Validating the classification results.
(6) Repeating steps (2–5) for different parameter values to optimize the classification performance. Recording the model with the best performance.
(7) Visualizing the trained SNN and analyzing its connectivity and spiking activity for a better understanding of the data and the involved brain processes.

We present the application of this method on a benchmarking experiment with the MNIST-DVS dataset for spike-based dynamic visual recognition and go into further detail about the tuning of parameters and analysis of the SNN.

BENCHMARKING ON THE MNIST-DVS DATASET

Description of the MNIST-DVS Dataset

The MNIST dataset of handwritten digits (Lecun et al., 1998) has been one of the most popular benchmarking datasets for image recognition for over 20 years. With the advent of spiking neural networks, MNIST has naturally been used as a benchmark for spike-based visual recognition systems (Brader et al., 2007; Querlioz et al., 2013; Diehl and Cook, 2015; Zhao et al., 2015; Kheradpisheh et al., 2017). However, these works only account for the recognition of the static MNIST pictures and do not aim toward dynamic visual recognition of moving objects. An important part of the functioning of spiking neural networks is the dimension of time within the spike trains and on datasets that also have such a temporal dimension, spiking neural networks might be superior to classical artificial neural networks.

The NE15-MNIST database (Neuromorphic Engineering 2015 on MNIST, Serrano-Gotarredona and Linares-Barranco, 2015; Liu et al., 2016) that we used for our study is based on the original MNIST dataset. NE15-MNIST consists of four subsets that all aim to provide a benchmark for spike-based visual recognition. While the *Poissonian* and the *FoCal* subsets are synthetically generated from static MNIST images, the other two subsets are based on 128 × 128 pixel DVS recordings of the MNIST images. The MNIST-FLASH-DVS subset contains DVS recordings of MNIST digits that are flashed on a screen. Because we were interested in dynamic visual recognition of moving objects, we decided to work on the MNIST-DVS subset that consists of DVS recordings of MNIST digits that move back and forth across a screen and thereby produce temporal contrast and DVS events on the digits' edges.

The MNIST-DVS dataset is available online (Yousefzadeh et al., 2015). It consists of 30,000 recordings of 10,000 original MNIST digits recorded at three different scales each (scale-4, scale-8, and scale-16). Each recording has a time length of about 2.5 s, during which the digit moves twice from a position at the bottom left of the middle of the screen to the top right and back. The files are provided in the jAER format (Delbruck, 2008) and the dataset includes Matlab scripts for a conversion to Matlab arrays and three kinds of data preprocessing: removal of a 75 Hz timestamp harmonic produced by the LCD screen, stabilization of the digits on the center of the screen and removal of the event polarity information.

Previous classification results on the MNIST-DVS dataset are shown in **Table 1**. Henderson et al. (2015) derive a new event-based learning scheme and apply it to a layered feedforward spiking neural network, which is trained self-supervised for classification of the MNIST-DVS digits. Zhao et al. (2015) use a composite system, consisting of a convolutional spiking neural network for feature extraction and a network of tempotron neurons for spike-based classification. While these two systems are fully event-driven, Stromatias et al. (2017) use a combination of a spiking neural network and a conventional artificial neural network. A convolutional SNN is used to capture the temporal dynamics of the DVS data and create a new, frame-based dataset, which is fed into a fully-connected artificial neural network. The supervised learning itself then takes place in this non-spiking network, using a stochastic gradient descent algorithm. In our concluding remarks we suggest how this approach could be combined with our model to maintain the high classification accuracies while providing greater biological plausibility.

Model Design and Implementation

The only preprocessing we applied to the data was the removal of the 75 Hz timestamp harmonic. Stabilizing the video data would have been contrary to our intention to develop a system for dynamic visual recognition, and in fact, preliminary experiments suggested that the system would perform better on the original unstabilized videos. To run our spike encoding algorithm on the data, we used the script provided with the dataset to convert the jAER files into Matlab arrays.

The pooling of the DVS spikes into 128 input spike trains (ganglion cells) for the SNN, as described within section The Proposed System Architecture, remained the same throughout all experiments. Inside the spike encoding algorithm, only those four thresholds were changed that determine how many pixels within the receptive field of a ganglion cell must fire within one time step to make the ganglion cell itself emit a spike. As a first step, we wanted to find out how the system would perform differently when these thresholds and, thus, the average spike rate of the input data for the SNN, were changed. As described in section Firing Mechanism, the ganglion cells' receptive fields decrease from the periphery toward the center. Starting from the periphery, ganglion cells in group 1 integrate the signal of $32 \times 32 = 1.024$ DVS pixels, cells in group 2 from $16 \times 16 = 256$ pixels, cells in group 3 from $8 \times 8 = 64$ pixels, and cells in group 4 from $4 \times 4 = 16$ pixels. Assigning the same percentage threshold to all four groups would result in very low or no activity in the peripheral ganglion cells, e.g., with a threshold of 10% it would take only two DVS events within the receptive field of a ganglion cell in group 4 to trigger a spike, but 103 DVS events within the receptive field of a ganglion cell in group 1. Especially with the MNIST-DVS dataset, where DVS events only occur at the edges of the moving digits and not in larger blobs, this would make the peripheral ganglion cells redundant. On the other hand, increasing the thresholds too much from group to group toward the center would put more emphasis on the peripheral parts of the video than intended.

We carefully watched the MNIST-DVS videos and compared the distribution of DVS events with the average spike rates for the groups of ganglion cells that were produced by different spiking thresholds. We found that increasing the percentage thresholds by a factor of two from group to group toward the center would preserve the distribution of DVS events relatively well and not put too much emphasis on any single group. **Figure 6** shows the average spike rates for 1,000 scale-8 videos (100 per digit), produced by thresholds of 0.5% for group 1, 1% for group 2, 2% for group 3 and 4% for group 4. Since time is discrete in our model, we measure the average spike rates in %, dividing the number of time steps in which a cell fired by the total number of time steps. Most spikes occur in groups 2 and 3, consistent with the general distribution of DVS events in the scale-8 videos. The total spike average of the samples shown in **Figure 6** is 27.57%.

TABLE 1 | Previous classification results on the MNIST-DVS dataset.

	Network type	Learning algorithm	Total number of samples used	Train-test-ratio	Classification on test set (%)
Henderson et al., 2015	Feedforward SNN	A new scheme for spike-based learning	10.000 (scale not mentioned)	90–10	87.41
Zhao et al., 2015	Composite system, including convolution, motion detector, feature spike conversion, and SNN classifier	Tempotron learning	10.000 (scale-4)	90–10	88.14
Stromatias et al., 2017	Composite system, including convolutional SNN, non-spiking fully connected classifier, and spiking output layer	Stochastic Gradient Descent (inside the non-spiking classifier)	10.000 (scale-16)	80–20	97.95

We altered the thresholds to get clearly distinguishable total spike averages. **Table 2** shows four different choices of thresholds, resulting in average spike rates of roughly 7, 14, 26, and 32% (exact numbers vary between different video scales). The last row represents the maximal achievable average spike rate with a threshold of 0% for each group. In that case, every ganglion cell fires if there is at least one DVS event in its receptive field at a given time step.

The mapping of the input spikes into the SNN NeuCube was done according to the proposed retinotopic mapping and it remained the same throughout all experiments. In all experiments NeuCube was initialized with 732 leaky integrate and fire neurons (LIF), representing the primary visual cortex. For future experiments with higher video resolutions and more input neurons, NeuCube can easily be extended to include neurons that represent the secondary and the tertiary visual cortex. Initial connections are formed following "small-world" connectivity with random connections within a predefined maximum distance from each neuron. This maximum distance was set to 2.5 in all experiments.

As described previously, unsupervised learning using STDP is performed first to learn spatio-temporal patterns by forming new connections between neurons, before the output classifier is trained in a supervised manner using the dynamic evolving SNN (deSNN) algorithm (Kasabov et al., 2013). The NeuCube architecture is a stochastic model and, therefore, sensitive to parameter settings. To find the best values for the major parameters that influence the system's performance, we applied a grid search method that tests the system on different combinations of parameters within a predefined range and used those parameter values that resulted in the best classification accuracy. For the firing threshold, the refractory time and the potential leak rate of the LIF neurons we used values of 0.5, 6, and 0.002, respectively. The STDP learning parameter was set to 0.01. The variables *Mod* and *Drift* of the deSNN classifier were set to 0.8 and 0.005. See Kasabov and Capecci (2015) for a more detailed explanation of these parameters.

Experimental Results

To compare the system's performance, we performed 10-fold cross-validation on 1,000 videos (first 100 of each digit), with 900 videos used for training and 100 for testing in each fold, for different video scales and average spike rates. **Table 3** summarizes the results. As a general trend, with few exceptions, the classification accuracy increased together with the average spike rate of the input neurons. For all video scales, the classification accuracy also increased when the system was run on all 10,000 videos of a given scale. The best classification results were achieved with all 10,000 videos of one scale, encoded with the highest possible spike rate (0% as spike encoding threshold for all four groups). Classification accuracies were 90.56, 92.03, and 86.09% % for scale-4, scale-8, and scale-16, respectively. The best accuracy in a single run with 90% of randomly selected data samples for training and the remaining 10% for testing was 92.90% for 10,000 scale-8 videos with the highest possible spike rate. This result is comparable to previous results on the MNIST-DVS dataset, presented in **Table 1**.

The lower accuracies on the scale-4 and the scale-16 samples reflect the fact that in these videos, the MNIST digits fill out either the whole screen (scale-16) or only a very little region in the center (scale-4). For the scale-4 digits, the signals transmitted by ganglion cells from groups 1, 2, and 3 are mostly noise and do not contain much information about the digits. In the scale-16 videos, there is almost no activity in the central region of the screen and, thus, no information is transmitted by the 64 foveal ganglion cells. Since our method puts heavy emphasis on the center of the video (50% of the input neurons represent data from only the central 64 pixels), performance on the scale-16 videos is lower.

Model Interpretation for a Better Understanding of the Processes Inside the Visual Cortex

The main purpose of the above experiments, carried out on the MNIST-DVS dataset, is to confirm the system's classification performance on a benchmark dataset, and the moving digits do not represent a real-life scene. However, we want to show how the SNN can be analyzed after being trained, to see how its connectivity changes in response to the data. **Figure 7** compares the connectivity of the SNN before and after unsupervised training on 1,000 scale-4 videos with the highest possible spike rate. Blue and red lines represent positive and negative connections, respectively. We can notice that some of the randomly created initial connections disappear during the

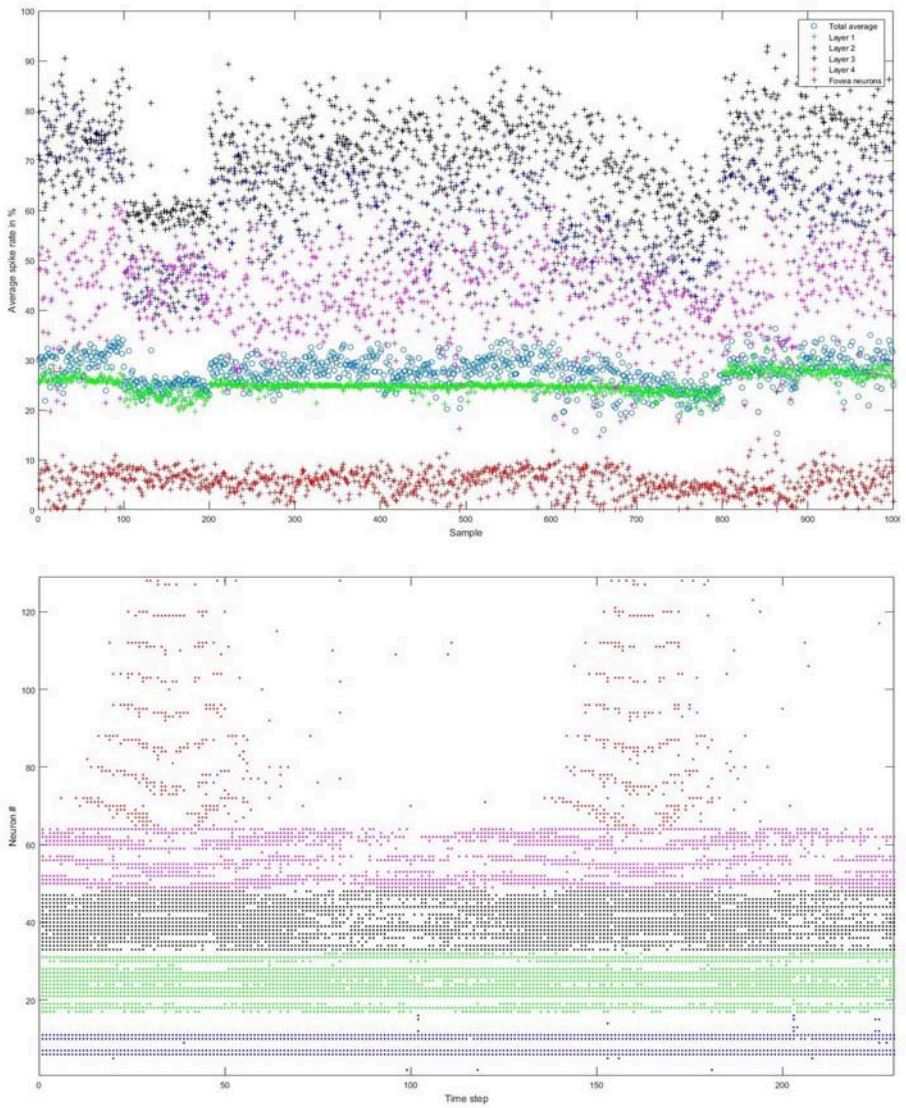


FIGURE 6 | Top: Average spike rates of 1,000 scale-8 videos (100 per digit) resulting from encoding-thresholds of 0.5, 1, 2, and 4% for the four groups of retinal ganglion cells, respectively (from periphery to center). Average spike rates are measured in %, dividing the average number of time steps in which the cells of a given group fired by the total number of time steps. The total average spike rate is 27.57%. **Bottom**: Example of encoded spike trains for one sample (digit 0, scale-8, sample #1). Neurons 1–16, 17–32, 33–48, and 49–64 represent the four groups of ganglion cells from the periphery to the center; neurons 65–128 represent the foveal ganglion cells. The spike pattern of the foveal ganglion cells clearly represents the two times that the digit moves across the center of the screen.

training process. Instead, many new negative connections are created, mostly between neurons in the region that represents the posterior part of the primary visual cortex, where signals from the foveal ganglion cells arrive. Some of the new connections connect neurons over a long distance, especially in the very posterior part of the SNN, where a gap between neurons prevents the initial formation of "small-world" connections. As can be seen in **Figure 5**, the neurons on both sides of this gap represent adjacent DVS pixels, and by bridging this gap, the new connections allow for communication between these neurons. A comparison with the connectivity after training the SNN on 1,000 scale-16 videos shows that slightly fewer connections are formed between neurons processing foveal information since the scale-16 videos contain less DVS events in the foveal region. This effect is due to the acquisition hardware used and could be compensated for by the simulation of saccadic eye movements inside the encoding algorithm. In a biological retina, these rapid eye movements ensure that the fovea centralis focuses on salient features instead of constantly covering a less important area of the visual field. We discuss this possible improvement of the encoding algorithm in the next section.

There is also a visible difference between connections created for different digits. **Figure 8** shows the status of the network after

TABLE 2 | Different choices of spike thresholds within the spike encoding algorithm and corresponding average spike rates.

Spike threshold in %				Approximate average spike rate (%)
Group 1	Group 2	Group 3	Group 4	
5	10	20	40	5
2.5	5	10	20	13
0.5	1	2	4	26
0	0	0	0	32

TABLE 3 | Results of 10-fold cross validation for different video scales and average spike rates.

Video scale	Number of samples	Average spike rate (%)	Classification accuracy (%)
Scale-4	1,000	7.85	63.80
"	1,000	13.94	77.10
"	1,000	25.77	75.50
"	1,000	31.77	83.40
"	10,000	31.98	90.56
Scale-8	1,000	5.29	66.40
"	1,000	13.49	83.00
"	1,000	27.57	84.20
"	1,000	32.96	86.20
"	10,000	32.93	92.03
Scale-16	1,000	3.81	60.50
"	1,000	12.64	82.90
"	1,000	26.94	78.60
"	1,000	31.72	77.50
"	10,000	31.79	86.09

unsupervised training using only digits 1, 5, and 8, respectively. Interestingly, the connections created for digits 5 and 8 look similar, just like the digits themselves have a similar shape. The connections created after training on digit 1, on the other hand, look distinctly different. We can, therefore, conclude that the visual characteristics of the digits are preserved in our system, just like they are in the human visual cortex.

DISCUSSION OF THE SYSTEM'S ADVANTAGES AND LIMITATIONS

The proposed system achieves a classification performance on the benchmark MNIST-DVS dataset that can keep up with previous works on this dataset and is superior to those works that used a spiking neural network classifier. Every part of the system, the DVS sensor, the algorithm for encoding the DVS output into spike trains, and the SNN NeuCube adopt features from the human visual system. This allows for future experiments where the same stimuli are presented to humans and the proposed system and brain processes visualized by neuroimaging methods can be compared to the network processes of the SNN, which can be easily visualized within the NeuCube architecture.

Another advantage of the proposed system is the high flexibility of the SNN's three-dimensional structure. The NeuCube architecture is not restricted to consist of neurons that represent only the visual cortex. For example, one could map aural stimuli to input neurons representing the auditory cortex, to obtain a model that processes aural and visual information at the same time in a brain-like way. The integration of other kinds of data, such as tactile or olfactory information, within a multimodal model is conceivable as well.

We found that the system's classification performance increases together with the average spike rates of the 128 input neurons. To account for the findings of Berry et al. (1997) in retinal ganglion cells of rabbits and salamanders, we started our experiments with low spike rates of approximately 5%, but the classification accuracies were very low in these cases. However, the reported firing rates of rabbit and salamander ganglion cells were measured during the presentation of random flicker, which might yield very different firing behavior than stimuli like the moving digits. Single cell recordings of retinal ganglion cells could provide more evidence about the firing rates under specific stimuli. The parameters of the spike encoding algorithm that determine the average spike rates can then easily be tuned to mimic the behavior of real retinal ganglion cells and it would be interesting to see if classification accuracy increases when the average spike rates conform to the biological evidence.

Since so much is known about the human visual system and we aimed to develop a biologically plausible, yet computationally feasible implementation, there are many details not included in our model. There already exist very advanced mathematical models for the function of retinal ganglion cells (Wei and Ren, 2013) and our spike encoding algorithm has by far not touched every detail of them. The receptive field of each ganglion cell, for example, is split into a center region and a surrounding region with opposite behavior toward light (Nelson, 1995). In so-called on-center cells, the center region is stimulated, whereas the surrounding region is inhibited when exposed to light. So-called off-center cells exhibit converse behavior. Including the function of on- and off-center ganglion cells inside the spike encoding algorithm would highly increase the model's biological plausibility, but also its computational complexity. Another computational restriction of our model is that the random initial creation of excitatory and inhibitory connections causes a violation of Dale's Principle, which states that all axonal branches of a neuron perform the same chemical reaction.

One shortcoming of the DVS when compared to the human retina is its inability to process colors. The DVS only encodes temporal changes in brightness that signal motion (Delbruck, 2008), similar to the rod photoreceptors on the retina and the functionality of the magnocellular fibers in the optical nerve (Purves, 2012). However, the cone photoreceptors on the retina as well as the comparatively large amount of parvocellular fibers in the optic nerve are not modeled by the DVS despite their importance for detecting and transmitting information about color and details of the perceived objects (Purves, 2012). This means that all object recognition approaches using DVS input are

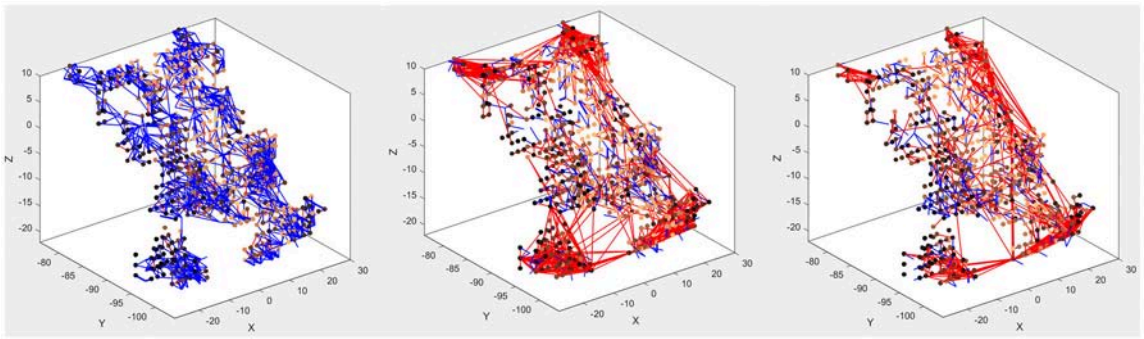


FIGURE 7 | Connectivity of the SNN before **(left)** and after training on 1,000 scale-4 samples **(middle)** and 1,000 scale-16 samples **(right)**. During training, new connections are created while others vanish, representing relations between spiking neurons that evolve as a response to the spatio-temporal patterns of the data.

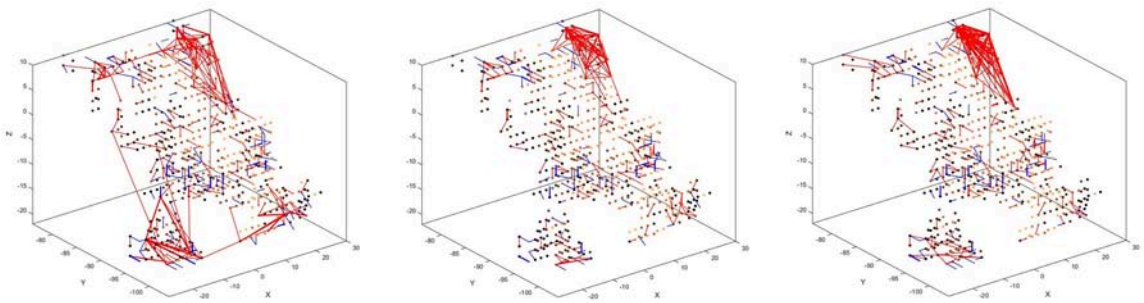


FIGURE 8 | Connectivity of the SNN after unsupervised training on 1,000 scale-8 samples each for the three digits 1 **(left)**, 5 **(center)**, and 8 **(right)**. There is a visible difference between the connections, corresponding to the visual characteristics of the digits.

somewhat limited because the DVS only captures signals that the human visual system would use to detect motion and distances to objects, but not those signals necessary for recognizing objects and details.

The proposed system puts strong emphasis on the central part of the videos in both the encoding of DVS events to spike trains and the representation inside the SNN. This is justified by analogous features of the fovea centralis in the center of the human retina, responsible for focused vision. However, there is no evidence that there exist retinal ganglion cells with large receptive fields in the human retina that cover the fovea centralis in a redundant manner as in our system. Further, our system does not account for the very fast and simultaneous movement of human eyes, called saccades. Saccades help to scan a broader part of the visual field with the fovea and integrate this information into a detailed map (Purves, 2012). Human eye movement is also controlled by the visual grasp reflex that directs the eyes toward salient events in the periphery of the visual field (Monsell and Driver, 2000). These mechanisms for eye movement could be implemented in the spike encoding algorithm by changing the coordinates for the pooling of DVS pixels for each time step, and thereby virtually moving the center of the visual field. However, this would require additional features to save the movement and integrate it into the SNN.

CONCLUSION

This paper presents a new methodology for dynamic visual recognition, inspired by different features of the human visual system. The proposed system is designed to take data from a DVS silicon retina and encodes them into spike trains using an algorithm that mimics the organization and function of retinal ganglion cells. The spike trains are then fed into the brain-like SNN NeuCube, following the retinotopic mapping of photoreceptors from the retina into their neural representations in the primary visual cortex. Two stages of learning, unsupervised and supervised, are performed by NeuCube to extract spatio-temporal patterns from the data and perform a classification task. Results on the benchmark MNIST-DVS dataset have shown that the system can keep up with the classification performance of other methods for dynamic visual recognition. Furthermore, it is possible to dynamically visualize and analyze the activity inside the SNN for a better understanding of the data and the process of their deep learning in the model.

Due to the promising benchmark results and the benefit of the visualization tools for an in-depth understanding of the data and the network processes, we endorse further research on the system. In particular, we suggest the exploration of new learning methods inside NeuCube and of different algorithms for the encoding of DVS data into spike trains.

To date, the highest classification accuracy on the MNIST-DVS dataset has been achieved by Stromatias et al. (2017), who used a spiking convolutional neural network to create a new frame-based dataset, which captures the dynamics of the DVS output and serves as input for a fully-connected classifier that uses stochastic gradient descent. The non-spiking classifier is then mapped to a spiking output layer of LIF neurons. As they mention in their paper, the non-spiking classifier and the spiking output layer can be used with any spiking neural network that has already extracted features from the data in an unsupervised manner. We propose to explore how the connectivity or spiking activity of the NeuCube after the unsupervised learning stage could be used to create a similar frame-based dataset, and how the classifier used by Stromatias et al. (2017) would perform on such a dataset. This way, the biological plausibility of our model could be combined with current state-of-the-art classification algorithms.

We also encourage the development of further benchmark datasets for spike-based visual recognition, e.g., spiking versions of the KTH and the Weizmann datasets of human actions (Laptev and Caputo, 2005; Gorelick et al., 2007). Since the NeuCube architecture is not bound to only consist of neurons representing the visual cortex, future directions can include the integration of our system for visual recognition inside a broader, multimodal methodology, e.g., for the biologically plausible processing of visual and aural data at the same time within the same system. The used DVS format for visual data encoding into spike trains is not a restriction for the proposed SNN method for retinotopic mapping. Learning and other encoding methods for different types of visual data are envisaged to be explored in the future.

AUTHOR CONTRIBUTIONS

LP the main author, contributes to the spike encoding algorithm, the retinotopic mapping into NeuCube, the choice of MNIST-DVS as a benchmarking dataset, performance evaluation, and paper writing. AW contributes to the initial design of the NeuCube model and partial implementation, and takes part in discussions and reviewing the paper. NK originated the initial idea of this project, and takes part in discussions and reviewing the paper.

ACKNOWLEDGMENTS

The authors thank the reviewers for the useful comments and suggestions. NK acknowledges his discussions with Giacomo Indiveri, Tobi Delbrück and other colleagues from INI, ETH/UZH during his Marie Curie visit in 2011/2012 and the contacts afterwards. AW is funded by a scholarship from Auckland University of Technology, and LP by a scholarship from the Baden-Württemberg Foundation for his visit to the Knowledge Engineering and Discovery Research Institute (KEDRI) at Auckland University of Technology. The authors would further like to thank Dr. Josafath Israel Espinosa Ramos for his valuable support.

REFERENCES

Berry, M. J., Warland, D. K., and Meister, M. (1997). The structure and precision of retinal spike trains. *Proc. Natl. Acad. Sci. U.S.A.* 94, 5411–5416.

Bichler, O., Querlioz, D., Thorpe, S. J., Bourgoin, J. P., and Gamrat, C. (2012). Extraction of temporally correlated features from dynamic vision sensors with spike-timing-dependent plasticity. *Neural Netw.* 32, 339–348. doi: 10.1016/j.neunet.2012.02.022

Born, R. T., Trott, A. R., and Hartmann, T. S. (2015). Cortical magnification plus cortical plasticity equals vision? *Vision Res.* 111, 161–169. doi: 10.1016/j.visres.2014.10.002

Brader, J. M., Senn, W., and Fusi, S. (2007). Learning real-world stimuli in a neural network with spike-driven synaptic dynamics. *Neural Comput.* 19, 2881–2912. doi: 10.1162/neco.2007.19.11.2881

Croner, L. J., and Kaplan, E. (1995). Receptive fields of P and M ganglion cells across the primate retina. *Vision Res.* 35, 7–24.

Curcio, C. A., Sloan, K. R., Kalina, R. E., and Hendrickson, A. E. (1990). Human photoreceptor topography. *J. Comp. Neurol.* 292, 497–523. doi: 10.1002/cne.902920402

Delbruck, T. (2008). "Frame-free dynamic digital vision," in *Proceedings of International Symposium on Secure-Life Electronics* (Tokyo: Advanced Electronics for Quality Life and Society, University of Tokyo), 21–26.

Diehl, P. U., and Cook, M. (2015). Unsupervised learning of digit recognition using spike-timing-dependent plasticity. *Front. Comput. Neurosci.* 9:69. doi: 10.3389/fncom.2015.00099

Gorelick, L., Blank, M., and Shechtman, E. (2007). *Actions as Space-Time Shapes*. Available online at: http://www.wisdom.weizmann.ac.il/~vision/SpaceTimeActions.html (Accessed Aug 29, 2017).

Henderson, J. A., Gibson, T. A., and Wiles, J. (2015). *Spike Event Based Learning in Neural Networks*. Available online at: http://arxiv.org/pdf/1502.05777.

Jaygandhi786 (2015). *Visual Cortex*. Available online at: https://commons.wikimedia.org/wiki/User:OgreBot/Uploads_by_new_users/2015_May_02_15:00 (Accessed Aug 29, 2017).

Jimenez-Fernandez A., Lujan-Martinez C., Paz-Vicente R., Linares-Barranco A., Jimenez G., and Civit A. (2009) "From vision sensor to actuators, spike based robot control through address-event-representation," in *Bio-Inspired Systems: Computational and Ambient Intelligence*. IWANN 2009. Lecture Notes in Computer Science, Vol. 5517, eds J. Cabestany, F. Sandoval, A. Prieto, and J.M. Corchado (Berlin; Heidelberg: Springer).

Kasabov, N., and Capecci, E. (2015). Spiking neural network methodology for modelling, classification and understanding of EEG spatio-temporal data measuring cognitive processes. *Inf. Sci.* 294, 565–575. doi: 10.1016/j.ins.2014.06.028

Kasabov, N., Dhoble, K., Nuntalid, N., and Indiveri, G. (2013). Dynamic evolving spiking neural networks for on-line spatio- and spectro-temporal pattern recognition. *Neural Netw.* 41, 188–201. doi: 10.1016/j.neunet.2012.11.014

Kasabov, N. K. (2014). NeuCube: a spiking neural network architecture for mapping, learning and understanding of spatio-temporal brain data. *Neural Netw.* 52, 62–76. doi: 10.1016/j.neunet.2014.01.006

Kheradpisheh, S. R., Ganjtabesh, M., Thorpe, S. J., and Masquelier, T. (2017). *STDP-Based Spiking Deep Neural Networks for Object Recognition*. Available online at: http://arxiv.org/pdf/1611.01421.

Krantz, J. (2012). *Experiencing Sensation and Perception*. Upper Saddle River, NJ: Prentice Hall.

Krieger, B., Qiao, M., Rousso, D. L., Sanes, J. R., and Meister, M. (2017). Four alpha ganglion cell types in mouse retina: function, structure, and molecular signatures. *PLoS ONE* 12:e0180091. doi: 10.1371/journal.pone.0180091

Laptev, I., and Caputo, B. (2005). *Recognition of Human Actions*. Available online at: http://www.nada.kth.se/cvap/actions/ (Accessed Aug 29, 2017).

Lecun, Y., Bottou, L., Bengio, Y., and Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proc. IEEE* 86, 2278–2324. doi: 10.1109/5.726791

Lichtsteiner, P., Posch, C., and Delbruck, T. (2008). A 128x128 120 dB 15 μs latency asynchronous temporal contrast vision sensor. *IEEE J. Solid State Circ.* 43, 566–576. doi: 10.1109/JSSC.2007.914337

Liu, Q., Pineda-García, G., Stromatias, E., Serrano-Gotarredona, T., and Furber, S. B. (2016). Benchmarking spike-based visual recognition: a dataset and evaluation. *Front. Neurosci.* 10:496. doi: 10.3389/fnins.2016.00496

Masquelier, T., Guyonneau, R., and Thorpe, S. J. (2009). Competitive STDP-based spike pattern learning. *Neural Comput.* 21, 1259–1276. doi: 10.1162/neco.2008.06-08-804

Masquelier, T., and Thorpe, S. J. (2007). Unsupervised learning of visual features through spike timing dependent plasticity. *PLoS Comput. Biol.* 3:e31. doi: 10.1371/journal.pcbi.0030031

Monsell, S., and Driver, J. (2000). *Control of Cognitive Processes: Attention and Performance XVIII*. Cambridge, MA; London: MIT Press.

Mozafari, M., Kheradpisheh, S. R., Masquelier, T., Nowzari-Dalini, A., and Ganjtabesh, M. (2017). First-spike based visual categorization using reward-modulated STDP. Available online at: http://arxiv.org/pdf/1705.09132

Nelson, R. (1995). "Visual responses of ganglion cells," in *Webvision: The Organization of the Retina and Visual System*, eds H. Kolb, E. Fernandez, and R. Nelson (Salt Lake City, UT: University of Utah Health Sciences Center).

Perez-Carrasco, J.-A., Serrano, C., Acha, B., Serrano-Gotarredona, T., and Linares-Barranco, B. (2010). "Spike-Based Convolutional Network for Real-Time Processing," in *20th International Conference on Pattern Recognition (ICPR), 2010: 23-26 Aug. 2010, Istanbul, Turkey; proceedings* (Piscataway, NJ: IEEE), 3085–3088.

Perez-Peña, F., Morgado-Estevez, A., Linares-Barranco, A., Jimenez-Fernandez, A., Gomez-Rodriguez, F., Jimenez-Moreno, G., et al. (2013). Neuro-inspired spike-based motion: from dynamic vision sensor to robot motor open-loop control through spike-VITE. *Sensors (Basel)* 13, 15805–15832. doi: 10.3390/s131115805

Purves, D. (ed.). (2012). *Neuroscience*. Sunderland, MA: Sinauer.

Querlioz, D., Bichler, O., Dollfus, P., and Gamrat, C. (2013). Immunity to device variations in a spiking neural network with memristive nanodevices. *IEEE Trans. Nanotechnol.* 12, 288–295. doi: 10.1109/TNANO.2013.2250995

Rosa, M. G. P. (2002). Visual maps in the adult primate cerebral cortex: some implications for brain development and evolution. *Braz. J. Med. Biol. Res.* 35, 1485–1498. doi: 10.1590/S0100-879X2002001200008

Serrano-Gotarredona, T., and Linares-Barranco, B. (2015). Poker-DVS and MNIST-DVS. Their history, how they were made, and other details. *Front. Neurosci.* 9:481. doi: 10.3389/fnins.2015.00481

Stromatias, E., Soto, M., Serrano-Gotarredona, T., and Linares-Barranco, B. (2017). An event-driven classifier for spiking neural networks fed with synthetic or dynamic vision sensor data. *Front. Neurosci.* 11:350. doi: 10.3389/fnins.2017.00350

Thorpe, S., and Gautrais, J. (1999). "Rank Order Coding," in *Computational Neuroscience: Trends in Research, 1998*, ed J. M. Bower (Boston, MA: Springer US), 113–118.

Uzzell, V. J., and Chichilnisky, E. J. (2004). Precision of spike trains in primate retinal ganglion cells. *J. Neurophysiol.* 92, 780–789. doi: 10.1152/jn.01171.2003

van Rossum, M. C., Bi, G. Q., and Turrigiano, G. G. (2000). Stable Hebbian learning from spike timing-dependent plasticity. *J. Neurosci.* 20, 8812–8821. doi: 10.1523/jneurosci.20-23-08812.2000

van Rullen, R., and Thorpe, S. J. (2001). Rate coding versus temporal order coding: what the retinal ganglion cells tell the visual cortex. *Neural Comput.* 13, 1255–1283. doi: 10.1162/08997660152002852

Wei, H., and Ren, Y. (2013). A mathematical model of retinal ganglion cells and its applications in image representation. *Neural Process Lett* 38, 205–226. doi: 10.1007/s11063-012-9249-6

Yousefzadeh, A., Serrano-Gotarredona, T., and Linares-Barranco, B. (2015). *MNIST-DVS and FLASH-MNIST-DVS Databases*. Available online at: http://www2.imse-cnm.csic.es/caviar/MNISTDVS.html (Accessed Aug 21, 2017).

Zhao, B., Ding, R., Chen, S., Linares-Barranco, B., and Tang, H. (2015). Feedforward Categorization on AER Motion Events Using Cortex-Like Features in a Spiking Neural Network. *IEEE Trans. Neural Netw. Learn. Syst.* 26, 1963–1978. doi: 10.1109/TNNLS.2014.2362542

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 17 August 2018
doi: 10.3389/fnins.2018.00559

A Model for a Filling-in Process Triggered by Edges Predicts "Conflicting" Afterimage Effects

Hadar Cohen-Duwek and Hedva Spitzer*

Vision Research Laboratory, School of Electrical Engineering, Tel-Aviv University, Tel-Aviv, Israel

The goal of our research was to develop a compound computational model that predicts the "opposite" effects of the alternating aftereffects stimuli, such as the "color dove illusion" (Barkan and Spitzer, 2017), and the "filling in the afterimage after the image" (van Lier et al., 2009). The model is based on a filling-in mechanism, through a diffusion equation where the color and intensity of the perceived surface are obtained through a diffusion process of color from the stimulus edges. The model solves the diffusion equation with boundary conditions that takes the locations of the chromatic edges of the chromatic inducer (chromatic stimulus) and the achromatic remaining contours into account. These contours (edges) trigger the diffusion process. The same calculations are done for both types of afterimage effects, with the only difference related to the location of the remaining contour. While a gradient toward the inducing color produces a perception of the complementary color, an opposite gradient yields the perception of the same color as that of the chromatic inducer. Furthermore, we show that the same computational model can also predict new alternating aftereffects stimuli, such as the spiral stimulus, and the averaging of colors in alternating afterimage stimuli described by Anstis et al. (2012). The suggested model is able to predict most of the additional properties related to the "conflicting" phenomena that have been recently described in the literature, and thus supports the idea that a shared visual mechanism is responsible for both the positive and the negative effects.

Keywords: afterimage effects, filling-in, diffusion, visual system mechanism, computational model

OPEN ACCESS

Edited by:
Qasim Zaidi,
University at Buffalo, United States

Reviewed by:
Greg Francis,
Purdue University, United States
Jihyun Yeonan-Kim,
National Institutes of Health (NIH), United States

****Correspondence:***
Hadar Cohen-Duwek
hadarli@gmail.com

Specialty section:
This article was submitted to Perception Science, a section of the journal Frontiers in Neuroscience

Received: *04 January 2018*
Accepted: *25 July 2018*
Published: *17 August 2018*

Citation:
Cohen-Duwek H and Spitzer H (2018) A Model for a Filling-in Process Triggered by Edges Predicts "Conflicting" Afterimage Effects. Front. Neurosci. 12:559. doi: 10.3389/fnins.2018.00559

INTRODUCTION

This study concerns two non-classical afterimage illusions, both involving a chromatic stimulus i.e., a chromatic inducer that is presented for a short duration of time, and is then followed by the presentation of an achromatic remaining contour that may overlap with the inner or outer border of the chromatic region of the inducer. The location of this remaining contour, can determine whether the perceived filling-in color will be the same as, or complementary to, the chromatic inducer. Two famous examples of these phenomena are: the "Filling-in the Afterimage after the image effect" (van Lier et al., 2009), and the color dove illusion (Barkan and Spitzer, 2009, 2017; Macknik and Martinez-Conde, 2010). Both phenomena involve a filling-in process of surfaces between edges, and the effects are obtained with a narrow spatial inducing area and relatively short induction time. Since these two phenomena yield complementary perceived colors, derived from the very same inducer, we refer to them as "conflicting" effects.

In the "Filling in the Afterimage after the image" (van Lier et al., 2009) illusion, the inducing stimulus is a chromatic shape that may have two or more colors. After the chromatic inducing stimulus is removed, an outline contour matching one of the shape colors is presented. The complementary afterimage color perceived depends on the shape and the location of the drawn outline contour (van Lier et al., 2009), (**Figure 1**, second column). Since the color inside the contour in the perceived afterimage is complementary to the color of the inducing stimulus, we henceforth, refer to this illusion as a "negative effect."

It should be noted that this negative effect is not a simple variation of the "classical" negative afterimage, where, when a stimulus is removed after a relatively long (20-30 seconds) exposure, the observer perceives the opposite chromaticity (complementary color DeValois and Webster, 2011). It should also be noted that the colors in the classical afterimage are perceived only in the retinotopic area that was induced.

In the color dove illusion (Barkan and Spitzer, 2009, 2017), the inducing stimulus is a shape surrounded by a colored area or strip (red in **Figure 1**, first row). After the chromatic inducing stimulus is removed, an outline contour matching the original inducing stimulus is presented (**Figure 1**, second row). This gives rise to the perception of an afterimage (**Figure 1**, third row) filled with a color similar to that in the inducing stimulus (although weaker), and not the complementary color as in the negative effect. Such an effect has also been reported with objects of different shapes (Hazenberg and van Lier, 2013). Since the perceived color inside the shape is similar to that presented in the inducing stimulus, we henceforth refer to this illusion as a "positive effect," (**Figure 1**, first column).

A similar positive aftereffect was previously investigated by Anstis et al. (1978) who suggested that the positive chromatic afterimage effect is a result of the synergy of two known visual mechanisms: simultaneous contrast (Gerrits and Vendrik, 1970; Anstis et al., 1978) and colored afterimage (Daw, 1962; Wyszecki, 1986; Shimojo et al., 2001).

The alternating effects differ from a classical afterimage in their temporal and spatial properties. A classical afterimage requires a relatively long exposure time and a large spatial area of induction, in order to obtain a filling-in effect in a small region with the complementary color (Anstis et al., 1978). In the phenomena described here, preliminary results indicate that the positive effect is not abolished even if the area of the chromatic inducer is spatially thin (Hazenberg and van Lier, 2013; Barkan and Spitzer, 2017. This is in contrast to the explanation given by Anstis et al. (1978), since psychophysically, when the area of a chromatic inducer is thin, the effect of simultaneous contrast is not manifested (preliminary results). The positive and the negative effects are also distinguished from the classical aftereffect (Anstis et al., 1978), in their temporal properties. The duration of the alternating stimuli can be very short (500 ms), a period of time that is insufficient to obtain the classical afterimage effect (Anstis et al., 1978; van Lier et al., 2009; Barkan and Spitzer, 2017).

A further distinguishing characteristic of these phenomena is that, in addition to the temporal and spatial differences from the classic afterimage effect, the color in both the positive and negative effects is perceived in new areas that have not been induced or adapted previously (van Lier et al., 2009). It has to be noted that even though the positive and the negative effects share several common properties, they are still phenotypically different and therefore they can be seen as "conflicting effects."

Hazenberg and van Lier (2013) investigated "alternating watercolors," which have the spatial and the chromatic structure as of the classical watercolor stimuli. These types of stimuli can be considered as the positive and the negative stimuli, while the same classical watercolor stimulus is used as the chromatic inducer stimuli for both positive and negative aftereffects. In this case, the remaining contours are located at the inner or the outer contours of the chromatic edges of the inducer stimulus. The reported results (Hazenberg and van Lier (2013) indicated that the positive and negative effects were affected differently by a number of parameters including the luminance of the area inside the shape and the luminance of the remaining contour.

At present, the visual mechanisms responsible for the recently described positive and negative effects are still unknown and there are no successful computational models for the phenomena. This is less surprising in view of the fact that there remains a lack of consensus concerning the mechanism of even the classical afterimage, despite the wealth of research in the literature. The physiological mechanisms commonly proposed as responsible for the classical negative afterimages range from bleaching of cone photo-pigments to cortical adaptation (Williams and Macleod, 1979; Shimojo et al., 2001; Clair et al., 2007; van Lier et al., 2009; Zaidi et al., 2012; Webster, 2015; Zeki et al., 2017). A recent paper suggested a different mechanism to the van Lier et al. (2009) effect and attributed the filling-in process to the perception of transparency cue and cortical mechanisms (On and van Boxtel, 2017).

Additional recent research (Zaidi et al., 2012) has suggested that the classical and the negative afterimage effects are derived from the retinal ganglion mechanism, which yields the neuronal rebound effect. According to this mechanism, the ganglion

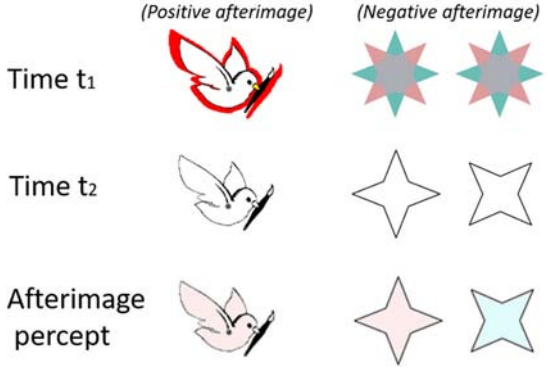


FIGURE 1 | Demonstration of two "conflicting" alternating aftereffects, "The Color Dove illusion" (Positive effect, left column) and the Von Lier et al. Illusion (Negative effect, right columns). The first row shows the inducing stimulus, the second row shows the drawn contours presented at time t2, and the third row represents the resulting perceived afterimage.

neurons can fire bursts if inhibited and then released from inhibition (Spitzer et al., 1993; Grunfeld and Spitzer, 1995; Francis, 2010; Zaidi et al., 2012). It should be noted that while the rebound effect may modulate the creation of complementary colors, it cannot be responsible for the either the negative or positive effects in their entirety.

Previous computational models have been reported to describe both the complementary perceived color and the filling-in components (Grossberg and Todorovic, 1988; Francis and Rothmayer, 2003; Francis and Ericson, 2004; Francis and Schoonveld, 2005; Wede and Francis, 2006, 2007; Van Horn and Francis, 2008) . These models were based on the original "Form And Color And Depth" FACADE) model (Grossberg and Mingolla, 1985), which described two main visual processing systems: a boundary contour system (BCS) that processes boundary or edge information, and a feature contour system (FCS) that uses information from the BCS to control the spreading (filling-in) of surface properties, such as color and brightness. According to the FACADE model, the filling-in stage requires the FCS networks to diffuse signals containing feature information about color and brightness across the surface, while boundaries in the BCS block the spreading.

The FACADE model and its variations succeed in predicting the afterimage effects of the MacKay modal complementary afterimages (MCAI) phenomena (MacKay, 1957; Vidyasagar et al., 1999). This effect involves sequential viewing of two orthogonally related patterns (the first one a constant pattern and the second one a flickering contrast reversal pattern). The result is an afterimage percept that is related to the first pattern (Francis and Rothmayer, 2003; Francis and Ericson, 2004; Francis and Schoonveld, 2005; Wede and Francis, 2006, 2007; Van Horn and Francis, 2008). A number of studies have examined the different spatial and temporal properties of the MCAI effect, for example the spatial and temporal frequency of the two gratings from the first and second presentations (Francis and Rothmayer, 2003), the gap width (Francis and Ericson, 2004), the split gratings (Francis and Schoonveld, 2005), duration between the two grating presentations and the blank presentation (Wede and Francis, 2006), attentional properties (Wede and Francis, 2007), and the role of the difference orientations of the constant and the flickering grating (Van Horn and Francis, 2008). Francis and colleagues confronted their computational model's prediction with the perceived results.

It should be noted that the MCAI and its variations discussed in these Francis papers are not necessarily related to the positive and negative aftereffects phenomena described in our current report. The main differences between the MCAI (MacKay, 1957; Vidyasagar et al., 1999) phenomena and the positive and the negative effects concern the different types of the stimulus components, at these two groups of effects. The stimulus differences related to the orientation gratings and contrast reversal flickering patterns used to produce the MCAI effect versus the chromatic shape of inducer and remaining contour that trigger the positive and negative effects. These differences in the type of stimuli might imply distinct mechanisms that involve additional different components, even though both models can basically be attributed to diffusion processes.

Francis (2010) applied a similar diffusion model to that described previously in Francis and Rothmayer (2003) in order to address the negative effect of van Lier's illusion (van Lier et al., 2009), and succeeded with the model's predictions. At a later stage, Kim and Francis (2011) conducted a series of psychophysical experiments designed to prove that a simple diffusion model (Francis, 2010) cannot account for the additional properties characterize the negative after effect. They tested the hypothesis, for example, that a contour traps the perceived afterimage color, by adding additional remaining contours. Their model simulations predicted that these additional remaining contours would block the spread of a color to the middle of the surface, **Figure 4**.

However, contrary to Francis's predictions (Francis, 2010), the results of the psychophysical experiments showed that additional remaining contours blocked color spreading only when they overlapped with the inducer edges, but not when they were drawn away from the inducer edges (Kim and Francis, 2011), **Figure 4**. More important to our discussion is the fact that FACADE model did not and cannot model the positive effect. In this study, we present a computational model that can predict both the negative and the positive effects, and postulate that these effects are derived from the same mechanism. We also test whether the model can predict additional afterimage phenomena beyond the two described effects.

MODEL

The following sections describe a unified computational model that can predict the two known "conflicting" (opposite) phenomena, the positive "color dove illusion" and the negative "filling-in afterimage after the image" illusion. The model is also able to predict additional variations of the positive and the negative effects. We suggest here, that despite differences in their spatial and temporal properties, these two types of phenomena are produced by a very similar (mutual) mechanism. The model considers several crucial factors for the perceived temporal effects and these are presented in **Figure 2**.

Model Assumptions

The model is based on the following assumptions: (a) An edge triggers a diffusion process in its complementary color. (b) A contour can be a perceived contour and not necessarily a physical spectral gradient. (c) The diffusion process depends on the correspondence between the chromatic stimulus gradients and the remaining contours. (d). The positive and the negative effects are always present, while the dominant perceived color is determined by the location of the remaining contours.

The Stimulus: The Chromatic Inducer and the Remaining Contours

The input of the model is composed of two temporal components, the first one is a chromatic stimulus, I_0 in **Figure 2**, and the second one relates to the remaining contours I_{1a} and I_{1b} in **Figure 2**. The remaining contours can appear in different possible locations, and these locations determine whether the perceived result will be a positive effect or a negative effect.

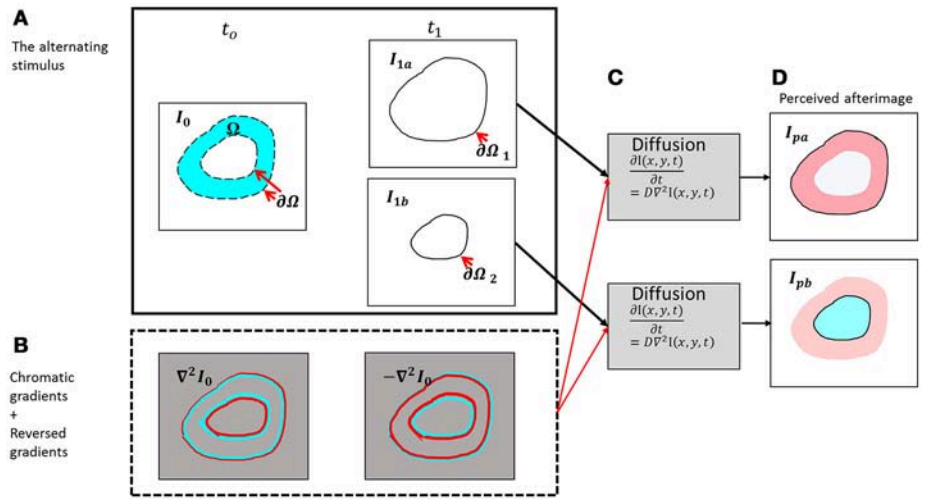


FIGURE 2 | Schematic diagram of the presented model. **(A)** The chromatic stimulus (I_0) at t_0 and the two options for the remaining contour [the outer contour - I_{1a} (upper box) and the inner contour – I_{1b} (lower box)]. **(B)** The chromatic gradients of the chromatic stimulus and their reversed chromatic gradients. **(C)** The diffusion process for the outer contour (upper box) and the diffusion process for the inner contour (lower box). **(D)** The perceived afterimages according to the outer remaining contour (negative effect, I_{pa}) and to the inner remaining contour (positive effect, I_{pb}).

Chromatic Gradients

The building blocks of the model are designed to simulate components of the visual system, and in this case, the opponent and double-opponent receptive fields. The color coding opponent receptive fields encode color contrast, but not spatial contrast. In other words, the color opponent receptive fields are able to differentiate between colors, but cannot detect spatial gradients or edges (Barkan et al., 2008). The double opponent receptive fields, however, are sensitive to both spatial and chromatic gradients and have color opponent receptive fields both at the center and in the surround receptive field regions (Shapley and Hawken, 2011). This opponency in both spatial and chromatic properties produces a spatio-chromatic edge detector.

For the sake of simplicity, we compute the opponent response of the opponent receptive fields as color-opponent only, where, in this simplified case, each chromatic encoder contains the same spatial resolution. This is computed by an opponent color-transformation (Sande et al., 2010), Equation (1). This transformation converts each pixel of the image I_0, in each chromatic channel R,G, and B into opponent color-space, via the transformation matrix O (Sande et al., 2010). $I_{OPPONENT} = OPPONENT\{RGB\}$ as follows:

$$I_{OPPONENT} = \begin{pmatrix} O_{RG} \\ O_{YB} \\ O_{BW} \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} & \frac{-2}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \end{pmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix} \quad (1)$$

where O_{RG}, O_{YB}, O_{BW} are the new channels of the transformed image $I_{OPPONENT}$. R, G, and B are the red, green and blue channels of I, respectively.

In order to implement the double-opponent response, DO, on an image, we subtract the surround, $O_{surround}$, region of the receptive fields from its center, O_{center}, at the same spatial location:

$$DO = O_{center} - O_{surround}$$

The structure of the double opponent receptive field can be seen as a filter which performs as a second derivative in both spatial and chromatic domains (Conway, 2001; Conway and Livingstone, 2006). For the sake of simplicity and clarity of the calculations, we use a discrete Laplace operator, L, which is commonly used as an approximation to the Difference of Gaussian (DOG) function (Marr, 1982). The discrete Laplace operator, L is (Weickert, 1998):

$$\mathrm{L} = \begin{pmatrix} 0 & \frac{-1}{4} & 0 \\ \frac{-1}{4} & 1 & \frac{-1}{4} \\ 0 & \frac{-1}{4} & 0 \end{pmatrix} \quad (2)$$

The responses of the relevant receptive fields, $DO_{response}$, of the color coding receptive fields to the aftereffect stimuli are presented in Equation (3). The double-opponent $DO_{response}$ response is calculated as a convolution of each opponent channel of $I_{OPPONENT}$ with the discrete Laplace operator Equation (2).

$$DO_{response}(stimulus\ on) = \nabla^2 I_{OP} \approx I_{OP} * L \quad (3)$$

Figure 2B demonstrates the responses of the receptive fields to the original stimulus (**Figure 2A**) at time t_0, Equation (3).

The Perceived Gradients—The Responses of the Receptive Fields to the Aftereffects

The model suggests that after the chromatic stimulus disappears, the chromatic gradients obtain the opposite sign. We refer to this condition as "off response," a term commonly used

in electrophysiology (Kandel et al., 2012). The physiological mechanism behind this behavior is still a matter of discussion (Williams and Macleod, 1979; Spitzer et al., 1993; Shimojo et al., 2001; Clair et al., 2007; van Lier et al., 2009; Francis, 2010; Zaidi et al., 2012; Webster, 2015; Zeki et al., 2017). This response has also been termed the rebound response and a variety of models and mechanisms have been suggested to explain how this rebound phenomenon yields a reversed type of response (Spitzer et al., 1993; Grunfeld and Spitzer, 1995; Francis, 2010; Zaidi et al., 2012). **Figure 2B** demonstrates the responses of the simulated receptive fields before and after the chromatic stimulus is removed at times t_0 and t_1, Equation (4).

$$DO_{response}(stimulus\ off) = I_{OP} * (-L) \tag{4}$$

In other words, in this case, the sign of the chromatic gradient, $DO_{response}$, is reversed. Note that the disappearance of the chromatic stimulus, which causes the sign of the edge to be reversed, is in accordance to the model's assumption (section Model Assumptions, A). There are also experimental results that support this assumption (Zaidi et al., 2012).

This operation of edge reversal is realized in the model through reversing the sign of the DO receptive field responses, Equation (3). This reversed chromatic gradient triggers the diffusion process, **Figure 2C**, Equation (5).

Filling-in as a Diffusion Process

The diffusion process is expressed by the diffusion (or heat) Equation (5), (Weickert, 1998). The model assumes that the suggested diffusion of the filling in process is similar to the physical diffusion where the signals spread in all directions, until "blocked" by contours or edges. This type of filling-in process is referred in the literature as the "isomorphic filling-in theory" (von der Heydt et al., 2003). The choice of such a type of filling-in infers that the borders (chromatic or achromatic) do not function primarily as blockers, but instead that the borders play a role as heat sources for the diffusion. When the direction of the diffusion spread is in the opposite direction (colliding) to that of an additional heat source, the spread will actually be blocked by the heat source. These principles are applied in our model through the famous diffusion equation (Weickert, 1998), as in the following equation:

$$\frac{\partial I(x,y,t)}{\partial t} - D\nabla^2 I\left(x,y,t\right) = h_c \tag{5}$$

where $I\left(x,y,t\right)$ denote the image in a space-time location $\left(x,y,t\right)$, D is the diffusion (or heat) coefficient, and h_c represents a heat source. The time course of the perceived image is assumed to be very fast, in accordance with previous reports (van Lier et al., 2009; Barkan and Spitzer, 2017). This time course is also termed "immediate filling-in" (von der Heydt et al., 2003).

Following this assumption, for the sake of simplicity, we can ignore the fast dynamic stages of the diffusion equation, and therefore compute only the steady-state stage of the diffusion process. Consequently, the diffusion (heat) Equation (5) is reduced to the Poisson Equation (6).

$$\nabla^2 I\left(x,y,t\right) = -h_c \tag{6}$$

THE CHROMATIC EDGES AND THE REMAINING CONTOURS

In order to maintain and enhance and/or byproduct to trap this diffusion effect there is a "requirement" for a border. The model suggests that the chromatic diffusion can be "trapped" only when the achromatic remaining contour, $\partial\Omega_1$ **Figure 2A**, overlaps the original edges of the chromatic stimulus, $DO_{response}$. Support for this assumption is also provided from the psychophysical results of Kim and Francis (2011).

Whether the reminding contour $\partial\Omega_1$, is an inner or an outer contour, for example (**Figure 2**), determines the perceived color of the effect. When the remaining contour is the outer contour, the reversed contour, i.e.; the complementary contour, [**Figure 2A**, Equation (4)] triggers a diffusion color that is complementary to the color of the inducer, i.e. red in the specific case of **Figure 2B**. The outer contour, $\partial\Omega_1$, determines that the fill-in color will be complementary to the inducer (negative effect), whereas the inner contour, $\partial\Omega_2$, determines that the fill-in color will be the same color as that of the inducer (positive effect). It has to be noted that the mechanism detects the chromatic edges, and does not treat the inner or outer edges separately. The configuration and the locations of the remaining contours, and not the model, determine the predicted perceived colors.

It is clear that a remaining contour that overlaps the chromatic gradient plays a role as a diffusion trigger and at the same time as a "blocker." However, our preliminary results suggest that the original chromatic gradient, $DO_{response}$, also plays a role as a diffusion trigger and "blocker," even though it has a weaker effect when it does not overlap the remaining contours. This observation is also supported by findings of Hazenberg and van Lier (2013). They concluded that the chromatic border in the negative effect "apparently prevented the colored afterimage of the chromatic contour from spreading."

This minor effect of additional blockage, derived from the chromatic edges, has been integrated into the model by applying different weight functions to each chromatic and achromatic border. The model assumes that the remaining contour also plays a role as an enhancer to the reversed chromatic edges, $-DO_{response}$. Therefore, if the remaining edge, $\partial\Omega_1$, overlaps the original gradient edge (the chromatic gradients of the inducing stimulus, $-DO_{response}$), it will enhance these chromatic edges. The mathematical expression of this role is expressed by the weight functions α and β:

$$\nabla^2 O_p = -DO_{response} \cdot (\alpha\partial\Omega_1 + \beta),\ where\ \alpha > \beta \tag{7}$$

where O_p is the perceived image in the opponent color-space (Sande et al., 2010) and α and β are constants, but can be further extended to be functions.

Solving Equation (7) yields a response to the perceived afterimage O_p given the reversed gradients $-DO_{response}\left(x,y,i\right)$, Equation (4), according to specific initial constraint. **Figure 2D**

represents the perceived afterimage, O_p, but with an additional technical stage of transforming the opponent color space O_p to the RGB color space, $I_{P(rgb)}$, Equation (11).

The interpretation of the solution as suggested above is that a very similar mechanism is responsible for both the negative and the positive effects, although it is possible that the two phenomena do not stem from the exact same visual mechanism. The model may separate the positive and the negative effects to two channels. One channel is for the chromatic area, where the negative effect is more dominant, while the other channel serves the achromatic area, where the positive effect is more dominant. Since the negative effect is given by a response from the chromatic induced region, whereas with positive effect there is a perceived response to an area that has not been induced with color, we assumed that the weight function of the negative effect should be higher than the positive effect (Equation 10). This separation can be justified by analogy to the visual system. The existence of separated Magno, Parvo, and Konio visual pathways in the visual system suggests that separating chromatic and achromatic calculations in this way may be a true reflection of the visual system processing (Shevell, 2003).

We implanted the two separated channels for the positive and negative effects by calculating the diffusion Equation (5), separately for the chromatic and achromatic zones in the original image (I_0). The positive effect $O_{p,positive}$ occurs in the achromatic zones of the initial image I_0, **Figure 2A** and the negative effect $O_{p,negative}$ occurs in the chromatic zones of the initial image I_0, **Figure 2A.** Accordingly, the equation is solved separately for the negative effect $O_{p,negative}$ and for the positive effect , $O_{p,positive}$, (see the section above). The simulation result is calculated as:

$$\nabla^2 O_{p,negative}(\mathrm{x,y,i}) = -\mathrm{DO}_{response}(\mathrm{x,y,i}) \cdot (\alpha \partial\Omega_1 + \beta) \,\mathrm{in}\, \Omega \tag{8}$$

$$\nabla^2 O_{p,negative}(\mathrm{x,y,i}) = -\mathrm{DO}_{response}(\mathrm{x,y,i}) \cdot (\alpha \partial\Omega_1 + \beta) \,\mathrm{in}\, \overline{\Omega} \tag{9}$$

i = RG,YB,WB, where each opponent channel is solved separately.

$$O_p = \frac{O_{p,positive} + O_{p,negative}}{max_{all\_channels}\left\{I_{p,positive}\right\} + max_{all\_channels}\left\{I_{p,negative}\right\}} \tag{10}$$

$$I_{P(rgb)} = OPPONENT^{-1}\left\{O_p\right\} \tag{11}$$

where $\max_{all\_channels}\{I\}$ is the maximum value of all channels in the image I ($\max\{I\}$ *is a scalar*). α and β present the weights of the remaining contours and the chromatic stimulus edges, accordingly.

In order to calculate the perceived afterimage from both the negative $O_{p,negative}$ and the positive $O_{p,positive}$ effects, Equations (8–10), we need to define (a) boundary conditions, and (b) the initial values. We shall henceforth denote the inducing stimulus (the original color image) by I_0, where Ω is an area in the image I_0, and $\partial\Omega$ is the border of Ω, **Figure 2A**. I_1 is the remaining contour image and $\partial\Omega_1 or\ \partial\Omega_2$ are the remaining contours (the remaining boundaries, although the boundaries in I_1 might be different from those in I_0. Therefore, the chromatic edges, $\partial\Omega$, do not necessarily overlap the remaining contours $\partial\Omega_1 or\ \partial\Omega_2$). The boundary conditions of the perceived image I_p and the initial state (initial conditions) are chosen to be an achromatic white color on the output image border. Thus, the boundary condition is $O_p|_{border} = 1$, **Figure 2A** and the initial image is a blank white image (R = G = B = 1). These conditions are selected in order to enable the generation of the perceived afterimage on a white image as in the original stimuli (Barkan and Spitzer, 2009; van Lier et al., 2009), **Figure 2D**.

RESULTS

Simulation Details

The simulations are produced by assigning the conditions (boundary conditions and initial values) as described above, and applying the Gauss-Seidel method. The simulations are solved in a similar way to that reported in "Methods for Solving Equations" (Simchony et al., 1990) or "Poisson Image Editing" (Pérez et al., 2003). The simulations are implemented by MATLAB software.

The only parameters in the model are α and β, which present the weights of the remaining contours and the chromatic stimulus edges, accordingly. The Parameters were chosen as following: α = 1.3 and β = 0.1, as results of trial and error. These parameters are constant for all the simulations; beside in the **Supplementary Figure 1** where we intended to slightly enhance the result for demonstration.

Model's Simulation and Predictions

The simulation results are divided into three parts. The first part presents the model predictions for both the negative and the positive afterimage phenomena, (Barkan and Spitzer, 2009; van Lier et al., 2009). The second part presents the predictions of the model for two remaining edge variations, as presented in previous studies (Francis, 2010; Kim and Francis, 2011). The third part presents the model predictions for additional aspects of the afterimage phenomenon, where one relates to the color perceived when the remaining edge of the image is not complete (open boundaries, spiral image), and the second relates to spatial averaging of colors, (Anstis et al., 2012).

Negative and Positive Afterimages

We tested the model on the same stimuli as in the study of van Lier et al. (2009) (**Figure 3** first row), and for the general case of the chromatic stimuli I_0, **Figure 2**. **Figure 3** presents the model's predictions for a single colored ring as inducer (second and third rows). It can be seen that the model correctly predicts that the remaining contours can generate a negative or a positive effect depending on their location. Of note, the model correctly predicted the filling-in process of the achromatic area with respect to both negative and positive effects, with the results in accordance with the psychophysical findings reported previously (van Lier et al., 2009; Hazenberg and van Lier, 2013). Having different weight functions for the positive and negative

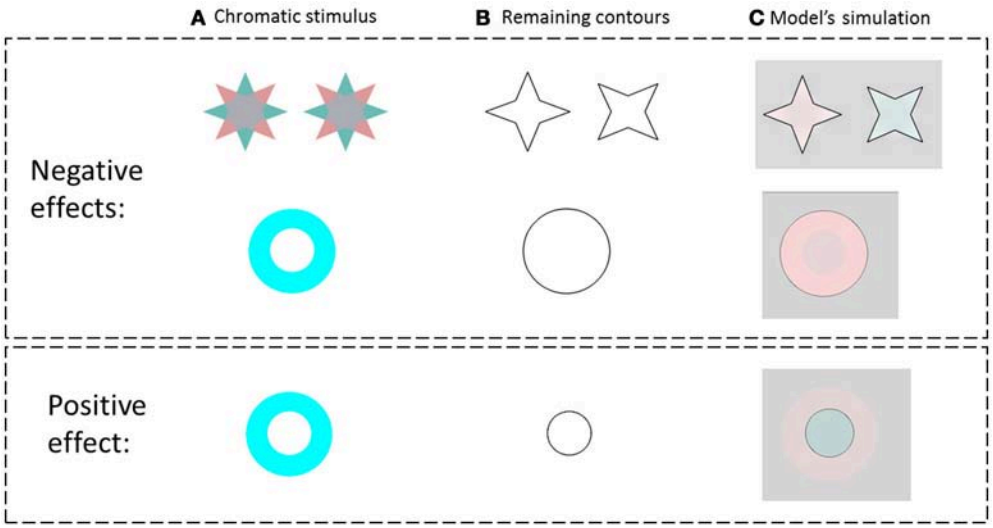


FIGURE 3 | The model predictions for the negative and the positive effects. **(A)** The chromatic stimulus. **(B)** The remaining contours. **(C)** The simulation results. The first row presents the van Lier et al. (2009) stimuli, e.g., the negative effect (stars) and the model prediction for these stimuli. The second row, in the negative effect block, presents a general case of the negative effect, which displays only one chromatic inducer. The third row represents the model predictions for a general case of the positive effect (instead of the "color dove illusion").

effects (Equation 11) enables us to control the predicted effect of a stronger diffusion in the inner than in the outer region of the remaining contours (**Figure 3**). These studies showed that the perceived afterimage has the complementary color when the outer contour is remained, (**Figure 3**, second row), while the same color is perceived when the inner contour is remained (**Figure 3**, third row).

The Role of the Remaining Edges

Comparison to Previous Results

We also tested our model on the same variations of the van Lier et al. (2009), stars stimulus that were tested by Francis (2010) and Kim and Francis (2011). These variations are related to the location and shape of the remaining contour. **Figure 4** presents a comparison between the predictions of our model and that of Francis for two possibilities of drawn remaining contours, (**Figures 4C,D**, respectively). In one case, the remaining edges overlap the chromatic gradients (**Figure 4**, First row), which exist in the inducing stimuli, while in the second case, there is no overlap (**Figure 4**, second row).

The predictions of both models yielded the same results when the boundaries overlapped (**Figure 4**, first row, **C,D**), and these results agree with the experimental perceived results reported previously (Kim and Francis, 2011). However, the predictions of the models differed when the boundaries were non-overlapping. **Figure 4** second row shows that the inner rectangle is reddish (**Figure 4D**) according to our model, but gray according to the predictions of Francis' model's (**Figure 4C**). Notably, the psychophysical findings (Kim and Francis, 2011) support our model, which predicts that remaining contours that do not overlap the chromatic gradient, do not block the diffusion process.

Model Predictions for a New Stimulus With Different Variations of Remaining Edges

Having successfully tested our model on previously described stimuli, we proceeded to further challenge the simulations with new spiral stimuli, which have not been described previously or experimentally tested. The new stimuli can simultaneously generate both positive and negative effects because they have both inner and outer borders. This type of stimulus enables us to test a critical property regarding the effect of closed or open remaining edges, on the relevant aftereffects.

The model's results for the spiral stimuli, indicated that the dominant color perceived in the afterimage depends on whether the remaining edges are the inner or outer edge, (first and second rows of **Figure 5**, respectively). The dominant color, predicted by our model, can therefore be either complementary or similar to that of the inducer color, where the outer border produces a dominant positive effect, while the inner border produces a dominant negative effect, (**Figure 5C**). These predictions are supported by preliminary psychophysical results (Manuscript in preparation).

As a further test, we examined the ability of our model to predict the psychophysical results of the aftereffects that can be perceived from performing spatial averaging within the remaining contours (Anstis et al., 2012). This question was tested by our model simulation under two configurations representing variations of the negative and positive effects (**Figures 6**, **7**). While the negative stimuli are as previously reported (Anstis et al., 2012), the positive stimuli are new and are designed to induce the positive effect. **Figures 6**, **7** demonstrate the model's predictions for the negative and positive versions of averaging effect, respectively. Note that only the positive configuration (**Figure 7**) induces a classical filling-in, since this is the only

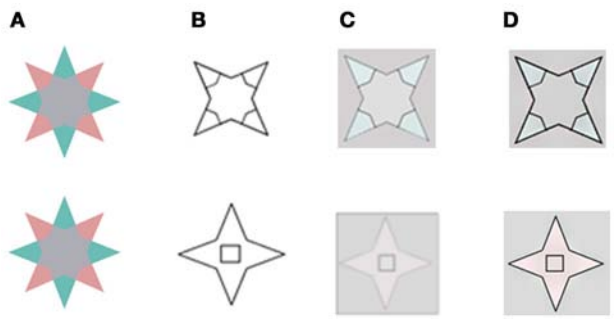


FIGURE 4 | A comparison of our model's predictions to that of Francis for the two locations for the remaining edge that Francis tested experimentally. **(A)** The chromatic stimulus. **(B)** The remaining contours. **(C)** The simulation results as reported in (Francis, 2010; Kim and Francis, 2011). **(D)** The simulation results of the suggested model. In the first row, the inner drawn contours **(B)** overlap the chromatic gradients that exist in the inducing stimulus **(A)**. In the second row, the inner drawn contours do not overlap the chromatic gradients of the inducing stimulus. The results in **(D)** are in agreement with psycho-physical experiments (Kim and Francis, 2011).

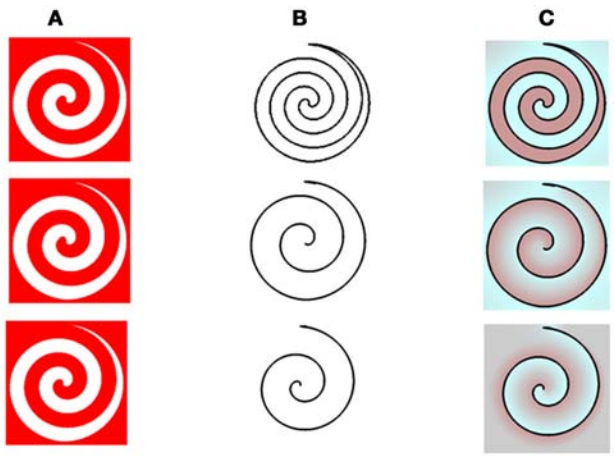


FIGURE 5 | The model predictions for the spiral stimuls with variations in the remaining contours. **(A)** The chromatic stimulus. **(B)** The remaining contours. **(C)** Model's predictions. In this figure, the chromatic stimulus is the same in all the rows, column **(A)**, but the remaining contours are different. In the first row, the drawn contour is a full spiral. In the second row, the outer edge of the spiral shape is presented and in the third row the drawn contour is the inner edge of the spiral shape, column **(B)**. Our model predicts that both cyan and red colors are dominant in the full spiral (first row). When the remaining contour is the outer one, the dominant percieved color is reddish (second row), while, when the remaining contour is on the inner side, the dominant color is cyan (third row).

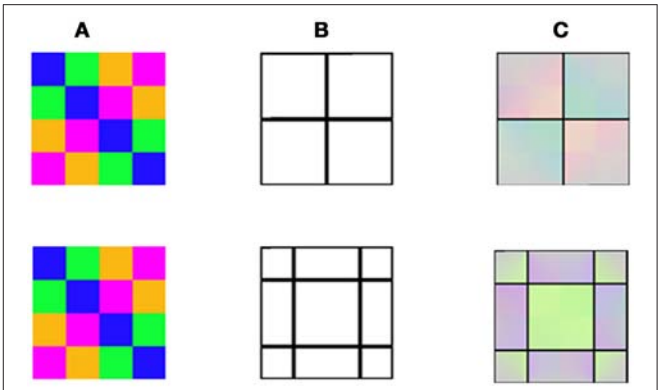


FIGURE 6 | Model predictions for averaging of negative afterimage colors (Anstis et al., 2012). **(A)** The chromatic stimulus. **(B)** The remaining contours with two different locations. **(C)** The model's predictions. It appears that colors of the remaining contours determine the role of color averaging.

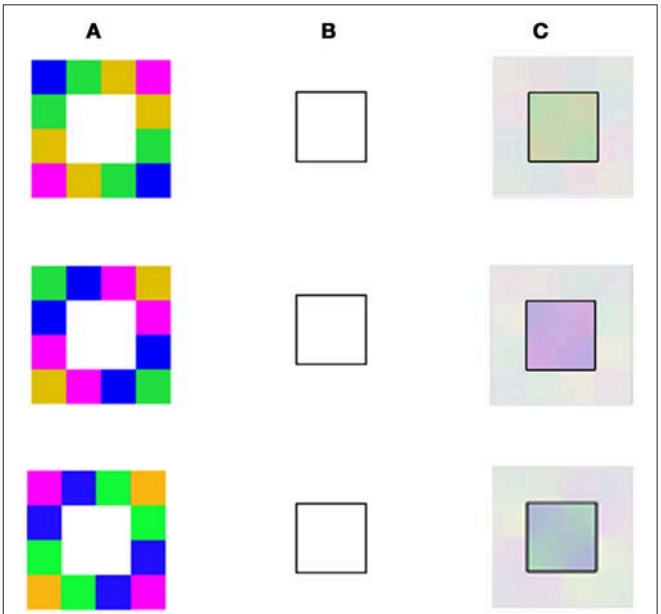


FIGURE 7 | Model predictions for averaging of negative and positive afterimage colors. **(A)** The chromatic stimuli with different color combinations (different rows). **(B)** The remaining contour (identical in all the rows). **(C)** The model's predictions which show that a different "perceived" color is obtained as a result of the chromatic combination of the inducer. It can be seen that there is also an averaging of colors in the positive effect with these new averaging color stimuli.

configuration where there is an achromatic area that can be filled with color.

DISCUSSION

The suggested model involves several stages that can be regarded as a cascade of component mechanisms and responses, i.e., a short duration chromatic stimulus, cessation of this stimulus, creation of complementary chromatic edges which trigger a diffusion process. The suggested model predicts afterimage phenomena, which some of them might appear as "opposite ("conflicting") effects," through the same mechanism and therefore the same equations.

We present here a model that is able to predict both the negative and the positive effects, i.e., where the illusionary filled-in color is either the same color or is complementary to that of the

inducer. The model, therefore can also predict both the famous "filling-in the afterimage after the image" illusion and the "color dove illusion" (van Lier et al., 2009; Barkan and Spitzer, 2017). In addition, the model can also predict both the positive and the negative versions of the effect in shapes that possess non-closed remaining edges and successfully predicted a recently reported predominantly negative afterimage effect related to averaging of colors (Anstis et al., 2012), **Figure 6**.

It might be claimed that diffusion models have been previously suggested to predict the aftereffect in general, and also to predict the alternating aftereffect (Grossberg and Mingolla, 1985; Grossberg and Todorovic, 1988; Francis and Rothmayer, 2003; Francis and Ericson, 2004; Francis and Schoonveld, 2005; Wede and Francis, 2006; Van Horn and Francis, 2008). However, in contrast to previous models, such as FACADE, in our model the trigger for the diffusion mechanism is a "heat source," which implements the diffusion (or heat) equation with a "heat source," Poisson equation (Weickert, 1998). In other words, in our model, the edges are the only trigger for the diffusion process, and have no other role, for example as direct blockers to the diffusion process, as presented in the FACADE model (Grossberg and Mingolla, 1985; Grossberg and Todorovic, 1988; Francis and Rothmayer, 2003; Francis and Ericson, 2004; Francis and Schoonveld, 2005; Wede and Francis, 2006; Van Horn and Francis, 2008). This difference in rationale between FACADE and our model leads to a different structure of diffusion models, (Equation 7). While the FACADE model is composed of two separated components 1) Boundary contour system (BCS) 2) Feature contour system (FCS), our model is consisted of a single component. This component includes both borders and diffusion mechanism, which are computed in the same process (Equation 7). It is not surprising that such differences give rise to different model predictions in the two types of models, as will be described below.

The model described by Francis (2010) succeeded in predicting the negative effect (van Lier et al., 2009), in which the visual afterimage could spread across regions that were not colored in the inducing stimulus. He also could show, by the application of the FACADE model (Grossberg and Mingolla, 1985), that the perceived color and shape of the afterimage could be manipulated by remaining contours that apparently trapped the spread of afterimage color signals. However, this model also mistakenly predicts that a remaining edge will block the spread of color even if there is no overlap with the chromatic gradient edge border (**Figure 1B** in: Francis, 2010). This prediction is in disagreement with the psychophysical findings of the experiments conducted by Kim and Francis (2011). In contrast, our simulations indicate that the diffusion process is not blocked when the achromatic remaining contours do not overlap the chromatic contours.

In addition, as already claimed in the introduction, Francis's model cannot predict the positive effect, since his model assumes that the spread of complementary color across a surface will be blocked by the remaining contour. According to the Francis model (Francis, 2010), the positive effect is predicted to be negated, due to the role of the remaining contour which prevents diffusion of the color to the inner part of the shape. Consequently, the model cannot predict the possibility of obtaining result that shows perception of the same color as of the inducer at a different spatial location. Our model, on the other hand, can predict the positive effect (**Figure 3**), since it assumes that the main role of the contours is to trigger the diffusion process and not primarily aimed to block the diffusion process.

It should be clarified that the FACADE model has been implemented with a number of different diffusion algorithms. Francis, for example, implemented the filling-in process by using a Connected-Component algorithm (Francis and Rothmayer, 2003; Francis, 2010). In the FACADE models the diffusion process is implemented with iterative algorithm, whereas each pixel is averaged with adjacent pixels only if the neighbors are not edges (Grossberg and Todorovic, 1988; Francis and Ericson, 2004; Francis and Schoonveld, 2005; Wede and Francis, 2006; Van Horn and Francis, 2008). In additional studies (Francis and Ericson, 2004; Francis and Schoonveld, 2005) the diffusion model was extended in order to predict additional properties that are related to the MCAI effect. Consequently, the investigators suggested a "non-diffusion" filling-in mechanism, built from directional operations. It has to be noted that in order to predict the MCAI effect a special component was added to the FACADE model, which express the inhibition between orthogonal oriented grids (Francis and Ericson, 2004; Francis and Schoonveld, 2005; Wede and Francis, 2006; Van Horn and Francis, 2008). One important question is whether any of these previous diffusion implementations of the FACADE model (Grossberg and Mingolla, 1985; Grossberg and Todorovic, 1988; Francis and Rothmayer, 2003; Francis and Ericson, 2004; Francis and Schoonveld, 2005; Wede and Francis, 2006; Van Horn and Francis, 2008; Francis, 2010) can successfully predict the positive effect and its variations.

Since the FACADE models mentioned above share the same BCS, which trap the diffusion process and prevent diffusion of the color to the inner part of the shape, they wrongly predict the blockage of the diffusion process in the inner shape, as described experimentally (Kim and Francis, 2011). They also cannot predict the possibility of obtaining the same color as the inducer at different spatial locations and thus cannot predict the positive effect.

While both types of model (ours and the FACADE) assume that the filling-in process is performed by the isomorphic diffusion mechanisms, other groups have suggested that the symbolic mechanism might determine the diffusion process (von der Heydt et al., 2003; Komatsu, 2006; On and van Boxtel, 2017). According to the symbolic theory, "early visual areas extract only the contrast information at the surface border, while the color and shape of the surface are reconstructed in higher areas on the basis of this information" (Komatsu, 2006). Komatsu (2006) reported, however, that neuronal activity of V1 and V2 plays a role in most of the filling-in phenomena such as filling-in at the blind spot, the Craik–O'Brien–Cornsweet illusion, or neon color spreading.

A recent experimental study (On and van Boxtel, 2017) suggested a symbolic mechanism for the negative effect seen in the "stars" of van Lier et al. (2009)." They hypothesized that transparency cues play an important role in the filling-in process of the negative effect and attempted to validate this suggestion

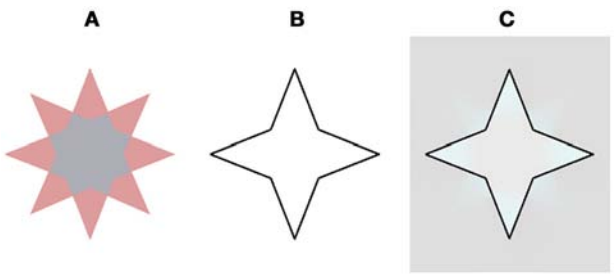


FIGURE 8 | Our model prediction for a star stimulus that does not contains transparency cues and does not yield a filling effect (On and van Boxtel, 2017). This stimulus figured in an experimental study that claimed that the filling does not play a role in the negative aftereffect (Discussion). **(A)** The chromatic stimulus. **(B)** The remaining contour. **(C)** The model's prediction. Note that the complementary color is evident only at the vertices of the star and does not diffuse to the central hexadecagon of the star.

through psychophysical experiments. Their results indicated that transparency clues are a prerequisite for the perceived filling-in effect. When the transparency cues were eliminated by removing one color from the star, the new stimulus contained only one color (**Figure 1B**, in: On and van Boxtel, 2017, **Figure 8**), and the filling-in effect indeed vanished. However, there is a different and even simpler explanation that can explain their psychophysical results.

Figure 8 demonstrates our model's prediction for this specific star stimulus. The rationale for this correct prediction is based on the fact that if a combination of the negative and the positive effects act on the same spatial location they cancel each other out, as a result of the simultaneous induction of complementary colors in the same spatial location, **Figure 8**, (Hazenberg and van Lier, 2013). The original star stimulus of van Lier et al. (2009) consisted of a similar combination of negative and the positive effects, although in this case the two effects enhanced each other. This enhancement was due to the fact that the stars contain two complementary colors (cyan and reddish). When the cyan four-point-star is located inside the remaining contour, the negative effect is produced and the perceived complementary color is reddish. In this case, however, because this reddish four-point-star is located outside the remaining contour, it gives rise to the positive effect, where the perceived color would also be reddish. As a result, the perceived reddish color is enhanced, as a result of the combination of the positive and the negative effects.

It is interesting to consider the stages of analysis of the proposed model as related to components of the visual system. The formation of a complementary or opponent chromatic edge following the cessation of chromatic stimulus (**Figure 2**) has recently been described in the literature as being attributable to a rebound response (Off response), evoked as a burst of spikes from neurons released from the period of inhibition (Spitzer et al., 1993; Grunfeld and Spitzer, 1995; Francis, 2010; Zaidi et al., 2012). The mechanism by which this produces the perception of the complementary color was suggested to be through cross inhibition between opponent channels (Grossberg, 1972; Francis, 2010), or through fast adaptation from the first order (Spitzer and Semo, 2002; Spitzer and Barkan, 2005). The mechanism suggested for the rebound model of Grunfeld and Spitzer (1995) includes the parameters required for the rebound effect, such as the duration of adaptation, the rate and the intensity of the offset of the stimulus. The current model does not include these additional stimulus parameters, but we plan to include these parameters in future.

The development of a further stage of the model has to be discussed in relation to the visual system and to other models. After the rebound response creates the complementary color, the diffusion process is triggered by different components in each model. According to the FACADE model (Grossberg and Mingolla, 1985; Grossberg and Todorovic, 1988; Francis and Rothmayer, 2003; Francis and Ericson, 2004; Francis and Schoonveld, 2005; Wede and Francis, 2006; Van Horn and Francis, 2008), the trigger for the diffusion process is the color of the surface at each location. This was described as "color spreads all across the surface within the boundary" (Kim and Francis, 2011). In contrast, in our model, the borders (the chromatic edges, i.e., double opponent, in the chromatic stimulus and the remaining contours, as a modulation to the chromatic edges) are the trigger for the diffusion process (Equation 7).

The experimental results of Hazenberg and van Lier (2013) appear to support our model with regard to the trigger for the diffusion process. These researchers demonstrated experimentally that the location of remaining contour that overlaps the chromatic edge can determine whether the result will be a positive or a negative effect. In fact, our model suggests that the perceived chromatic edge triggers an isomorphic filling-in process, according to isomorphic filling-in theory (von der Heydt et al., 2003). It should be noted that the idea that an afterimage of the chromatic contours triggers the isomorphic diffusion process has been raised previously by Hazenberg and van Lier (2013). It has also been suggested that the color signals in this type of filling-in process, spread in all directions except across borders formed by contour activity (Gerrits and Vendrik, 1970; Cohen and Grossberg, 1984; Arrington, 1994; von der Heydt et al., 2003). The role of the remaining contour is therefore in agreement with the previous suggestion that the contours act as diffusion barriers (Cohen and Grossberg, 1984; von der Heydt et al., 2003). However, according to the current model, this remaining contour is effective as a barrier only when it overlaps with the original chromatic edge of the inducer stimulus. Our model therefore suggests that the remaining contour fulfills two functions: a. enhancing the effect of the inverted chromatic edge Equation (4), b. trapping the diffusion. This dual role is supported by the isomorphic filling-in theory of von der Heydt et al. (2003) who suggested that the chromatic or achromatic receptive field plays a role in the filling-in process. The chromatic-edge receptive fields receive additional activation through horizontal connections (Gilbert and Wiesel, 1979), which keep the border activity high. Their suggestion is general and was not specifically related to the visual effects discussed here (the positive and negative effects).

In addition to the crucial role of the remaining contour, which overlap the chromatic gradients, the chromatic edges (by themselves) also play a role in the perceived afterimage, (Equation 11). This assumption was supported by the findings

of Hazenberg and van Lier (2013), who reported that the filling-in process, (in their version for the positive effect), should be influenced less by the chromatic gradients (Anstis et al., 1978; Hazenberg and van Lier, 2013).

Since the model takes into account the role of the chromatic edges, albeit with less weight than the remaining contour, it predicts that the diffusion at the negative effect will be partially blocked by the original chromatic gradient of the inducing stimulus. As a result, it predicts that the diffusion will not spread to the central area in the negative effect stimuli, **Figure 3**.

Our model predicts that if a border does not exist in the original inducing stimulus, it will not block the diffusion process, as found psychophysically (Kim and Francis, 2011). After conducting psychophysical experiments, Kim and Francis (2011) formulated a qualitative rule that additional contours block color spreading when these contours overlap the inducer edges, but not when they are separated (**Supplementary Figure 1**). It has to be noted that our model's predictions of these results also agree with the qualitative arguments of Hazenberg and van Lier (2013) that there has to be a match (or overlap) between the chromatic edges and the remaining contours. This is derived from a repeated activation of orientation selective neurons that also code for color (von der Heydt et al., 2003).

We also investigated the question of whether it is necessary for the remaining contour to be closed or whether an open spiral stimulus, (**Figure 5**) can produce the effect. Preliminary results are in agreement with our model predictions that the effect can exist in open boundary conditions (**Figure 5**). It should be noted that Francis's simulations cannot predict the negative effect on open boundary conditions, such as in the spiral stimulus (**Figure 5**), because his model depends on a boundary that traps the spread of color (Francis, 2010). However, by applying a previous diffusion model as in Grossberg and Todorovic (1988), a correct prediction can be achieved, but only for the negative parts of the spiral illusion (i.e., only the configuration where the inner border of the spiral is displayed, third row of **Figure 5**). This is because this case involves a difusion process rather than a Connected Component algorithm as in the Francis implementation (Francis and Rothmayer, 2003; Francis, 2010). However, this modification still cannot predict the positive effect in the spiral illusion (second row of **Figure 5**).

A further question was whether the aftereffects can be perceived from spatial averaging within the area of remaining contours. Anstis et al. (2012) showed that colors can undergo spatial averaging within, but not across, contours but tested this effect only on the negative aftereffect. Our model's simulations (**Figures 6**, **7**) are with agreement to the experiments conducted by Anstis et al. (2012). We believe that even if the Francis model was able to predict this averaging effect, it could only work on the negative configuration of the effect.

Our results thus far suggest that the same basic mechanism is responsible for both the negative and the positive effects, but there remains a question as to whether there are additional mechanism's components that differentiate between these two mechanisms. The recent study of Hazenberg and van Lier (2013) can shed a light on this issue, since they investigated several properties of the positive and the negative effects on the afterimage watercolor stimuli. Specifically, they examined the role of the intensity of the inner area of the inducer stimulus and the remaining contour with reference to the positive and the negative effects.

The results of their study indicated that the filling-in effect was stronger in the negative effect under conditions where the inner area of the inducer stimulus was gray (iso-luminance with the chromatic borders) rather than white. This preference was not found in the positive effect. Hazenberg and van Lier (2013) interpreted these findings as the result of the influence of the luminance border between the inner chromatic contour and the interior area. This luminance border was presumed to prevent the colored afterimage of the chromatic contour from spreading. However, under iso-luminance conditions, the luminance borders do not exist, and indeed, the filling-in process is more prominently perceived. Our model can be modified, by taking into account a combination of the chromatic and the achromatic gradients of the chromatic stimulus, in order to predict this influence on the inner area intensity. Due to the differences related to the positive and negative effects, our model predicts that the negative effect will be more prominent with regard to the degree of saturation, while the positive effect will be more prominent in its ability to perform a filling-in task. This prediction should be confirmed by psychophysical experiments.

In order to test the role of the intensity of the remaining contour Hazenberg and van Lier (2013) used thick contours colored either light or dark gray as the remaining contours. They reported that the filling-in effect was perceived only when the contours were gray and not black, and only in case of the positive effect (i.e., where the perceived color is the same as the inducer).

We now suggest that according to our model (**Figure 3**), both gray and black contours can create a complementary color effect, but only in the near vicinity of the chromatic border in the original chromatic stimulus. It is possible that the lack of filling-in color in the positive effect (**Figure 8** in: Hazenberg and van Lier, 2013), was a consequence of the contour thickness of the remaining contours, since in the positive effect, the color has to diffuse through the remaining contour. The border contrast with a gray contour is weaker, and therefore reveals a partial filling-in effect. We suggest that the negative effect was not observed in the reported experiments (Hazenberg and van Lier, 2013) because they were looking mainly at the central area of the stimulus. Such a filling-in color is not expected in the inner white area (**Figure 7** in: Hazenberg and van Lier, 2013) because it is blocked by the luminance border, which contributes to the blockage of the filling-in process [Equation (7), **Figure 3**].

Additional factors that might affect the degree of the aftereffect e.g., include the size of the inducer and induced area, the shape curvature of the chromatic edge, and the exposure duration of the chromatic stimulus. These factors should be separately investigated experimentally for their influence on the positive and the negative effects. Psychophysical experiments are important in order to detect differences in the mechanisms acting in these two types of effects. In addition, psychophysical experiment are required for cases where the remaining contours that trigger the filling-in effect are illusory contours, such as those in the Kanizsa effect and the Neon color spreading effects

(Van Tuijl, 1975; Kanizsa, 1976). This should be tested separately for the positive and negative effects. Our model predicts that for an illusory contour stimulus (which replaces the achromatic remaining contour), the chromatic and the illusory remaining contour have to overlap. However, we believe that the mechanism which creates the illusory contour (as produced in a Kanizsa illusion) is different from the filling-in mechanism. [Different computational models suggested in the literature for the Kanizsa illusion (Grossberg and Mingolla, 1985, 1987; Heitger et al., 1998; Ron and Spitzer, 2011)]. In order to include the prediction of the filling-in effect triggered by illusory contours, we will need to combine the different mechanisms of the illusory contours and the filling-in mechanism, and will therefore need to add an additional model component to detect the illusory contours.

The MCAI Effect (MacKay, 1957; Vidyasagar et al., 1999) is an alternating aftereffect but it differs from the positive and the negative aftereffects, as it contains an additional component, which relates to a different mechanism. This component enables oriented adaptation in the MCAI oriented stimulus (more specifically, of the flickering grid in the relevant stimulus). We expect that our filling-in model will predict this MCAI effect, but only if an additional component, which describes such oriented adaption mechanism (of the MCAI effect), will be added to the model.

Even though the present model does not permit predictions of the behavior of all the free parameters that play a role in the negative and positive effects, this is the first time that a computational model has been able to make crucial predictions on both the positive and the negative effects. In other words, our model succeeds in predicting apparently conflicting phenomena, i.e., those producing the complementary or same color aftereffect, and implies that the same mechanisms function in both effects despite the different manifestations. An important conclusion of this study is that a different appearance does not necessarily infer a difference in the causative mechanisms and driving forces.

The proposed model has several possible applications with the potential to be an applicable algorithm for the restoration of corrupted old images and videos, for example. Such an algorithm may be able to make an educated guess for filling-in color, based on partial information, such as having only remaining contours.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

ACKNOWLEDGMENTS

A short version of this model was presented in a conference (Cohen-Duwek and Spitzer, 2017).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2018.00559/full#supplementary-material

REFERENCES

Anstis, S., Rogers, B., and Henry, J. (1978). Interactions between simultaneous contrast and coloured afterimages. *Vis. Res.* 18, 899–911. doi: 10.1016/0042-6989(78)90016-0

Anstis, S., Vergeer, M., and Van Lier, R. (2012). Luminance contours can gate afterimage colors and "real" colors. *J. Vis.* 12:2. doi: 10.1167/12.10.2

Arrington, K. F. (1994). The temporal dynamics of brightness filling-in. *Vis. Res.* 34, 3371–3387. doi: 10.1016/0042-6989(94)90071-X

Barkan, Y., and Spitzer, H. (2009). *Color Dove Illusion | Best Illusion of the Year Contest*. Available online at: http://illusionoftheyear.com/2009/05/color-dove-illusion/

Barkan, Y., and Spitzer, H. (2017). "The color dove illusion- chromatic filling in effect following a spatial-temporal edge," in *The Oxford Compendium of Visual Illusions*, eds A. G. Shapiro and D. Todorovic (Oxford; New York, NY: Oxford University Press), 752–755.

Barkan, Y., Spitzer, H., and Einav, S. (2008). Brightness contrast-contrast induction model predicts assimilation and inverted assimilation effects. *J. Vis.* 8, 27–27. doi: 10.1167/8.7.27

Clair, R. S., Hong, S. W., and Shevell, S. (2007). Misbinding of color to form in afterimages. *J. Vis.* 7, 366–366. doi: 10.1167/7.9.366

Cohen, M. A., and Grossberg, S. (1984). Neural dynamics of brightness perception: features, boundaries, diffusion, and resonance. *Percept. Psychophys.* 36, 428–456. doi: 10.3758/BF03207497

Cohen-Duwek, H., and Spitzer, H. (2017). "A new diffusion computational model predicts both the positive and the negative short afterimage effects," in *Color and Imaging Conference* (Springfield, VA: Society for Imaging Science and Technology), 103–107.

Conway, B. R. (2001). Spatial structure of cone inputs to color cells in alert macaque primary visual cortex (V-1). *J. Neurosci.* 21, 2768–2783. doi: 10.1523/JNEUROSCI.21-08-02768.2001

Conway, B. R., and Livingstone, M. S. (2006). Spatial and temporal properties of cone signals in alert macaque primary visual cortex. *J. Neurosci.* 26, 10826–10846. doi: 10.1523/JNEUROSCI.2091-06.2006

Daw, N. W. (1962). Why after-images are not seen in normal circumstances. *Nature* 196, 1143–1145. doi: 10.1038/1961143a0

DeValois, K., and Webster, M. (2011). Color vision. *Scholarpedia* 6:3073. doi: 10.4249/scholarpedia.3073

Francis, G. (2010). Modeling filling-in of afterimages. *Atten. Percept. Psychophys.* 72, 19–22. doi: 10.3758/APP.72.1.19

Francis, G., and Ericson, J. (2004). Using afterimages to test neural mechanisms for perceptual filling-in. *Neural Netw.* 17, 737–752. doi: 10.1016/j.neunet.2004.01.007

Francis, G., and Rothmayer, M. (2003). Interactions of afterimages for orientation and color: experimental data and model simulations. *Percept. Psychophys.* 65, 508–522. doi: 10.3758/BF03194579

Francis, G., and Schoonveld, W. (2005). Using afterimages for orientation and color to explore mechanisms of visual filling-in. *Percept. Psychophys.* 67, 383–397. doi: 10.3758/BF03193319

Gerrits, H. J. M., and Vendrik, A. J. H. (1970). Simultaneous contrast, filling-in process and information processing in man's visual system. *Exp. Brain Res.* 11, 411–430.

Gilbert, C. D., and Wiesel, T. N. (1979). Morphology and intracortical projections of functionally characterised neurones in the cat visual cortex. *Nature* 280, 120–125. doi: 10.1038/280120a0

Grossberg, S., (1972). A neural theory of punishment and avoidance, II: quantitative theory. *Math. Biosci.* 15, 253–285. doi: 10.1016/0025-5564(72)90038-7

Grossberg, S., and Mingolla, E. (1985). Neural dynamics of form perception: boundary completion, illusory figures, and neon color spreading. *Psychol. Rev.* 92, 173–211. doi: 10.1037/0033-295X.92.2.173

Grossberg, S., and Mingolla, E. (1987). "Neural dynamics of perceptual grouping: textures, boundaries, and emergent segmentations," in *The Adaptive Brain, II*, (Elsevier), 143–210.

Grossberg, S., and Todorovic, D. (1988). Neural dynamics of 1-D and 2-D brightness perception: a unified model of classical and recent phenomena. *Percept. Psychophys.* 43, 241–277. doi: 10.3758/BF03207869

Grunfeld, E. D., and Spitzer, H. (1995). Spatio-temporal model for subjective colours based on colour coded ganglion cells. *Vis. Res.* 35, 275–283. doi: 10.1016/0042-6989(94)00119-7

Hazenberg, S. J., and van Lier, R. (2013). Afterimage watercolors: an exploration of contour-based afterimage filling-in. *Front. Psychol.* 4:707. doi: 10.3389/fpsyg.2013.00707

Heitger, F., von der Heydt, R., Peterhans, E., Rosenthaler, L., and Kübler, O. (1998). Simulation of neural contour mechanisms: representing anomalous contours. *Image Vis. Comput.* 16, 407–421. doi: 10.1016/S0262-8856(97)00083-8

Kandel, E. R., Schwartz, J. H., Jessell, T. M., Siegelbaum, S. A., and Hudspeth, A. J. (2012). *Principles of Neural Science, 5th Edn.* New York, NY: McGraw-Hill Education/Medical.

Kanizsa, G. (1976). Subjective contours. *Sci. Am.* 234, 48–52. doi: 10.1038/scientificamerican0476-48

Kim, J., and Francis, G. (2011). Color selection, color capture, and afterimage filling-in. *J. Vis.* 11, 23–23. doi: 10.1167/11.3.23

Komatsu, H. (2006). The neural mechanisms of perceptual filling-in. *Nat. Rev. Neurosci.* 7, 220–231. doi: 10.1038/nrn1869

MacKay, D. M. (1957). Moving visual images produced by regular stationary patterns. *Nature* 180, 849–850. doi: 10.1038/180849a0

Macknik, S. L., and Martinez-Conde, S. (2010). The Neuroscience of Yoricks's Ghost and Other Afterimages. *Sci. Am.* 20, 12–15.

Marr, D. (1982). *Vision: A Computational Approach.* New York, NY: Freeman.[aAC].

On, Z. X., and van Boxtel, J. J. (2017). The role of transparency cues in afterimage color perception. *Sci. Rep.* 7:9183. doi: 10.1038/s41598-017-09186-1

Pérez, P., Gangnet, M., and Blake, A., (2003). "Poisson image editing," in *ACM SIGGRAPH 2003 Papers* (New York, NY: ACM), 313–318.

Ron, E., and Spitzer, H., (2011). Is the Kanizsa illusion triggered by the simultaneous contrast mechanism? *J. Opt. Soc. Am. A Opt. Image Sci. Vis.* 28, 2629–2641. doi: 10.1364/JOSAA.28.002629

Sande, K., van de, Gevers, T., and Snoek, C. (2010). Evaluating color descriptors for object and scene recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* 32, 1582–1596. doi: 10.1109/TPAMI.2009.154

Shapley, R., and Hawken, M. (2011). Color in the cortex—single- and double-opponent cells. *Vis. Res.* 51, 701–717. doi: 10.1016/j.visres.2011.02.012

Shevell, S. K. (2003). *The Science of Color.* New York, NY: Elsevier.

Shimojo, S., Kamitani, Y., and Nishida, S. (2001). Afterimage of perceptually filled-in surface. *Science* 293, 1677–1680. doi: 10.1126/science.1060161

Simchony, T., Chellappa, R., and Shao, M. (1990). Direct analytical methods for solving poisson equations in computer vision problems. *IEEE Trans. Pattern Anal. Mach. Intell.* 12, 435–446. doi: 10.1109/34.55103

Spitzer, H., Almon, M., and Sandler, V. M. (1993). A model for detection of spatial and temporal edges by a single X cell. *Vis. Res.* 33, 1871–1880. doi: 10.1016/0042-6989(93)90178-Y

Spitzer, H., and Barkan, Y. (2005). Computational adaptation model and its predictions for color induction of first and second orders. *Vis. Res.* 45, 3323–3342. doi: 10.1016/j.visres.2005.08.002

Spitzer, H., and Semo, S. (2002). Color constancy: a biological model and its application for still and video images. *Pattern Recognit.* 35, 1645–1659. doi: 10.1016/S0031-3203(01)00160-1

Van Horn, D. R., and Francis, G. (2008). Orientation tuning of a two-stimulus afterimage: implications for theories of filling-in. *Adv. Cogn. Psychol.* 3, 375–387. doi: 10.2478/v10053-008-0002-7

van Lier, R., Vergeer, M., and Anstis, S. (2009). Filling-in afterimage colors between the lines. *Curr. Biol.* 19, R323–R324. doi: 10.1016/j.cub.2009.03.010

Van Tuijl, H. (1975). A new visual illusion: neonlike color spreading and complementary color induction between subjective contours. *Acta Psychol.* 39, 441–445. doi: 10.1016/0001-6918(75)90042-6

Vidyasagar, T. R., Buzás, P., Kisvárday, Z. F., and Eysel, U. T. (1999). Release from inhibition reveals the visual past. *Nature* 399:422. doi: 10.1038/20836

von der Heydt, R., Friedman, H. S., and Hong, Z. (2003). "Searching for the neural mechanisms of color filling-in," in *Filling-In : From Perceptual Completion to Cortical Reorganization: From Perceptual Completion to Cortical Reorganization,* eds L. Pessoa, and P. Weerd (Oxford, UK: Oxford University Press), 106–127.

Webster, M. A. (2015). Visual adaptation. *Annu. Rev. Vis. Sci.* 1, 547–567. doi: 10.1146/annurev-vision-082114-035509

Wede, J., and Francis, G. (2007). Attentional effects on afterimages: theory and data. *Vision Res.* 47, 2249–2258. doi: 10.1016/j.visres.2007.04.024

Wede, J., and Francis, G., (2006). The time course of visual afterimages: data and theory. *Perception* 35, 1155–1170. doi: 10.1068/p5521

Weickert, J. (1998). *Anisotropic Diffusion in Image Processing.* Stuttgart: Teubner.

Williams, D. R., and Macleod, D. I. (1979). Interchangeable backgrounds for cone afterimages. *Vis. Res.* 19, 867–877. doi: 10.1016/0042-6989(79)90020-8

Wyszecki, G., (1986). "Color appearance," in *Handbook of Perception and Human Performance*, eds K. R. Boff, L. Kaufman, and J. P. Thomas (New York, NY: John Wiley & Sons), 29–30.

Zaidi, Q., Ennis, R., Cao, D., and Lee, B. (2012). Neural locus of color afterimages. *Curr. Biol.*. 22, 220–224. doi: 10.1016/j.cub.2011.12.021

Zeki, S., Cheadle, S., Pepper, J., and Mylonas, D. (2017). The constancy of colored after-images. *Front. Hum. Neurosci.* 11:229. doi: 10.3389/fnhum.2017.00229

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Computational Neuroscience

ORIGINAL RESEARCH
published: 18 January 2019
doi: 10.3389/fncom.2018.00106

Predicting Illusory Contours Without Extracting Special Image Features

Albert Yankelovich[1]* *and Hedva Spitzer*[2]

[1] Department of Biomedical Engineering, Faculty of Engineering, Tel Aviv University, Tel Aviv, Israel, [2] Faculty of Engineering, School of Electrical Engineering, Tel Aviv University, Tel Aviv, Israel

Boundary completion is one of the desired properties of a robust object boundary detection model, since in real-word images the object boundaries are commonly not fully and clearly seen. An extreme example of boundary completion occurs in images with illusory contours, where the visual system completes boundaries in locations without intensity gradient. Most illusory contour models extract special image features, such as L and T junctions, while the task is known to be a difficult issue in real-world images. The proposed model uses a functional optimization approach, in which a cost value is assigned to any boundary arrangement to find the arrangement with minimal cost. The functional accounts for basic object properties, such as alignment with the image, object boundary continuity, and boundary simplicity. The encoding of these properties in the functional does not require special features extraction, since the alignment with the image only requires extraction of the image edges. The boundary arrangement is represented by a border ownership map, holding object boundary segments in discrete locations and directions. The model finds multiple possible image interpretations, which are ranked according to the probability that they are supposed to be perceived. This is achieved by using a novel approach to represent the different image interpretations by multiple functional local minima. The model is successfully applied to objects with real and illusory contours. In the case of Kanizsa illusion the model predicts both illusory and real (pacman) image interpretations. The model is a proof of concept and is currently restricted to synthetic gray-scale images with solid regions.

Keywords: figure ground segregation, illusory contours, functional minimization, multiple perceptions, computational Gestalt

OPEN ACCESS

Edited by:
Jonathan D. Victor,
Weill Cornell Medicine, Cornell University, United States

Reviewed by:
Mikhail Katkov,
Weizmann Institute of Science, Israel
Leila Montaser-Kouhsari,
Columbia University, United States

***Correspondence:**
Albert Yankelovich
alberovich@gmail.com

Received: *14 July 2018*
Accepted: *13 December 2018*
Published: *18 January 2019*

Citation:
Yankelovich A and Spitzer H (2019) Predicting Illusory Contours Without Extracting Special Image Features. Front. Comput. Neurosci. 12:106. doi: 10.3389/fncom.2018.00106

INTRODUCTION

An important and non-trivial task in process of image understanding is the detection of *object boundaries*, also termed *figure-ground segregation* or *image segmentation*. This task is especially difficult in conditions where the object boundary is not fully visible. The human visual system, in many cases, is able to construct the whole object boundary (Kanizsa, 1955). An extreme example of such a completion is demonstrated by illusory contours (**Figures 1A,B**), where the visual system "creates" object boundaries in locations without any intensity gradient (Schumann, 1900; Ehrenstein, 1925; Kanizsa, 1955; Gregory, 1972; Kennedy and Lee, 1976; Day and Jory, 1980; Prazdny, 1983; Bradley, 1987; Kennedy, 1988).

While numerous models for performing image segmentation have been reported (Leclerc, 1989; Nitzberg and Mumford, 1990; Pal and Pal, 1993), relatively few are designed to incorporate illusory contours. Most of the models are capable of generating illusory contours by

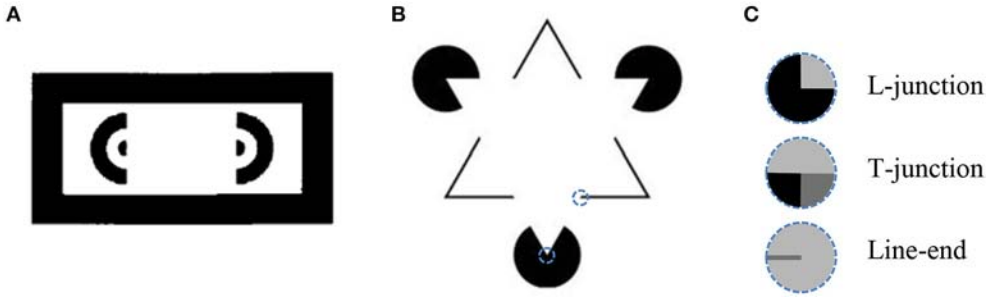


FIGURE 1 | (A) An illusory rectangle is perceived between the halves of the donut (Schumann, 1900). **(B)** An illusory triangle that occludes three solid circles is perceived (Kanizsa, 1955). **(C)** Some special image features that are found at specific points in input image and are used for generating illusory contours, see text. The dashed line circles represent a piece of image and the special feature location is at the center of the circle. Example of L-junction and line-end are shown in Kanizsa triangle B. The features location is indicated by dash line circle.

extracting special image features, such as L-junctions, T-junctions, and line-ends (**Figure 1C**), and using them as key-points to create the illusory contours (Finkel and Edelman, 1989; Guy and Medioni, 1993; Williams and Hanson, 1994; Gove et al., 1995; Williams and Jacobs, 1995; review: Lesher, 1995; Kumaran et al., 1996; Heitger et al., 1998; Kogo et al., 2002; Ron and Spitzer, 2011). This approach is supported by psychophysical evidence that the existence of special image features are required for illusory contours to emerge (Rubin, 2001). Many of these models exploit neurophysiological knowledge about neuronal mechanisms of the visual system. For example, in the model of Heitger et al. (1998), the responses of *end stopped cells* that detect L-junctions and line-ends are grouped and added to the responses of *simple cells,* which detect image edges (image intensity gradient) to produce the illusory contour.

The special features extraction is a difficult task in real world images, since in order to decide which junctions are significant relative to others, the structure of the scene in the image needs to be understood (Nitzberg and Mumford, 1990). In addition, the fact that only a small fraction of the image is exploited for special feature extraction (image region around the special feature point) makes this approach less robust.

A widely accepted explanation of illusory contours is the perception of relative depth, where the illusory contour represents the boundary of an object located at an other depth than the region around it (Kanizsa, 1955; Coren, 1972; Gregory, 1972; Lesher, 1995). According to this point of view, the illusory contours are just regular object boundaries, with the object intensity being the same as that of the background. The object with the illusory contour is revealed by the objects that are being occluded behind it, as in **Figure 1B**. The special image features, such as L-junctions and line ends, can provide a clue for object occlusion. Extracting special features, however, means making a specific effort for illusory contours detection. In this case the illusory contours are not treated as the regular contours. We prefer not to extract special features and to use instead a common way to detect both real and illusory object boundaries. Detection of illusory contours without using special image features is very challenging, since it requires the prediction of contours *ex nihilo*, without using the *occlusion clues*.

An approach that has the potential of not extracting special image features is the functional optimization, used by some boundary detection models capable of generating illusory contours (Kass et al., 1988; Madarasmi et al., 1994; Williams and Hanson, 1994; Geiger et al., 1996; Saund, 1999; Gao et al., 2007). The functional is used to give a score for each contour configuration, and the final contours are not "constructed" by the model, but rather "come out" as the minimizer of the functional. Special features extraction is not necessarily required in these models, since the demand that the resultant boundaries will match the input image can be expressed in the functional without the special features extraction. An additional significant advantage of functional optimization approach is that giving a preference score to a given contour configuration is much simpler than constructing the correct contour configuration. The optimization approach is a computational realization of the Gestalt psychology (Koffka, 1935), since it derives the contours from some contour configuration preference rules ("grouping rules" in Gestalt psychology). By this it accounts for both real and illusory contours based on a general unified approach.

Kass et al. (1988) applied *snakes* algorithm of energy minimizing splines to track image edges. The continuity and elasticity properties of the snakes enable the illusory contours to emerge. This model indeed does not extract special image features, however, it is not fully automatic, since user interaction is required to draw the initial contour. One might argue that some automatic initial contours such as small circles matrix can be used, however in this case illusory contours will be extracted even for images that actually lack them. For example, the model will predict illusory contours for a Kanizsa illusion configuration with solid circle inducing elements, although in this case the illusory contour is not perceived. Currently there is no fully automatic boundary detection model that does not require special features extraction for illusory contours generation.

The proposed model is a proof of concept and is restricted to gray scale images with solid non-textured regions and without lines. The stress in the model is not on the way used to encode the Gestalt rules, nor on the rules themselves, but on the mere

possibility by predicting real and illusory boundaries based solely on general boundary formation rules.

METHODS

Model Rational

The basic idea of the model was inspired by the assumption that object detection is one of the intelligent tasks performed by the visual system. This task uses a set of simple assumptions, based on our natural perception of an object's appearance, to provide the most reasonable "explanation" of what is presented in the image. With several possible perceptions of what we see, a critical question is what makes us prefer one perception over another? Especially we are interested to reveal the reason for perception of illusory contours. As an example, let us consider the Kanizsa triangle (Kanizsa, 1955) in **Figure 2** and examine the factors responsible for the perception of an illusory contour in this case.

The perception in **Figure 2A** is that of three "pacman" objects and the perception in **Figure 2B** is of a triangular object above three circular objects. In the pacman perception the boundary of each pacman has three corners (or bends)–two convex and one concave. On the other hand, in the triangle perception instead of three bends per pacman there is only one, since the circle is perceived as continuing under the triangle. Moreover, the concave bend in the circle center is replaced by a convex bend of the triangle vertex. The conclusion is that in the illusory interpretation the object's boundary is less bent and the bends are more convex. Both criteria can be derived from preference of simplest description (van Tuijl, 1975). The preference for convex bends also explains why in the image containing a square (**Figure 7D**), we perceive a square object more readily than a square hole.

Although the functional optimization approach enables us to avoid special features extraction, it has the drawback of having a tremendous search space of the possible solutions. To overcome this issue, we use an "economic" boundary representation called a *border ownership map*, holding boundary segments in discrete locations and discrete directions. Our representation is inspired by the neural findings of Zhou et al. (2000) who discovered V1 visual cortical cells that respond to an edge only when the object is located on one of the edge sides. This ability was already termed *border ownership* by Nakayama and Shimojo (1990). Using the border ownership map makes the free variable of the problem much smaller than using, for example, contour parametrization.

An additional difficulty is that the functional that accounts for several object boundary properties and depends on many variables has a large number of local minima. To overcome this, the functional was smoothed and the functional minimizers were found by gradual relaxation technique (Lee, 1995). This reduces the number of minima by smoothing out the shallow minima and finding only prominent stable minima.

In the proposed model we define a functional that accounts for basic object properties, such as boundary continuity and convexity, and demands the object boundaries to match the input image. The object boundaries are found as the minimizer of the functional. The illusory contours are predicted in same way as the real contours, by being the most probable object boundaries matching the input image. This is the first time that the perception of illusory contours from a general object boundary detection task is shown computationally.

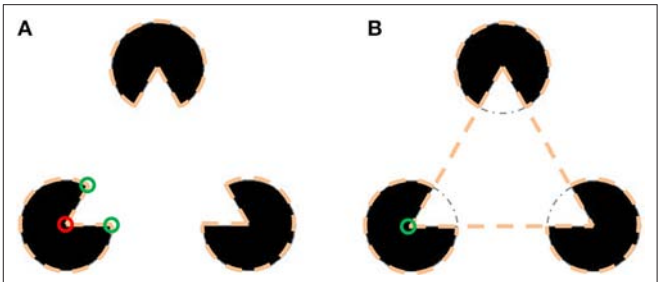


FIGURE 2 | Illustration of two interpretations of the Kanizsa triangle. **(A)** Three objects having shape of a pacman. The dashed lines represent the perceived objects. Boundary bends are marked by small circles, red for concave and green for convex. Only one of the pacmans is marked to avoid burden. **(B)** Illusory triangle occluding three circular objects. The circular objects are perceived as being occluded by the triangular object. The dash-dot arc marks the circular object part behind the triangular object.

The minimizers of the functional are compared to the expected perception, known from psychophysical evidence. Due to the suggestion that the visual system is actually finding the best solution to object formation rules, we are not necessarily obliged to use the mechanisms of the visual system (which are also not fully known), to find that solution. It has to be noted that in spite of this the model exploits some of the physiological knowledge of low-level mechanisms of the visual system, such as simplification of visual cell receptive fields that perform edge detection [section Border ownership at image edges (F^{A})], logical "and" operation (Appendix section 1.2) and cell response grouping (Appendix section 1.1). In addition, the model includes the crucial component of the border ownership map, section Boundary Representation.

Using functional minimization in the model has an additional important benefit. Usually, there are several possible object configurations that can explain a single image (**Figure 2**). Multiple image interpretations are present even in a simplest image of a white square on black background (**Figure 7D**). This image can be interpreted as a white square object over a black background, or as a black frame with a square hole through which a white background is seen. The illusory Kanizsa triangle (**Figure 1B**), also has several possible interpretations. The most prominent is the illusory interpretation of a white triangle occluding three black solid circles and a black boundary triangle (Ringach and Shapley, 1996). An additional easily perceived interpretation does not include an illusory triangle, but consists of three cut-out circles, "pacmans", and three V-shaped figures. For real-world images there may be numerous plausible configurations of objects. The desired interpretation may be chosen, for example, by applying a higher level knowledge, like object recognition. The ability to predict multiple possible perceptions of the image is therefore a desired property of a robust boundary detection model. The multiple possible image interpretations, that are described above, are represented in our model by multiple minima of the functional.

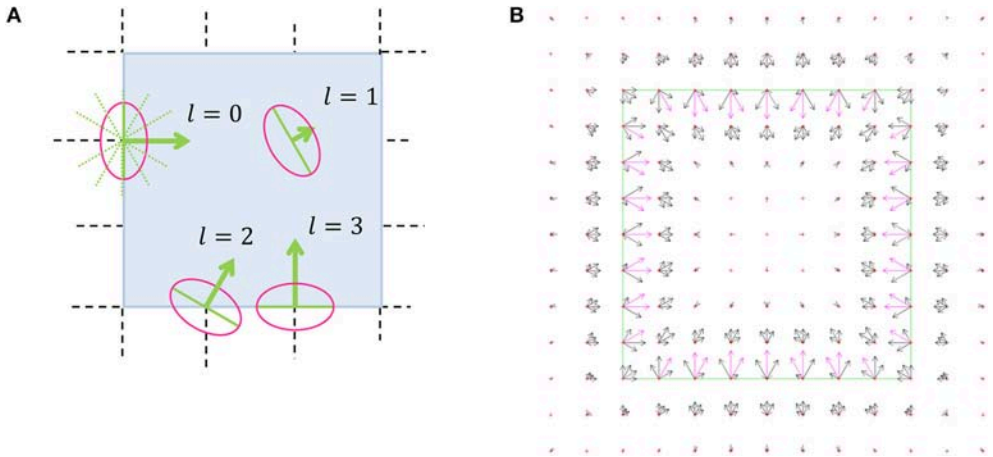


FIGURE 3 | Border ownership map illustration. **(A)** Schematic illustration. The bold arrows are the border ownership vectors, while the vector length indicates the edge strength and the vector direction indicates the edge direction. The length of the ownership vector represents the probability that there is an object edge that passes through the vector origin in orientation perpendicular to the vector. The object is located on the side pointed by the border ownership vector. *l* is the discrete direction index, ranging from 0 to $(L-1)$. Here and in all other border ownership maps $L = 12$. At each coordinate there is a border ownership vector for all possible directions, although for sake of clarity only some of the vectors are shown here. The dashed lines represent the discrete grid over the image of a square (the dark area). The dotted diagonal lines going out of the origin of $l = 0$ vector which show all the possible discrete directions. The ellipse around the vector origin illustrates the area in which the object edge is represented by the border ownership vector that is relevant. This resembles the receptive field of a V1 neuron (Hubel and Wiesel, 1959). **(B)** The output of border ownership map of the model for an input of a square object image. The border ownership vectors point inside the square (solid line). The vector with the greatest length at a point is directed perpendicular to the square edge.

Model Overview

The model consists of four main parts:

1. Encoding of object boundaries.
2. The cost functional, specifying a cost value for each object boundaries configuration.
3. A method to identify object boundaries with minimal cost.
4. A method of finding multiple functional minima, corresponding to different image perceptions.

The main challenge of identifying illusory contours as a solution of a minimization problem is occupying the huge size of the solutions space. We attacked this problem by choosing a compact boundaries representation method and by applying various types of smoothing to the functional, in order to reduce the number of local minima. The smoothing leaves only the stable minima. A method was invented to find different local minima of the cost functional, section Finding Multiple Local Minima. Each local minimum corresponds to a possible image interpretation, with a lower cost for a more probable (pop-out) interpretation.

The variables notation below is that the subscript of a variable describes the discrete coordinate on which this variable is measured. For example, f_{xy} is a filter intensity at coordinate (x, y), for integer x and y. There are no continuous coordinates in the model. We omit the comma between the coordinates for brevity. The superscript of a variable is part of the variable name. For example, σ^X is a constant. In the following we describe the model parts in more detail.

Boundary Representation

The *border ownership map* (**Figure 3A**), represents the probability that an object edge passes through a discrete coordinate in some discrete direction. The orientation of the object edge is perpendicular to the discrete direction, and the object resides on the side that is pointed by the pointed direction. As an example, **Figure 3B** represents the border ownership map of a square object. At each discrete coordinate, the border ownership is specified for a discrete set of equally distributed L orientations (**Figure 3A**). Note that for opposite directions there are two different border ownership values. The border ownership is not strictly a probability value. Only the relative values of border ownership are important. We choose to interpret positive and negative values of border ownership in the same way, since in the minimization process additional effort is required to avoid negative values. To achieve this interpretation, the border ownership always appears squared in the functional.

Cost Functional

The functional that depends on the border ownership map is designed to measure to what extent the expected properties of the object boundaries configuration are followed. Each property is allocated a *cost functional component* and the overall cost functional is a weighed sum of all the components.

$$F\left(\vec{b}\right) = \alpha^A F^A\left(\vec{b}\right) + \alpha^R F^R\left(\vec{b}\right) + \alpha^V F^V\left(\vec{b}\right) + \alpha^N F^N\left(\vec{b}\right) + \alpha^C F^C\left(\vec{b}\right) + \alpha^E F^E\left(\vec{b}\right) \quad (1)$$

Where F^{type} are the cost functional components that are dependent on the border ownership map

$$\vec{b} = \left\{b_{xyl}\right\}_{x,y,l} \quad (2)$$

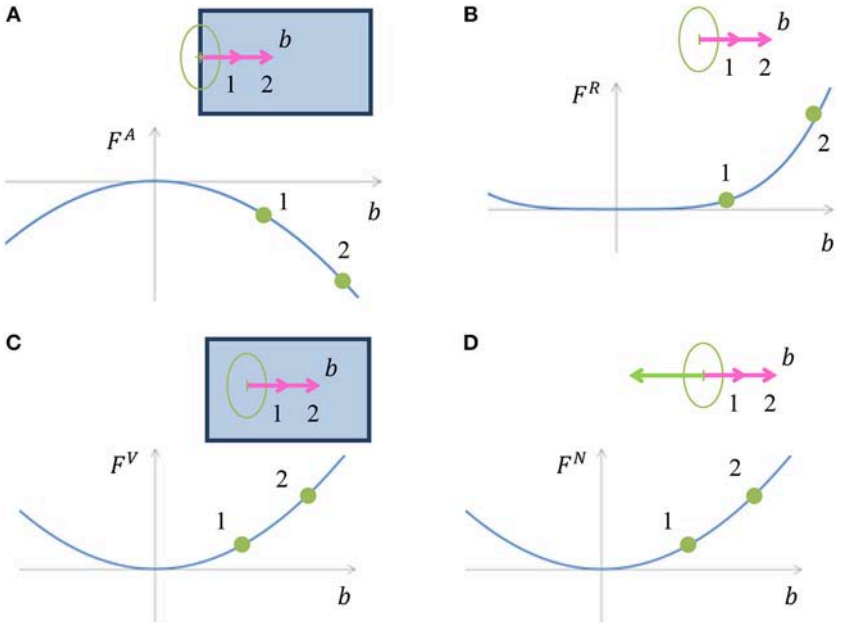


FIGURE 4 | (A) Illustration of cost component F^A, which is inducing border ownership in a direction perpendicular to the image edge. When a border ownership is of length denoted by 1 (pink arrow) in the image, the cost is as pointed in point 1 on the chart. For bigger border ownership denoted by 2 (pink arrow) in the image, the cost is as pointed in point 2, and is lower than the cost at point 1. **(B)** Illustration of border ownership limitation cost component F^R. The cost increases with increasing the vector value of the border ownership in order to limit the infinite growth of the vector value, due to cost component F^A. The polynomial degree of F^A in **(A)** is 2, while the polynomial degree of F^R is 4, which makes sure that the border ownership value will be limited. **(C)** Illustration of cost component F^V, which gives penalty to border ownership in places with no edge in the image. The cost increases with increasing border ownership at a location with no edge in the image. **(D)** Cost component F^N discourages border ownership in opposite directions, since an object is expected to be only on one side of the edge.

and α^{type} are weight parameters. x, y are discrete coordinates and l is the discrete direction index. The first three components F^A, F^R and F^V are responsible for appearance of border ownership at image edges. The other components are responsible for encoding the expected object boundary properties, and therefore depend only on the border ownership map and not on the input image. The component F^N is designed to make sure that the object is located only on one side of a boundary. F^C is responsible for object boundary continuity. F^E gives penalty for bending in the object boundary, while concave bends receive a greater penalty, section Model Rational. The cost components are visualized in **Figures 4**, **5** and are described in the following paragraph. Since the full definition of the components F^C and F^E is more complicated and occupy larger volume, their details are provided in **Appendix** in Supplementary Material.

Border Ownership at Image Edges (F^A)

This chapter describes how border ownership is induced from image edges. In the case of an intensity edge in the input image with a specific orientation, the border ownership in the perpendicular direction is encouraged. Since we do not know on which side of the edge the object is situated, the border ownerships are encouraged in both directions which are perpendicular to the edge. The cost component sums multiplication of the border ownership b_{xyl} by the intensity of edge in the image in an orientation perpendicular to l, termed A_{xyl}. This "encourages" border ownership perpendicular to the edge in input image (**Figure 4A**).

$$F^A = \frac{1}{T}\sum_{x,y,l} -{A_{xyl}}^2 {b_{xyl}}^2 \tag{3}$$

where

$$A_{xyl} = I_{xy} * f^A_{xyl} \tag{4}$$

The operation marked by * is a discrete cross-correlation (or filtering), given by:

$$I_{xy} * f^A_{xyl} = \sum_{x',y'} I_{(x+x')(y+y')} f^A_{x'y'l} \tag{5}$$

The filter f^A_{xyl} detects an image edge at point (x, y) and orientation perpendicular to l. It is defined by rotation of function f^A_{xy} by $2\pi\frac{l}{L}$.

$$f^A_{xy} = \frac{1}{2\pi{\sigma^A}^2} s(x)\, e^{-\frac{x^2+y^2}{2{\sigma^A}^2}} \tag{6}$$

where function $s(x)$ is a sign function, giving zero for values close to zero

$$s(x) = \begin{cases} 0, & |x| \le 0.001 \\ \frac{x}{|x|}, & else \end{cases} \tag{7}$$

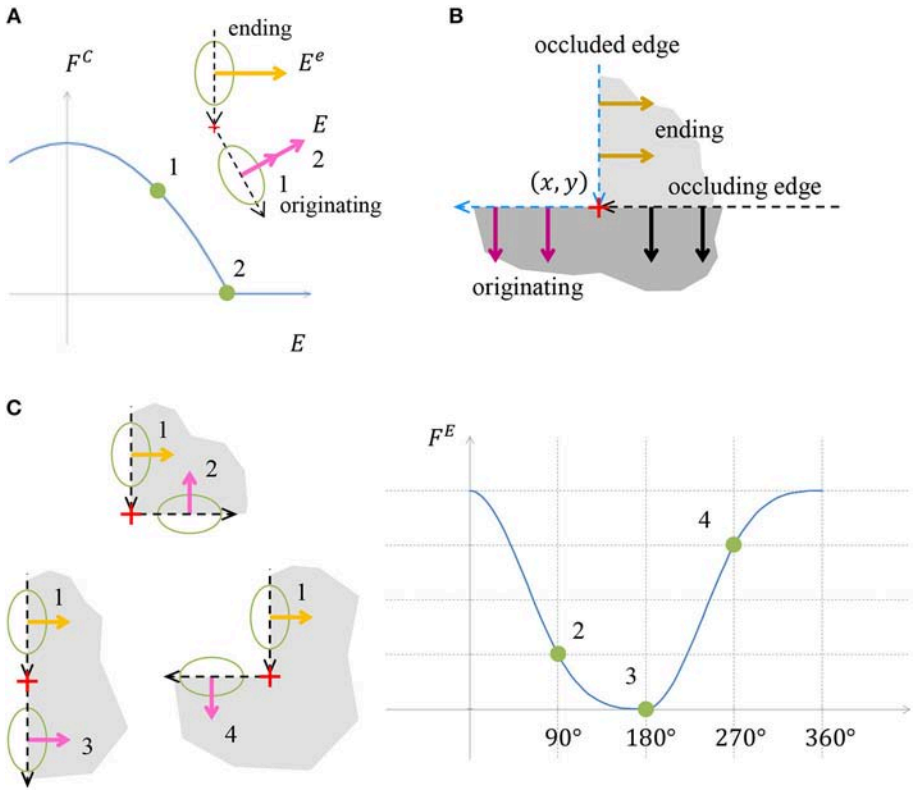


FIGURE 5 | (A) Illustration of object edge continuity component F^C. The value E is the strength of the object boundary edge originating from a specific location point. The chart presents the cost component value as a function of the originating edge strength E. The originating edge strength 1 is less than the strength of the ending edge, E^e, hence a positive cost is assigned. Edge strength 2 is the same as the ending edge strength, thus the cost is zero. **(B)** Illustration of how the continuity is preserved in case of object boundary occlusion by additional object. The vertical edge is occluded below by an object with a horizontal edge. The occluding edge serves as the originating edge of the occluded ending edge. In this case no discontinuity is indicated by the continuity cost component $\boldsymbol{F^C}$. **(C)** Illustration of the cost component accounting for object edge bending, F^E. The object edge defined by vector marked 1 can continue by one of the edges marked by 2, 3, or 4. The costs for these three continuations are depicted in the chart. Note that the contribution of the convex continuation 2 is smaller than of the concave continuation 4, although both deviate by 90° from the straight continuation 3. The contribution of the straight continuation 3 is zero, since there is no boundary bending in this case.

and σ^A is a constant. The constant T is used to normalize the cost to be per coordinate and orientation and is given by:

$$T = I^X I^Y L \tag{8}$$

where I^X and I^Y are the width and height of the input image. The border ownership value b_{xyl} in (3) is squared in order to have same cost for positive and negative values of border ownership, section Boundary Representation.

Border Ownership Is Limited (F^R)

If F^A was the only component of the functional, the border ownership at image edges would grow infinitely to make the cost lower. The following cost component is added to ensure that the value of border ownership is limited:

$$F^R = \frac{1}{T}\sum_{x,y,l} {b_{xyl}}^4 \tag{9}$$

The reason for taking the border ownership to power 4 is to make F^R stronger than F^A at high border ownership values. The cost component F^R is illustrated in **Figure 4B**.

Suppress Border Ownership in the Absence of Image Edge (F^V)

An illusory contour introduces border ownership also at places with no intensity gradient in the image. To avoid spurious illusory contours, this component adds a penalty for boundary ownership in places with no edge in the input image (**Figure 4C**).

$$F^V = \frac{1}{T}\sum_{x,y,l} \frac{\varepsilon^V}{{A_{xyl}}^2 + \varepsilon^V} {b_{xyl}}^2 \tag{10}$$

where ε^V is a small constant and A_{xyl} is intensity of edge in the image (4), used in component F^A. Note that the equation and the rational of F^A (3) and F^V are similar, but have opposite trends, such that a large edge leads to lower cost, while a small edge causes a higher cost. The only functional components that depend on input image are F^A and F^V. They depend only on

image edges and not on special image features, as required in previous models, section Introduction.

Object on One Side (F^N)

The model assumes that the object usually resides on only one side of an edge. Hence, if there is border ownership in a specific direction, the border ownership in the opposite direction is discouraged (**Figure 4D**). If there is a significant border ownership in direction l, border ownership in opposite direction $l+\frac{L}{2}$ is not expected, section Boundary Representation. Border ownership is also not expected in directions close to $l+\frac{L}{2}$, therefore, we add a cost for border ownership vectors with deviation m from $l+\frac{L}{2}$. We also consider border ownership in spatial vicinity to the border ownership vector origin (x,y) by filtering the border ownership map in space. The filtered border ownership map is termed B^N_{xyl}.

$$F^N = \frac{1}{TT^N}\sum_{x,y,l}\sum_{m=-\left(\frac{L}{4}-1\right)}^{\frac{L}{4}-1}\cos^2\left(2\pi\frac{m}{L}\right)B^N_{xyl}B^N_{xy\left(l+\frac{L}{2}+m\right)} \quad (11)$$

where

$$B^N_{xyl} = {b_{xyl}}^2 * f^N_{xy} \quad (12)$$

and

$$f^N_{xy} = \frac{1}{2\pi{\sigma^N}^2}e^{-\frac{x^2+y^2}{2{\sigma^N}^2}} \quad (13)$$

For a larger deviation m, the cost increase should be smaller, thus a weight factor $\cos^2\left(2\pi\frac{m}{L}\right)$ is added accordingly. The maximum deviation considered is $\frac{L}{4}-1$, since this is the maximum angle which is less than $\frac{\pi}{2}$. The term T^N (11) is used to normalize the contributions from all deviations and is given by

$$T^N = \sum_{m=-\left(\frac{L}{4}-1\right)}^{\frac{L}{4}-1}\cos^2\left(2\pi\frac{m}{L}\right) \quad (14)$$

Object Boundary Continuity (F^C)

One of the basic properties of an object is the continuity of its boundary, thus the boundary is not expected to end abruptly, unless it is occluded by the boundary of another object. To encourage object boundary continuity, we require that when an object edge ends at a coordinate, there should be an object edge originating from the same coordinate (**Figure 5A**). The occluding object edge plays the role of the originating edge to the occluded object ending edge, in case of occlusion (**Figure 5B**). The main innovation of the model is the mere possibility to predict illusory contours without special features extraction, following the functional optimization approach. Since the full details of this component are quite lengthy and the exact functional definition is not the main aim of the model, this component details are provided in Appendix section 1.1.

Object Boundary Bending (F^E)

We concluded in section Model Rational that the preferred perception is the one with fewer bends, and if there are bends, then convex bends are preferable. Taking this preference into account, we will assign a positive cost for bends in the object boundary, with an increased penalty for concave bends (**Figure 5C**). The details of this component are also lengthy, hence they are provided in Appendix section 1.2.

Cost Functional Smoothing

The cost functional (1), accounting for several object boundary properties and depending on many variables, has a large number of local minima, while not all of them represent expected image interpretations. The problem is then how to "get rid" of these redundant local minima. We assume that the redundant local minima are shallower than desirable ones. To avoid trapping in shallow local minima, four types of smoothing methods are applied, as described in the following sections.

Border Ownership Map Smoothing in Angle and Space

To make the cost functional less sensitive to small changes in border ownership, the border ownership map $\vec{b}$ is smoothed in angle and space. The result $\vec{b}^S$ is used as input to the cost functional (1).

$$b^S_{xyl} = \left[\sum_{j=-\left(\frac{L}{2}-1\right)}^{\frac{L}{2}} b_{xy(l+j)}f^{SA}_j\right] * f^{SX}_{xy} \quad (15)$$

where f^{SA}_j and f^{SX}_{xy} are Gaussians in angle (A) and space (X) coordinates, respectively:

$$f^{SA}_j = \frac{1}{\beta^{SA}}e^{-\frac{j^2}{2{\sigma^{SA}}^2}} \quad (16)$$

$$f^{SX}_{xy} = \frac{1}{2\pi{\sigma^{SX}}^2}e^{-\frac{x^2+y^2}{2{\sigma^{SX}}^2}} \quad (17)$$

with σ^{SA} and σ^{SX} constants, and β^{SA} is a normalization constant:

$$\beta^{SA} = \sum_{m=-\left(\frac{L}{2}-1\right)}^{\frac{L}{2}} e^{-\frac{m^2}{2{\sigma^{SA}}^2}} \quad (18)$$

Spatial Filters Smoothing

The cost functional calculation uses various spatial filters. To make the cost smoother and less dependent on the discrete grid step, we sum up the cost components on multiple spatial scales.

$$G^{type} = \frac{1}{N}\sum_{n=0}^{N-1}F^{type}_n \quad (19)$$

where N is the number of scales and F^{type}_n is the same as F^{type} (1), except that it uses spatial filters derived by scaling the original filters by factor

$$\mu^n \quad (20)$$

where $\mu > 1$ is a scaling constant. The smoothed components G^{type} (17) are used in the functional instead of the components F^{type} (1).

Ramp Function Smoothing

The ramp function

$$r(x) = \begin{cases} 0, & x \leq 0 \\ x, & x > 0 \end{cases} \tag{21}$$

is used in components F^C and F^E to account for positive and not negative values. There are two benefits in smoothing the ramp function $r(x)$. The first is that the smoothed function is differentiable at $x = 0$ and the second is that the cost functional also becomes smoother, which reduces the number of local minima. The smoothed function is obtained by filtering $r(x)$ through a Gaussian function:

$$\frac{1}{\sqrt{2\pi\sigma^{RP2}}} e^{-\frac{x^2}{2\sigma^{RP2}}} \tag{22}$$

Where σ^{RP} is a constant.

Gradual Relaxation-Find the Minimum at Coarse to Fine Scale

In order to avoid trapping into shallow local minima, the minimum is found first on a coarse and then at a finer scale, a method called gradual relaxation (Lee, 1995). This is done by first finding the minimum of the functional on a broad scale. Then, the border ownership found is used as the initial point for finding the minimum on a finer scale. This process is repeated until the desired detailed scale is reached. The details of this process are as follows. A *scale parameter s* is initially set to $s^0 > 0$. To proceed to a more detailed scale, the scale parameter s is multiplied by constant s^R with $0 < s^R < 1$. The process is finished when the desired resolution of s^M is reached. For the scale s^M the smoothed functional is close to the functional without smoothing. The scale parameter s influences the model as follows.

The border ownership smoothing scale σ^{SX} (17) is multiplied by:

$$s^{B0} + s^{BS}s \tag{23}$$

where s^{B0} and s^{BS} are constants. The scale μ^n (20) of spatial filters smoothing, is multiplied by:

$$s^{X0} + s^{XS}s \tag{24}$$

where s^{X0} and s^{XS} are constants. The width of Gaussian (22) used for the ramp function smoothing is multiplied by:

$$s^{R0} + s^{RS}s \tag{25}$$

where s^{R0}, s^{RS} are constants.

Finding the Local Minimum

The search for a minimum starts from a random border ownership map $\vec{b}^{\,R}$, with component values selected from a uniform random distribution, in the range $[0.01, 0.02]$. The reason for starting with a random border ownership rather than a zero vector is to avoid being trapped in a saddle point. For each scale parameter s, section Gradual Relaxation-Find the Minimum at Coarse to Fine Scale, the method used to search for the local minimum is a variant of a gradient descent (Curry, 1944). Suppose that at gradient descent iteration i, the current border ownership map is $\vec{b}^{\,i}$. We find the derivative of cost functional at $\vec{b}^{\,i}$ with respect to each border ownership component b_{xyl}:

$$\vec{D} = \frac{\partial F}{\partial \vec{b}}\left(\vec{b}^{\,i}\right) = \left\{\frac{\partial F}{\partial b_{xyl}}\left(\vec{b}^{\,i}\right)\right\}_{x,y,l} \tag{26}$$

$\vec{D}$ is a matrix pointing in the direction of the greatest increase of F (1). To move toward the minimum of F, we need to move in the opposite direction $-\vec{D}$. The functional F near the minimum is roughly second order, see Appendix section 1.3. Based on this, we approximate the values of F along $-\vec{D}$ by a parabola and move to its minimum. The details of this process are specified in Appendix section 1.3.

Finding Multiple Local Minima

The multiple local minima of the cost functional correspond to different interpretations of the image, section Introduction. Although there are several well established methods for finding a single minimum of a functional, there are relatively few studies on how to find multiple minima. The main question is how to escape from the first local minimum, in which the minimization process stopped. We attack this problem by positioning a "repulsive particle" in the location of the first local minimum. Here by location we mean the border ownership map of the minimum. The repulsive particle acts like an electric charge that repulses the border ownership map that is being searched and prevents it from coming too close to the repulsive particle location. This is achieved by adding to the cost functional (1) a component that increases for border ownership maps that are close to the first local minimum. This component is described in details in Appendix section 1.4, and it resembles an electric potential. The process of finding multiple local minima is performed as follows.

The gradient descent starts from some random border ownership $\vec{b}^{\,R}$, section Finding the Local Minimum, to obtain a local minimum for border ownership $\vec{b}^{\,1}$, (**Figure 6**). To find additional local minimum we place a repulsive particle at the $\vec{b}^{\,1}$ position (red $\vec{b}^{\,1}$ in **Figure 6**) and reinitiate the search for new local minimum from $\vec{b}^{\,R}$. Suppose that now the new local minimum is $\vec{b}^{\,2'}$ (magenta $\vec{b}^{\,2'}$). The repulsive particle at $\vec{b}^{\,1}$ causes $\vec{b}^{\,2'}$ to be pulled out further from

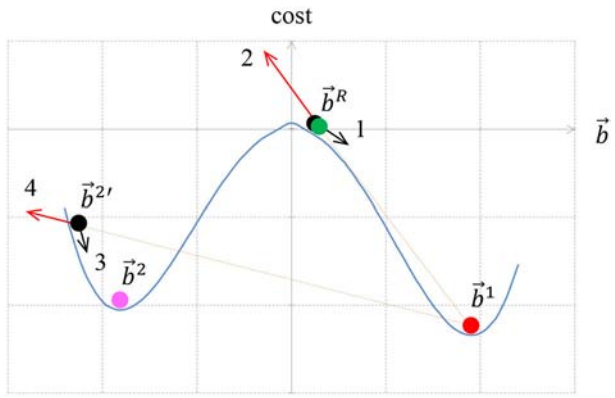


FIGURE 6 | The method for finding multiple local minima of a functional. We start from a random particle $\vec{b}^R$ (green) and reach a minimum at $\vec{b}^1$ (red). The force that is responsible for moving the particle from $\vec{b}^R$ is caused by the cost functional (arrow 1 black). The particle $\vec{b}^1$ is then replaced by an immovable repulsive particle at the same location. When the process is restarted with a new particle from $\vec{b}^R$ (black), the repulsive particle at $\vec{b}^1$ pushes the particle at $\vec{b}^R$ (red arrow 2) toward a new minimum $\vec{b}^2$. However, due to the repulsive force (red arrow 4) the particle is pushed beyond the new minimum $\vec{b}^2$, and arrives at location $\vec{b}^{2\prime}$ (black). To find a local minimum, which is not influenced by the repulsive particle, the repulsive particle at $\vec{b}^1$ is removed and a new minimum search starts from $\vec{b}^{2\prime}$. This particle is pushed to a new minimum at $\vec{b}^2$ (magenta) by the cost functional force (arrow 3 black).

the actual local minimum of the cost functional. To find the actual local minimum, we start a new search for the minimum of the functional without repulsive particle component from location $\vec{b}^{2\prime}$. Suppose the search reached the minimum for $\vec{b}^2$.

If $\vec{b}^2$ is sufficiently far from $\vec{b}^1$, then $\vec{b}^2$ is added as a new interpretation and a repulsive particle is added at $\vec{b}^2$. To measure how close $\vec{b}^1$ is to $\vec{b}^2$, the following simple distance measure is used:

$$\frac{1}{T}\sum_{x,y,l}\left|b^1_{xyl}{}^2 - b^2_{xyl}{}^2\right| \tag{27}$$

where T is defined in (8). If this distance is above a specific threshold level d^T, the particles are considered different. If $\vec{b}^2$ is close to $\vec{b}^1$ (27), then the optimization is trapped into a local minimum that has been already identified. Since the search was trapped twice in the same local minimum, we try to increase the force of the repulsive particle. This is achieved by multiplying the repulsive term by a constant factor $\tau > 1$. In order to avoid the same location $\vec{b}^{2\prime}$ again, an additional repulsive particle is added at the $\vec{b}^{2\prime}$ location, and the search for the minimum is repeated from a start point at $\vec{b}^R$ (**Figure 6**). After finding this minimum we perform a new search, but without the repulsive particle component, in order to find the actual local minimum of the original functional. If a new particle is found, then the new particle is added as additional interpretation. The repulsive force is returned to its initial strength (without multiplication by τ) and a search for a new particle is performed. If, on the other hand, no new particle is found, the repulsive force factor is multiplied again by τ. The repulsive force multiplication factor is increased until a maximum factor $\tau^{\max}$ is reached. If even for the maximum multiplication factor no new particle is found, then the process of finding multiple local minima is stopped.

Retrieving Object Shape by Contour Evolution

At this stage, the output of the model is a border ownership map (2) that assigns border ownership strength values to each discrete location and direction. To show that the actual object shape can be easily and automatically retrieved from the border ownership map, we designed a simple contour evolution algorithm that finds the top-most object in the scene. The contour evolution method finds a contour which maximizes a given functional that depends on the contour. The way to find the maximizing contour is by moving some initial contour toward the contour that brings the functional to maximum. In the level set approach, the contour is represented by the intersection of a two dimensional function ψ with x-y plane, that is by the zero-level of the function ψ. The contour motion is described and performed in terms of the function ψ. For further details see Osher and Sethian (1988).

We start with a simple small object (e.g., circular contour) which is adjacent to the border ownership vector with the biggest value. The contour representing the object boundary is then moved to maximize the border ownership vectors having direction perpendicular to the contour. Following Malladi et al. (1995), the contour dynamics is defined by:

$$\vec{C}_t = (k - \nu)\, g \vec{N} \tag{28}$$

$\vec{C}_t$ is the velocity of moving the contour $\vec{C}$. $\vec{N}$ is the contour normal vector, pointing toward the inner area of the object. The contour is moved in direction of the normal. The velocity magnitude is defined by $(k - \nu)\, g$, (28). This function is designed to cause the contour to grow until it reaches the highest value of border ownership vectors and to keep the contour as simple as possible. The term k is the contour curvature and the operation of including this term makes the contour tend to be as straight as possible. This is because a point with positive curvature, that is a convex point, the contour is "encouraged" to move inside, which decreases the curvature. For negative curvature the contour is "encouraged" to move outwards, decreasing the absolute curvature and again making the contour more straight. ν is a constant called the balloon force, giving the contour the tendency to grow. The contour friction term g causes the contour to stop when it reaches a high value of border ownership vectors in the direction perpendicular to the contour. g is a threshold of another function h:

$$g_{xy} = \begin{cases} 0, & h_{xy} < g^T \\ h_{xy}, & else \end{cases} \tag{29}$$

$$h_{xy} = \frac{1}{\left(1 + \frac{q_{xy}}{R^2}\right)} \tag{30}$$

where R is a constant and q_{xy} measures the strength of the border ownership in a direction roughly perpendicular to the contour. h_{xy} is designed such that it will be small in locations where the value of the border ownership perpendicular to the contour is high. Since h_{xy} is small in this locations, g_{xy} will be zero and the contour evolution will stop. q_{xy} is given by:

$$q_{xy} = \sum_{l=1}^{L} w_l {b_{xyl}}^2 \tag{31}$$

The weighting factor w_l measures how close the direction l is to the direction of the contour normal:

$$w_l = e^{-\frac{{\beta_l}^2}{2\sigma^{Q^2}}} \tag{32}$$

where σ^Q is a constant, and β_l is the angle between the direction of index l and the contour normal, pointing toward the inner area of the object:

$$\beta_l = \cos^{-1}\left(\overrightarrow{u}_l \cdot \overrightarrow{N}\right) \tag{33}$$

And $\overrightarrow{u}_l$ is the unit vector in direction of index l:

$$\overrightarrow{u}_l = (\cos\alpha_l, \sin\alpha_l), \quad \alpha_l = 2\pi\frac{l}{L} \tag{34}$$

Further details of the approach in field of level set curve evolution can be supplied from Osher and Sethian (1988).

RESULTS

The model was tested on various simple synthetic gray scale images with non-textured regions. The same set of model parameters were used for all tests and stimuli. The parameters were chosen by trial and error.

The first image contains two adjacent regions separated by a straight line (**Figure 7A**). Two local minima were found for this image, one corresponding to a black object on the right side over white background (**Figure 7B**), and the other one found relating to a white object on the left side over the black background (**Figure 7C**). Note that the two interpretations have equal cost −53.1, since there is no preference for the object to be on the right or on the left side.

The next tested image was a square (**Figure 7D**), also having two interpretations. The first interpretation was of a square object (**Figure 7E**), and the second interpretation was of a frame with a square hole (**Figure 7F**). The square object interpretation has cost −117, while the square hole in a frame interpretation has a higher cost −102. This is consistent with the fact that the square interpretation is perceived more readily than the square hole interpretation, section Model Rational . In all results the interpretations are presented ordered from lower to higher cost. The model behaves in the same manner for a larger square with size of 20 pixels (results are not shown). For a more complex image of an object with both convex and concave vertexes (**Figure 7G**), the model identifies two interpretations, the first corresponding to a C-shaped object (**Figure 7H**), and the second to a frame with a C-shaped hole (**Figure 7I**).

The main goal of the study was to show the possibility to detect objects with illusory contours without extracting special image features. To show this, the model was applied on Kanizsa squares with different sizes. One of the essential factors that determines the strength of the illusory contour is the ratio between the visible edge length and the total edge length, termed *support ratio* (Shipley and Kellman, 1992; **Figure 8A**). The illusory object is perceived when the support ratio values are close to 1. The model was tested on images corresponding to a broad range of support ratios. The first example is of a prominent illusory contour image (**Figure 8A**), with a relatively high support ratio of 0.67. The first interpretation, having the smallest cost −67.3, is the interpretation of an illusory square (**Figures 8B,C**). The second interpretation, having a higher cost −64.6, is of four pacemans (**Figure 9**). These two interpretations are consistent with our expectations from the model.

Additional higher cost interpretations have been found, and are not presented here. The smallest support ratio for which the illusory square is still detected for this pacman radius is 0.57. **Figure 10** shows the first interpretation for this support ratio. For a smaller support ratio of 0.53 the first interpretation is of four pacmans (the border ownership map is not shown, but has the same structure as the interpretation in **Figure 9**). For this support ratio there is no illusory interpretation at all, as expected.

To ensure that the illusory square border ownership map (**Figure 10**), can be interpreted as a square over four circles we applied a level set optimization method to extract the nearest object, section Retrieving object shape by contour evolution. The result of object extraction is shown in **Figure 11**. It shows detection of the square object with a partially illusory boundary.

DISCUSSION

The proposed model successfully extracts both real and illusory contours in various synthetic images (**Figures 7–10**). The model is generic and was not specifically designed to detect illusory contours, while special image features are not extracted. The illusory contour detection was achieved by introducing only simple desired object properties, and the illusory parts of the object boundary were generated as the most reasonable image "description" obtained by the functional minimization. The model shows the possibility to view the illusory contours as derived from general object detection task, performed by the visual system. Although this idea is not new (Gregory, 1972), this is the first time that the possibility to derive illusory contours from general object boundary detection task has been proved computationally.

Moreover, the multiple possible image perceptions were predicted here and ranked by perception probability. In case of the Kanizsa square illusion image, the most probable perception predicted by the model is of an illusory square (**Figures 8B,C**), and the second perception is of four pacman objects (**Figure 9**).

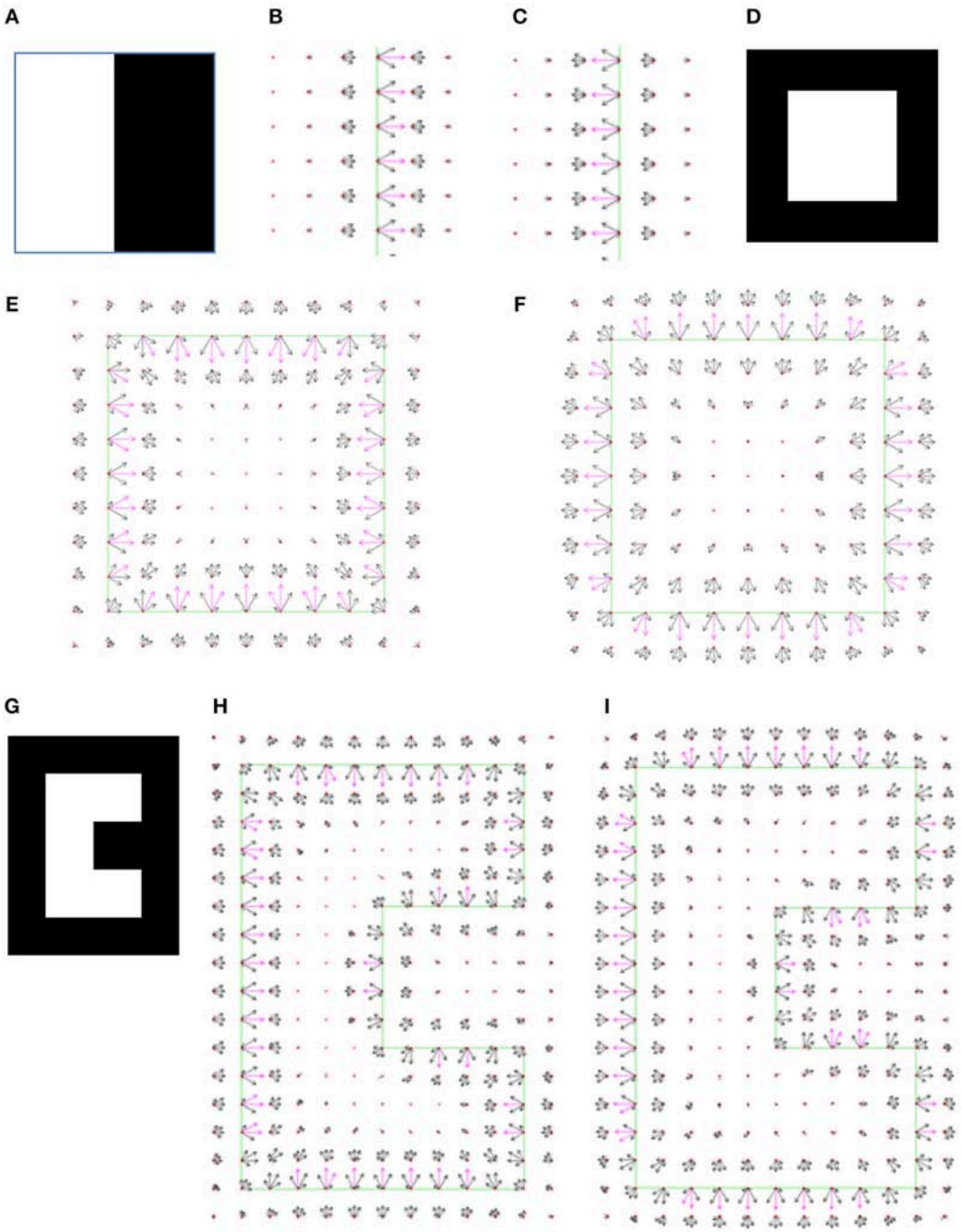


FIGURE 7 | (A) The simplest input image, with size 20 × 20 pixels. **(B)** The border ownership map of the first model interpretation of the image **(A)**. The object is located on the right side of the edge that is between the white and the black area in the input image. In all border ownership maps shown in following figures, the edges in the input image are marked by green lines for reference. The border ownership vectors with a value above 80% of the maximum border ownership vector value in the current map are colored magenta. Other border ownership vectors are black. The small red crosses depict the discrete grid of the input image. Note that only part of the border ownership map is shown, in order to make the view clearer. **(C)** The second model interpretation represents an object on the left side of the boundary between the white and the black regions in the input image **(A)**. **(D)** Input image with white square 8 × 8 pixels on black background. **(E)** The first model interpretation of the image in **(D)** represents a white square object on black background. The interpretation has a lowest cost −117. **(F)** The second model interpretation of the image in **(D)** represents a black frame with a square hole through which a white background is seen. This interpretation has cost−102, higher than the first interpretation, meaning it is less probable. **(G)** Input image of a C-shaped object. A similar image was applied in the original study of border ownership neurons (Zhou et al., 2000). **(H)** The first model interpretation of image **(G)** represents a C-shaped object. **(I)** The second model interpretation of image **(G)** represents a C-shaped hole in a frame.

Both predictions are consist with psychophysical findings (Rubin, 2001). Detecting different plausible solutions of a problem by finding multiple local minima of the functional is a novel approach.

There are numerous models that predict illusory contours in the Kanizsa square image (Williams and Hanson, 1994; Heitger et al., 1998; Kogo et al., 2002; Ron and Spitzer, 2011). The presented model approach, however, is essentially different from most of the models, since it is not oriented to detect illusory contours or locations of object occlusion. The model defines general preference rules of object boundaries and finds a stable minimizer to these rules. The illusory contours come out "by the way" as the minimizer of the problem. Since the essential approach of the model is the prediction of illusory contours based on general boundary detection approach, the model results cannot be compared to models that use specific mechanism of constructing illusory contours. The fact that the model does not use a general boundary detection approach is manifested by extraction of special image features.

Most of the existing models do extract special image features. For example, Madarasmi et al. (1994) use stochastic minimization of a functional to predict real and illusory contours of objects at different depth planes. The model is successfully applied to Kanizsa square illusion, where it detects both the illusory square and the overlapped inducer objects. The model, however, extracts special image features, namely L and T junctions, and only a single image interpretation is predicted. On the other hand, the model of Kass et al. (1988) detects real and illusory contours using energy minimizing splines. The model does not require special features extraction and both edge induced and line-end induced illusory contours are detected. However, the model is not fully automatic, since user interaction is required to draw the initial contour, section Introduction. In addition, only a single image interpretation is predicted in their model.

The functional optimization is usually used to obtain the best solution to a problem and only the global minimum is considered important (Figueiredo et al., 2003). Local minima are often considered to be disruptive and efforts are made to avoid them (Lee, 1995). The idea of a functional that has multiple minima is strongly related to the Gestalt psychology concept of Pragnantz: a *simple* and *stable* grouping (Koffka, 1935). Since the simplicity is measured by the cost functional, a local minimum of the functional indeed represents a simple and stable interpretation. Moreover, the values of the functional achieved at the different minima provide a general method, to compare the solutions at these minima. The multiple interpretations of the image are found in our model as the multiple stable minima of a functional. Thus, expressing multiple plausible solutions of a problem as multiple local minima of a functional is a new approach in the framework of functional optimization.

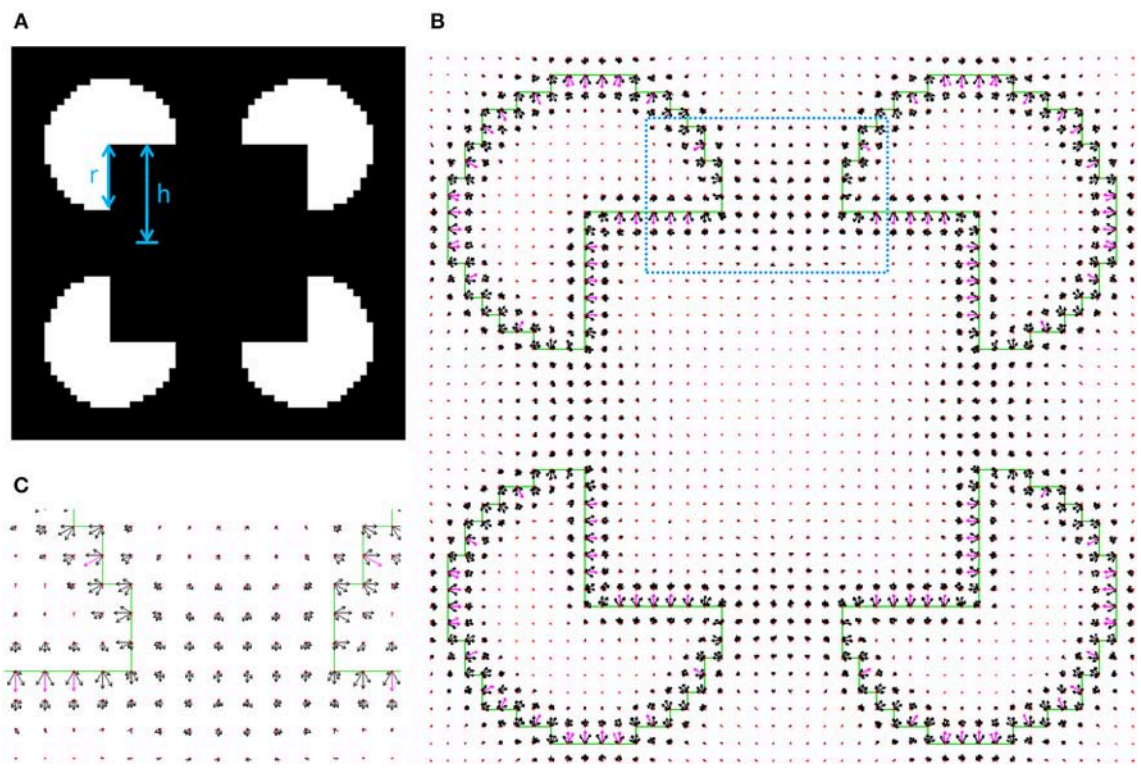


FIGURE 8 | (A) The Kanizsa illusory square. In this image the support ratio is 0.67. (The support ratio is defined as r/h, where r is the radius of the pacman and h is half the size of the illusory square). **(B)** The first model interpretation representing the square object, with partially illusory contours, occluding four circular objects. **(C)** Zoom-in into illusory boundary region between two pacmans, marked with dotted square in **(B)**.

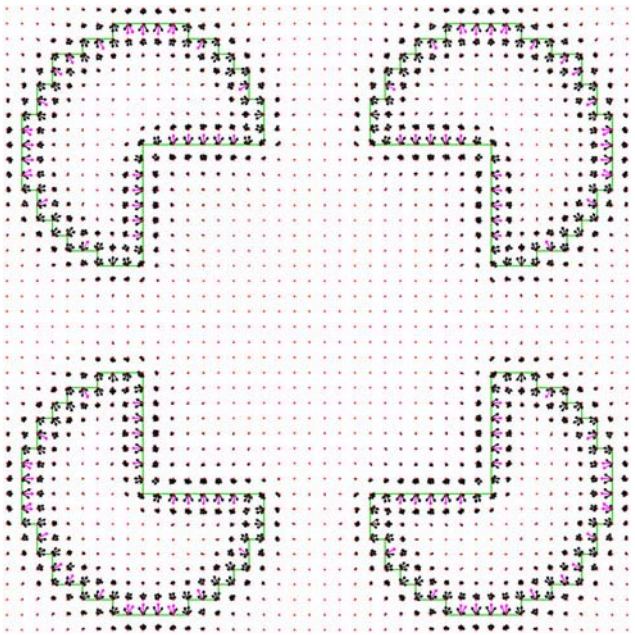


FIGURE 9 | The second model interpretation of image in **Figure 8A** represents four pacman objects.

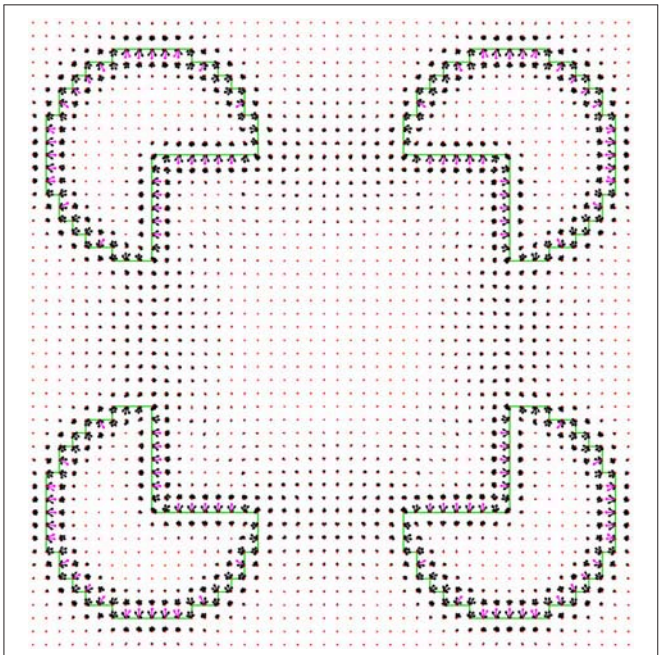


FIGURE 10 | The first illusory interpretation of Kanizsa square image with support ratio 0.57.

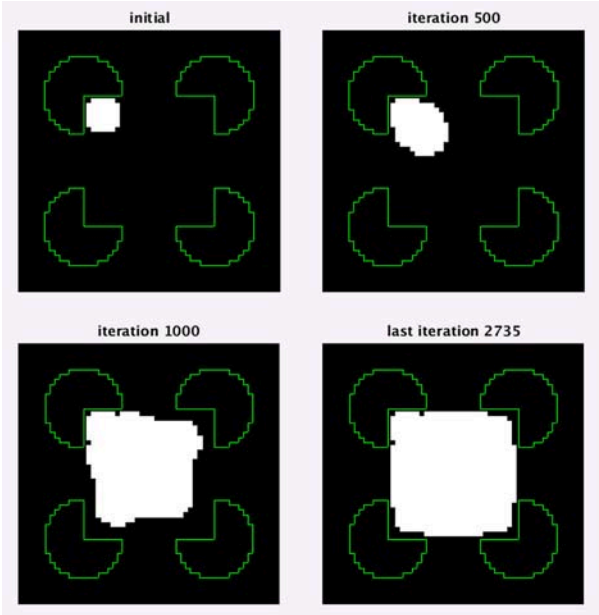


FIGURE 11 | An optimization test showing that a square object can be determined from the border ownership map, found by the model. The object extraction is for the first interpretation of Kanizsa square with support ratio 0.57 (**Figure 10**). Four optimization stages at different number of iterations are shown. In the images, the white region is the object at the depicted iteration. The green lines show the input image edges, which are shown for reference.

The method used to avoid minima that were already found in a functional section Finding multiple local minima, is related to the filled function method (Renpu, 1990), which has been used to find the global minimizer of a functional. In their method, an identified local minimum is replaced by a maximum in the functional. The main difference between the methods is the nature of the change in the function. The filled function depends on the functional in a complicated way, while in the proposed method the repulsive term is just added to the cost functional. In addition, our minimization is always initiated from the same point, while according to their method it requires trial over a set of directions, which is less efficient computationally.

The level set approach method section Retrieving object shape by contour evolution, can be used not only to find the top-most object boundary, but also the boundary of additional objects. To perform this, the initial small object should be placed adjacent to part of the boundary of the other object. This can enable us, for example, to complete the boundary of an occluded object.

The constants in the model were chosen by trial and error. Since the presented model proposes new a approach to the boundary detection task and contains a lot of complexity at this stage already, it is hard to also make it a fully robust model. Previous new conception models also did not supply a parameter sensitivity test at the first stage (Geiger et al., 1996). In any case, the same set of parameters were used for all experiments, hence we assume and experienced that the model is not very sensitive to parameter choice.

The proposed proof of concept model is restricted to gray scale images with solid non-textured regions and without lines. The model in its current version is not applicable yet for contour integration and detection of illusory lines such as defined by abutted gratings, since the model does not include

components dealing with lines or texture. Dealing with such type of images will require us to extend the measure of "description length" in the functional (van Tuijl, 1975) to include textured regions. It is very interesting to compare the model to available psychophysical data, like classification images obtained from human participants (Murray et al., 2005), however this is currently out of scope of the presented preliminary model.

Future work is planned to develop a robust model for object detection in real-world images. For this purpose, the object boundary based approach of current model should probably be replaced by an area based approach. We expect that this change will make the model much simpler, since, for example, matching the image by regions does not require even extraction of edges in the image. This change can also enable us to account for region based effects in the Kanizsa illusion (Kanizsa, 1976; Grossberg and Mingolla, 1987; Spehar, 2000; Ron and Spitzer, 2011).

AUTHOR CONTRIBUTIONS

AY developed and tested the model. HS supervised the work, made contributions to the model and reviewed the paper.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fncom.2018.00106/full#supplementary-material

REFERENCES

Bradley, D. R. (1987). "Cognitive contours and perceptual organization," in *The Perception of Illusory Contours*, eds S. Petry and G. E. Meyer (New York. NY: Springer), 201–212. Available online at: http://link.springer.com/chapter/10.1007/978-1-4612-4760-9_22 (Accessed Aug 18, 2015).

Coren, S. (1972). Subjective contours and apparent depth1. *Psychol. Rev.* 79, 3S9–367.

Curry, H. B. (1944). The method of steepest descent for nonlinear minimization problems. *Q. Appl. Math.* 2, 250–261.

Day, R. H., and Jory, M. K. (1980). A note on a second stage in the formation of illusory contours. *Percept. Psychophys.* 27, 89–91. doi: 10.3758/BF03199910

Ehrenstein, W. (1925). Versuche über die Beziehungen zwischen Bewegungs- und Gestaltwahrnehmung. *Z. Für Psychol.* 96, 305–352.

Figueiredo, M., Zerubia, J., and Jain, A. (2003). "Energy minimization methods in computer vision and pattern recognition," in *Third International Workshop, EMMCVPR 2001, Sophia Antipolis France, September 3-5, 2001* (Heidelberg: Springer).

Finkel, L. H., and Edelman, G. M. (1989). Integration of distributed cortical systems by reentry: a computer simulation of interactive functionally segregated visual areas. *J. Neurosci.* 9, 3188–3208.

Gao, R.-X., Wu, T.-F., Zhu, S.-C., and Sang, N. (2007). "Bayesian Inference for Layer Representation with Mixed Markov Random Field," in *Energy Minimization Methods in Computer Vision and Pattern Recognition Lecture Notes in Computer Science*, eds A. L. Yuille, S.-C. Zhu, D. Cremers, and Y. Wang (Berlin; Heidelberg: Springer), 213–224. doi: 10.1007/978-3-540-74198-5_17

Geiger, D., Kumaran, K., and Parida, L. (1996). "Visual organization for figure/ground separation," in *Proceedings CVPR '96, 1996 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (San Francisco, CA), 155–160. doi: 10.1109/CVPR.1996.517068

Gove, A., Grossberg, S., and Mingolla, E. (1995). Brightness perception, illusory contours, and corticogeniculate feedback. *Vis. Neurosci.* 12, 1027–1052.

Gregory, R. L. (1972). Cognitive contours. *Nature* 238, 51–52. doi: 10.1038/238051a0

Grossberg, S., and Mingolla, E. (1987). "Neural dynamics of form perception: boundary completion, illusory figures, and neon color spreading," in *Advances in Psychology The Adaptive Brain II*, ed S. Grossberg (Elsevier), 82–142. doi: 10.1016/S0166-4115(08)61758-6

Guy, G., and Medioni, G. (1993). "Inferring global perceptual contours from local features," in *Proceedings CVPR '93 1993 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1993* (New York, NY), 786–787. doi: 10.1109/CVPR.1993.341175

Heitger, F., von der Heydt, R., Peterhans, E., Rosenthaler, L., and Kübler, O. (1998). Simulation of neural contour mechanisms: representing anomalous contours. *Image Vis. Comput.* 16, 407–421. doi: 10.1016/S0262-8856(97)00083-8

Hubel, D. H., and Wiesel, T. N. (1959). Receptive fields of single neurones in the cat's striate cortex. *J. Physiol.* 148, 574–591.

Kanizsa, G. (1955). Margini quasi-percettivi in campi con stimolazione omogenea. *Riv. Psicol.* 49, 7–30.

Kanizsa, G. (1976). Subjective contours. *Sci. Am.* 234, 48–53.

Kass, M., Witkin, A., and Terzopoulos, D. (1988). Snakes: active contour models. *Int. J. Comput. Vis.* 1, 321–331. doi: 10.1007/BF00133570

Kennedy, J. M. (1988). Line endings and subjective contours. *Spat. Vis.* 3, 151–158. doi: 10.1163/156856888X00104

Kennedy, J. M., and Lee, H. (1976). A figure-density hypothesis and illusory contour brightness. *Perception* 5, 387–392. doi: 10.1068/p050387

Koffka, K. (1935). *Principles of Gestalt Psychology*. New York, NY: Psychology Press.

Kogo, N., Strecha, C., Fransen, R., Caenen, G., Wagemans, J., and Gool, L. V. (2002). "Reconstruction of Subjective Surfaces from Occlusion Cues," in *Biologically Motivated Computer Vision* Lecture Notes in Computer Science, eds H. H. Bülthoff, C. Wallraven, S.-W. Lee, and T. A. Poggio (Berlin; Heidelberg: Springer), 311–321. doi: 10.1007/3-540-36181-2_31

Kumaran, K., Geiger, D., and Gurvits, L. (1996). Illusory surface perception and visual organization. *Netw. Comput. Neural Syst.* 7, 437–437. doi: 10.1088/0954-898X_7_2_023

Leclerc, Y. G. (1989). Constructing simple stable descriptions for image partitioning. *Int. J. Comput. Vis.* 3, 73–102. doi: 10.1007/BF00054839

Lee, T. S. (1995). A Bayesian framework for understanding texture segmentation in the primary visual cortex. *Vision Res.* 35, 2643–2657. doi: 10.1016/0042-6989(95)00032-U

Lesher, G. W. (1995). Illusory contours: toward a neurally based perceptual theory. *Psychon. Bull. Rev.* 2, 279–321. doi: 10.3758/BF03210970

Madarasmi, S., Pong, T.-C., and Kersten, D. (1994). "Illusory contour detection using MRF models," in *IEEE World Congress on Computational Intelligence 1994 IEEE International Conference on Neural Networks, 1994,* Vol. 7 (Orlando, FL), 4343–4348. doi: 10.1109/ICNN.1994.374966

Malladi, R., Sethian, J. A., and Vemuri, B. C. (1995). Shape modeling with front propagation: a level set approach. *IEEE Trans. Patt. Anal. Mach. Intell.* 17, 158–175. doi: 10.1109/34.368173

Murray, R. F., Bennett, P. J., and Sekuler, A. B. (2005). Classification images predict absolute efficiency. *J. Vis.* 5, 5–5. doi: 10.1167/5.2.5

Nakayama, K., and Shimojo, S. (1990). Toward a neural understanding of visual surface representation. *Cold Spring Harb. Symp. Quant. Biol.* 55, 911–924. doi: 10.1101/SQB.1990.055.01.085

Nitzberg, M., and Mumford, D. (1990). "The 2.1-D sketch," in *[1990] Proceedings Third International Conference on Computer Vision*, 138–144. doi: 10.1109/ICCV.1990.139511

Osher, S., and Sethian, J. A. (1988). Fronts propagating with curvature-dependent speed: algorithms based on Hamilton-Jacobi formulations. *J. Comput. Phys.* 79, 12–49. doi: 10.1016/0021-9991(88)90002-2

Pal, N. R., and Pal, S. K. (1993). A review on image segmentation techniques. *Patt. Recognit.* 26, 1277–1294. doi: 10.1016/0031-3203(93)90135-J

Prazdny, K. (1983). Illusory contours are not caused by simultaneous brightness contrast. *Percept. Psychophys.* 34, 403–404. doi: 10.3758/BF03203054

Renpu, G. (1990). A filled function method for finding a global minimizer of a function of several variables. *Math. Program.* 46, 191–204. doi: 10.1007/BF01585737

Ringach, D. L., and Shapley, R. (1996). Spatial and temporal properties of illusory contours and amodal boundary completion. *Vis. Res.* 36, 3037–3050. doi: 10.1016/0042-6989(96)00062-4

Ron, E., and Spitzer, H. (2011). Is the Kanizsa illusion triggered by the simultaneous contrast mechanism? *J. Opt. Soc. Am. A* 28:2629. doi: 10.1364/JOSAA.28.002629

Rubin, N. (2001). The role of junctions in surface completion and contour matching. *Perception* 30, 339–366. doi: 10.1068/p3173

Saund, E. (1999). "Perceptual organization of occluding contours generated by opaque surfaces," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1999*, Vol. 2 (Fort Collins, CO). doi: 10.1109/CVPR.1999.784988

Schumann, F. (1900). Beiträge zur Analyse der Gesichtswahrnehmungen. Erste Abhandlung. Einige Beobachtungen über die Zusammenfassung von Gesichtseindrücken zu Einheiten. *Z. Für Psychol. Physiol. Sinnesorgane* 23, 1–32.

Shipley, T. F., and Kellman, P. J. (1992). Strength of visual interpolation depends on the ratio of physically specified to total edge length. *Percept. Psychophys.* 52, 97–106.

Spehar, B. (2000). Degraded illusory contour formation with non-uniform inducers in Kanizsa configurations: the role of contrast polarity. *Vis. Res.* 40, 2653–2659. doi: 10.1016/S0042-6989(00)00109-7

van Tuijl, H. F. (1975). A new visual illusion: neonlike color spreading and complementary color induction between subjective contours. *Acta Psychol.* 39, 441-IN1. doi: 10.1016/0001-6918(75)90042-6

Williams, L. R., and Hanson, A. R. (1994). "Perceptual completion of occluded surfaces," in *Proceedings CVPR '94 1994 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1994* (Seattle, WA), 104–112. doi: 10.1109/CVPR.1994.323803

Williams, L. R., and Jacobs, D. W. (1995). "Stochastic completion fields: a neural model of illusory contour shape and salience," in *Proceedings of IEEE International Conference on Computer Vision* (Cambridge, MA), 408–415. doi: 10.1109/ICCV.1995.466910

Zhou, H., Friedman, H. S., and von der Heydt, R. (2000). Coding of border ownership in monkey visual cortex. *J. Neurosci. Off. J. Soc. Neurosci.* 20, 6594–6611. doi: 10.1523/JNEUROSCI.20-17-06594.2000

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 18 February 2019
doi: 10.3389/fnins.2019.00008

In Praise of Artifice Reloaded: Caution With Natural Image Databases in Modeling Vision

Marina Martinez-Garcia[1,2], *Marcelo Bertalmío*[3] and *Jesús Malo*[1]*

[1] Image Processing Lab, Universitat de València, Valencia, Spain, [2] CSIC, Instituto de Neurociencias, Alicante, Spain, [3] Departamento de Tecnologías de la Información y las Comunicaciones, Universidad Pompeu Fabra, Barcelona, Spain

Subjective image quality databases are a major source of raw data on how the visual system works in *naturalistic environments*. These databases describe the sensitivity of many observers to a wide range of distortions of different nature and intensity seen on top of a variety of natural images. Data of this kind seems to open a number of possibilities for the vision scientist to check the models in realistic scenarios. However, while these natural databases are great benchmarks for models developed in some other way (e.g., by using the well-controlled *artificial stimuli* of traditional psychophysics), they should be carefully used when trying to fit vision models. Given the high dimensionality of the image space, it is very likely that some basic phenomena are under-represented in the database. Therefore, a model fitted on these large-scale natural databases will not reproduce these under-represented basic phenomena that could otherwise be easily illustrated with well selected artificial stimuli. In this work we study a specific example of the above statement. A standard cortical model using wavelets and divisive normalization tuned to reproduce subjective opinion on a large image quality dataset fails to reproduce basic cross-masking. Here we outline a solution for this problem by using artificial stimuli and by proposing a modification that makes the model easier to tune. Then, we show that the modified model is still competitive in the large-scale database. Our simulations with these artificial stimuli show that when using steerable wavelets, the conventional unit norm Gaussian kernels in divisive normalization should be multiplied by high-pass filters to reproduce basic trends in masking. Basic visual phenomena may be misrepresented in large natural image datasets but this can be solved with model-interpretable stimuli. This is an additional argument *in praise of artifice* in line with Rust and Movshon (2005).

Keywords: natural stimuli, artificial stimuli, subjective image quality databases, wavelet + divisive normalization, contrast masking

OPEN ACCESS

Edited by:
Hedva Spitzer,
Tel Aviv University, Israel

Reviewed by:
Sophie Wuerger,
University of Liverpool,
United Kingdom
Kendrick Norris Kay,
University of Minnesota Twin Cities,
United States

***Correspondence:**
Jesús Malo
jesus.malo@uv.es

Specialty section:
This article was submitted to Perception Science, a section of the journal Frontiers in Neuroscience

Received: *21 December 2017*
Accepted: *07 January 2019*
Published: *18 February 2019*

Citation:
Martinez-Garcia M, Bertalmío M and Malo J (2019) In Praise of Artifice Reloaded: Caution With Natural Image Databases in Modeling Vision. Front. Neurosci. 13:8. doi: 10.3389/fnins.2019.00008

1. INTRODUCTION

In the age of *big data* one may think that machine learning applied to representative databases will automatically lead to accurate models of the problem at hand. For instance, the problem of modeling the perceptual difference between images showed up in the discussion of eventual challenges at the NIPS-11 *Metric Learning* Workshop (Shakhnarovich et al., 2011). However, despite its interesting implications in visual neuroscience, the subjective metric of the image space

was dismissed as a *trivial* regression problem because there are subjectively-rated image quality databases that can be used as training set for supervised learning.

Subjective image and video quality databases (such as VQEG, LIVE, TID, CID, CSIQ)[1] certainly are a major source of raw data on how the visual system works in *naturalistic environments*. These databases describe the sensitivity of many observers to a wide range of distortions (of different nature and with different suprathreshold intensities) seen on top of a variety of natural images. These databases seem to open a number of possibilities to check the models in realistic scenarios.

Following a tradition that links the image quality assessment problem in engineering with human visual system models (Sakrison, 1977; Watson, 1993; Wang and Bovik, 2009; Bodrogi et al., 2016), these subjectively rated image databases have been used to fit models coming from classical psychophysics or physiology (Watson and Malo, 2002; Laparra et al., 2010; Malo and Laparra, 2010; Bertalmio et al., 2017). Given the similarity between these biological models (Carandini and Heeger, 2012) and feed-forward convolutional neural nets (Goodfellow et al., 2016), an interesting analogy is possible. Fitting the biological models to reproduce the opinion of the observers in the database is algorithmically equivalent to the learning stage in deep networks. This deep-learning-like use of the databases is a convenient way to train a physiologically-founded architecture to reproduce a psychophysical goal (Berardino et al., 2017; Laparra et al., 2017; Martinez-Garcia et al., 2018). When using these biologically-founded approaches, the parameters found have a straightforward interpretation as for instance the frequency bandwidth of the system or the extent of the interaction between sensors tuned to different features.

On the other hand, pure machine-learning (data-driven) approaches have also been used to predict subjective opinion. In this case, after extracting features with reasonable statistical meaning or perceptual inspiration, generic regression techniques are applied (Moorthy and Bovik, 2010, 2011; Saad et al., 2010, 2012, 2014), even though this regression has no biological grounds.

1.1. Eventual Problems With Databases

The problem with the above uses of naturalistic image databases is the conventional concern about training sets in machine learning: *is the training set a balanced representation of the range of behaviors to be explained?*

If it is not the case, the resulting model may be biased by the dataset and it will have generalization problems. This overfitting risk has been recognized by the authors of image quality metrics based on generic regression (Saad et al., 2012). Perceptually meaningful architectures impose certain constraints on the flexibility of the model, as opposed to generic regressors. These constraints could be seen as a sort of *Occam's Razor* in favor of lower-dimensional models. However, even in the biologically meaningful cases, there is a risk that the model found by fitting the naturalistic database misses well-known texture perception facts.

Accordingly, Laparra et al. (2010) and Malo and Laparra (2010) used artificial stimuli after the learning stage to check the Contrast Sensitivity Function and some properties of *visual masking*. Similarly, in Ma et al. (2018) after training the deep network in the dataset they have to show model-related stimuli to human observers to check if the results are meaningful (and discard eventual over-fitting).

1.2. The Regression Hypothesis Questioned

In this work we question the hypothesis suggested at the NIPS Metric Learning Workshop (Shakhnarovich et al., 2011) that assumes that pure regression on naturalistic databases will lead to sensible vision models.

Of course, training whatever regression model with subjectively rated natural images to predict human opinion is a *perfectly fine* approach to tackle the restricted image quality problem. Actually, sometimes disregarding any prior knowledge about how the visual system works is seen as a plus (Bosse et al., 2018): the quantitative solution to this specific problem may gain nothing from understanding the elements of a successful regression model in terms of properties of actual vision mechanisms.

However, from a broader perspective, models intended to understand the behavior of the visual system should be more ambitious: they should be interpretable in terms of the underlying mechanisms and be able to reproduce other behavior. Our message here is that large-scale naturalistic databases should not be the only source of information when trying to fit *vision models*. Given the high dimensionality of the image space, it is very likely that some basic phenomena (e.g., the visibility of certain distortions in certain environments) are under-represented in the database. As a result, the model is not forced to reproduce these under-represented phenomena. And more importantly, the use of model-interpretable artificial stimuli is useful to determine the values of specific parameters in the model.

In particular, we study a specific example of the generalization risk suggested above and the benefits of model-based artificial stimuli. We show that a wavelet+divisive normalization layer of a standard cascade of linear+nonlinear layers fitted to maximize the correlation with subjective opinion on a large image quality database (Martinez-Garcia et al., 2018), fails to reproduce basic cross-masking. Here we point out the problem and we outline a solution using well selected artificial stimuli. Then, we show that the model corrected to account for these extra artificial tests is also a competitive explanation for the large-scale naturalistic database. This example is interesting because showing convincing Maximum Differentiation stimuli, as done in Berardino et al. (2017), Martinez-Garcia et al. (2018), and Ma et al. (2018), may not be enough to guarantee that the model reproduces related behaviors and points out the need to explicitly check with artificial stimuli.

[1] A non exhaustive list of references and links to subjective quality databases includes (Webster et al., 2001; Ponomarenko et al., 2009, 2015; Larson and Chandler, 2010; Pedersen, 2015; Ghadiyaram and Bovik, 2016).

1.3. In Praise of Artifice: Interpretable Models and Interpretable Stimuli

In line with Rust and Movshon (2005), our results in this work, namely pointing out the misrepresentation of basic visual phenomena in subjectively-rated natural image databases and the proposed procedure to fix it, are additional arguments *in praise of artifice*: the artificial model-motivated stimuli in classical visual neuroscience are helpful to (a) point out the problems that remain in models fitted to natural image databases, and (b) to suggest intuitive modifications of the models.

Regarding interpretable models, we propose a modification for the considered Divisive Normalization (Carandini and Heeger, 2012) that stabilizes its behavior. As a result of this stabilization, the model is easy to tune (even by hand) to qualitatively reproduce cross-masking. Interestingly, as a consequence of this modification and analysis with artificial stimuli, we show that the conventional unit-norm kernels in divisive normalization may have to be re-weighted depending on the selected wavelets.

It is important to note that the observations made in this work are not restricted to the specific image quality problem. Following seminal ideas based on information theory (Attneave, 1954; Barlow, 1959), theoretical neuroscience considers explanations of sensory systems based on statistical learning as alternative to physiological and psychophysical descriptions (Dayan and Abbott, 2005). Therefore, the points made below on natural image datasets, artificial stimuli from interpretable models, and optimization goals in statistical learning, also apply to a wider range of computational explanations.

The paper is organized as follows: section 2 describes the visual stimuli and introduces the cortical models considered in the work. First it illustrates the intuition that can be obtained from proper artificial stimuli as opposed to the not-so-obvious interpretation of natural stimuli. Then, it presents the structure of wavelet-like responses in V1 cortex and two standard neural interaction models: **Model A** (intra-band), and **Model B** (inter-band). Section 3 shows that despite **Model A** is tuned to maximize the correlation with subjective opinion in a large-scale naturalistic image quality database it fails to reproduce basic properties of visual masking. Simulations with artificial stimuli allow intuitive tuning of **Model B** to get the correct contrast response curves while preserving the success on the large-scale naturalistic database. Finally, as suggested by the failure-and-solution example considered in this work, in section 4 we discuss the opportunities and precautions of the use of natural image databases to fit vision models, and the relevance of artificial stimuli based on interpretable models.

2. MATERIALS AND METHODS

Here we present the visual stimuli and the cortical interaction models considered throughout the work. The use of model-inspired artificial stimuli is critical to point out the limitations of simple models and to tune the parameters of more general models.

2.1. Natural vs. Artificial Stimuli

Figure 1 shows a representative subset of the kind of patterns subjectively rated in image quality databases. This specific example comes from the TID2008 database (Ponomarenko et al., 2008). In these databases, natural scenes (photographic images with uncontrolled content) are corrupted by noise sources of different nature. Some of the noise sources are stationary and signal independent, while others are spatially variant and depend on the background. Ratings depend on the visibility of the distortion seen on top of the natural background. The considered distortions come in different suprathreshold intensities. In some cases these intensities have controlled (linearly spaced) energy or contrast, but in general, they come from arbitrary scales. Examples include different compression ratio or color quantization coarseness with no obvious psychophysical meaning. This is because the motivation of the original databases (e.g., VQEG or LIVE) was the assessment of distortions occurring in *image processing* applications (e.g., transmission errors in digital communication) and not necessarily to be a tool for *vision science*. More recent databases include more accurate control of luminance and color of both the backgrounds and the distortions (Pedersen, 2015), or report the intensities of the distortions in JND units (Alam et al., 2014). Perceptual ratings in such diverse sets certainly provide a great ground truth to check vision science models in naturalistic conditions.

However, the result of such variety is that the backgrounds and the tests seen on top have no clear interpretation in terms of specific perceptual mechanisms or controlled statistics in a representation with physiological meaning. Even though not specifically directed against subjectively rated databases, this was also the main drawback pointed out in Rust and Movshon (2005) against the use of generic natural images in vision science experiments.

In this work we go a step further in that criticism: due to the uncontrolled nature of the natural scenes and the somewhat arbitrary distortions found in these databases, the different aspects of a specific perceptual phenomenon are not fully represented in the database. Therefore, these databases should be used carefully when training models because this misrepresentation will have consequences when fitting the models.

For instance, let's consider pattern masking (Foley, 1994; Watson and Solomon, 1997). It is true that some distortions in the databases introduce relatively more noise in high contrast regions, which seems appropriate to illustrate masking. This is the case of the JPEG or JPEG2000 artifacts, or the so called *masked noise* in the TID database. See for instance the third example in the first row of **Figure 1**. These deviations on top of high contrast regions are less visible than equivalent deviations of the same energy on top of flat backgrounds. This difference in visibility is due to the inhibitory effect of surround in *masking* (Foley, 1994; Watson and Solomon, 1997). Actually, perceptual improvements of image coding standards critically depend on using better masking models that allow using less bits in those regions (Malo et al., 2000a, 2001, 2006; Taubman and Marcellin, 2001). Appropriate prediction of the visibility of these distortions in the database should come from an accurate

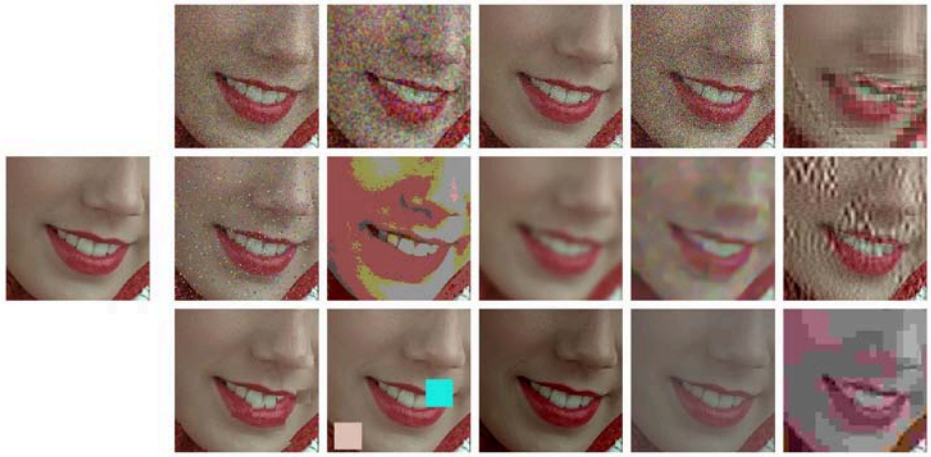


FIGURE 1 | Natural scenarios and complex distortions. The isolated image at the left is an example of a natural background (uncontrolled scene) to be distorted by a variety of degradations of different nature. The images in the array illustrate the kind of stimuli rated by the observers in image quality databases. The score of the degraded images is related to the visibility of the corresponding distortion (the test) seen on top of the original image (the background). The reported subjective ratings constitute the ground truth that should be predicted by vision models from the variation of the responses due to the distortions.

model of texture masking. However, a systematic set of examples illustrating the different aspects of masking is certainly not present in the databases. For example, there are no stimuli showing crossmasking between different frequencies in different backgrounds. Therefore, this phenomenon is under-represented in the database.

Such basic texture perception facts can be easily illustrated using artificial stimuli. Artificial stimuli can be designed with a specific perceptual phenomenon in mind, and using patterns which have specific consequences in models, e.g., stimulation of certain sensors of the model. Model/phenomenon-based stimuli is the standard way in classical psychophysics and physiology. **Figure 2** is an example of the power of well controlled artificial stimuli: it represents a number of major texture perception phenomena in a single figure.

This figure shows two basic tests (low-frequency vertical and high-frequency horizontal) of increasing contrast from left to right. These series of tests are, respectively, shown on top of (a) no background, and (b) on top of backgrounds of controlled frequency and orientation.

First, of course we can see that the visibility of the tests (or the response of the mechanisms that mediate visibility) increases with contrast, from left to right. This is why even the trivial Euclidean distance between the original and the distorted images is positively correlated with subjective opinion of distortion.

Second, the visibility, or the responses, depend(s) on the frequency of the test. Note that the lower frequency test is more visible than the high frequency test at reading distance. This illustrates the effect of the Contrast Sensitivity Function (Campbell and Robson, 1968).

Third, the response increase is non-linear with contrast. Note that for lower contrasts (e.g., from the second picture to the third in the series) the increase in visibility is bigger than for higher contrasts (e.g., between the pictures at the right-end). This means that the slope of the mechanisms mediating the response is high for lower amplitudes and saturates afterwards. This sort of Weber-like behavior for contrast is a distinct feature of contrast masking (Legge, 1981).

Finally, the visibility (or response) decreases with the background energy depending on the spatio-frequency similarity between test and background. Note for instance that the low frequency test is less visible on top of the low frequency background than on top of the high frequency background. Important for the example considered throughout this paper, note that the visibility of the high frequency test behaves *the other way around*: it is bigger on top of the low frequency test. Moreover, this *masking* effect is bigger for bigger contrasts of the background. This adaptivity of the nonlinearity is a distinct feature of the *masking* effect (Foley, 1994; Watson and Solomon, 1997), and more importantly, it is a distinct feature of real neurons (Carandini and Heeger, 1994, 2012) with regard to the simplified neurons used in deep learning (Goodfellow et al., 2016).

As a result, just by looking at **Figure 2**, one may imagine how the visibility (or response) curves vs. the contrast of the test should be for the series of stimuli presented. **Figure 3** shows an experimental example of the kind of response curves obtained in actual neurons in masking situations. Note the saturation of the response curves and how they are attenuated when the background is similar to the test. Even this qualitative behavior highlighted in green (saturation and attenuation) may be used to discard models that do not reproduce the expected behavior, i.e., that do not agree with what we are seeing.

More importantly, the relative visibility of these artificial stimuli can also be used to intuitively tune the parameters of a model to better reproduce the visible behavior. This can be done because these artificial stimuli were crafted to have a clear interpretation in a standard model of texture vision: a set of V1-like wavelet neurons (oriented receptive fields tuned to different frequency scales). **Figure 4** illustrates this fact: note how the test patterns considered in the figure mainly stimulate a specific

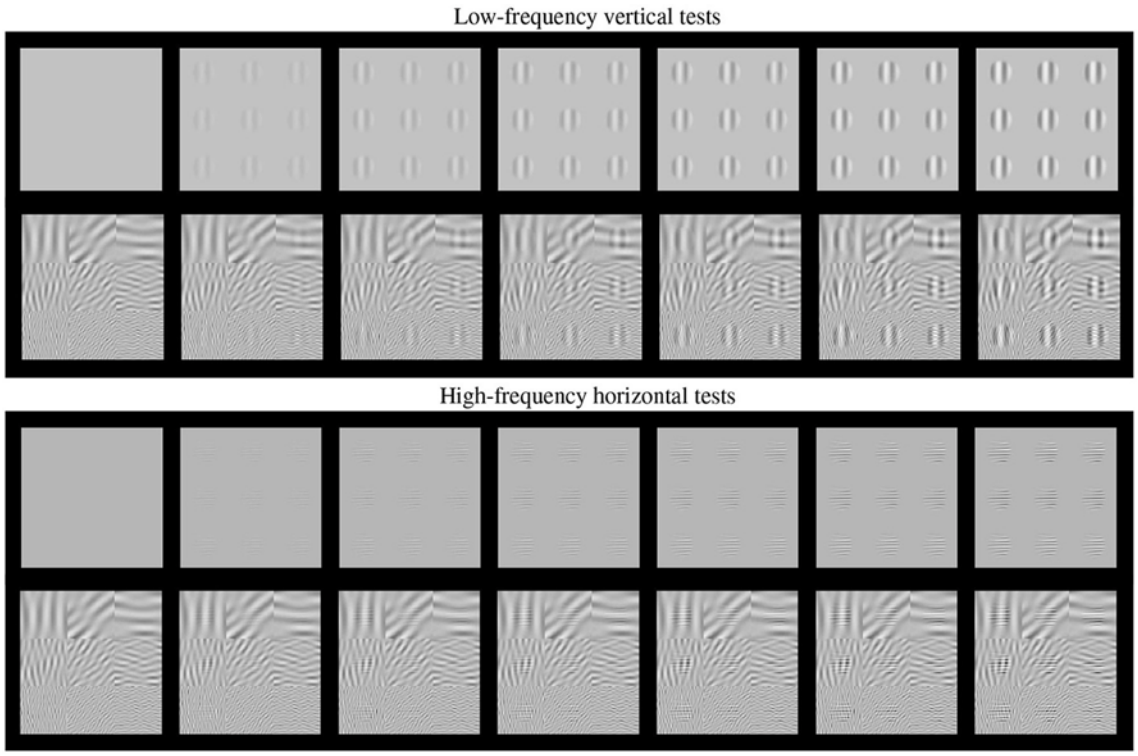


FIGURE 2 | Artificial stimuli. Several texture phenomena illustrated in a single figure (see text for details). Here the tests are the 9 patterns in the gray frames. These tests increase in contrast from the frame at the left to the frame at the right. The visibility of the tests (a) nonlinearly increases with the contrast from left to right; (b) the visibility depends on the frequency of the tests, low frequency at the top panel and high frequency at the bottom panel; and (c) the visibility of the tests depends on the background (cross-masking).

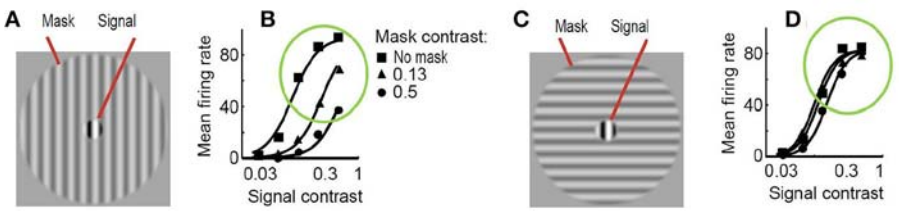


FIGURE 3 | Experimental response of V1 neurons (mean firing rate) in masking situations. Adapted from (Cavanaugh, 2000; Schwartz and Simoncelli, 2001). At the left **(A)** test and mask do have the same spatio-frequency characteristics. At the right **(C)** test is substantially different from the mask. Note the decay in the responses, compare the curves in green circles, when test and background share properties **(B)** as opposed to the case where they do not **(D)**.

subband of a 3-scale 4-orientation steerable wavelet pyramid (Simoncelli et al., 1992), which is a commonly used model of V1 sensors. As a result, it is easy to select the set of sensors that will drive the visibility descriptor in the model: see the highlighted wavelet coefficients in the diagrams at the right of **Figure 4**.

The same intuitive energy distribution over the pyramid is true for the backgrounds, which stimulate the corresponding subband (scale and orientation). As a result, given the distribution of test and backgrounds in the pyramid, it is easy to propose intuitive cross-band inhibition schemes to lead to the required decays in the response.

The intuitions obtained from artificial model-oriented stimuli about response curves and eventual-crossmasking schemes are fundamental both to criticize the results obtained from *blind learning from a database*, and to propose intuitive improvements of the model.

2.2. Cortical Interaction Models: Structure and Response

In this work we analyze the behavior of standard retina-cortex models that follow the program suggested in Carandini and Heeger (2012) i.e., cascades of isomorphic linear+nonlinear layers, each focused on a different psychophysical factor:

Layer $S^{(1)}$ linear spectral integration to compute luminance and opponent tristimulus channels, and nonlinear brightness/color response.

Layer $S^{(2)}$ definition of local contrast by using linear filters and divisive normalization.

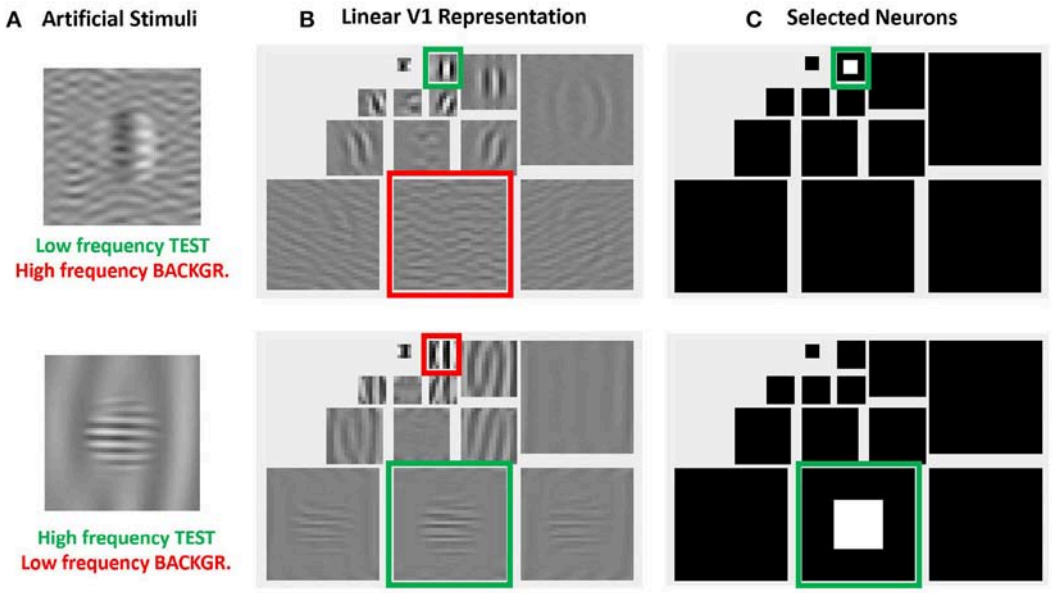


FIGURE 4 | Advantages of artificial stimuli. Model-related construction of stimuli simplifies the reproduction of results form model outputs and the interpretation of results. In this example, frequency and orientation of tests and backgrounds in the artificial stimuli at the left **(A)** are selected to stimulate specific subbands of the model, see highlighted regions in the wavelet transform at the central panel **(B)**. Therefore, it is easy to select the sensors that mediate the visibility of the tests, see the coefficients in white in the wavelet diagram at the right panel **(C)**.

Layer $S^{(3)}$ linear LGN-like contrast sensitivity filter and nonlinear local energy masking in the spatial domain.

Layer $S^{(4)}$ linear V1-like wavelet decomposition and nonlinear divisive normalization to account for orientation and scale-dependent masking.

This family of models represents a system, S, that depends on some parameters, $\boldsymbol{\Theta}$, and applies a series of transforms on the input radiance vector, $\boldsymbol{x}^0$, to get a series of intermediate response vectors, $\boldsymbol{x}^i$,

$$S(\boldsymbol{x}^0, \boldsymbol{\Theta}) \qquad \boldsymbol{x}^0 \xrightarrow{S^{(1)}} \boldsymbol{x}^1 \xrightarrow{S^{(2)}} \boldsymbol{x}^2 \xrightarrow{S^{(3)}} \boldsymbol{x}^3 \xrightarrow{S^{(4)}} \boldsymbol{x}^4 \tag{1}$$

Each layer in this sequence accounts for the corresponding psychophysical phenomenon outlined above and is the concatenation of a linear transform $\mathcal{L}$ and a nonlinear transform $\mathcal{N}$:

$$\cdots \boldsymbol{x}^{i-1} \xrightarrow{\mathcal{L}^{(i)}} \boldsymbol{y}^i \xrightarrow{\mathcal{N}^{(i)}} \boldsymbol{x}^i \cdots \qquad S^{(i)} \tag{2}$$

Here, in each layer we use convolutional filters for the linear part and the canonical Divisive Normalization for the nonlinear part. The mathematics of this type of models required to set their parameters are detailed in Martinez-Garcia et al. (2018).

In this kind of models the psychophysical behavior (visibility of a test) is obtained from the behavior of individual units (increment of responses) through some sort of *summation*. The visibility of a test, $\Delta\boldsymbol{x}^0$, seen on top of a background, $\boldsymbol{x}^0$, is given by the perceptual distance between *background* and *background+test*. Specifically, this perceptual distance, d_p, may be computed through the q norm of the vector with the increment of responses in the last neural layer (Watson and Solomon, 1997; Laparra et al., 2010; Martinez-Garcia et al., 2018). In the 4-layer model of Equation 1, we have $\|\Delta\boldsymbol{x}^4\|_q$:

$$d_p(\boldsymbol{x}^0, \boldsymbol{x}^0 + \Delta\boldsymbol{x}^0) = \|\Delta\boldsymbol{x}^4\|_q = \left(\sum_j |\Delta x_j^4|^q\right)^{\frac{1}{q}} \tag{3}$$

There is a variety of summation schemes: one may choose to use different summation exponents for different features (e.g., splitting the sum over j in space, frequency, and orientation), and order of summation matters if the exponents for the different features are not the same. Besides, there is no clear consensus on the value of the summation exponents either (Graham, 1989): the default quadratic summation choice, $q = 2$ (Teo and Heeger, 1994; Martinez-Garcia et al., 2018), has been questioned proposing bigger (Watson and Solomon, 1997; Laparra et al., 2010) and smaller (Laparra et al., 2017) summation exponents.

More important than all the above technicalities, the key points in Equation (3) are: (a) it clearly relates the visibility with the response of the units, and (b) for $q \geq 2$, the visibility is

driven by the response of the units that undergo bigger variation, $|\Delta x_j^4|$, such as the ones highlighted in **Figure 4**. Therefore, in this kind of models, analyzing the visibility curves or the response curves of the units tuned to the test is qualitatively the same. In the simulations we do the latter since we are interested in direct observation of the effect of the interaction parameters on the curves; and this is more clear when looking at the response of selected subsets of units as those highlighted in **Figure 4**.

In this work we compare two specific examples of this family of models. These two models will be referred to as **Model A** and **Model B**. They have identical layers 1–3, and they only differ in the nonlinear part of the fourth layer: the stage describing the interaction between cortical oriented receptive fields. In **Model A** we only consider interactions between the sensors tuned to the same subband (scale and orientation) because we proved that this simple scheme is appropriate to obtain good performance in subjectively rated databases (Laparra et al., 2010; Malo and Laparra, 2010). In **Model B** on top of the intra-band relation we also considered inter-band relations according to a standard unit-norm Gaussian kernel over space, scale and orientation (Watson and Solomon, 1997). Additionally to the classical inter-band generalization we also included extra weights and a stabilization constant that makes the model easier to understand. The software implementing **Model A** and **Model B** is available at "http://isp.uv.es/docs/BioMultiLayer_L_NL_a_and_b.zip".

Let's consider the differences between the models in more detail. Assuming that the output of the wavelet filter-bank is the vector $\boldsymbol{y}$, and assuming that the vector of energies of the coefficients is obtained by coefficient-wise rectification and exponentiation, $\boldsymbol{e} = |\boldsymbol{y}|^\gamma$, the vector of responses after divisive normalization in the last layer of **Model A** is:

$$\boldsymbol{x} = sign(\boldsymbol{y}) \odot \frac{\boldsymbol{e}}{\boldsymbol{b} + H \cdot \boldsymbol{e}} \tag{4}$$

where $\odot$ stands for element-wise Hadamard product and the division is also an element-wise Hadamard quotient where the energy of each linear response is divided by a linear combination of the energies of the neighboring coefficients in the wavelet pyramid. This linear combination (that attenuates the response) is given by the matrix-on-vector product $H \cdot \boldsymbol{e}$. Note that, for simplicity, in Equation 4 we omitted the indices referring to the 4th layer [as opposed to the more verbose formulation in the Appendix (**Supplementary Material**)].

The i-th row of this matrix, H, tells us how the responses of neighbor sensors in the vector $\boldsymbol{e}$ attenuate the response of the i-th sensor in the numerator, e_i. The attenuating effect of these linear combinations is moderated by the semisaturation constants in vector $\boldsymbol{b}$.

The structure of these vectors and matrices is relevant to understand the behavior on the stimuli. First, one must consider that all the vectors, $\boldsymbol{y}$, $\boldsymbol{e}$, and $\boldsymbol{x}$, have wavelet-like structure. **Figure 4** shows this subband structure for specific artificial stimuli and **Figure 5** shows it for natural stimuli.

The i-th coefficient has a 4-dimensional spatio-frequency meaning, $i \equiv (\boldsymbol{p}_i, f_i, \phi_i)$, where $\boldsymbol{p}$ is a two-dimensional location, f is the modulus of the spatial frequency, and ϕ is orientation.

In **Model A** we only consider Gaussian intra-band relations. This means that interactions in H decay with spatial distance and it is zero between sensors tuned to different frequency and orientation. This implies a block-diagonal structure in H with zeros in the off-diagonal blocks. In Martinez-Garcia et al. (2018) the norm or each Gaussian neighborhood (or row) in H was optimized to maximize the correlation with subjective opinion.

It is important to stress that the specific distribution of responses of natural images over the subbands of the response vector (green line in **Figure 5**) is critical to reproduce the good behavior of the model on the database. Note that this is not a regular (linear) wavelet transform, but the (nonlinear) response vector. Therefore, this distribution tells us *both* about the statistics of natural images and about the behavior of the visual system. On the one hand, natural images have relatively more energy in the low-frequency end. But, on the other hand, it is visually relevant that the response of sensors tuned to the high frequency details is much lower than the response of the sensors tuned to the low frequency details. The latter is in line with the different visibility of the artificial stimuli of different frequency shown in **Figure 2**, and it is probably due to the effect of the Contrast Sensitivity Function (CSF) in earlier stages of the model. This is important because keeping this relative magnitude between subbands is crucial to have good alignment with subjective opinion in the large-scale database.

In the case of **Model B**, we consider (a) a more general interaction kernel in the divisive normalization, and (b) a constant diagonal matrix to control the dynamic range of the responses. Specifically, the vector of responses is:

$$\boldsymbol{x} = sign(\boldsymbol{y}) \odot \left[\boldsymbol{\kappa} \odot \frac{\boldsymbol{b} + H_G \cdot \boldsymbol{e}^\star}{\boldsymbol{e}^\star}\right] \odot \frac{\boldsymbol{e}}{\boldsymbol{b} + H_G \cdot \boldsymbol{e}}. \tag{5}$$

Here the response still follows a nonlinear divisive normalization because $\boldsymbol{e}^\star$ is just a fixed vector (not a variable), and hence the term in brackets is just another constant vector. In **Model B**, following Watson and Solomon (1997), we consider a generalized interaction kernel H_G that consists of separable Gaussian functions which depend on the distance between the location of the sensors, $H_{\boldsymbol{p}}$, and on the difference between their scales and orientations, H_f and H_ϕ. Moreover, we extend the unit-norm Gaussian kernel already proposed in Watson and Solomon (1997) with additional weights in case extra inter-band tuning is needed:

$$H_G = \mathbb{D}_{\boldsymbol{c}} \cdot \left[H_{\boldsymbol{p}} \odot H_f \odot H_\phi \odot C_{\text{int}}\right] \cdot \mathbb{D}_{\boldsymbol{w}}, \tag{6}$$

where C_{int} is a subband-wise full matrix, $\mathbb{D}_{\boldsymbol{w}}$ is a diagonal matrix with vector $\boldsymbol{w}$ in the diagonal, and the normalization of each row of the kernel is controlled by a diagonal matrix $\mathbb{D}_{\boldsymbol{c}}$, which contains the vector of normalization constants, $\boldsymbol{c}$, in the diagonal. This means that the elements c_i normalize each interaction neighborhood, and the elements w_j control the relative relevance of the energies e_j before these are considered for the interaction.

In addition to the generalized kernel, the other distinct difference of **Model B** is the extra constant $K(\boldsymbol{e}^\star) = \left[\boldsymbol{\kappa} \odot \frac{\boldsymbol{b}+H_G \cdot \boldsymbol{e}^\star}{\boldsymbol{e}^\star}\right]$. This constant has a relevant qualitative rationale:

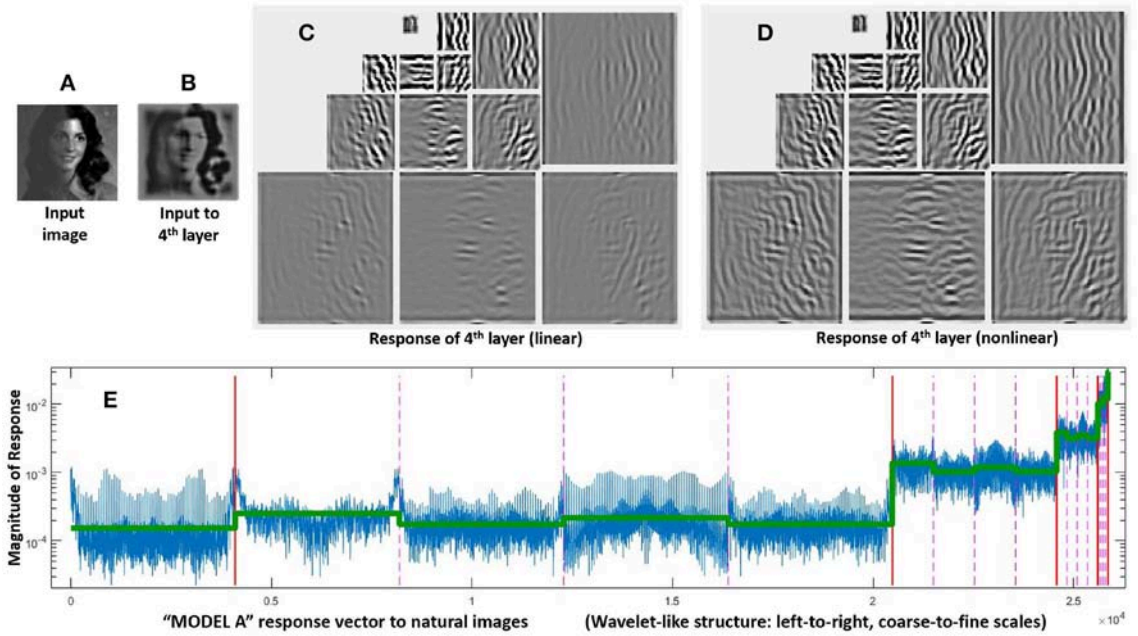


FIGURE 5 | Response of Model A to natural images. Given a luminance distribution, input image **(A)**, the initial layers of the model (retina-to-LGN) compute a filtered version of brightness contrast with adaptation to lower contrasts due to divisive normalization. That is why the contrast in the input to the 4th layer, image **(B)**, is more uniform than in the input image. Finally, the linear part of the 4th layer, wavelet diagram **(C)**, computes a multi-scale / multi-orientation decomposition and then, these responses nonlinearly interact as given by Equation (4), final responses in wavelet diagram **(D)**. The structure of a representative vector of responses depicted at the bottom is relevant to understand the assumed interactions and the eventual modifications that may be required. As usual in the wavelet literature (Simoncelli and Adelson, 1990), data in the vector are organized from high-frequency (fine scales at the left) to low-frequency (coarse scales at the right), wavelet vector **(E)**. Abscissas indicate the wavelet coefficient. The specific scale of the ordinate axis is not relevant. Solid vertical lines in red indicate the limits of the different scales. Within each scale, the dashed lines in pink indicate the limits of the different orientations. The different coefficients within each scale/orientation block correspond to different spatial locations. The line in green shows the average amplitude per subband for a set of natural images. As discussed in the text, this specific energy distribution per scale is relevant for the good performance of the model.

it keeps the response bounded *regardless of the choice for the other parameters*.

Note that, when the input energy, $\boldsymbol{e}$, arrives to the *reference value*, $\boldsymbol{e}^{\star}$, the response of **Model B** reduces to the vector $\boldsymbol{\kappa}$ regardless of model parameters. This simplifies the qualitative control of the dynamic range of the system because one may set a desired output $\boldsymbol{\kappa}$ (e.g., certain amplitudes per subband) for some relevant reference input $\boldsymbol{e}^{\star}$ regardless of the other parameters. This stabilization constant, $K(\boldsymbol{e}^{\star})$, does not modify the qualitative effect of the relevant parameters of the divisive normalization, but, as it constraints the dynamic range, it allows the modeler to freely play with the relevant parameters γ, $\boldsymbol{b}$, and H_G, and still preserve the relative amplitude of the subbands. And this freedom is particularly critical to understand the kind of modifications needed in the parameters to reproduce certain experimental trend.

Here we propose that $\boldsymbol{e}^{\star}$ is related to the average energy of the *input* to this nonlinear neural layer. Similarly, we propose to set the global scaling factor, $\boldsymbol{\kappa}$, according to a desired dynamic range in the *output* of this neural layer. These stabilization settings simplify the use of the model thus allowing to get the desired qualitative behavior even modifying the parameters *by hand*. Interestingly, this freedom to explore will reveal the modulation required in the conventional unit-norm Gaussian kernel.

3. RESULTS

In this section we show the performance of **Model A** and **Model B** in two scenarios: (a) reproducing subjective opinion in large-scale naturalistic databases using quadratic summation in Equation 3, and (b) obtaining meaningful contrast response curves for artificial stimuli.

The parameters of **Model A** are those obtained in Martinez-Garcia et al. (2018) to provide the best possible fit to the mean opinion scores on a large natural image database. These parameters of **Model A** are kept fixed throughout the simulations in this section. On the contrary, in the case of **Model B**, we start from a base-line situation in which we import the parameters from **Model A**, but afterwards, this naive guess is fine tuned to get reasonable response curves for the artificial stimuli considered above. Our goal is checking if the models account for the trends of masking described in **Figures 2**, **3**: we are not fitting actual experimental data but just refuting models that do not follow the qualitative trend.

In this model verification context, the fine tuning of **Model B** is done *by hand*: we just want to stress that while **Model A** cannot account for specific inter-band interactions, the interpretability of **Model B** when using the proper artificial stimuli makes it very easy to tune. And this intuitive tuning is

possible thanks to the stabilization effect of the constant $K(\boldsymbol{e}^\star)$ proposed above.

Nevertheless, it is important to stress that the Jacobian with regard to the parameters of **Model B** given in appendix (**Supplementary Material**) are implemented in the code associated to the paper. Therefore, despite the exploration of the responses in this section will be just qualitative, the code of **Model B** is ready for gradient descent tuning if one decides to measure the contrast incremental thresholds for the proper artificial stimuli.

Accurate control of spatial frequency, luminance, contrast and appropriate rendering of artificial stimuli can be done using the generic routines of `VistaLab` (Malo and Gutiérrez, 2014). In order to do so, one has to take into account a sensible sampling frequency (e.g., bigger than 60 cpd to avoid aliasing at visible frequencies) and the corresponding central frequencies and orientations of the selected wavelet filters in the model. The specific software used in this paper to generate the stimuli and to compute the response curves is available at: "http://isp.uv.es/docs/ArtificeReloaded.zip".

3.1. Success of "*Model A*" in Naturalistic Databases

Optimization of the width and amplitude of the Gaussian kernel, H, in each subband as well as the semisaturation parameters $\boldsymbol{b}$ in each subband of **Model A** led to the results in **Figure 6**. This was referred to as *optimization phase I* in Martinez-Garcia et al. (2018). Even though *optimization phase II* using the full variability in $\boldsymbol{b}$ led to higher correlations, here we restrict ourselves to *optimization phase I* because we want to keep the number of parameters small. Note that $\boldsymbol{b}$ has $2.5{\cdot}10^4$ elements but restricting to a single semisaturation per subband we only have 14 free parameters. In the *optimization phase I* only 1/25 of the TID database was used in the training.

As stated above, spatial-only intra-band relations leads to symmetric block diagonal kernels. Optimization acted on the width and amplitude of these kernels per subband. Similarly, optimization lead to bigger semisaturation for low frequencies except for the low-pass residual.

The performance of the resulting model on the naturalistic database is certainly good: compare the correlation of **Model A** with subjective opinion in **Figure 6** as opposed to the widely used Structural SIMilarity index (Wang et al., 2004), in red, considered here just as useful reference. Given the improvement in correlation with regard to SSIM, one can certainly say that **Model A** is *highly successful* in predicting the visibility of uncontrolled distortions seen on naturalistic backgrounds.

3.2. Relative Failure of "*Model A*" With Artificial Stimuli

Despite the reasonable formulation of **Model A** and its successful performance in reproducing subjective opinion in large-scale naturalistic databases, a simple simulation with the kind of artificial stimuli presented in section 2.1 shows that it does not reproduce all the aspects of basic visual masking.

Specifically, we computed the response curves of the highlighted neurons in **Figure 4** for low-frequency and high-frequency tests like those illustrated in **Figure 2** as a function of their contrast. We considered four different contrasts for the background. Different orientations of the background (vertical, diagonal and horizontal) were also considered.

Figure 7 presents the results of such simulation. This figure highlights some of the good features of **Model A**, but also its shortcomings.

On the positive side we have the following. First, the response increases with contrast as expected. Second, the response for the low frequency test is bigger than the response for the high frequency test (see the scale of the ordinate axis for the high frequency response). This is in agreement with the CSF. Third, the response saturates with contrast as expected. And also, increasing the contrast of the background decreases the responses.

However, *contrarily to what we can see when looking at the artificial stimuli*, the response for the high frequency test *does not* decay more on top of high frequency backgrounds. While the decay behavior is qualitatively ok for the low-frequency test, definitely it is not ok for the high-frequency test. Compare the decays of the signal at the circles highlighted in red in **Figure 7**: the response of the sensors tuned to high-frequency test decays by the same amount when they are presented on top of low-frequency backgrounds than when the background also has high-frequency. The model is failing here despite its good performance in the large database.

3.3. Success of "*Model B*" With Natural and Artificial Stimuli

The starting point of our heuristic exploration with **Model B** is a straightforward translation of **Model A** into **Model B**. We will refer to this as **Model B naive**. This starting point consists of importing the values of the parameters from **Model A** except for the modulations depending on the scale and orientation. Following Watson and Solomon (1997) we assumed reasonable interaction lengths of one octave (for scales) and 30 degrees (for orientation). We used no extra weights to break the symmetry ($C_{\text{int}} = \mathbb{1}$ is an all-ones matrix, and $C_{\boldsymbol{w}} = I$ is the identity). And the values for $\boldsymbol{c}$ and $\boldsymbol{b}$ also come from **Model A**. The parameters of this **Model B naive** are shown in **Figure 8** (left panels). The idea of this starting point, **Model B naive**, is reproducing the behavior of **Model A** to build on from there.

Results in **Figure 9** (top) and **Figure 10** (left) show that **Model B naive** certainly reproduces the behavior of **Model A**: both the success in the natural image database and the relative failure with artificial stimuli.

On top of kernel generalization, there is a second relevant intuition: modifications in the kernel may be ineffective if the semisaturation constants are too high. Note that the denominator of Divisive Normalization, Equation 4, is a balance between the linear combination $H \cdot \boldsymbol{e}$ and the vector $\boldsymbol{b}$. This means that some elements of $\boldsymbol{b}$ should be reduced for the subbands where we want to act. Increasing the corresponding elements of vector $\boldsymbol{c}$, leads to a similar effect.

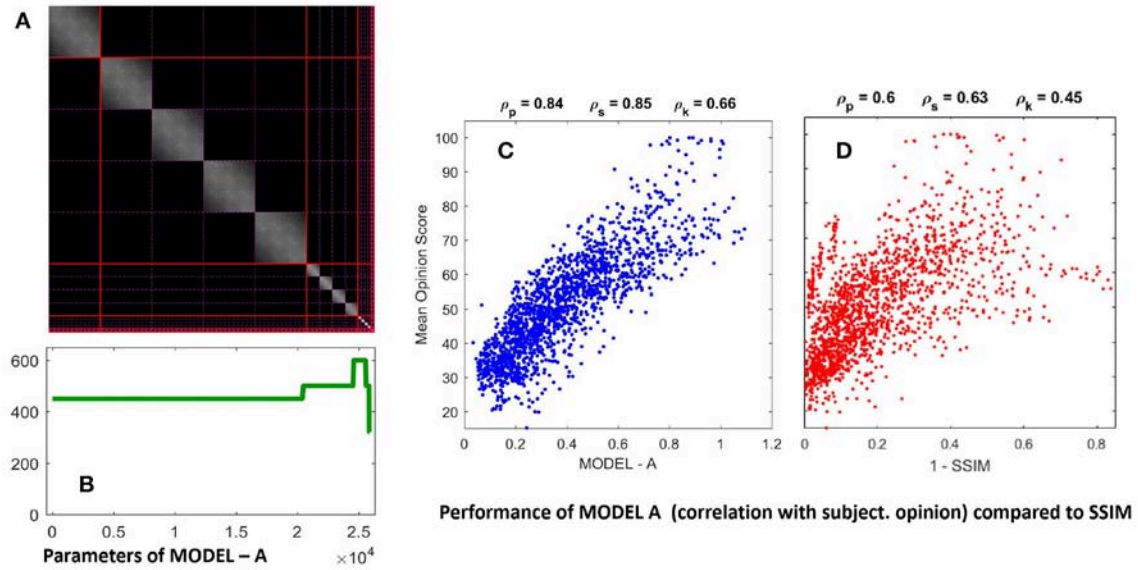


FIGURE 6 | Parameters of MODEL-A (left, **A,B**) and performance on large scale naturalistic database (right, **C,D**). The parameters are: the interaction kernel *H* (matrix on top, **A**), and the semisaturation per subband vector, $\boldsymbol{b}$. The structure of $\boldsymbol{b}$ is the same as a wavelet vectors in **Figure 5**. The kernel *H* describes how each wavelet coefficient interacts with the others, therefore, we included the solid and dashed lines in red and pink to highlight the limits between the subbands. The resulting submatrices describe the intra- and inter-subband interactions. The figures on top of the scatter plots are the Pearson, Spearman, and Kendall correlations. Here performance of Model A in plot **(C)** is compared with SSIM (Wang et al., 2004) in **(D)** just because it is the de-facto standard in image quality assessment.

With these intuitions one can start playing with H_G and $\boldsymbol{b}$. However, while the effect of the low-frequency is easy to reduce using the above ideas (thus solving the problem highlighted in red in **Figure 7**), the relative amplitude between the responses to low and high frequency inputs is also easily lost. This quickly ruins the low-pass CSF-like behavior and reduces the performance on the large-scale database. We should not lose the relative amplitudes of the responses of **Model A** to natural images (i.e., green lines in **Figure 5**) to keep its good performance. Unfortunately **Model A** is unstable under this kind of modifications making it difficult to tune. That is why it is necessary to include the constant $\left[\boldsymbol{\kappa} \odot \frac{\boldsymbol{b}+H_G \cdot \boldsymbol{e}^\star}{\boldsymbol{e}^\star}\right]$ in **Model B** to control the dynamic range of the responses.

Figure 8 (right panel) shows the fine-tuned parameters according to the heuristic suggested above: reduce semisaturation in certain bands and control the amplitude of the kernel in certain bands. This heuristic comes from the meaning of the blocks in the kernel and from the subbands that are activated by the different artificial stimuli. Note that we strongly reduced $\boldsymbol{b}$ and we applied bigger reductions for the high-frequency bands (which corresponds to the sensors we want to fix). In the same vein we increased the values for the global scale of the kernels of high frequencies $\boldsymbol{c}$ while reducing substantially these amplitudes for low-frequencies to preserve previous behavior, which was ok for low-frequencies. Finally, and more importantly, we moderated the effect of the low-frequencies in masking by using small weights for the low-frequency scales in $\boldsymbol{w}$, while increasing the values for high frequency. Note how this reduces the columns corresponding to the low-frequency subbands in the final kernel H_G, and the other way around for the high-frequency scales. This implies a bigger effect of high-frequency backgrounds in the attenuation of high-frequency sensors and reduces the effect of the low-frequency.

Results in **Figure 9** show that this fine-tuning fixes the qualitative problem detected in **Model A**, which was also present in **Model B naive**. We successfully modified the response of high-frequency sensors: see the decay in the green circles compared to the behavior in the red circles. Moreover, we introduced no major difference in the low-frequency responses, which already were qualitatively correct.

Moreover, **Figure 10** shows that the fine-tuned version of **Model B** not only works better for artificial stimuli but it also preserves the success in the natural image database. The latter is probably due to the positive effect of setting the relative magnitude of the responses in **Model B** as in **Model A** using the appropriate $K(\boldsymbol{e}^\star)$ (setting the output $\boldsymbol{\kappa}$ for the average input $\boldsymbol{e}^\star$).

It is interesting to stress that the solution to get the right qualitative behavior in the responses didn't require any extra weight in C_{int}, which remained an all-ones matrix. We only operated row-wise and column-wise with the diagonal matrices $\mathbb{D}_{\boldsymbol{c}}$ and $\mathbb{D}_{\boldsymbol{w}}$, respectively.

In summary, in order to fix the qualitative problems of **Model A** with masking of high-frequency patterns, the obvious use of generalized unit-norm inter-band kernels, as in Watson and Solomon (1997), was not enough: we had to consider the activation of the different subbands due to controlled artificial stimuli to tune the weights in the left- and right- diagonal matrices that modulate the unit-norm Gaussian kernels $H_G = \mathbb{D}_{\boldsymbol{c}} \cdot \left[H_{\boldsymbol{p}} \odot H_f \odot H_\phi\right] \cdot \mathbb{D}_{\boldsymbol{w}}$. It was necessary to include high-pass filters in $\boldsymbol{c}$ and $\boldsymbol{w}$ (see **Figure 8**, fine-tuned) to moderate the effect

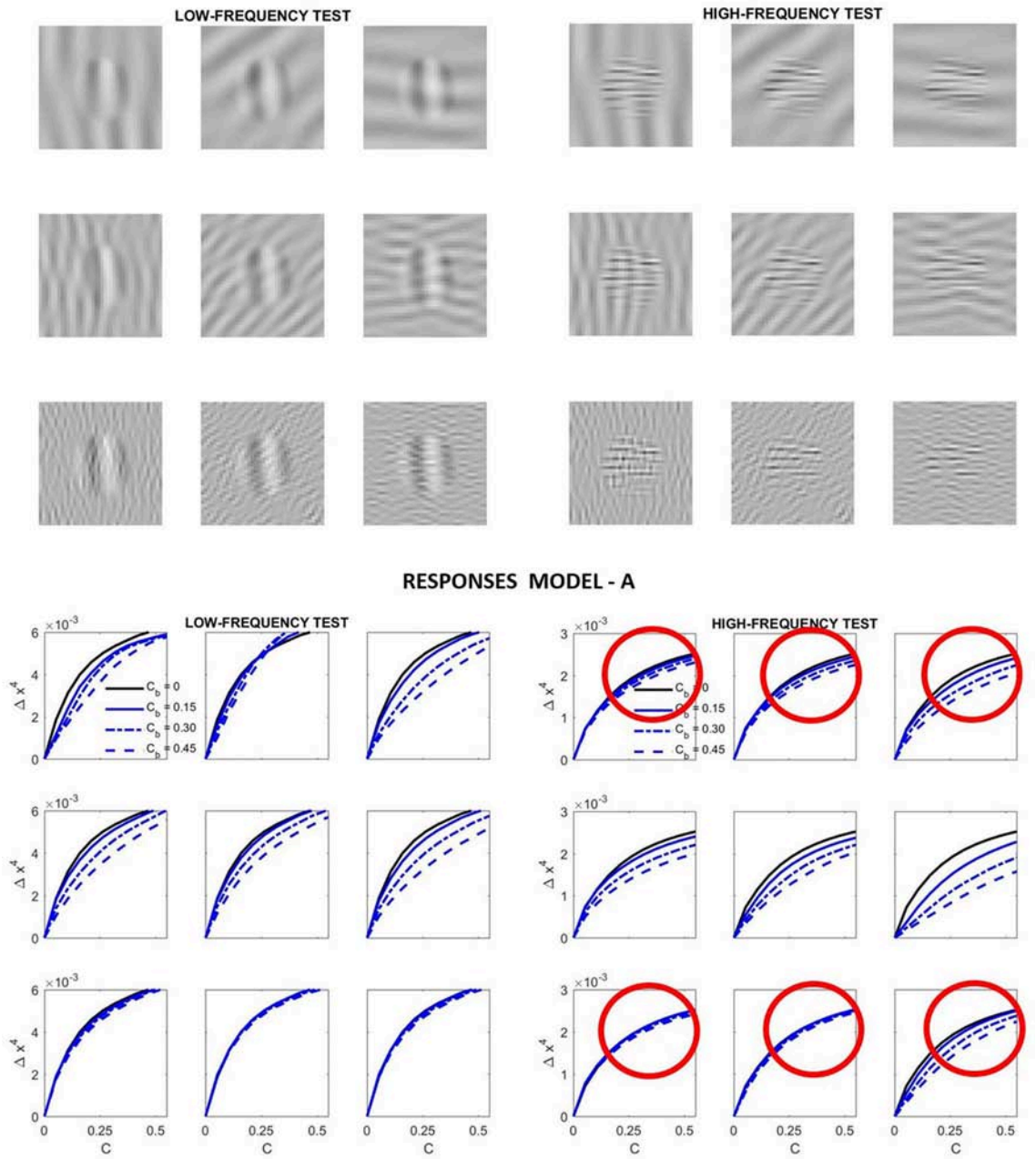


FIGURE 7 | Relative success and failures of **Model A** optimized on the large-scale database. Model-related stimuli such as the low-frequency and high-frequency tests shown on the top panel simplify the reproduction of results form model outputs and allow simple visual interpretation of results. In this simulation the response curves at the bottom panel are computed from the variation of the responses of the low-frequency and high-frequency sensors of the 4th layer highlighted in green in **Figure 4**. In each case, the variation of the response is registered as the contrast of the corresponding stimulus is increased. That is why we plot $\Delta \boldsymbol{x}^4$ vs. the contrast of the input, C. The different line styles represent the response for different contrast of the background, C_b. Simple visual inspection of the stimuli is enough to discard some of the predicted curves (e.g., those in red circles): the low frequency backgrounds *do not* mask the high frequency test more than the high frequency backgrounds.

of the low-frequency backgrounds on the masking of sensors tuned to high-frequencies.

The need of these extra filters can be interpreted in a interesting way: there should be a *balanced correspondence* between the linear filters and the interaction neighborhoods in the nonlinearity. Note that different choices for the filters to model the linear receptive fields in the cortex imply different energy distributions over the subbands[2]. In this situation, if the

[2]For instance, analyzing images by choosing Gabors or different wavelets, and by choosing different ways to sample the retinal and the frequency spaces, definitely leads to different distributions of the energy over the subbands.

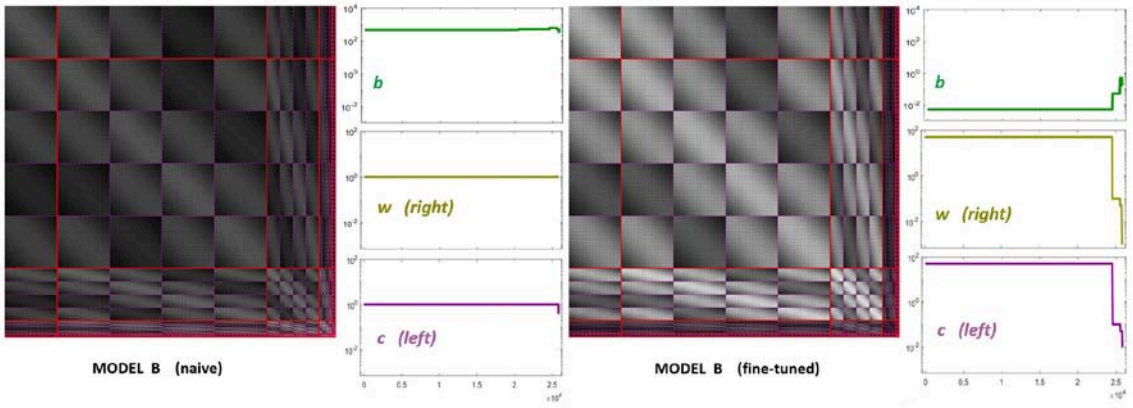


FIGURE 8 | Parameters of diferent versions of **Model B**. Left panel shows the parameters for the first guess of Model B: the interaction matrix, the semisaturation vector, $\boldsymbol{b}$, and the right and left filters (vectors $\boldsymbol{w}$ and $\boldsymbol{c}$ in Equation 6). This first-guess is called *naive* because the semisaturation and amplitudes of the kernel were imported directly from **Model A** in Martinez-Garcia et al. (2018). The panel at the right shows the corresponding parameters for the fine-tuned version of Model B, once we explored a range of values to fix the response curves. See text for details on how the parameters were tuned to get the desired responses.

energy in certain subband is overemphasized by the choice of the filters, the interaction neighborhoods should discount this fact.

Of course, more accurate tuning of **Model B** on actual exhaustive contrast incremental data of different tests+backgrounds may lead to more sophisticated weights in C_{int}. However, the simple toy simulation presented here using artificial stimuli with clear interpretation was enough to (a) discard **Model A**, (b) to point out the *balance problem* between the assumed linear cortical filters and the assumed interaction kernel in divisive normalization, and even (c) to propose an intuitive solution for the problem.

4. DISCUSSION

The relevant question is: *is the failure of **Model A** something that we could have expected?* And the unfortunate answer is, *yes*: the failure is not surprising given the (almost necessarily) imbalanced nature of large-scale databases. Note that it is not only that **Model A** is somewhat rigid[3], the fundamental problem is that the specific phenomenon is not present in the database with enough frequency or intensity to force the model to reproduce it in the learning stage.

Of course, this problem is hard to solve because it is not obvious to decide in advance the kind of phenomena (and the right amount of each one) that should be present in the database(s): as a result, databases are almost necessarily imbalanced and biased by the original intention of the creators of the database.

Here we made a full analysis (problem and route-to-solution) on texture masking, but note that focus on masking was just one important but arbitrary example to stress the main message. There are equivalent limitations affecting other parts of the optimized model that may come from the specific features of the database. For instance, the luminance-to-brightness transform (first layer in models A and B) is known to be strongly nonlinear and highly adaptive (Wyszecki and Stiles, 1982; Fairchild, 2013). It can be modeled using the canonical divisive normalization (Hillis and Brainard, 2005; Abrams et al., 2007) but also other alternative nonlinearities (Cyriac et al., 2016), and this nonlinearity has been shown to have relevant statistical effects (Laughlin, 1983; Laparra et al., 2012; Laparra and Malo, 2015; Kane and Bertalmio, 2016). However, when fitting layers 1st and 4th simultaneously to reproduce subjective opinion over the naturalistic database in Martinez-Garcia et al. (2018), even though we found a consistent increase in correlation, in the end, the behavior for the first layer turned out to be almost linear. The constant controlling the effect of the anchor luminance turned out to be very high. As a result, the nonlinear effect of the luminance is small. Again, one of the reasons for this result may be that the low dynamic range of the database did not require a stronger nonlinearity at the front-end given the rest of the layers. Similar effects could be obtained with the nonlinearities of color channels if the statistics is biased (MacLeod, 2003; Laparra and Malo, 2015).

The case studied here is not only a praise of artificial stimuli, but also a praise of *interpretable models*. When models are interpretable, it is easier to fix their problems from their failures on synthetic model-interpretable stimuli. For example, the solution we described here based on considering extra interaction between the sensors is not limited to *divisive* models of adaptation. Following Bertalmio et al. (2017), it may be also applied to other interpretable models such as the *subtractive* Wilson-Cowan equations (Wilson and Cowan, 1972; Bertalmio and Cowan, 2009). In this subtractive case one should tune the matrix that describes the relations between sensors. This kind of intuitive modifications in the architecture of the models

[3]It is true that **Model A** only included intra-band relations, but note also that, even though we wanted to introduce more general kernels in **Model B** for future developments, the solution to the qualitative problem considered here basically came from including $\mathbb{D}_{\boldsymbol{w}}$ in H (not from sophisticated cross-subband weights). The other ingredients, $\boldsymbol{b}$ and $\boldsymbol{c}$ were already present in **Model A**.

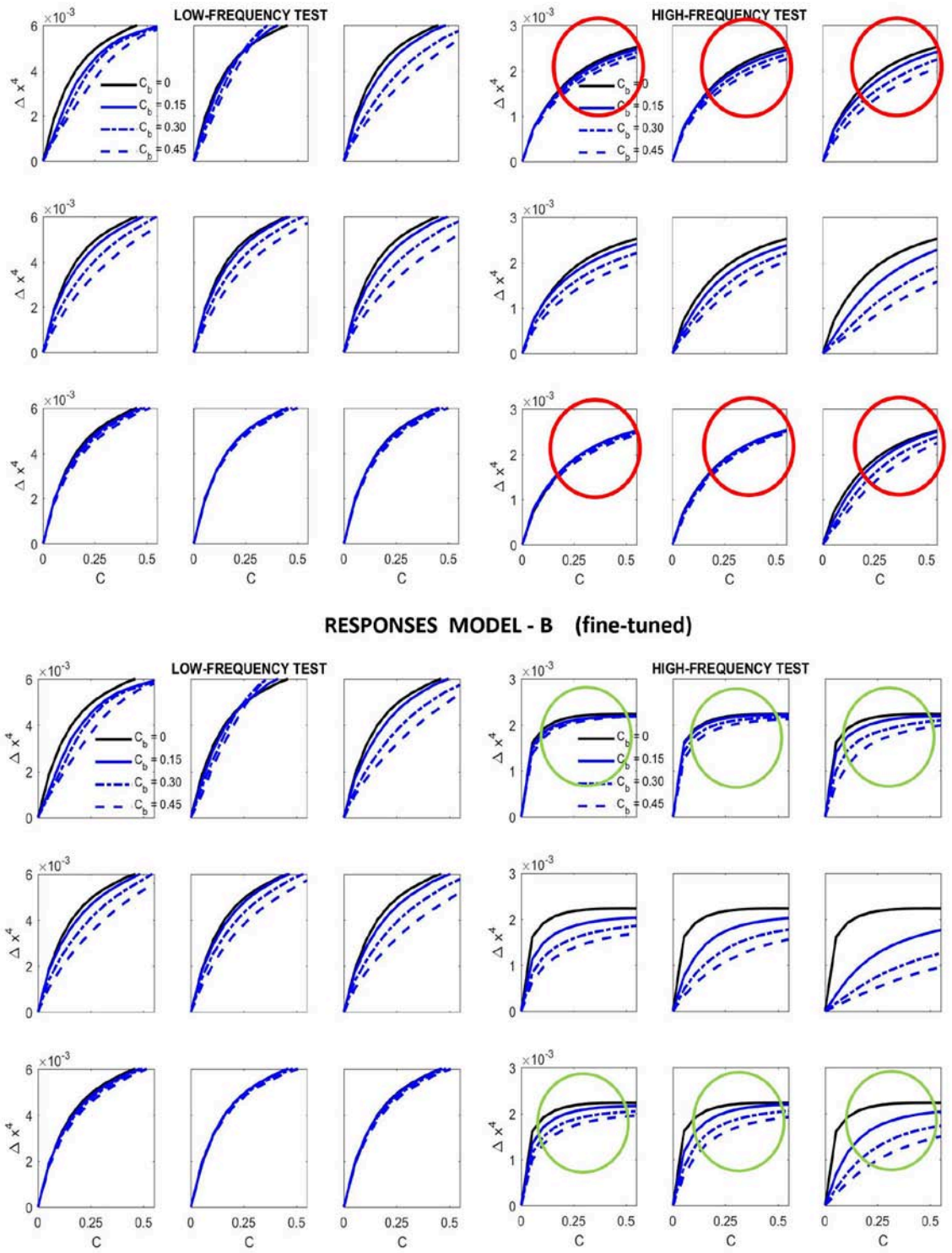


FIGURE 9 | Responses of different versions of **Model B** for the artificial stimuli. Curves correspond to the same stimuli considered in **Figure 7**.

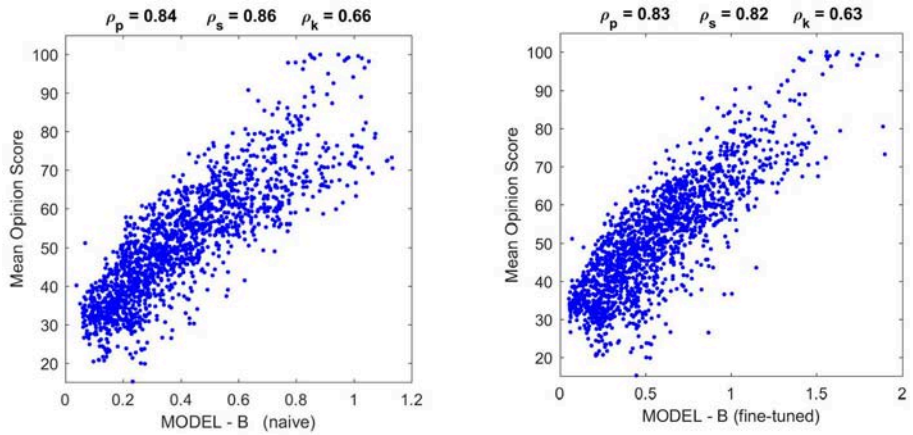


FIGURE 10 | Performance of different versions of **Model B** on the natural image database. The difference in correlation is not statistically significant according to the *F*-test used in Watson and Malo (2002), and the trend of the scatter plot is qualitatively the same.

would have been more difficult, if possible at all, with non-parametric data-driven methods. In fact, there is an active debate about the actual scientific gain of non-interpretable models, such as blind regression (Castelvecchi, 2016; Bohannon, 2017).

Finally, the masking curves considered in this paper also illustrate the fact that beyond the limitations of the database or the limitations of the architecture, the learning goal is also an issue. Note that, even using the same database and model, different learning goals may have different predictive power. For instance, other learning goals applied to natural images also give rise to cross-masking. Examples include information maximization (Schwartz and Simoncelli, 2001; Malo and Gutiérrez, 2006), and error minimization (Laparra and Malo, 2015). A systematic comparison between these different learning goals on the same database for a wide range of frequencies is still needed.

4.1. Consequence for Linear + Nonlinear Models: The Filter-Kernel Balance

Related to *model interpretability*, the results of our exploration with artificial stimuli suggests an interesting conclusion when dealing with linear+nonlinear models: *matching linear filters and non-linear interaction is not trivial*. Remember the *wavelet-kernel balance problem* described at the end of the results. Therefore, in building these models, one should not take filters and kernels off the shelf.

One may take this *balance problem* as another routinary parameter to tune. However, this *balance problem* may actually question the nature of divisive normalization in terms of other models. For instance, in Malo and Bertalmio (2018) we show that the divisive normalization may be seen as the stationary solution of *lower-level* Wilson-Cowan dynamics that do use a sensible unit-norm Gaussian interaction between units. This kind of questions are only raised, and solutions may be proposed, when testing interpretable models with model-related stimuli.

4.2. Using Naturalistic Databases Is Always a Problem?

Our criticism of naturalistic databases because their eventual imbalance and the problem in interpreting complicated stimuli in terms of models does not mean that we claim for an absolute rejection of these naturalistic databases. The case we studied here only suggests that one should not use the databases *blindly* as the only source of information, but in appropriate combination with well-selected artificial stimuli.

The use of carefully selected artificial stimuli may be considered as a safety-check of biological plausibility. Of course, our intention with the case studied here was not exhausting the search possibilities to claim that we obtained some sort of optimal solution. Instead, we just wanted to stress the fact that using the appropriate stimuli it is easy to propose modifications of the model that go in the right (biologically meaningful) direction, and still represent a competitive solution for the naturalistic database. This is an intuitive way to jump to other local minima which may be more biologically plausible in a very different region of the parameter space.

A sensible procedure would be alternating different learning epochs using natural and artificial data: while the large-scale naturalistic databases coming from the *image processing* community may enforce the main trends of the system, the specific small-scale artificial stimuli coming from the *vision science* community will fine-tune that first order approximation so that the resulting model has the appropriate features revealed by more specific experiments. In this context, standardization efforts such as those done by the CIE and the OSA organizations are really important to make this double-check. Examples include the data supporting the standard color observer (Smith and Guild, 1931; Stockman, 2017) and the standard spatial observer (Ahumada, 1996).

From a more general perspective, *image processing* applications do have a fundamental interest in *visual neuroscience* because these applications put into a broader context the relative relevance of the different phenomena described by classical

psychophysics or physiology. For instance, one can check the variations in performance by testing vision models of different complexity, e.g., with or without this or that nonlinearity accounting for some specific perceptual effect/ability. This approach oriented to check different perceptual modules in specific applications has been applied in image quality databases (Watson and Malo, 2002), but also in other domains such as perceptual image and video compression (Malo et al., 2000a,b, 2001, 2006), or in perceptual image denoising and enhancement (Gutiérrez et al., 2006; Bertalmio, 2014). These different applications show the relative relevance of improvements in masking models with regard to better CSFs or including more sensible motion estimation models in front of better texture perception models.

4.3. Are All the Databases Created Equal?

The case analyzed in this work illustrates the effect of (naively) using a database where texture masking is probably under-represented. The lesson to learn is that one has to take into account the phenomena for which database was created, or, equivalently, the absence of specific phenomena to address.

With this in mind, one could imagine what kind of artificial stimuli are needed to improve the results. Or alternatively, which other naturalistic databases are required as complementary check since they are more focused on other kind of perceptual behavior.

Some examples to illustrate this point: databases with controlled observation distance or accurate chromatic calibration such as Pedersen (2015) are more appropriate to set the spatial frequency bandwidth of the models in achromatic and chromatic channels. Databases with spectrally controlled illumination pairs (Laparra et al., 2012; Gutmann et al., 2014; Laparra and Malo, 2015) are appropriate to address chromatic adaptation models. Databases with high-dynamic range (Korshunov et al., 2015; Cerda-Company et al., 2016) will be more appropriate to point out the need of the nonlinearity of brightness perception. Finally, databases where visibility of incremental patterns was carefully controlled in contrast terms (Alam et al., 2014) are the best option to fit masking models as opposed to generic subjectively-rated image distortion databases.

4.4. Final Remarks

Previous literature (Rust and Movshon, 2005) criticized the use of too complex natural stimuli in vision science experiments because the statistics of such stimuli are difficult to control and conclusions may be biased by the interaction between this poorly controlled input and the complexities of the neural model under consideration.

In line with such precautions on the use of natural stimuli, here we make a different point: the general criticism to blind use of machine learning in large-scale databases (related to the proper balance in the data) also applies when using subjectively rated image databases to fit vision models. Using a variety of natural scenarios and distortions cannot guarantee that specific behaviors are properly represented, thus remaining hidden in the vast amount of data. In such situation, models that seem to have the right structure may miss these basic phenomena. Instead of trying to explicitly include model-oriented artificial stimuli in the large database to fix the unbalance, it is easier to address the issue by using the model-oriented artificial stimuli in illustrative experiments specifically intended to test some parameters of the model.

The case study considered here suggests that artificial stimuli, motivated by specific phenomena or by features of the model, may help both to (a) stress the problems that remain in models fitted to imbalanced natural image databases, and (b) to suggest modifications in the models. Incidentally, this is also an argument in favor of interpretable parametric models as opposed to data-driven pure-regression models. A sensible procedure to fit general purpose vision models would be alternating different fitting strategies using (a) uncontrolled natural stimuli, but also (b) well-controlled artificial stimuli to check the biological plausibility at each point.

In conclusion, predicting subjective distances between images may be a trivial regression problem, but using these large-scale databases to fit plausible models may take more than that: for instance, a vision scientist in the loop doing the proper fine-tuning of interpretable models using the classical artificial stimuli.

AUTHOR CONTRIBUTIONS

JM conceived the work, prepared the data and code for the experiments, and contributed to the interpretation of the results and manuscript writing. MM-G ran the experiments. MB contributed to the manuscript writing and to the criticism of blind machine-learning-like approaches.

FUNDING

This work was partially funded by the Spanish and EU FEDER fund through the MINECO/FEDER/EU grants TIN2015-71537-P and DPI2017-89867-C2-2-R; and by the European Union's Horizon 2020 research and innovation programme under grant agreement number 761544 (project HDR4EU) and under grant agreement number 780470 (project SAUCE).

ACKNOWLEDGMENTS

This work was conceived in La Fabrica de Hielo (Malvarrosa) after the reaction of Dr. C.A. Parraga to VanRullen (2017): scientists cannot be easily substituted by machines.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2019.00008/full#supplementary-material

REFERENCES

Abrams, A. B., Hillis, J. M., and Brainard, D. H. (2007). The relation between color discrimination and color constancy: When is optimal adaptation task dependent? *Neural Comput.* 19, 2610–2637. doi: 10.1162/neco.2007.19.10.2610

Ahumada, A. E. A. (1996). *OSA Modelfest Dataset*. Available online at: https://visionscience.com/data/modelfest/index.html

Alam, M. M., Vilankar, K., Field, D., and D.M., C. (2014). Local masking in natural images: a database and analysis. *J. Vis.* 14:22. doi: 10.1167/14.8.22

Attneave, F. (1954). Some informational aspects of visual perception. *Psychol. Rev.* 61, 183–193. doi: 10.1037/h0054663

Barlow, H. (1959). "Sensory mechanisms, the reduction of redundancy, and intelligence," in *Proceedings of the National Physical Laboratory Symposium on the Mechanization of Thought Process* (London, UK), 535–539.

Berardino, A., Laparra, V., Ballé, J., and Simoncelli, E. (2017). "Eigen-distortions of hierarchical representations," in *Advances in Neural Information Processing Systems*, Vol. 30, 3533–3542. Available online at: https://papers.nips.cc/

Bertalmio, M. (2014). From image processing to computational neuroscience: a neural model based on histogram equalization. *Front. Comput. Neurosci.* 8:71. doi: 10.3389/fncom.2014.00071

Bertalmio, M., and Cowan, J. (2009). Implementing the retinex algorithm with wilson-cowan equations. *J. Physiol. Paris* 103, 69–72. doi: 10.1016/j.jphysparis.2009.05.001

Bertalmio, M., Cyriac, P., Batard, T., Martinez-Garcia, M., and Malo, J. (2017). The wilson-cowan model describes contrast response and subjective distortion. *J. Vision* 17:657. doi: 10.1167/17.10.657

Bodrogi, P., Bovik, A., Charrier, C., Fernandez-Maloigne, C., Hardeberg, J., Larabi, M., et al. (2016). *A Survey About Image and Video Quality Evaluation Metrics*. Technical report of division 8: Image technology, Commission Internationale de l'Eclariage (CIE), Viena.

Bohannon, J. (2017). The cyberscientist. *Science* 357, 18–21. doi: 10.1126/science.357.6346.18

Bosse, S., Maniry, D., Mller, K., Wiegand, T., and Samek, W. (2018). Deep neural networks for no-reference and full-reference image quality assessment. *IEEE Trans. Image Process.* 27, 206–219. doi: 10.1109/TIP.2017.2760518

Campbell, F. W., and Robson, J. (1968). Application of Fourier analysis to the visibility of gratings. *J. Physiol.* 197, 551–566. doi: 10.1113/jphysiol.1968.sp008574

Carandini, M., and Heeger, D. (1994). Summation and division by neurons in visual cortex. *Science* 264, 1333–1336. doi: 10.1126/science.8191289

Carandini, M., and Heeger, D. J. (2012). Normalization as a canonical neural computation. *Nat. Rev. Neurosci.* 13, 51–62. doi: 10.1038/nrn3136

Castelvecchi, D. (2016). Can we open the black box of AI? *Nature* 538, 20–23. doi: 10.1038/538020a

Cavanaugh, J. R. (2000). *Properties of the Receptive Field Surround in Macaque Primary Visual Cortex*. Ph.D. Thesis, Center for Neural Science, New York University.

Cerda-Company, X., Parraga, C., and Otazu, X. (2016). Which tone-mapping operator is the best? A comparative study of perceptual quality. arXiv:1601.04450.

Cyriac, P., Kane, D., and Bertalmio, M. (2016). Optimized tone curve for in-camera image processing. *IST Electron. Imaging Conf.* 13, 1–7. doi: 10.2352/ISSN.2470-1173.2016.13.IQSP-012

Dayan, P., and Abbott, L. F. (2005). *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems*. Cambridge, MA: The MIT Press.

Fairchild, M. (2013). *Color Appearance Models*. Sussex, UK: The Wiley-IS&T Series in Imaging Science and Technology.

Foley, J. (1994). Human luminance pattern mechanisms: masking experiments require a new model. *J. Opt. Soc. Am. A* 11, 1710–1719. doi: 10.1364/JOSAA.11.001710

Ghadiyaram, D., and Bovik, A. C. (2016). Massive online crowdsourced study of subjective and objective picture quality. *IEEE Trans. Image Process.* 25, 372–387. doi: 10.1109/TIP.2015.2500021

Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning*. MIT Press. Available online at: http://www.deeplearningbook.org

Graham, N. (1989). *Visual Pattern Analyzers*. Oxford, UK: Oxford University Press.

Gutiérrez, J., Ferri, F. J., and Malo, J. (2006). Regularization operators for natural images based on nonlinear perception models. *IEEE Trans. Image Process.* 15, 189–200. doi: 10.1109/TIP.2005.860345

Gutmann, M. U., Laparra, V., Hyvärinen, A., and Malo, J. (2014). Spatio-chromatic adaptation via higher-order canonical correlation analysis of natural images. *PLoS ONE* 9:e86481. doi: 10.1371/journal.pone.0086481

Hillis, J. M., and Brainard, D. (2005). Do common mechanisms of adaptation mediate color discrimination and appearance? *JOSA A* 22, 2090–2106. doi: 10.1364/JOSAA.22.002090

Kane, D., and Bertalmio, M. (2016). System gamma as a function of image-and monitor-dynamic range. *J. Vis.* 16:4. doi: 10.1167/16.6.4

Korshunov, P., Hanhart, P., Richter, T., Artusi, A., Mantiuk, R., and Ebrahimi, T. (2015). "Subjective quality assessment database of HDR images compressed with JPEG XT," in *Proceedings of the 7th International Workshop Qual. Multimed. Exp. (QoMEX)* (Pilos).

Laparra, V., Berardino, A., Balle, J., and Simoncelli, E. (2017). Perceptually optimized image rendering. *JOSA A* 34, 1511–1525. doi: 10.1364/JOSAA.34.001511

Laparra, V., Jiménez, S., Camps-Valls, G., and Malo, J. (2012). Nonlinearities and adaptation of color vision from sequential principal curves analysis. *Neural Comput.* 24, 2751–2788. doi: 10.1162/NECO_a_00342

Laparra, V., and Malo, J. (2015). Visual aftereffects and sensory nonlinearities from a single statistical framework. *Front. Hum. Neurosci.* 9:557. doi: 10.3389/fnhum.2015.00557

Laparra, V., Muñoz-Marí, J., and Malo, J. (2010). Divisive normalization image quality metric revisited. *JOSA A* 27, 852–864. doi: 10.1364/JOSAA.27.000852

Larson, E. C., and Chandler, D. M. (2010). Most apparent distortion: full-reference image quality assessment and the role of strategy. *J. Electron. Imaging* 19:011006. doi: 10.1117/1.3267105

Laughlin, S. B. (1983). "Matching coding to scenes to enhance efficiency," in *Physical and Biological Processing of Images*, eds O. J. Braddick and A. C. Sleigh (Berlin: Springer), 42–52.

Legge, G. (1981). A power law for contrast discrimination. *Vis. Res.* 18, 68–91. doi: 10.1016/0042-6989(81)90092-4

Ma, K., Liu, W., Zhang, K., Duanmu, Z., Wang, Z., and Zuo, W. (2018). End-to-end blind image quality assessment using deep neural networks. *IEEE Trans. Image Process.* 27, 1202–1213. doi: 10.1109/TIP.2017.2774045

MacLeod, D. A. (2003). "Colour discrimination, colour constancy, and natural scene statistics," in *Normal and Defective Colour Vision*, eds J. Mollon, J. Pokorny, and K. Knoblauch (Oxford, UK: Oxford University Press), 189–218.

Malo, J., and Bertalmio, M. (2018). Appropriate kernels for divisive normalization explained by Wilson-Cowan equations. arXiv:1804.05964.

Malo, J., Epifanio, I., Navarro, R., and Simoncelli, E. P. (2006). Nonlinear image representation for efficient perceptual coding. *IEEE Trans. Image Process.* 15, 68–80. doi: 10.1109/TIP.2005.860325

Malo, J., Ferri, F., Albert, J., Soret, J., and Artigas, J. (2000a). The role of perceptual contrast non-linearities in image transform quantization. *Image Vision Comput.* 18, 233–246. doi: 10.1016/S0262-8856(99)00010-4

Malo, J., Ferri, F., Gutierrez, J., and Epifanio, I. (2000b). Importance of quantiser design compared to optimal multigrid motion estimation in video coding. *Electr. Lett.* 36, 807–809. doi: 10.1049/el:20000645

Malo, J., and Gutiérrez, J. (2006). V1 non-linear properties emerge from local-to-global non-linear ICA. *Network* 17, 85–102. doi: 10.1080/09548980500439602

Malo, J., and Gutiérrez, J. (2014). *VistaLab: The Matlab Toolbox for Spatio-temporal Vision Models*. Available online at: http://isp.uv.es/code/visioncolor/vistalab.html

Malo, J., Gutiérrez, J., Epifanio, I., Ferri, F. J., and Artigas, J. M. (2001). Perceptual feedback in multigrid motion estimation using an improved dct quantization. *IEEE Trans. Im. Proc.* 10, 1411–1427. doi: 10.1109/83.951528

Malo, J., and Laparra, V. (2010). Psychophysically tuned divisive normalization approximately factorizes the pdf of natural images. *Neural Comput.* 22, 3179–3206. doi: 10.1162/NECO_a_00046

Martinez-Garcia, M., Cyriac, P., Batard, T., Bertalmio, M., and Malo, J. (2018). Derivatives and inverse of cascaded linear+nonlinear neural models. *PLoS ONE* 13:e0201326. doi: 10.1371/journal.pone.0201326

Moorthy, A. K., and Bovik, A. C. (2010). A two-step framework for constructing blind image quality indices. *IEEE Signal Process. Lett.* 17, 513–516. doi: 10.1109/LSP.2010.2043888

Moorthy, A. K., and Bovik, A. C. (2011). Blind image quality assessment: from natural scene statistics to perceptual quality. *IEEE Trans. Image Process.* 20, 3350–3364. doi: 10.1109/TIP.2011.2147325

Pedersen, M. (2015). "Evaluation of 60 full-reference image quality metrics on the cid:iq," in *2015 IEEE International Conference on Image Processing (ICIP)* (Quebec, QC), 1588–1592. doi: 10.1109/ICIP.2015.7351068

Ponomarenko, N., Carli, M., Lukin, V., Egiazarian, K., Astola, J., and Battisti, F. (2008). "Color image database for evaluation of image quality metrics," in *Proceedings of the international Workshop on Multimedia Signal Processing* (Cairns, QLD), 403–408.

Ponomarenko, N., Jin, L., Ieremeiev, O., Lukin, V., Egiazarian, K., Astola, J., et al. (2015). Image database TID2013: peculiarities, results and perspectives. *Signal Process.* 30(Suppl. C):57–77.

Ponomarenko, N., Lukin, V., Zelensky, A., Egiazarian, K., Astola, J., Carli, M., et al. (2009). TID2008 - a database for evaluation of full-reference visual quality assessment metrics. *Adv. Mod. Radioelectr.* 10, 30–45.

Rust, N. C., and Movshon, J. A. (2005). In praise of artifice. *Nat. Neurosci.* 8, 1647–1650. doi: 10.1038/nn1606

Saad, M. A., Bovik, A. C., and Charrier, C. (2010). A dct statistics-based blind image quality index. *IEEE Signal Process. Lett.* 17, 583–586. doi: 10.1109/LSP.2010.2045550

Saad, M. A., Bovik, A. C., and Charrier, C. (2012). Blind image quality assessment: a natural scene statistics approach in the DCT domain. *IEEE Trans. Image Process.* 21, 3339–3352. doi: 10.1109/TIP.2012.2191563

Saad, M. A., Bovik, A. C., and Charrier, C. (2014). Blind prediction of natural video quality. *IEEE Trans. Image Process.* 23, 1352–1365. doi: 10.1109/TIP.2014.2299154

Sakrison, D. J. (1977). On the role of the observer and a distortion measure in image transmission. *IEEE Trans. Commun.* 25, 1251–1267. doi: 10.1109/TCOM.1977.1093773

Schwartz, O., and Simoncelli, E. (2001). Natural signal statistics and sensory gain control. *Nat. Neurosci.* 4, 819–825. doi: 10.1038/90526

Shakhnarovich, G., Batra, D., Kulis, B., and Weinberger, K. (2011). "Beyond mahalanobis: supervised large-scale learning of similarity," in *NIPS Workshop on Metric Learning* (Granada: Sierra Nevada).

Simoncelli, E., and Adelson, E. (1990). "Subband transforms," in *Subband Image Coding* (Norwell, MA: Kluwer Academic Publishers), 143–192.

Simoncelli, E. P., Freeman, W. T., Adelson, E. H., and Heeger, D. J. (1992). Shiftable multi-scale transforms. *IEEE Trans. Inform. Theory* 38, 587–607. doi: 10.1109/18.119725

Smith, T., and Guild, J. (1931). The C.I.E. colorimetric standards and their use. *Trans. Opt. Soc.* 33:73. doi: 10.1088/1475-4878/33/3/301

Stockman, A. (2017). *Colour and Vision Research Laboratory Databases*. Available online at: http://www.cvrl.org/

Taubman, D. S., and Marcellin, M. W. (2001). *JPEG 2000: Image Compression Fundamentals, Standards and Practice*. Norwell, MA: Kluwer Academic Publishers.

Teo, P., and Heeger, D. (1994). Perceptual image distortion. *Proc. SPIE* 2179, 127–141. doi: 10.1117/12.172664

VanRullen, R. (2017). Perception science in the age of deep neural networks. *Front. Psychol.* 8:142. doi: 10.3389/fpsyg.2017.00142

Wang, Z., and Bovik, A. C. (2009). Mean squared error: love it or leave it? A new look at signal fidelity measures. *IEEE Signal Process. Mag.* 26, 98–117. doi: 10.1109/MSP.2008.930649

Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P. (2004). Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Im. Proc.* 13, 600–612. doi: 10.1109/TIP.2003.819861

Watson, A. B. (ed.). (1993). *Digital Images and Human Vision*. Cambridge, MA: MIT Press.

Watson, A. B., and Malo, J. (2002). "Video quality measures based on the standard spatial observer," in *IEEE Proceedings of the International Conference Im. Proc.* Vol. 3 (Rochester, NY), III–41.

Watson, A. B., and Solomon, J. (1997). A model of visual contrast gain control and pattern masking. *JOSA A* 14, 2379–2391. doi: 10.1364/JOSAA.14.002379

Webster, A., Pinson, M., and Brunnstrm, K. (2001). *Video Quality Experts Group Database*. Available online at: https://www.its.bldrdoc.gov/vqeg/downloads.aspx

Wilson, H. R., and Cowan, J. (1972). Excitatory and inhibitory interactions in localized populations of model neurons. *Biophys. J.* 12, 1–24. doi: 10.1016/S0006-3495(72)86068-5

Wyszecki, G., and Stiles, W. (1982). *Color Science: Concepts and Methods, Quantitative Data and Formulae*. New York, NY: John Wiley & Sons.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Computational Neuroscience

ORIGINAL RESEARCH
published: 26 February 2019
doi: 10.3389/fncom.2019.00003

An Extreme Value Theory Model of Cross-Modal Sensory Information Integration in Modulation of Vertebrate Visual System Functions

Sreya Banerjee[1], *Walter J. Scheirer*[1] *and Lei Li*[2]*

[1] Department of Computer Science and Engineering, University of Notre Dame, Notre Dame, IN, United States, [2] Department of Biological Sciences, University of Notre Dame, Notre Dame, IN, United States

We propose a computational model of vision that describes the integration of cross-modal sensory information between the olfactory and visual systems in zebrafish based on the principles of the statistical extreme value theory. The integration of olfacto-retinal information is mediated by the centrifugal pathway that originates from the olfactory bulb and terminates in the neural retina. Motivation for using extreme value theory stems from physiological evidence suggesting that extremes and not the mean of the cell responses direct cellular activity in the vertebrate brain. We argue that the visual system, as measured by retinal ganglion cell responses in spikes/sec, follows an extreme value process for sensory integration and the increase in visual sensitivity from the olfactory input can be better modeled using extreme value distributions. As zebrafish maintains high evolutionary proximity to mammals, our model can be extended to other vertebrates as well.

Keywords: cross-modal sensory integration, statistical extreme value theory, classification, olfaction, vision, zebrafish

OPEN ACCESS

Edited by:
Hagit Hel-Or,
University of Haifa, Israel

Reviewed by:
Yin Tian,
Chongqing University of Posts and Telecommunications, China
Timothy Matthew Otchy,
Boston University, United States

***Correspondence:**
Lei Li
li.78@nd.edu

Received: *29 August 2018*
Accepted: *16 January 2019*
Published: *26 February 2019*

Citation:
Banerjee S, Scheirer WJ and Li L (2019) An Extreme Value Theory Model of Cross-Modal Sensory Information Integration in Modulation of Vertebrate Visual System Functions. Front. Comput. Neurosci. 13:3. doi: 10.3389/fncom.2019.00003

1. INTRODUCTION

The brain perceives the external world through an integration of stimuli received from different sensory modalities like vision, olfaction, and audition via the centrifugal pathway. A recent study taking inspiration from Cajal's original work on brain mapping (Gire et al., 2013) describes current knowledge of the centrifugal olfactory and visual pathways in mammalian species as being incomplete. While, for instance, the signaling pathways mediating brain feedback in human olfaction have been characterized, the origins and effects of signals to visual system functions remain to be examined. In this work, we seek to understand the modulation of the circuits between sensory modalities. A crucial observation, yielding from our own work, points to how due to olfacto-visual sensory integration, measures of visual performance or behavior in response to multi-sensory input are enhanced, when a stimulus in one modality is ambiguous or undetermined. In fact, in all vertebrate species (e.g., teleost, reptiles, birds, rodents, primates) examined thus far, the retina receives brain feedback through the centrifugal visual pathways (Harter and Aine, 1984; Mick et al., 1993; Gastiner et al., 2004). Depending on the species under consideration, the centrifugal pathways may originate from different parts of brain, such as the pre-tectal cortex, isthmo-optic nucleus, thalamus, or olfactory bulb.

In zebrafish (*Danio rerio*), the olfacto-retinal centrifugal (ORC) pathway originates from terminalis neurons (TNs) in the olfactory bulb (OB) and terminates in retina. TNs (**Figure 1A**) synthesize gonadotropin-releasing hormone (GnRH) as a major neurotransmitter. In the retina, TN fibers synapse with dopaminergic interplexiform cells (DA-IPCs), retinal ganglion cells (RGCs), and possibly other retinal cell types. Insights from relatively recent research (Li and Dowling, 2000; Huang et al., 2005) have shown that the function of the ORC pathway is directly regulated by the olfactory input. TN input alters GnRH signaling transduction and decreases dopamine release in the retina, thereby increasing outer retinal sensitivity and inner retinal activity (e.g., firing of ganglion cells). Specifically, the olfactory input mediated by the ORC pathway decreases the light threshold (i.e., the minimum light intensity required to fire evoked action potentials) of retinal ganglion cells, and thereby increases retinal sensitivity. Together, the olfactory input amplifies behavioral visual sensitivity (Maaswinkel and Li, 2003).

Zebrafish maintain high evolutionary proximity to mammals, and their retinas share great similarities to humans (e.g., structure, cellular organization, neural circuitry and signaling transmission) (Li, 2001; Vacaru et al., 2014). While much progress has been made to understand the anatomy of cross-modal circuitry in zebrafish, our knowledge of the underlying regulatory mechanism and physiological roles of centrifugal input to the retina is still in its nascent stage. Interestingly, Huang et al. (2005) demonstrate how the visual sensitivity in zebrafish is increased in the presence of olfactory signals whereas disrupting the ORC pathway impairs visual function. An important observation found in that work reveals the importance of olfactory signals for vision. According to Huang et al. (2005), under normal conditions the minimum threshold light intensity to invoke a retinal ganglion cell response (measured in spikes/sec) in a dark-adapted zebrafish embryo may decrease 1–2 log units after olfactory stimulation. This demonstrates the dramatic impact of olfactory signals on vision.

Such a sudden gain in visual sensitivity through olfactory stimulation is an intriguing target for a computational model. We argue that visual sensitivity follows the statistical Extreme Value Theory (EVT). The mean visual sensitivity does not clearly explain the increased sensitivity due to olfactory signals since that scenario is able to sense a stimulus that is an extreme aberration from the norm, i.e., retinal ganglion cell responses without any olfactory stimulation. EVT lays solid groundwork for modeling as it is independent of the underlying distribution of data (all of the cell responses) and is only applicable to the tails of the distribution (the extremes) such that samples which have the least, or no possible, probability of occurrence under a central tendency model are distinguished, providing greater discrimination while requiring few statistical assumptions.

At a deeper level, one can ask the following question: is there a theoretical justification for using EVT for neural modeling? Our key insight is that the characterization of the firing behavior of a neuron as repeated integration/thresholding within a circuit suggests positive answers to these questions. Neurons are generally modeled as an electro-chemical process integrating input (ions) and eventually crossing a threshold whereby they fire and release ions. We posit that this inherently leads to an EVT-based model because the distribution of samples that exceed a threshold T likely yields an extreme value distribution (EVD). If all neurons use a fixed threshold T, the inputs to subsequent neurons in the circuit must follow an EVD, with each neuron integrating data from such a distribution and thresholding it. Thus, EVT can provide a plausible consistent multi-layer neuron model.

Beyond the merits of cultivating a better understanding of the operation of cross-modal sensory information integration in vertebrates, there is the possibility that an accurate computational model for this phenomenon could translate into a general algorithm for pattern recognition tasks in computer science. A direct application of this method lies in the development of novel information fusion algorithms that leverage inputs from multiple sensory modalities, i.e., vision and audition (Nagrani et al., 2018). Another practical application is the invention of innovative sensors capable of detecting changes in the environment and then re-configuring on the fly to change operational parameters and power consumption requirements. Currently, sensors are typically designed to sense a single type of physical property such as temperature, pressure, radiation, motion or proximity. But with a biologically-consistent model they could be remodeled to use multiple observations from the environment for more agile operation. The work presented in this article is in this spirit of leveraging biological observations to forward engineer algorithms that can operate in a general context.

In the following sections, we provide a detailed explanation of our work. Section 2 describes the single unit cell recording procedure from which our analysis is derived and the definition of EVT from which the proposed model is based. Section 3 goes on to describe the exact specification of that model. Section 4 describes our experiments and Section 5 presents the corresponding results. Finally Section 6 concludes by putting this research into a larger biological and computational context.

2. MATERIALS AND METHODS

In this section, we explain the methods we use that are crucial for understanding our computational model of cross-modal sensory information integration. This includes the physical experiments that were conducted to collect the source data, as well as the formal elements of EVT.

2.1. Single-Unit Recordings and Odor Stimulation

This research builds upon the previous work of Huang et al. (2005). An overview is provided in **Figure 1B**. Traces of RGC are recorded before and after odor stimulation (the sites of odor treatment are indicated by numbers 1 and 2 in **Figure 1B**), or when dopamine and/or GnRH signaling transduction is manipulated by the application of receptor agonists or antagonists (indicated by numbers 3–8 in **Figure 1B**). For electrophysiological recordings, zebrafish were anesthetized with 0.04% 3-amino benzoic acid and immobilized by intraperitoneal injections of 3 5 μl of 0.5 mg ml^{-1} gallamine triethiodide

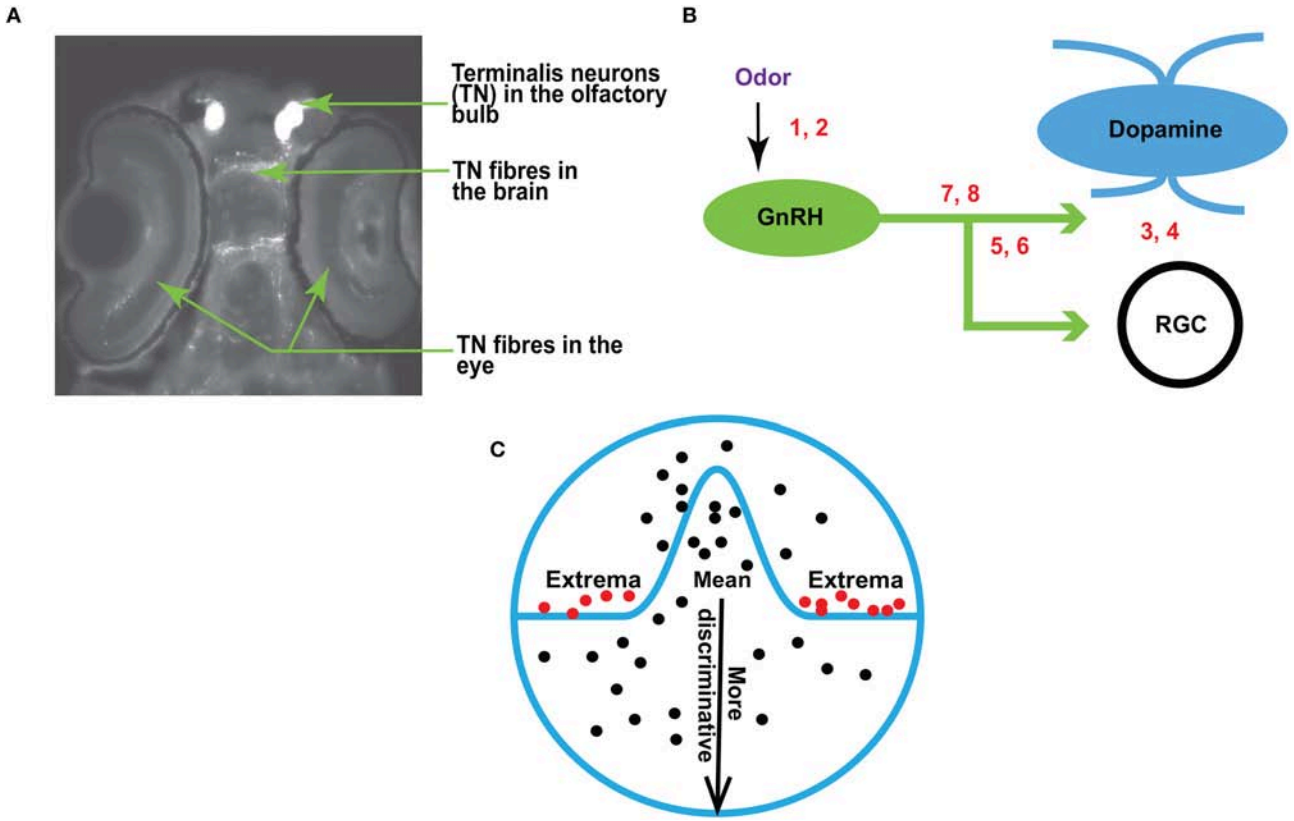


FIGURE 1 | (A) A fluorescent image of zebrafish brain (dorsal view, anterior is up) showing the terminalis neurons and axons (arrows). **(B)** A schematic diagram showing the experimental setup for RGC recordings in response to olfactory and TN stimulation. The numbers correspond with the following conditions: 1, 2–sham or odor stimulation; 3, 4–activation or inhibition of dopamine receptors; 5, 6–activation or inhibition of GnRH receptors; 7, 8–manipulation of both dopamine and GnRH receptors. **(C)** An overview of EVT. Prior work (Tanaka, 1996; Leopold et al., 2006; Freiwald et al., 2009) suggests that it may be the extremes (red dots), and not the mean (black dots near the center of the circle), that produce strong responses in the brain.

dissolved in phosphate-buffered saline (PBS), and then placed on a wet sponge with most of the body covered by a wet paper towel. A slow stream of system water (distilled water with ocean salt added, 3 g gal^{-1}, pH 7.0) was directed into the mouth to keep the fish oxygenized. The eye was slightly pulled out of its socket and held in place by glass rods, thus exposing the optic nerve. Single-unit RGC responses (determined by the spike waveform) were recorded from the optic nerve by using a Tungsten microelectrode (resistance, 5 10 MΩ). Electrical signals were filtered with a band pass filter between 30 and 3,000 Hz.

To test the effect of olfactory stimulation on visual sensitivity, we measured the light threshold required to evoke RGC responses before and after olfactory stimulation. Each fish was dark adapted for 30 min before the first RGC recording was made. The light stimuli (full-field dim white light, generated by a halogen bulb) were directed to the fish eye via a mirror system. The intensity of the unattenuated light beam ($\log I = 0$) measured in front of the fish eye was 670 $\mu W\ cm^{-2}$ (Optical Power Meter, UDT Instruments, MD, USA). To determine the threshold, the light intensity was first set below threshold level (e.g., $\log I = -6.0$) and then increased by 0.5 log-unit steps until the first light-evoked RGC responses were recorded (criteria, 20% above or below the rate of spontaneous firing). This light intensity was noted as the threshold. For each recording, 10 stimuli (600 ms flashes) were delivered at 3 s intervals.

Amino acids (methionine) were chosen to stimulate the olfactory neurons to activate the ORC pathway. Previous studies have demonstrated that amino acids are strong odors for zebrafish (Edwards and Michel, 2002). Among the amino acids tested in zebrafish, methionine produced the most obvious and dose-dependent responses on visual function (Maaswinkel and Li, 2003). In this study, odors (methionine, 0.5, 2, and 5 mM; total 8 10 μl per stimulation) were delivered to the nostril through a glass pipette. The light threshold required to evoke RGC responses was measured before the application of methionine, and was measured again within 10 s following the application of methionine. Thereafter, the measurement was repeated at 1 min intervals for 10 min. In total, 24 cells were recorded. 24 animals were used in this process with 1 cell/animal for the recordings. Among these 24 animals, in response to odor stimulation, 17 showed increased visual sensitivity. In the remaining 7 animals, 6 showed no changes in visual sensitivity and 1 showed decreased visual sensitivity.

2.2. Extreme Value Theory

The extreme value theorem (Coles, 2001) that underpins EVT (**Figure 1C**) is very similar to the central limit theorem (Jaynes, 2003). Both theorems involve limiting behaviors of distributions of independent and identically distributed random variables as n, the number of random variables, tends to ∞. However while the central limit theorem is concerned with the behavior of entire distributions of random variables, the extreme value theorem only applies to the random variables at the tails of those distributions.

To state this difference precisely, if $x_1, x_2, \ldots, x_n$ represent the i.i.d. random variables from a distribution, then the central limit theorem describes the limiting behavior of $x_1, x_2, \ldots, x_n$ while the extreme value theorem describes the limiting behavior of the extremes: $\max(x_1, x_2, \ldots, x_n)$ or $\min(x_1, x_2, \ldots, x_n)$ (Coles, 2001). It encompasses a number of distributions that apply to extrema.

An extreme value distribution is a limiting model for the maximums and minimums of a dataset. A limiting distribution simply models how large (or small) the data to be modeled will probably get. It is widely used in applications where there is interest in not only estimating the average, but also the maximum or minimum (Weibull, 1951, 1952; Galambos, 1994; Castillo et al., 2005). For example, when designing a dam, engineers might not only be interested in the average yearly flood which foretells the amount of water to be stored in the reservoir, but also in the maximum flood, the maximum intensity of earthquakes in the region during the past decade, or maximum strength of concrete to be used in building the dam to mitigate the possibility of a disaster. Castillo et al. (2005) list a number of applications where extreme value distributions can be used.

Now that the preliminaries have been covered, we can formally define an extreme value theorem (Fisher and Tippett, 1928):

Let $(s_1, s_2, \ldots, s_n)$ be a sequence of independent and identically distributed samples and let $M_n = max(s_1, s_2, \ldots, s_n)$. If a sequence of pairs of real numbers (a_n, b_n) exists such that each $a_n > 0$ and

$$\lim_{x\to\infty} P\left(\frac{M_n - b_n}{a_n} \leq x\right) = F(x) \tag{1}$$

then if $F(x)$ is a non-degenerate distribution function, it belongs to one of three extreme value distributions: Gumbel, Fréchet or Reverse Weibull.

In contrast to the Gumbel or Fréchet distributions which are used for unbounded data, the Weibull distribution applies to data that are bounded from below and when the shape (k) and scale (λ) parameters are positive (the Reverse Weibull is simply the opposite of the Weibull's non-degenerate distribution function). Moreover, the Weibull is used for modeling minima. In order to use it for modeling data that fall in the upper tail of a distribution, a minor adjustment needs to be made by flipping the data such that maxima become minima before applying the Weibull distribution. The probability distribution function of the two-parameter Weibull distribution is given as:

$$f(x;\lambda,k) = \begin{cases} \frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{k-1} e^{-\left(\frac{x}{\lambda}\right)^k}, & \text{if } x \geq 1 \\ 0, & \text{if } x < 0 \end{cases} \tag{2}$$

Note that there are other types of extreme value theorems one can make use of, such as the Pickands-Balkema-de Haam Theorem (Pickands, 1975). We limit ourselves to the theorem in Equation (1) in this work for the modeling of explicit tail data, but we will invoke the Pareto distribution, which is derived from the Pickands-Balkema-de Haam Theorem, in the modeling of the overall distribution. This is described below in the next section.

3. A MODEL FOR CROSS-MODAL SENSORY INFORMATION INTEGRATION

Now that the relevant background has been introduced, we formally define our computational model for cross-modal sensory information integration (**Figure 2**). It is motivated by the following hypothesis: *The tuning curves for RGC responses with and without olfactory signals are different. The extreme values in the tails of the distributions underlying those curves contribute to the determination of the visual sensitivity of zebrafish and should not be discarded as outliers.*

The single unit recordings that we used for our experiments can be regarded as samples from a large population. One way to infer more about the population statistics is to extrapolate from the available samples by fitting distributions to them and sampling additional data. However, fitting a known distribution to available data can be difficult because of limited sample sizes, leaving one to make a "best guess" based on prior information about the behavior of large sample statistics. The best guess can come from making an assumption (for example, a null hypothesis as a starting place), or a more rigorous method of model selection using some metric.

If n represents the sample size, $n \to \infty$ with the number of RGC responses acquired from an animal as it senses its environment over time. And the distribution of mean RGC responses calculated throughout an animal's entire lifecycle becomes Gaussian. This assumption directly follows from the central limit theorem. So perhaps the underlying distribution of measured responses is also Gaussian (a typical assumption in such modeling). Because our experiments involve two different sets of RGC responses, with and without olfaction, we can hypothesize that each set is normally distributed with varying parameters. This null hypothesis can be tested through commonly used measures of normality, failing which it can be rejected and we can look for alternative distributions using a model selection approach.

In statistical modeling, statisticians are often faced with the task of selecting a suitable model (a distribution, in our case) among a set of viable and finite candidates. There are several metrics or selection criteria one can use to determine the best explanatory model given the data. The Bayesian Information Criterion (BIC) (Schwarz et al., 1978; Neath and Cavanaugh, 2012) serves as a canonical method for model selection when priors are hard to state precisely. In a large sample setting the model found by BIC is equivalent to the candidate model that is *a posteriori* most probable, given the available data. It primarily amounts to maximizing the likelihood function separately for each candidate model and then choosing the one for which the

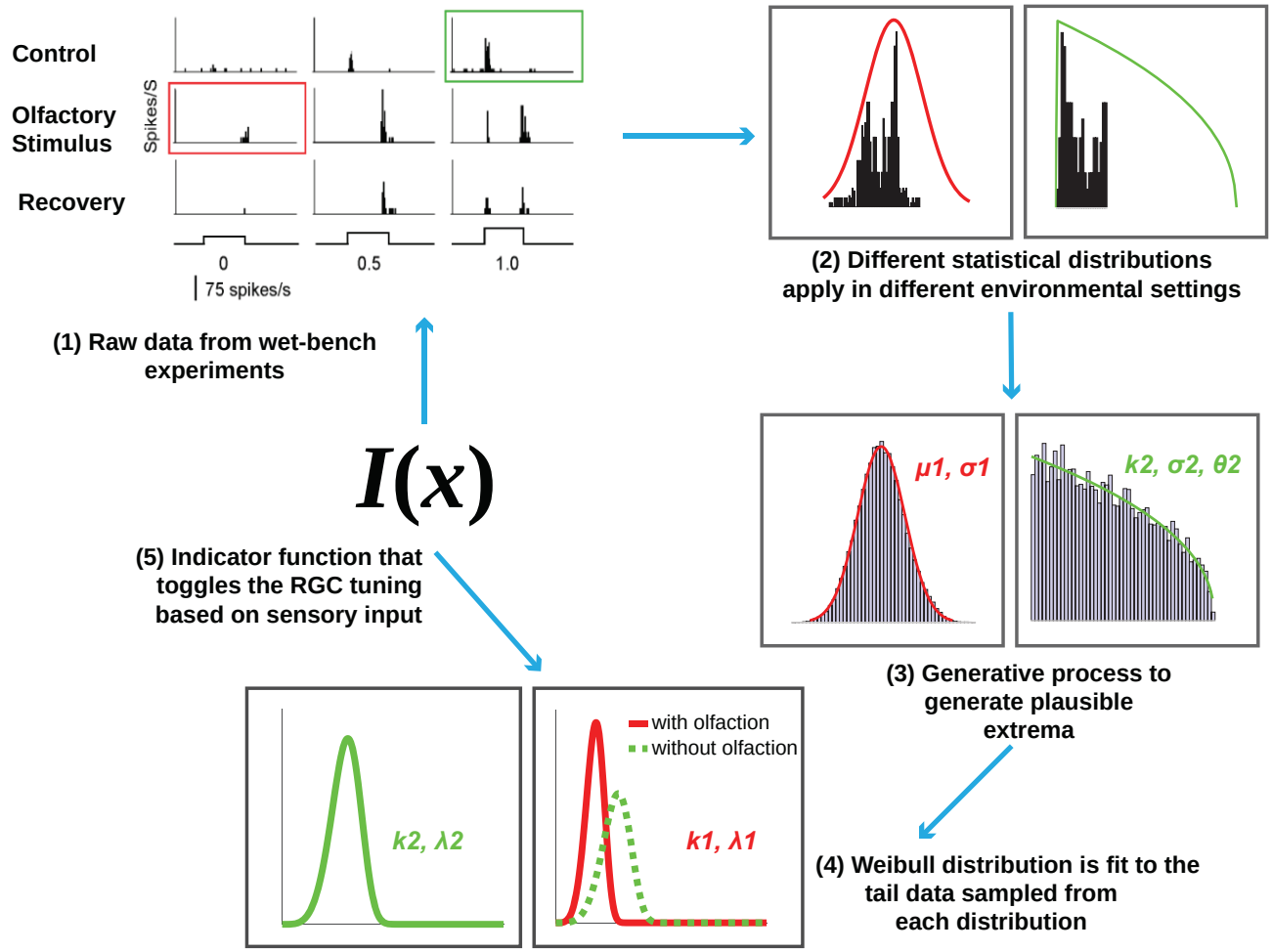


FIGURE 2 | An overview of the proposed computational model of cross-modal sensory information integration. The red curve represents the RGC responses when an olfactory stimulus is present and the green curve represents the responses when there is no olfactory stimulus present. The first step represents the data collection effort from the wet-bench experiments. The second and third steps include fitting distributions to the data collected in order to draw samples for further processing. The fourth step represents fitting a Weibull distribution in order to model the underlying difference between visual sensitivity when olfactory input (i.e., additional sensory information) is present as opposed to when it is not. The final step is identifying an indicator function, $I(x)$ that toggles between the two distributions based on the sensory input received.

log likelihood is the largest, with a fixed penalty term for guessing the wrong model.

To identify a good distribution to fit to non-normally distributed empirical data, we used a Matlab implementation of BIC[1]. A large set of valid parametric distributions were fit to the data and sorted using the output of the BIC metric to compare the goodness of the fits. The overall process returns a set of fitted distributions with their respective parameters. The list of distributions that were tried includes: Beta, Birnbaum-Saunders, Exponential, Extreme Value, Gamma, Generalized Extreme Value, Generalized Pareto, Inverse Gaussian, Logistic, Log-Logistic, Log-Normal, Nakagami, Rayleigh, Rician, t Location-Scale, and Weibull. It was assumed that all data were continuous.

[1]github.com/dcherian/tools/blob/master/misc/allfitdist.m

Our initial assumption that the overall data representing RGC responses without olfactory signals are normally distributed was rejected by the normality tests at the 1% significance level (a detailed description of the normality tests is given in section 4). Using the BIC method, the distribution that fit accurately to the overall RGC response data without olfactory stimulation was found to be the Generalized Pareto distribution (see **Supplementary Material**). Interestingly, this distribution is considered to be in the EVT family. The null hypothesis that the overall RGC responses with olfactory stimulus are normally distributed was not rejected at the 1% significance level by the normality tests, thus we fit a Gaussian distribution to that data.

Suppose we have n observations, or number of RGC responses. If x_i represents the i-th RGC response where $i \in (1, 2, \ldots, n)$, the population statistics (mean μ and variance σ^2) of the RGC response data with olfactory signal are found as the

unbiased estimates of the distribution parameters and are given by the following equations:

$$\mu = \sum_{i=1}^{n} \frac{x_i}{n} \tag{3}$$

$$\sigma^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \mu)^2 \text{ for all } i \in (1, 2, 3, ..., n) \tag{4}$$

The probability density function for the Generalized Pareto distribution with shape parameter k, scale parameter σ and threshold parameter τ is given by the following equation:

$$y = f\left(x \mid k, \sigma, \tau\right) = \left(\frac{1}{\sigma}\right)\left\{1 + k\frac{(x-\tau)}{\sigma}\right\}^{-1-\frac{1}{k}} \tag{5}$$

We used maximum likelihood to estimate the parameters k and σ from the two-parameter Generalized Pareto distribution by fitting RGC responses without olfaction[2].

Having access to a model of the entire population facilitates generative sampling, which in turn allows for better tail modeling, and support for heightened visual sensitivity under certain conditions. Such generative processes in the brain may be responsible for a number of different phenomena, as they facilitate generalization in learning from limited sampling (Rao et al., 2002). We use random sampling and the Metropolis-Hastings algorithm, a Markov chain Monte Carlo (MCMC) sampling method (Hastings, 1970) to generate in total 100, 000 simulated RGC responses with and without olfaction, respectively. The maximum (or the minimum) RGC response values within these samples follow an EVD. For our analysis, we concentrate only on the maximum RGC responses from the distributions described above because the lowest possible RGC response can be 0 spikes per second, indicating no response. Since the RGC responses (both with and without olfactory signals) can be assumed to be i.i.d samples from continuous distributions that are bounded from below, the Weibull distribution is the correct choice for modeling them. We expect the Weibull cumulative distribution curves (CDFs) for RGC responses with and without olfaction to be widely separated and the threshold RGC response value for an olfactory signal to shift sensitivity leftward (see **Figure 3** for an example), indicating that the cells are now more sensitive. This effect, replicated within the model, would confirm in a more rigorous sense that the presence of olfactory signals increases the fish's sensitivity toward its surrounding and almost endows it with night vision that would be otherwise impossible in absence of those signals.

This process is analogous to the super-additivity phenomenon in the multi-sensory superior colliculus of higher-order organisms like mammals, where the presence of two weak sensory signals from the environment enhances the animal's neural response toward that environment (Holmes and Spence, 2005). The RGC threshold value represents an average RGC response for visual sensitivity, which changes throughout an animal's entire life-cycle as it adapts to an ever-changing environment. However, the threshold varies (decreases or increases) in the presence or absence of a sensory stimulus other than visual input. This leads us to the possibility of the existence of some decision making mechanism in the fish's brain that toggles between two different distributions to adjust the tuning of the RGCs based on sensory input. Mathematically, this decision making procedure can be implemented as an indicator function $I(x)$. If θ represents the parameters of an RGC distribution, i.e., the prior information available for RGC responses with or without olfactory signals and x represents a new RGC response due to a stimulus from the environment such that $x \in R^n$ (here $n = 22$, as we successfully retrieved 22 dimensions representing RGC spikes over time after stimulation of the olfactory neurons from the wet-bench experiments of Huang et al. For further explanation, see section 4), then the indicator function $I(x)$ can be represented as:

$$I\left(x \mid \theta\right) = \begin{cases} 1, \text{ if olfactory signal is present} \\ 0, \text{ otherwise} \end{cases} \tag{6}$$

We speculate that the actual neural computation for the overall phenomenon is far more complex and is not restricted to just two modalities. However, given the recordings available for this study, we limit our model to just one particular circuit.

3.1. Choices for an Indicator Function

For the indicator function, we address the following problem: given a set of vectors representing RGC responses in spikes/s with and without olfactory signals, is it possible for an indicator function to identify whether a new RGC response has been triggered after an olfactory signal or not? Our intuition behind using an indicator function is that such a process exists in some capacity in the brain where the presence of one signal enhances the other signal, thereby eliciting responses much different from the situation when the signal is not present. In essence, this task can be formulated as a binary classification problem with two possible outcomes: presence or absence of olfactory signals. Ideally, any discrimitative supervised learning method can easily solve the problem. For our analysis, we examine the utility of support vector machines and an artificial neural network which, to some extent, mimics the functions of a biological neuron and is closer to the mechanism that the brain uses to process such signals. The motivation for choosing these particular classifiers is their simplicity—we desire an indicator function with an efficient training regime that can operate over thousands of multi-dimensional data points, such as a large collection of RGC responses. Other classifiers (e.g., decision trees, random forests, logistic regression) may also be suitable.

3.1.1. Support Vector Machine

The Support Vector Machine (SVM) is a supervised learning approach that is widely used for classification and regression

[2]For finding the maximum likelihood estimates of the Generalized Pareto distribution, we used the Matlab function *gpfit*, which only returns the estimates of the shape k and scale σ parameters of a two-parameter Generalized Pareto distribution. The function *makedist* was then used to create a probability distribution object reflecting where samples are taken from, using the parameters k and σ.

analysis (Cortes and Vapnik, 1995). Since our data is numeric and high-dimensional, SVM is a natural choice as it has been found to be extremely efficient in high-dimensional spaces for large-scale classification problems. SVMs use a subset of training points in the decision function, which form the "support vectors" that define the decision boundary between classes. As a consequence, it has been found to be memory efficient and has fast execution times if the data are normalized. For analysis, we assumed our data to be linearly separable and used a linear SVM formulation. We normalize all data using min-max normalization.

An SVM model with a set of labeled training data tries to find an optimal hyperplane for classifying new samples based on some constraints. Given a training dataset, $D = (x_i, y_i)$ of size m with $x_i = (x_1, x_2, ..., x_m)$, an n-dimensional feature/attribute vector, and label y_i = -1 or +1, formally the SVM classifier can be defined as a quadratic optimization problem solving the following equation:

$$min \|w\|^2 \text{ s.t } y_i(w^T x_i + b) \geq 1 \text{ for all } i \tag{7}$$

where $w = (w_1, w_2, ..., w_n)$ is a weight vector and b is the bias.

An important consideration when training an SVM model is the parameter C that dictates the trade-off between having a wide margin and correctly classifying training data.

$$min \|w\|^2 + C\sum_{1}^{m} \xi_i \text{ s.t } y_i(w^T x_i + b) \geq (1 - \xi_i), \xi_i \geq 0 \text{ for all } i \tag{8}$$

A larger value of C implies a smaller number of mis-classified training samples and is prone to overfitting.

3.1.2. Artificial Neural Network

We also consider a multi-layer perceptron (MLP) neural network as the indicator function. Similar to SVM, MLP is a supervised learning algorithm that learns a non-linear mapping from input $x \in R^n$, where n represents the number of dimensions, to $y \in R^m$ where m can be any number $m < n$, depending on the number of classes in the training dataset. However, unlike SVMs, a simple MLP includes one or more hidden layers consisting of artificial neurons. The hidden layers act as feature detectors and gradually discover the salient features of the training data through backpropagation (Rumelhart et al., 1986; Werbos, 1990). Each neuron includes a non-linear and differential activation function and is connected to every neuron in the previous layer exhibiting a high degree of connectivity between layers. As a result, due to the distributed nature of non-linearities, the learning process is difficult to visualize. However, neural networks are usually assumed to be non-parametric functions, i.e., they can be used as function approximators without having any prior information about the distribution of input or training dataset and hence are well suited to represent the indicator function. If x represents a p-dimensional input vector such that $x = (x_1, x_2, x_3, ..., x_p)$ with $y = (+1, -1)$ as labels and $g : \mathbb{R} \mapsto \mathbb{R}$ as the activation function, then the equation for a single neuron is given by:

$$y = g\left(b + \sum_{i=1}^{p} w_i x_i\right) \tag{9}$$

where $w = \left[w_1, w_2, w_3, ..., w_p\right]$ represents the weights learned through backpropagation.

4. EXPERIMENTS

4.1. Data Collection and Representation

As stated above, the first step in building a computational model of this nature is to attempt to define the underlying distribution of the data one is trying to explore. We use the data from a study by Huang et al. (2005) for our analysis. The data consists of single unit RGC responses measured in spikes/sec before and after olfactory stimulation under varying light intensity (see **Figure 2** from Huang et al.). In terms of raw data organization, it is primarily a histogram with the x-axis representing the visual sensitivity of fish binned into approximately 22 positions representing a timestamp and their corresponding frequency measured in spikes/sec on the y-axis. Under normal conditions, the minimum threshold light intensity to invoke a retinal ganglion cell response in a dark-adapted zebrafish embryo is 10^{-5}. However, with olfactory stimulation with methionine, the threshold light intensity decreases to 10^{-6}. We calculated the minimum RGC response threshold to be at 75 spikes/s. Hence, the data can be separated into two parts: one with olfactory stimulus and the other without it. In total, there were 22 RGC responses across time with olfactory stimulus and 29 without olfactory stimulation.

4.2. Experiment 1

The first experiment was to check whether the raw data we collected from the experiments confirms our hypothesis that the EVT can be applied to build an accurate model. We posit that since the RGC responses with olfactory stimulation represent extreme aberration from the baseline and are non-negative integers, the Weibull distribution is the right candidate for modeling our data. But how differently does our data fit with the Weibull distribution vs. a central tendency model like the Gaussian distribution? We explore this by comparing the CDFs of the Weibull and Gaussian distributions with parameters derived from our data.

4.3. Tests of Normality and Synthetic Data Generation

Using the data collected from wet-bench experiments as a basis, we simulated an expansive data space by fitting distributions over the original data. The goal was to generate as much evidence as possible for statistical inference. However, in order to fit distributions to generate more samples from the existing data, we need to make some assumptions about the underlying distribution. Initially, as described above in section 3, we assumed a null hypothesis that the distribution of RGC responses in a zebrafish throughout its entire lifecycle is Gaussian. Since our work involves two different sets of RGC responses—one with olfactory stimulus and the other without it—under this assumption the distributions underlying each should be Gaussian with different parameters. To test this, we performed several commonly used tests of normality: the Kolmogorov Smirnov test (Massey, 1951), the Shapiro-Wilk test (Shapiro and Wilk,

1965), and a Lilliefors test (Lilliefors, 1967, 1969; Conover and Conover, 1980)[3]. Due to the small sample size ($n = 22$ or 29), we preferred the Shapiro-Wilk test over Kolmogorov-Smirnov and Lilliefors. For datasets that failed the normality test, The BIC selection criterion was deployed to find another distribution with the best fit. Afterwards, we generated 100,000 non-negative samples of RGC responses from the respective distributions for further analysis.

4.4. Experiment 2

The second experiment was to check whether the points we sampled confirm our hypothesis that the EVT can be applied in a generative scenario. In order to verify this, we fit a Weibull distribution to the top n RGC responses to understand how the curves vary when olfactory input is present as opposed to when it is not. The value n was selected via empirical observation. The sampling methods used were: random sampling and MCMC sampling. Since EVDs like the Weibull only apply to samples at the tails of distributions, it is independent of the underlying distribution of the data as a whole. Hence, irrespective of the overall data distribution and sampling process, the results of Experiment 2 for the Weibull distributions for the top n responses should ideally be similar to Experiment 1. We expect the Weibull cumulative distribution functions for data with and without olfactory stimulus to be widely separated, with the curve for data with olfaction shifting leftward, giving higher probability scores to RGC responses that would be improbable under conditions where olfaction is not engaged.

4.5. Experiment 3

Additionally, we wanted to corroborate whether we can define a deterministic indicator function such that given some RGC response it is possible for the function to identify if an olfactory stimulus is present or not. In essence, this task becomes a binary classification problem where the presence of olfactory signals can be labeled as 1 and the absence as 0. As described above in section 3, we use a linear SVM or a multi-layer perceptron as our binary classifier. For consistency in the operation of the indicator function, we limit the dimensionality of all vectors to the dimensionality of RGC responses with olfactory stimulus ($n = 22$). We use the 100,000 samples we generated for each scenario (with olfactory stimulus and without olfactory stimulus), dividing the sets into 80% training and 20% testing partitions.

In summary, the entire modeling effort is encapsulated in the following steps (also depicted in **Figure 2**):

1. **Data collection and representation.** This step consists of collecting and representing data based on the wet-bench experiments for control (without any stimulation) and experimental (with olfactory stimulation) zebrafish as a histogram and collecting the statistics for further analysis.
2. **Experiment 1.** This first test consists of an experiment to evaluate our hypothesis that EVT applies with the raw data collected in step 1. We fit Gaussian distributions (to the entire collection of data with and without olfaction individually) and Weibull distributions (to the top-n RGC responses from the two datasets). The value n was selected via empirical observation.
3. **Tests of normality and synthetic data generation.** Here we begin by assuming that the distribution of RGC responses in a zebrafish throughout its entire life cycle is normal, and attempt to falsify that assumption via tests of normality. The appropriate distributions are subsequently fit to the data to generate 100,000 synthetic samples. The data with olfactory stimulus follows the Gaussian distribution, whereas the underlying distribution for data without olfactory stimulus is Generalized Pareto.
4. **Experiment 2.** Similar to Experiment 1 but instead uses 100,000 generated samples and only Weibull distributions fit to the top-n samples generated to examine how olfactory signals influence visual sensitivity as reflected by the CDF curves for the two scenarios. The value n is selected through empirical observation.
5. **Experiment 3.** This experiment involves identifying an indicator function $I(x)$ that can distinguish when an olfactory stimulus is present and when it is not. Here this function is a deterministic binary classifier, either a linear SVM or a multi-layer perceptron.

5. RESULTS

5.1. Experiment 1

Figure 3 depicts the result of Experiment 1, which was conducted to examine the difference between central tendency modeling and EVT modeling. The data for this experiment were what was directly collected from the wet-bench experiments for both control (without olfaction) and experimental (with olfaction) zebrafish.

As can be seen in the figure, with olfactory stimulation the visual sensitivity in zebrafish shifts leftward, making the RGC responses below the normal threshold of 75 spikes/s probable, as indicated by the physiology experiments of Huang et al. (2005). Moreover, if we look closely, the Weibull distributions (represented by the red and blue solid and dashed lines) are a better fit to the data because the RGC responses with olfactory stimulation represent a set of extreme responses as opposed to RGC responses without any stimulation. If we fix our attention at the threshold RGC response at 75 spikes/s, the Weibull curves provide a better explanation for getting an RGC response below 75 spikes/sec for olfactory stimulation in comparison to the normal distribution, which makes those values more improbable. In other words, the tuning becomes more sensitive if we use the Weibull distribution. We plotted the curves by varying n ($n = 3, 8$) of the top-n RGC responses. The tuning becomes more sensitive as n becomes smaller.

[3]We used the following Matlab implementations of the normality tests: *lillietest* (for the Lilliefors test), *swtest* (from Matlab central for the Shapiro-Wilk test), *kstest* (for the one-sample Kolmogorov-Smirnov test). Each of these tests returns a decision (1 or 0) for the null hypothesis that the data comes from a distribution in the normal family, against the alternative that it does not come from such a distribution. A result of 1 rejects the null hypothesis at the 5% significance level (default). For our experiments, we set the significance level to 1%.

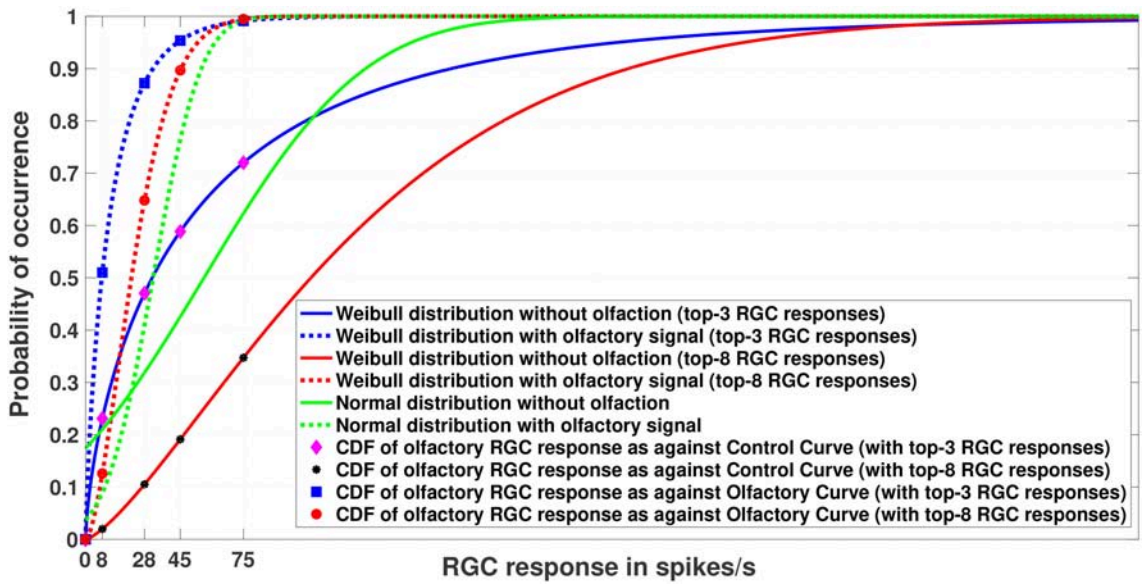


FIGURE 3 | Experiment 1. Cumulative Distribution Functions for zebrafish with and without olfactory stimulation at light intensity $10^{-}5$ and $10^{-}6$, respectively. The curves depict the difference between central tendency modeling (green) and EVT modeling (red and blue). As can be seen, tuning becomes more sensitive when the Weibull distribution is used. The number of maximal RGC responses taken is either 3 or 8 (indicated within the parentheses). Best viewed in color.

5.2. Tests of Normality and Synthetic Data Generation

The null hypothesis that the data without olfactory stimulus are normally distributed was rejected at the 1% significance level for all of the tests. However, the other assumption of normality for data with olfactory stimulus was not rejected at the 1% significance level. Based on these results, we fit a Gaussian distribution to the data with olfactory stimulus. Using the BIC selection criterion to find the best fit, the distribution for the data without olfactory stimulus was determined to be Generalized Pareto. We then collected non-negative samples simulating RGC responses via random sampling or MCMC sampling (100, 000 samples from each sampling method), to be used for fitting a Weibull distribution to the top n samples in order to understand how the curves vary when olfactory input is present (i.e., when the overall distribution is Gaussian) as opposed to when it is not (i.e., when the overall distribution is Pareto).

5.3. Experiment 2

Figures 4, **5** show the models of visual sensitivity calculated over the simulated data from random sampling and MCMC sampling[4]. Similar results are achieved for both sampling methods. An important observation to note here is that tuning is always more sensitive when olfactory stimulus is present. The values of n in this experiment are much larger ($n = 50, 250$) due to the increased availability of data, but still represent a small number of points from the tail of the overall distribution. The CDF curves for data with and without olfactory stimulation are widely separated and the width of separation increases as n grows larger. This reflects how the visual sensitivity threshold can change throughout a fish's life cycle as it is exposed to an ever-changing environment and acquires new RGC responses for modulating its internalized model of visual sensitivity. Note that zebrafish build new cells within their nervous systems via a neurogenesis process, meaning the number of responses available at a point in time can change in a non-stimulus dependent way. Our proposed model supports this phenomenon.

5.4. Experiment 3

With respect to testing the possible indicator functions $I(x)$, we began by considering a linear binary SVM classifier trained using 80, 000 generated samples and tested using 20, 000 generated samples. With random sampling, we achieved a testing accuracy of 95.5 ($\pm$ 0.163) percent, but with MCMC sampling accuracy decreased to 93.925 ($\pm$ 0.123) percent. With a multi-layer perceptron classifier, the accuracy dropped to 95.25 ($\pm$ 0.007) percent using the same training-testing split and data from MCMC sampling[5]. The success of this experiment establishes that the two different classes of RGC responses are separable. Thus it is possible, in a statistical learning sense, to have a mechanism to toggle between RGC tuning configurations when an olfactory stimulus is present and when it is not. One possibility for why the classification was successful in these experiments is that the indicator function implicitly learns that the data are distributed differently in the two classes (Generalized Pareto for data without olfactory stimulus and Gaussian for the data with olfactory stimulus). That the two classes of data are distributed

[4]We ran experiments 1 and 2 ten times. In each of those trials, the leftward shift of the distribution after olfactory stimulation was preserved.

[5]Each of these experiments was run ten times. The numbers in parentheses represent standard error.

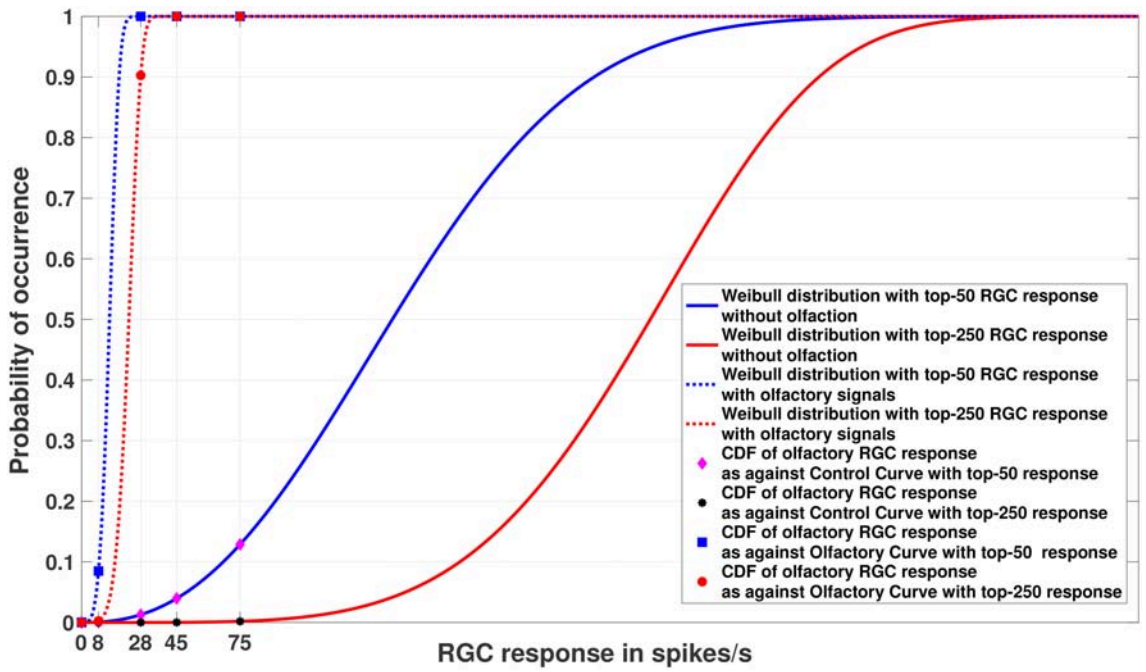


FIGURE 4 | Experiment 2. Cumulative Distribution Functions for zebrafish with and without olfactory stimulation at light intensity $10^{-}5$ and $10^{-}6$, respectively, with data points generated through random sampling. The curves labeled "Control" in the legend describe the Weibull distributions (as represented by the solid blue and red lines) without olfactory stimulus. As can be seen, tuning is most sensitive when an olfactory stimulus is involved. Best viewed in color.

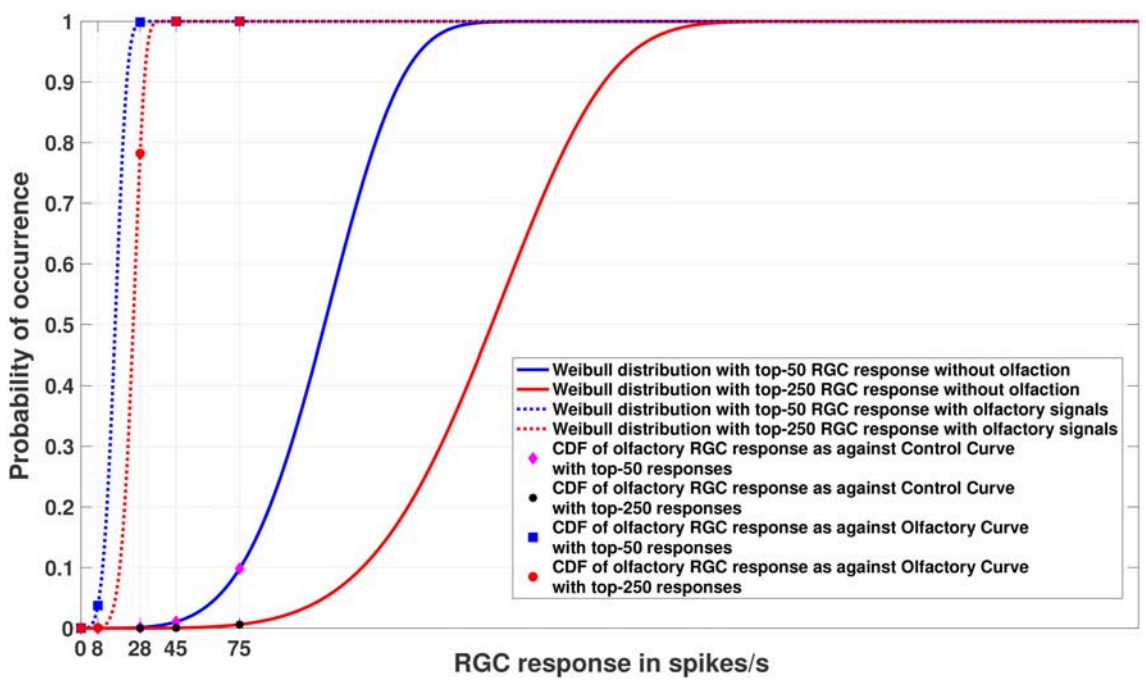


FIGURE 5 | Experiment 2. Cumulative Distribution Functions for zebrafish with and without olfactory stimulation at light intensity $10^{-}5$ and $10^{-}6$, respectively, with data points generated through MCMC sampling. The curves labeled "Control" in the legend describe the Weibull distributions (as represented by solid blue and red lines) without olfactory stimulus. The result is very similar to random sampling—the tuning is more sensitive when an olfactory stimulus is involved. Best viewed in color.

differently lends further support to our hypothesis that an indicator function is involved in the integration of cross-modal sensory information—the distributional difference facilitates a very straightforward pattern recognition process to separate the classes.

6. DISCUSSION

As vertebrates evolved over centuries, sensory organs adapted with the ever-changing environment. In many vertebrate species, at any given time the brain integrates and processes multi-sensory information. In humans, for example, the functions of the olfactory and visual systems are influenced by sensory input from each organ. Most mammals have specialized multimodal neurons in the superior colliculus that are capable of integrating multiple stimuli from the environment and providing a uniform reaction. In lower vertebrates such as fish, however, such advanced mechanisms are absent. In zebrafish, the integration of sensory information from the olfactory system facilitates signaling transduction in the visual pathway. As a consequence, retinal neural activities such as the firing of retinal ganglion cells are increased. This is particularly important for wild type animals that live under natural environmental conditions. For example, zebrafish normally mate in the early morning hours before the sun comes up, during which time the light illumination is low. It is conceivable that under such conditions stimulation of olfactory neurons may increase visual sensitivity and thereby facilitate the process of mating. While the system mechanisms underlying this olfacto-retinal sensory integration have been well characterized, statistical models that describe the phenomenon at the cellular level have not been described. In this paper, we have described a computational model that supports the research into how the visual system integrates information from other sensory modalities.

The idea of building computational models for multisensory input has been explored previously (Anastasio et al., 2000; Driver and Noesselt, 2008; Angelaki et al., 2009). When it comes to determining the statistical relationship between sensory responses among different sensory organs, the Bayesian model has been a preferred framework. However, almost all of the existing work focuses on higher vertebrates such as mammals. Angelaki et al. (2009) attempted to reconcile the difference between the traditional physiological studies on multisensory integration with computational and psychological studies using Bayesian inference on the visual-vestibular system for the perception of self-motion in macaques. They describe how the multimodal neurons represent probablistic information defined by multiple stimuli and propose that special neurons accomplish near optimal cue integration through a linear summation of input signals.

With respect to models of simpler animals, Wessnitzer and Webb (2006) explore multimodal sensory integration for navigation from the physiological perspective of the insect's nervous system. In zebrafish, using a similar linear model (Hughes et al., 1998) the contribution of different types of cone photoreceptor cells to photopic spectral visual sensitivity was determined. This was done by re-modeling the electroretinographic data recorded from the cornea, which include absorbance spectrum of four types of cone photoreceptor cells (cone cells that are sensitive to ultra-violet light, blue light, green light, and red light, respectively) given as the visual pigment template for the appropriate maximum absorption, neural signals obtained from different cone cell types, relative fraction of the individual cone cells across the retina, and linear gains for each cone type (Cameron, 2002). The model incorporates the first-order cellular and biophysical aspects of cone photoreceptor cells and thereby predicts the second-order physiological functions of cone cell-mediated visual sensitivity. Using this model, linear gains that represent the strength of four different types of cone cell-derived neural signals onto four different inferred cone processes in the whole retina can be assessed.

Turning to extreme value theory, the objective of nearly all extant models in computational neuroscience has been to discard the extreme values located at the tails of distributions as noise and concentrate on the mean or average. However, evidence suggests that extremes, and not means, of cell responses direct activity in the brain. For example, the ability of primates, like macaque monkeys, to identify individual faces can be localized to a group of special neurons that fire in response to specific regions of the face (Freiwald et al., 2009). An interesting finding that came out of that study was that neurons were tuned to the geometry of extreme facial features. Previous investigations along this line concentrated on how the brain fundamentally adapts itself to the statistics of the sensory world, extracting relevant information from sensory inputs by modeling the distribution of inputs that are encountered by the organism (Simoncelli and Olshausen, 2001; Simoncelli, 2003). This led to the advent of "sparse coding" which attempts to explain how neurons encode sensory information using a small number of active neurons at any given point in time (Olshausen and Field, 1997). A direct extension of this work suggests that sparse coding is an all-pervasive phenomenon used by all types of sensory neurons in different modalities across different species (Olshausen and Field, 2004). EVT builds upon these concepts but is more specialized.

Much prior work related to EVT modeling has focused on various non-biological applications from trend detection in ground-level ozone (Smith, 1989) to quantifying extreme precipitation levels using Generalized Pareto distributions (Cooley et al., 2007). Other applications of EVT include, but are not limited to, finance, telecommunications, the environment (Finkenstadt and Rootzén, 2003), and hydrology (Katz et al., 2002). Recent work in computer vision and machine learning has extensively used the concept of EVT (Shi et al., 2008; Broadwater and Chellappa, 2010; Scheirer et al., 2011, 2014; Fragoso et al., 2013). For instance, for biometric verification systems, Shi et al. (2008) used the Generalized Pareto Distribution to model the genuine and impostor scores and made a significant observation that the tails of each score distribution contain the most relevant information that helps in defining each distribution considered for prediction and the associated decision boundaries, which are often difficult to model.

Our research extends this theory to multi-sensory inputs through a model that demonstrates strong neural fidelity. With a biologically-consistent information fusion algorithm based on retinal circuits in the zebrafish, we believe that we have access to a better general solution to the problem at hand and possibly many other information processing problems of interest. In this article, we have developed a neural computation model that simulates the process of multi-organ sensory integration and predicts the consequence of sensory integration in higher-order brain functions. In contrast to Gaussian modeling, we propose that EVT models of the extrema found in the tails of the data can form a powerful basis for cross-modal sensory information integration, facilitating heightened sensitivity in targeted modalities that have been influenced by a stimulus in the environment. This resulted in the development of a computational EVT-based framework for multi-organ sensory integration in the zebrafish that is not only an explanatory model in neuroscience, but also shows promise for applications in machine learning and neuromorphic systems.

DATA AVAILABILITY

The datasets analyzed for this study and the source code used for modeling have been released for reproducability and can be downloaded from https://github.com/sbanerj2/Zebrafish_EVT.

ETHICS STATEMENT

An ethical review process was not required for our study. All data used in this article come from a previously published study (Huang et al., 2005). All experimental procedures in that paper adhered to the NIH guidelines for animals in research.

AUTHOR CONTRIBUTIONS

WS and LL initially conceived of the idea. LL was responsible for conducting the wet-bench experiments and preparing the source data. SB designed, analyzed, implemented the model and wrote the paper. WS supervised the entire modeling effort.

FUNDING

The research was funded in part by the Department of Defense (Army Research Laboratory) under the contracts W911NF-16-1-0316 and W911NF-18-1-0292.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fncom.2019.00003/full#supplementary-material

REFERENCES

Anastasio, T. J., Patton, P. E., and Belkacem-Boussaid, K. (2000). Using Bayes' rule to model multisensory enhancement in the superior colliculus. *Neural Comput.* 12, 1165–1187. doi: 10.1162/089976600300015547

Angelaki, D. E., Gu, Y., and DeAngelis, G. C. (2009). Multisensory integration: psychophysics, neurophysiology, and computation. *Curr. Opin. Neurobiol.* 19, 452–458. doi: 10.1016/j.conb.2009.06.008

Broadwater, J. B., and Chellappa, R. (2010). Adaptive threshold estimation via extreme value theory. *IEEE Trans. Signal Process.* 58, 490–500. doi: 10.1109/TSP.2009.2031285

Cameron, D. A. (2002). Mapping absorbance spectra, cone fractions, and neuronal mechanisms to photopic spectral sensitivity in the zebrafish. *Vis. Neurosci.* 19, 365–372. doi: 10.1017/S0952523802192121

Castillo, E., Hadi, A. S., Balakrishnan, N., and Sarabia, J.-M. (2005). *Extreme Value and Related Models With Applications in Engineering and Science*. Hoboken, NJ: Wiley.

Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*. Springer.

Conover, W. J., and Conover, W. J. (1980). *Practical Nonparametric Statistics*. New York, NY: Wiley.

Cooley, D., Nychka, D., and Naveau, P. (2007). Bayesian spatial modeling of extreme precipitation return levels. *J. Am. Stat. Assoc.* 102, 824–840. doi: 10.1198/016214506000000780

Cortes, C., and Vapnik, V. (1995). Support-vector networks. *Mach. Learn.* 20, 273–297. doi: 10.1007/BF00994018

Driver, J., and Noesselt, T. (2008). Multisensory interplay reveals crossmodal influences on sensory-specific brain regions, neural responses, and judgments. *Neuron* 57, 11–23. doi: 10.1016/j.neuron.2007.12.013

Edwards, J. G., and Michel, W. C. (2002). Odor-stimulated glutamatergic neurotransmission in the zebrafish olfactory bulb. *J. Compar. Neurol.* 454, 294–309. doi: 10.1002/cne.10445

Finkenstadt, B., and Rootzén, H. (2003). *Extreme Values in Finance, Telecommunications, and the Environment*. CRC Press.

Fisher, R. A., and Tippett, L. H. C. (1928). "Limiting forms of the frequency distribution of the largest or smallest member of a sample," in *Mathematical Proceedings of the Cambridge Philosophical Society*, Vol. 24 (Cambridge University Press), 180–190.

Fragoso, V., Sen, P., Rodriguez, S., and Turk, M. (2013). "EVSAC: accelerating hypotheses generation by modeling matching scores with extreme value theory," in *Proceedings of the IEEE International Conference on Computer Vision* (Sydney), 2472–2479.

Freiwald, W. A., Tsao, D. Y., and Livingstone, M. S. (2009). A face feature space in the macaque temporal lobe. *Nat. Neurosci.* 12, 1187–1196. doi: 10.1038/nn.2363

Galambos, J. (1994). "Extreme value theory for applications," in *Extreme Value Theory and Applications*, eds J. Galambos, J. Lechner, and E. Simiu (Boston, MA: Springer), 1–14. doi: 10.1007/978-1-4613-3638-9_1

Gastiner, M. J., Yusupov, R. G., Glickman, R. D., and Marshak, D. W. (2004). The effects of histamine on rat and monkey retinal ganglion cells. *Vis. Neurosci.* 21, 935–943. doi: 10.1017/S0952523804216133

Gire, D. H., Restrepo, D., Sejnowski, T. J., Greer, C., De Carlos, J. A., and Lopez-Mascaraque, L. (2013). Temporal processing in the olfactory system: can we see a smell? *Neuron* 78, 416–432. doi: 10.1016/j.neuron.2013.04.033

Harter, M. R. and Aine, C. J. (1984). Brain mechanisms of visual selective attention. *Variet. Attent.* 293–321.

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* 57, 97–109. doi: 10.1093/biomet/57.1.97

Holmes, N. P., and Spence, C. (2005). Multisensory integration: space, time and superadditivity. *Curr. Biol.* 15, R762–R764. doi: 10.1016/j.cub.2005.08.058

Huang, L., Maaswinkel, H., and Li, L. (2005). Olfactoretinal centrifugal input modulates zebrafish retinal ganglion cell activity: a possible role for dopamine-mediated Ca2+ signalling pathways. *J. Physiol.* 569, 939–948. doi: 10.1113/jphysiol.2005.099531

Hughes, A., Saszik, S., Bilotta, J., Demarco, P. J., and Patterson, W. F. (1998). Cone contributions to the photopic spectral sensitivity of the zebrafish ERG. *Vis. Neurosci.* 15, 1029–1037. doi: 10.1017/S095252389815602X

Jaynes, E. T. (2003). *Probability Theory: The Logic of Science*. Cambridge University Press.

Katz, R. W., Parlange, M. B., and Naveau, P. (2002). Statistics of extremes in hydrology. *Adv. Water Resour.* 25, 1287–1304. doi: 10.1016/S0309-1708(02)00056-8

Leopold, D. A., Bondar, I. V., and Giese, M. A. (2006). Norm-based face encoding by single neurons in the monkey inferotemporal cortex. *Nature* 442:572. doi: 10.1038/nature04951

Li, L. (2001). Zebrafish mutants: behavioral genetic studies of visual system defects. *Dev. Dyn.* 221, 365–372. doi: 10.1002/dvdy.1159

Li, L., and Dowling, J. E. (2000). Disruption of the olfactoretinal centrifugal pathway may relate to the visual system defect in night blindness bmutant zebrafish. *J. Neurosci.* 20, 1883–1892. doi: 10.1523/JNEUROSCI.20-05-01883.2000

Lilliefors, H. W. (1967). On the Kolmogorov-Smirnov test for normality with mean and variance unknown. *J. Am. Stat. Assoc.* 62, 399–402. doi: 10.1080/01621459.1967.10482916

Lilliefors, H. W. (1969). On the Kolmogorov-Smirnov test for the exponential distribution with mean unknown. *J. Am. Stat. Assoc.* 64, 387–389. doi: 10.1080/01621459.1969.10500983

Maaswinkel, H., and Li, L. (2003). Olfactory input increases visual sensitivity in zebrafish: a possible function for the terminal nerve and dopaminergic interplexiform cells. *J. Exp. Biol.* 206, 2201–2209. doi: 10.1242/jeb.00397

Massey, F. J. Jr. (1951). The Kolmogorov-Smirnov test for goodness of fit. *J. Am. Stat. Assoc.* 46, 68–78. doi: 10.1080/01621459.1951.10500769

Mick, G., Cooper, H., and Magnin, M. (1993). Retinal projection to the olfactory tubercle and basal telencephalon in primates. *J. Compar. Neurol.* 327, 205–219. doi: 10.1002/cne.903270204

Nagrani, A., Albanie, S., and Zisserman, A. (2018). "Seeing voices and hearing faces: Cross-modal biometric matching," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Salt Lake City, UT), 8427–8436.

Neath, A. A., and Cavanaugh, J. E. (2012). The bayesian information criterion: background, derivation, and applications. *Wiley Interdiscipl. Rev.* 4, 199–203. doi: 10.1002/wics.199

Olshausen, B. A., and Field, D. J. (1997). Sparse coding with an overcomplete basis set: a strategy employed by V1? *Vis. Res.* 37, 3311–3325. doi: 10.1016/S0042-6989(97)00169-7

Olshausen, B. A., and Field, D. J. (2004). Sparse coding of sensory inputs. *Curr. Opin. Neurobiol.* 14, 481–487. doi: 10.1016/j.conb.2004.07.007

Pickands, J. (1975). Statistical inference using extreme order statistics. *Ann. Stat.* 3, 119–131. doi: 10.1214/aos/1176343003

Rao, R. P., Olshausen, B. A., Lewicki, M. S., Jordan, M. I., and Dietterich, T. G. (2002). *Probabilistic Models of the Brain: Perception and Neural Function*. MIT Press.

Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature* 323:533. doi: 10.1038/323533a0

Scheirer, W. J., Jain, L. P., and Boult, T. E. (2014). Probability models for open set recognition. *IEEE Trans. Patt. Anal. Mach. Intell.* 36, 2317–2324. doi: 10.1109/TPAMI.2014.2321392

Scheirer, W. J., Rocha, A., Micheals, R. J., and Boult, T. E. (2011). Meta-recognition: the theory and practice of recognition score analysis. *IEEE Trans. Patt. Anal. Mach. Intell.* 33, 1689–1695. doi: 10.1109/TPAMI.2011.54

Schwarz, G. et al. (1978). Estimating the dimension of a model. *Ann. Stat.* 6, 461–464. doi: 10.1214/aos/1176344136

Shapiro, S. S., and Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika* 52, 591–611. doi: 10.1093/biomet/52.3-4.591

Shi, Z., Kiefer, F., Schneider, J., and Govindaraju, V. (2008). "Modeling biometric systems using the general pareto distribution (gpd)," in *SPIE Defense and Security Symposium* (International Society for Optics and Photonics), 69440.

Simoncelli, E. P. (2003). Vision and the statistics of the visual environment. *Curr. Opin. Neurobiol.* 13, 144–149. doi: 10.1016/S0959-4388(03)00047-3

Simoncelli, E. P., and Olshausen, B. A. (2001). Natural image statistics and neural representation. *Annu. Rev. Neurosci.* 24, 1193–1216. doi: 10.1146/annurev.neuro.24.1.1193

Smith, R. L. (1989). Extreme value analysis of environmental time series: an application to trend detection in ground-level ozone. *Stat. Sci.* 4, 367–377. doi: 10.1214/ss/1177012400

Tanaka, J. W. (1996). Caricature recognition in a neural network. *Vis. Cogn.* 3, 305–324. doi: 10.1080/135062896395616

Vacaru, A. M., Unlu, G., Spitzner, M., Mione, M., Knapik, E. W., and Sadler, K. C. (2014). *In vivo* cell biology in zebrafish – providing insights into vertebrate development and disease. *J. Cell Sci.* 127, 485–495. doi: 10.1242/jcs.140194

Weibull, W. (1951). A statistical distribution function of wide applicability. *J. Appl. Mech.* 18, 293–297.

Weibull, W. (1952). A survey of statistical effects in the field of material failure. *Appl. Mech. Rev.* 5, 449–451.

Werbos, P. J. (1990). Backpropagation through time: what it does and how to do it. *Proc. IEEE* 78, 1550–1560. doi: 10.1109/5.58337

Wessnitzer, J., and Webb, B. (2006). Multimodal sensory integration in insects towards insect brain control architectures. *Bioinspir. Biomimet.* 1:63. doi: 10.1088/1748-3182/1/3/001

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 22 March 2019
doi: 10.3389/fnins.2019.00225

A Compound Computational Model for Filling-In Processes Triggered by Edges: Watercolor Illusions

Hadar Cohen-Duwek and Hedva Spitzer*

Vision Research Laboratory, School of Electrical Engineering, Tel-Aviv University, Tel Aviv, Israel

The goal of our research was to develop a compound computational model with the ability to predict different variations of the "watercolor effects" and additional filling-in effects that are triggered by edges. The model is based on a filling-in mechanism solved by a Poisson equation, which considers the different gradients as "heat sources" after the gradients modification. The biased (modified) contours (edges) are ranked and determined according to their dominancy across the different chromatic and achromatic channels. The color and intensity of the perceived surface are calculated through a diffusive filling-in process of color triggered by the enhanced and biased edges of stimulus formed as a result of oriented double-opponent receptive fields. The model can successfully predict both the assimilative and non-assimilative watercolor effects, as well as a number of "conflicting" visual effects. Furthermore, the model can also predict the classic Craik–O'Brien–Cornsweet (COC) effect. In summary, our proposed computational model is able to predict most of the "conflicting" filling-in effects that derive from edges that have been recently described in the literature, and thus supports the theory that a shared visual mechanism is responsible for the vast variety of the "conflicting" filling-in effects that derive from edges.

Keywords: computational models, watercolor effect, filling-in, diffusion process, visual system mechanism

OPEN ACCESS

Edited by:
Haluk Ogmen,
University of Denver, United States

Reviewed by:
C. Alejandro Párraga,
Autonomous University of Barcelona, Spain
Greg Francis,
Purdue University, United States

****Correspondence:***
Hadar Cohen-Duwek
hadarli@gmail.com

Specialty section:
This article was submitted to Perception Science, a section of the journal Frontiers in Neuroscience

Received: *13 December 2018*
Accepted: *26 February 2019*
Published: *22 March 2019*

Citation:
Cohen-Duwek H and Spitzer H (2019) A Compound Computational Model for Filling-In Processes Triggered by Edges: Watercolor Illusions. Front. Neurosci. 13:225. doi: 10.3389/fnins.2019.00225

INTRODUCTION

One of the most important goals of the higher levels of visual system processing is to reconstruct an appropriate representation of a surface after edge detection is performed by early vision. Such tasks are attributed to the opponent receptive fields in the retina and in the lateral geniculate nucleus (LGN). The visual system processing involves the cortical double-opponent as well as the simple and complex receptive fields, which perform non-oriented and oriented edge detection of both chromatic and non-chromatic edges (von der Heydt et al., 2003).

There are a number of visual phenomena and illusions that can provide information about the mechanisms that enable the reconstruction of surfaces from their edges. These include the watercolor illusions (Pinna et al., 2001) and the Craik-O'Brien-Cornsweet illusion (Cornsweet, 1970). In this study we will concentrate mainly on developing a computational model for the watercolor illusions to include a prediction of "conflicting" watercolor effects.

The Watercolor Effect described in the literature refers to a phenomenon involving assimilative color spreading into an achromatic area, produced by a pair of heterochromatic contours

surrounding an achromatic surface area (Pinna et al., 2001; Pinna, 2008; Devinck and Spillmann, 2009). The coloration extends up to about 45° (visual degree) and is approximately uniform (Pinna et al., 2001).

There have been many studies that investigated the chromatic and the luminance parameters required for the two inducing contours and for the inducing contours and background of the watercolor effect (Pinna et al., 2001; Devinck et al., 2005, 2006, 2014; Pinna and Grossberg, 2005; Pinna and Reeves, 2006; Tanca et al., 2010; Cao et al., 2011; Devinck and Knoblauch, 2012; Hazenberg and van Lier, 2013; Coia and Crognale, 2014; Coia et al., 2014). The conclusion was that even though many color combinations can produce the effect, the strongest result is induced by a combination of complementary colors. The studies of Pinna et al. (2001),Devinck et al. (2005, 2006) characterized these findings as assimilation effects (i.e., the perceived color is similar to the color of the nearest inducer). Reversing the colors of the two inducing contours, reverses the resulting perceived colors accordingly (Pinna, 2008).

However, a non-assimilation effect of coloration has also been discussed (Pinna, 2006; Kitaoka, 2007). Pinna (2006) reported that if one of the inducers is achromatic, while the other is chromatic, the induced color can be complementary to that of the chromatic inducer. Kitaoka (2007) demonstrated that a combination of red-magenta or green-cyan can give rise to a yellowish coloration, indicating that the perceived effect may not be completely attributable to assimilation effects. Indeed, an achromatic watercolor effect has been recently proved to exist, albeit with a lower magnitude than the chromatic watercolor effect (Cao et al., 2011).

The only computational model that has been reported to explain the watercolor effect is called the "Form And Color And Depth" (FACADE) model (Grossberg and Mingolla, 1985) and is based on neurophysiological evidence from neurons in the cortical areas V1–V4 (Pinna and Grossberg, 2005). This model also attempts to explain a number of other visual phenomena including the Kaniza illusion (Kanizsa, 1976), neon color spreading (van Tuijl and Leeuwenberg, 1979), simultaneous contrast, and assimilation effects. FACADE describes two main visual processing systems: a boundary contour system (BCS) that processes boundary or edge information; and a feature contour system (FCS) that uses information from the BCS to control the spreading (filling-in) of surface properties such as color and brightness. According to this model, higher contrast boundaries in the BCS inhibit lower-contrast boundaries thereby enabling color to flow out through weaker boundaries.

A number of studies have proposed the FACADE model as a possible mechanism for predicting the watercolor effect since it explains some of the properties of the phenomenon (Grossberg et al., 2005; Pinna and Grossberg, 2005; Pinna, 2006; Tanca et al., 2010). However, neither the mathematical equations of the FACADE model nor other previous studies have succeeded in simulating and predicting all the experimental findings concerning the watercolor effect. Moreover, the FACADE model cannot predict the non-assimilative version of the watercolor effect (Pinna et al., 2001; Kitaoka, 2007; Hazenberg and van Lier, 2013; Kimura and Kuroki, 2014a). Kitaoka (2007) observed that in the non-assimilative watercolor effect, the induced color becomes more prominent when the outer contour has a higher luminance (and thus a lower-contrast with respect to the white background) than the inner contour. In this case, the BCS in the FACADE model would be expected to inhibit the boundaries of the lower-contrast outer contour and permit the color of the outer contour to spread out. This prediction is not supported by the actual perceived color as demonstrated in **Figure 5**, where a yellowish color spreads in and there is no perceived magenta color that spreads out, as the FACADE model would predict.

At present, the visual mechanisms responsible for the watercolor effect are still unknown and the watercolor effect "presents a significant challenge to any complete model of chromatic assimilation" (Devinck et al., 2014).

In their study on the watercolor effect, Knoblauch et al. (Devinck et al., 2014) summarized the requirements for a future computational model: "In a hierarchical model, two other steps need to be considered, surface detection then color filling-in."

In this study, we present a computational model, which detects edges through biological receptive fields, modifies them, and then applies them as a trigger for a diffusive filling-in process. The objective of the model is to predict both the assimilative and the non-assimilative configurations of the watercolor effect.

COMPUTATIONAL MODEL

The main building blocks of the model are: (A) The inducing stimulus (B) The chromatic and achromatic opponent receptive fields (RFs). (C) The oriented double-opponent RFs, which detect chromatic and achromatic edges. (D) Calculation of the modification value through determination of the dominant chromatic/achromatic stimulus edge among several edges, which have different spatial scales. (E) Calculation of the new modified edges that trigger a diffusive filling-in process. (F) The filling-in process, performed by solving the Poisson equation. (G) The perceived afterimage of both the assimilative and the non-assimilative watercolor effects (**Figures 1A-G**).

Model Assumptions

The model is based on the following assumptions: (A) The visual system needs to reconstruct surfaces that are not represented in the early vision stages, which perform chromatic and achromatic edge detection (in the retina and the cortical V1 and V2 areas). In addition, we assume that in cases such as the watercolor stimuli, the visual system performs filling-in processes in order to make an "educated guess" and to reconstruct surfaces. (B) Each edge triggers a diffusion process and determines its color (Cohen-Duwek and Spitzer, 2018). (C) The trigger for the diffusion process is determined by the interactions between the gradients of the image, i.e., the gradients between the inner contour (IC) and the outer contour (OC), the gradients between the IC and the background, and between the OC and the background. The exact contribution of each gradient is determined automatically according to the chromatic and achromatic stimulus. (D) The visual system uses separated chromatic opponent channels [L/M, (L+M)/S and achromatic], in order to process each contrast color pathway separately (Kandel et al., 2012). This assumption

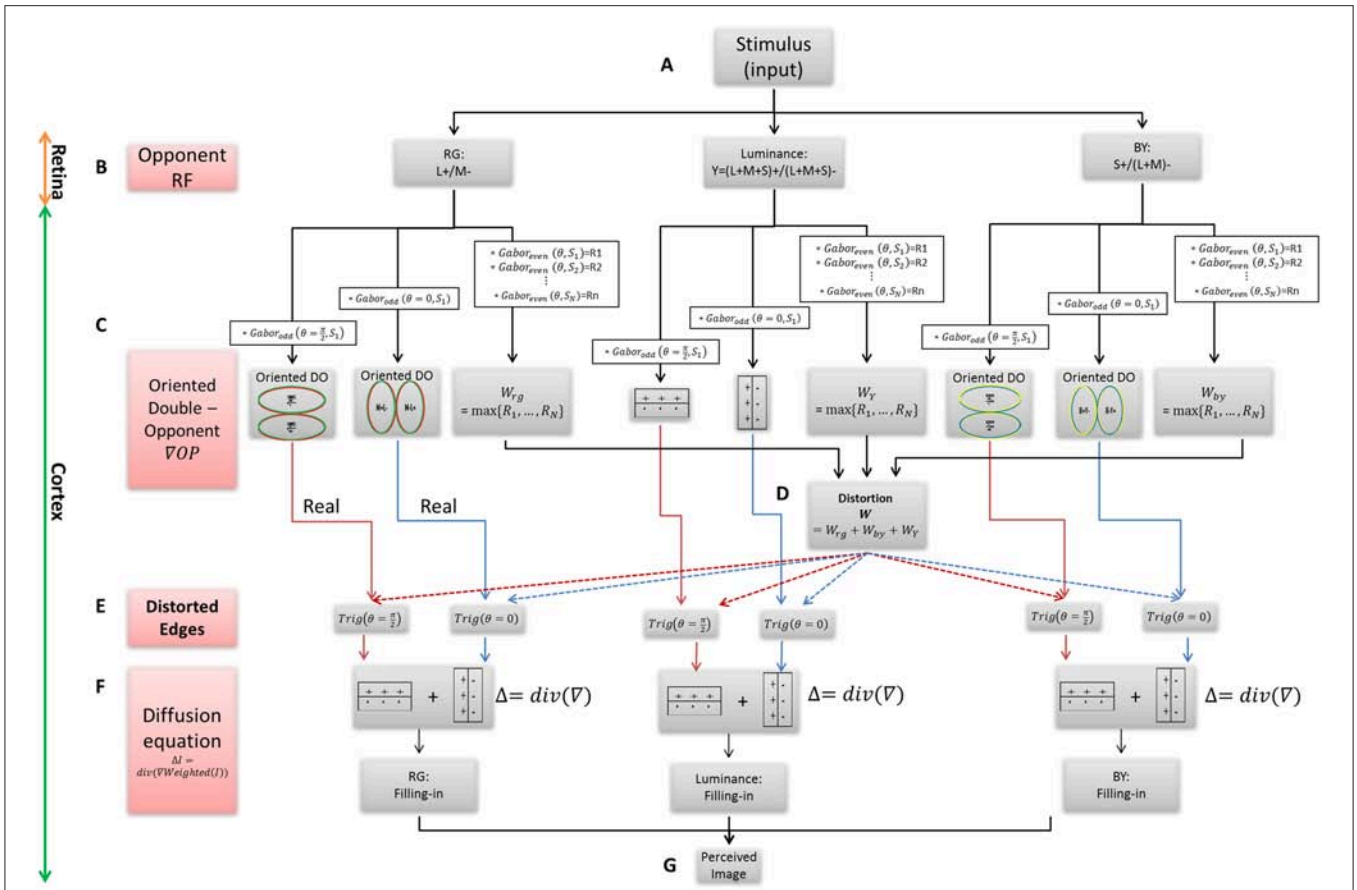


FIGURE 1 | The Flowchart of the suggested filling-in model. **(A)** The chromatic stimulus. **(B)** The opponent RFs, which are used as the first derivative in the chromatic and chromatic channels. **(C)** The oriented RFs, which represent the real chromatic gradients of the stimulus. **(D)** The calculation of weight function W for the modification of the gradients. **(E)** The calculation of the triggers for the filling-in process, i.e., the real and the modified edges where each of them exists in the x and y directions. **(F)** The filling-in process, calculated by the Poisson equation. **(G)** The solution of the diffusion equation yields to the perceived image.

is in agreement with experimental studies which claimed that the (L/M) and S-cones are regulated differently with respect to the watercolor effect (Devinck et al., 2005; Kimura and Kuroki, 2014a,b). (E) The chromatic channels are mediated by the Luminance channel (the achromatic channel). This assumption is supported by the observation that there is color spreading in response to a stimulus where both the IC and OC have the same color (hue) but a different luminance (Devinck et al., 2006).

Rationale for the Model

The early stages of the visual system, the retina, and the early visual areas V1 and V2, have receptive fields (RFs) that mainly detect edges. In the retina, for example, the opponent receptive fields perform a Difference of Gaussian (DOG) operation, which is approximately a second spatial derivative while the chromatic retinal opponent RFs performs derivatives on the color domain. The simple and complex RFs in the V1 and V2 areas perform oriented edge detection. It has been assumed that at higher visual processing levels, the system acts to reconstruct the surfaces that are not represented (lacked) by the early visual areas. In order to perceive the physical world and not only its edges/gradients, the system (visual system) needs to reconstruct the image from its edges (von der Heydt et al., 2003). To mimic the original surfaces, the system could use the image's original gradients (in a similar fashion to that used in the engineering world, i.e., by solving the Poisson equation or by any parallel method (Bertalmio et al., 2000; Pérez et al., 2003). However, we now believe that in addition, the visual system also performs additional tasks, which can be regarded as "educated guesses" in order to enhance important information in the scene. Examples of such "educated guesses" include: edge completion, detection of occluded objects in the image, and the interpretation of specific gradients as indicative of adjacent surfaces. The watercolor stimulus is such an example of specific edges, where the visual system supplies a guess regarding the chromatic surface. We suggest here, that this educated guess calculation is achieved by modifying the gradients and modifying the weights of the image gradients. In addition, we describe a set of rules that determine how the weights are calculated in the context of the stimulus.

In order to produce the chromatic (or the achromatic) diffusion process, the visual system needs to enhance or change the original gradients in order to obtain an image which creates

the perception and avoids a return to the original image. Based on psychophysical findings, the model assumes that the chromatic edges, which determine the filling-in effect, are significantly influenced by the intensity and by the chromaticity of the contours (IC and OC) (Pinna et al., 2001; Devinck et al., 2005, 2006; Pinna and Grossberg, 2005; Pinna and Reeves, 2006; Cao et al., 2011; Hazenberg and van Lier, 2013; Coia and Crognale, 2014; Kimura and Kuroki, 2014a,b).

The Watercolor Stimulus

The input of the model comprises the watercolor stimulus and its variations, which are composed of a pair of heterochromatic contours surrounding achromatic surface areas, **Figure 1A**.

Chromatic and Achromatic Opponent RF

The first component of the model (**Figure 1B**) is designed to simulate the opponent receptive fields (Nicholls et al., 2001). The spatial response profile of the retinal ganglion RF is expressed by the commonly used DOG. The "center" signals for the three spectral regions, *L*, *M*, and *S*, (Long, Medium, and Short wavelength sensitivity, respectively) that feed the retinal ganglion cells, are defined as the integral of the cone quantum catches, L_{cone}, M_{cone}, *and* S_{cone} with a Gaussian decaying spatial weight function (Shapley and Enroth-Cugell, 1984; Spitzer and Barkan, 2005):

$$\begin{aligned} i_c &= i_{cone} * f_c;\ i \in \{L, M, S\} \\ i_s &= i_{cone} * f_s;\ i \in \{L, M, S\} \\ f_j &= \frac{\exp\left(\frac{-(x^2+y^2)}{\rho_j^2}\right)}{\pi \rho_j^2}, \quad j \in \{c, s\} \end{aligned} \tag{1}$$

Where L_c, M_c *and* S_c represent the response of the center area of the receptive field of each cell type, Equation 1. L_s, M_s, *and* S_s represent the surround sub-region of these receptive fields. ρ_c and ρ_s represents the radius of the center and the surround regions, of the receptive field of the color-coding cells, respectively. f_c and f_s are the center and surround Gaussian profiles, respectively and * represents the convolution operation.

For the center-surround cells, the opponent responses are expressed as: $OP_{L^+M^-}$, $OP_{S^+(L+M)^-}$ and Y (for the summation of the L, M, and S channels) in order to express the Luminance channel.

$$\begin{aligned} &OP_{RG}: \ OP_{L^+M^-} = L_c - M_s \quad (\text{Red} - \text{Green channel}) \\ &OP_{BY}: \ OP_{S^+(L+M)^-} = S_c - (L+M)_s \quad (\text{Blue} - \text{YellowChannel}) \\ &Y = L_c + M_c + S_c \quad (\text{Luminance channel}) \end{aligned} \tag{2}$$

Where L_c, M_c, s_c, L_s, M_s, and S_c are the cell responses to the receptive filled sub-regions: center and surround, Equation (1).

Oriented Double-Opponent RF

The color coding of the opponent receptive fields, Equation (2), encodes color contrast, but not spatial contrast. In other words, the color opponent receptive fields are able to differentiate between colors, but cannot detect spatial gradients or edges (Conway, 2001; Spitzer and Barkan, 2005; Conway and Livingstone, 2006; Conway et al., 2010). The double opponent receptive fields, however, are sensitive to both spatial and chromatic gradients (Spitzer and Barkan, 2005) since they have color opponent receptive fields both at the center and in the surround RF regions (Shapley and Hawken, 2011). A large number of studies have reported that many double-opponent neurons are also orientation-selective (Thorell et al., 1984; Conway, 2001; Johnson et al., 2001, 2008; Horwitz et al., 2007; Solomon and Lennie, 2007; Conway et al., 2010). Accordingly, the model takes into account the oriented double opponent RF, ODO, to the three opponent RF channels, $OP_{L^+M^-}$, $OP_{S^+(L+M)^-}$, and Y (Conway and Livingstone, 2006), Equation (2). We modeled this chromatic RF structure, $ODO_{L^+M^-}$, $ODO_{S^+(L+M)^-}$ and OY by a convolution between the Gabor function and the opponent responses, Equation (3), **Figure 1C**. It should be noted that previous work indicates that by using the linear Gabor function, we neglect some non-linearities e.g., half wave rectification in the simple cells and full rectification in the complex cells, in the neuronal responses (Movshon et al., 1978; Spitzer and Hochstein, 1985).

$$\begin{aligned} ODO_{L^+M^-} &= OP_{L^+M^-} * Gabor_{odd,\theta,\sigma} \\ ODO_{S^+(L+M)^-} &= OP_{S^+(L+\ M)^-} * Gabor_{odd,\theta,\sigma} \\ OY &= Y * Gabor_{odd,\theta,\sigma} \end{aligned} \tag{3}$$

$$\begin{aligned} Gabor_{odd,\theta,\sigma} &= \exp\left(\frac{-\left(x'^2 + y'^2\right)}{2\sigma^2}\right)\sin(2\pi x') \\ Gabor_{even,\theta,\sigma} &= \exp\left(\frac{-\left(x'^2 + y'^2\right)}{2\sigma^2}\right)\cos(2\pi\ x') \\ \text{Where}: \quad x' &= x\cos(\theta) + y\sin(\theta) \\ y' &= -x\sin(\theta) + y\cos(\theta) \end{aligned} \tag{4}$$

This opponency in both spatial and chromatic properties produces a spatio-oriented-chromatic edge detector, Equation (3).

Where θ represents the orientation of the normal to the parallel stripes of a Gabor function and σ is the standard deviation of the Gaussian envelope of the Gabor function.

Gradient Weights

We chose to express this property of gradient modification by adding weighted functions to the Oriented-double-opponent RF (**Figure 1D**). The model modifies the original gradients (Equation 3) by multiplying the double-opponent responses by the weight function, Equation (6), **Figure 1D**. In order to calculate the weight functions, several Gabor-filters on different scales [different standard deviations, σ, Equation (5)] are calculated and the maximum response to a specific Gabor RF scale is chosen as the weight function for each channel separately, Equation (6). This maximum response represents the dominant gradient in the image, which is used by the model to determine the strongest effect on the diffusion process. This determination of the strongest effect (i.e., the strongest edge in the stimulus) is

in agreement with previously reported psychophysical findings (Pinna et al., 2001; Devinck et al., 2005, 2006; Kimura and Kuroki, 2014a,b). The multiplication operation of the chosen weight is done with a 2D Gabor filter, Equation (5). (It should be noted that we could also obtain good results by making a summation of the responses from all scales).

$$\begin{aligned} R_{RG,i} &= \left|OP_{RG} * Gabor_{even}(\theta, \sigma_i)\right| \\ R_{BY,i} &= \left|OP_{BY} * Gabor_{even}(\theta, \sigma_i)\right| \\ R_{Luminance,i} &= \left|Y * Gabor_{even}(\theta, \sigma_i)\right| \end{aligned} \tag{5}$$

Where σ_i represents different standard deviations of the Gaussian envelope (different scales).

$$\begin{aligned} W_{RG}(i,j) &= \max\{R_{RG,1}(i,j), R_{RG,2}(i,j), \ldots, R_{RG,N}(i,j)\} \\ W_{BY}(i,j) &= \max\{R_{BY,1}(i,j), R_{BY,2}(i,j), \ldots, R_{BY,N}(i,j)\} \\ W_Y(i,j) &= \max\{R_{Luminance,1}(i,j), R_{Luminance,2}(i,j), \ldots, \\ &\qquad R_{Luminance,N}(i,j)\} \end{aligned} \tag{6}$$

Where W_{RG}, W_{BY}, and W_{Lum} are the maximal responses among the several scales at each channel.

This calculation is done separately for both the chromatic channels and the achromatic channels (RG, BY, and Y). After determining which scale yields the strongest response at each channel, the three responses are summarized across the channels, Equation (7), to reflect a combination of all the edges in each spatial location. In other words, the weight function W, for each spatial location in the image (or stimulus), is taken as the normalized sum of the maxima, values from the strongest response scale, across all the channels, Equation (7).

$$W = W_{RG} + W_{BY} + W_Y \tag{7}$$

This calculation procedure can detect the middle chromatic (or achromatic) edge between the two contours (IC and OC), which are the triggers for the diffusion process. This detection is possible because in most cases, the dominant edge is a coarse edge, which contains the edge that is adjacent to the inner and the outer region. The center of this coarse region is often the edge between the two chromatic contours in the watercolor stimuli.

The Diffusion Triggers (Second Derivative)

The trigger for the diffusion process consists of the sum of two components: the modification component (β) and the "real" (α) oriented double-opponent RF component, Equation (8). These modification components are added separately for each orientation directions and then, the modified gradients are convolved again with an odd Gabor filter (in the same orientation, θ), Equation (10), in order to perform a second derivative. Both derivative direction (x and y axis, $\theta = 0$ and $\theta = \frac{\pi}{2}$) are then summarized in order to create the divergence, Equation (10), **Figure 1F**, which is then used as the trigger for the diffusion process in all the required directions, Equation (10), across each of the channels. The trigger for the diffusion process is the oriented-double-opponent response, Equation (3), multiplied by the weight function (W) in each individual channel, **Figure 1E**, Equation (8).

$$\begin{aligned} Trig_{RG} &= ODO_{RG} \cdot (\alpha + \beta W(x,y)) \\ Trig_{BY} &= ODO_{BY} \cdot (\alpha + \beta W) \\ Trig_Y &= OY \cdot (\alpha + \beta W) \end{aligned} \tag{8}$$

Where α and β are constants and $\alpha > \beta$. $Trig_{RG}$, $Trig_{BY}$, and $Trig_Y$ are the diffusion triggers in each channel.

Note that the results of the above equations change only the weights of the ODO (Equation 3) responses, and therefore their spatial properties and polarities are retained. According to the suggested model, the prominent gradient makes the strongest contribution to the filling-in process, Equation (7). However, the other two gradients also contribute to the filling-in process, due to the chromatic and achromatic strength of their gradients. This consideration of the different gradients is in agreement with the Weber contrast rule (Kimura and Kuroki, 2014a).

Filling-In Process

The filling-in process is expressed by the diffusion (or heat) Equation (10) (Weickert, 1998), and is determined according to the weighted triggers, Equation (8), **Figure 1E**. The model assumes that the filling-in process represents "isomorphic diffusion" (von der Heydt et al., 2003; Cohen-Duwek and Spitzer, 2018), although it does not necessarily negate other possible filling-in mechanisms, such as "edge integration" (Rudd, 2014). This filling-in process is reminiscent of the physical diffusion process, where the signals spread in all directions, until "blocked" by another heat source (image edges). We would like to emphasize that this type of filling-in infers that the borders (chromatic or achromatic) do not function primarily as blockers, but instead they act as heat sources that can trigger the diffusion. We would like to emphasize that this type of filling-in infers that the borders (chromatic or achromatic) do not function primarily as blockers, but instead they act as heat sources that can trigger the diffusion, and then spread in opposite directions and thus trap the diffused color. The diffusion spread, therefore, will be blocked by the heat source, in such a case. These principles are applied in our model through the well-known diffusion equation (Weickert, 1998):

$$\begin{aligned} &\frac{\partial \mathrm{I}(x,y,t)}{\partial t} - D\nabla^2 \mathrm{I}(x,y,t) = \mathrm{h}_s = -\mathrm{div}\left(Trig_c\right); \\ &\qquad \text{where } \mathrm{c} = \{L^+M^-, S^+(L+M)^-, \mathrm{Y}\} \end{aligned} \tag{9}$$

where $\mathrm{I}(x,y,t)$ denotes the image in a space-time location (x,y,t), D is the diffusion (or heat) coefficient, and h_s represents a heat source. The time course of the perceived image is assumed to be very fast, in accordance with previous reports (Pinna et al., 2001). This time course is also termed "immediate filling-in" (von der Heydt et al., 2003).

Following this assumption, for the sake of simplicity, we can ignore the fast-dynamic stages of the diffusion equation, and therefore compute only the steady-state stage of the diffusion

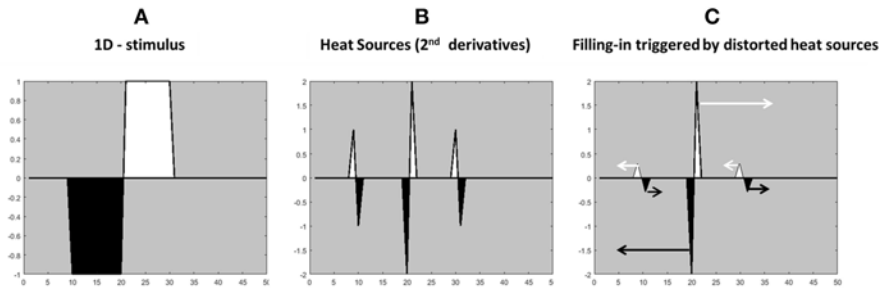


FIGURE 2 | Illustration of the calculation of the edges for the "heat sources" filling-in process from the stimulus gradients. **(A)** 1-D achromatic stimulus with white and black contours. **(B)** The second derivative of the stimulus **(A)**, with a negative sign. **(C)** The modified second derivative of the stimulus **(A)**. The arrows indicate the direction and the color of the diffusion process. The higher heat sources (the gradients in the middle) have a greater influence on the filling-in process.

process. Consequently, the diffusion (heat) Equation (5) is reduced to the Poisson Equation (10).

$$D\nabla^2 \mathrm{I} = -\mathrm{h}_s = \mathrm{div}\left(Trig_c\right); \quad \text{where } \mathrm{c} = \{\mathrm{RG}, \mathrm{BY}, \mathrm{Y}\} \tag{10}$$

$$D\nabla^2 \mathrm{I} = \mathrm{div}((\alpha + \beta W) \cdot \mathrm{ODO}) \tag{11}$$

The "heat sources" are the weighted second derivative of an opponent channel; **Figure 1E** (weighted oriented-double-opponent). The heat equation (diffusion equation) with heat sources requires second derivatives, reflecting the "heat generation rate" which is the second derivatives of a heat source. Because the edges are playing a role as heat sources, the values near the edges do not decay over time. Since the two adjacent edges operate as heat sources with opposite signs, the conclusion is that they are operating with opposite directions, and therefore the diffusion process of one color (one heat source) cannot diffuse to the "other" direction. This approach is not consistent with previous reports that the edges function as borders that prevent the colors from spreading (Cohen and Grossberg, 1984; Grossberg and Mingolla, 1985, 1987; Pinna and Grossberg, 2005). In the suggested model the derivatives trigger a positive diffusion process toward one side of the spatial derivative and a "negative diffusion" process to the other side of the spatial derivative, **Figure 2** demonstrates this type of diffusion, which is considered separately for each color channel.

METHODS

In this section we describe each stage of the model's implementation in detail.

Opponent RF

For the sake of simplicity, we compute the opponent response of the opponent receptive fields as color-opponent only, where each chromatic encoder has the same spatial resolution. This is computed by an opponent color-transformation (van de Sande et al., 2010), Equation (12). This transformation converts each pixel of the image I_0, in each chromatic channel R,G, and B into opponent color-space, via the transformation matrix O (van de Sande et al., 2010). In order to obtain more perceptual value in the luminance channel, we have slightly modified the transformation matrix O, and use a = 0.2989, b = 0.5870, and c = 0.1140, instead of using $a = b = c = 1/\sqrt{3}$ as originally reported (van de Sande et al., 2010). These values are taken from the Y channel in YUV (or YIQ) color space. The Y represents the Luma information: Y = 0.2989R + 0.5870B + 0.1140C. $\mathrm{I}_{\mathrm{OPPONENT}}$ = OPPONENT{RGB} as follows:

$$I_{OPPONENT} = \begin{pmatrix} O_{RG} \\ O_{YB} \\ O_Y \end{pmatrix} = \begin{pmatrix} 1/\sqrt{2} & -1/\sqrt{2} & 0 \\ 1/\sqrt{6} & 1/\sqrt{6} & -2/\sqrt{6} \\ a & b & c \end{pmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix} \tag{12}$$

Another perceptual option for the opponent transformation matrix is to use the transformation presented by Wandell (1995),

$$\begin{aligned} I_{OPPONENT} &= M_{OpponentW}\{M_{LMS}\{M_{XYZ}\{RGB\}\}\} \\ M_{XYZ} &= \begin{pmatrix} 0.4124 & 0.3576 & 0.1805 \\ 0.2126 & 0.7152 & 0.0722 \\ 0.0193 & 0.1192 & 0.9505 \end{pmatrix} \\ M_{LMS} &= \begin{pmatrix} 0.2430 & 0.8560 & -0.0440 \\ -0.3910 & 1.1650 & 0.0870 \\ 0.0100 & -0.0080 & 0.5630 \end{pmatrix} \\ M_{OpponentW} &= \begin{pmatrix} 1 & 0 & 0 \\ -0.59 & 0.80 & -0.12 \\ -0.34 & -0.11 & 0.93 \end{pmatrix} \end{aligned} \tag{13}$$

$$I_{OPPONENT} = \begin{pmatrix} O_Y \\ O_{RG} \\ O_{YB} \end{pmatrix} = \begin{pmatrix} 0.2814 & 0.6938 & 0.0638 \\ -0.0971 & 0.1458 & -0.0250 \\ -0.0930 & -0.2529 & 0.4665 \end{pmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix} \tag{14}$$

These matrix values are calculated from the linear conversion of the RGB color space to the XYZ color space, which is then converted to the LMS color space to which we apply the opponent transformation from Wandell (1995), Equation (13).

where O_{RG}, O_{YB}, and O_{Y}, Equations (12–14) are the new channels of the transformed image $\mathrm{I}_{\mathrm{OPPONENT}}$. R, G, and B are the red, green, and blue channels of the input image I, respectively.

Oriented Opponent and Double-Opponent RF

The oriented opponent RFs are modulated as convolution between each opponent channel and an odd Gabor function, Equation (4). For the sake of simplicity, we discretized the Gabor function and instead of computing the exact Gabor functions, we used a discrete derivative filter in two directions, vertical (y-axis, $\theta = 0$), and horizontal (x-axis, $\theta = \frac{\pi}{2}$), Equations (15–16) (Gonzalez and Woods, 2002).

$$Gabor_{odd,x} \approx G_{odd,x} = [-1, 1]\,;\; Gabor_{odd,y} \approx G_{odd,y} = \begin{bmatrix} -1 \\ 1 \end{bmatrix} \tag{15}$$

$$Gabor_{even,x} \approx G_{oeven,x} = [-1,\ 2, -1]\,;\; Gabor_{even,y} \approx G_{even,y} = \begin{bmatrix} -1 \\ 2 \\ -1 \end{bmatrix} \tag{16}$$

The above discretization of the Gabor filters: $G_{odd,x}$ and $G_{odd,y}$ also represent the discrete gradient operator ∇ :

$$\nabla I = (\nabla_x I, \nabla_y I) = (I * G_{odd,x}, I * G_{odd,y}) \tag{17}$$

The structure of the oriented-double-opponent receptive field can be seen as a filter which acts as a second derivative in both the spatial and chromatic domains.

Weights of Modified Edges

In order to calculate the response of an opponent channel to a Gabor RF on different scales, Equation (5), we use a Gaussian Pyramid (Adelson et al., 1984). In this way, the image is down-sampled instead of up-sampling the Gabor filter.

$$R_{c,i} = \left| GaussianPyramid\{OP_c\}_{\sigma_i} * Gabor_{even}\,(\theta) \right| \tag{18}$$

Filling-In Process

The divergence operator, div Equation (10), is computed as:

$$div\,(F) = \frac{\partial F}{\partial x} + \frac{\partial F}{\partial y} = F * G_{odd,x} + F * G_{odd,y} \tag{19}$$

Where F is an image input.

Therefore, Equation (10) can be written as:

$$\triangle I_{op} = \nabla^2 I_{op} = \mathrm{div(Trig)} = Trig_x * G_{odd,x} + Trig_y * G_{odd,y} \tag{20}$$

Parameters

We performed a set of simulations in order to determine the constants α and β . We found that increasing the β parameter (increasing the weight of the modified gradient, *ODO*, Equation 8) increases the saturation of the predicted result (since the level of the relevant gradient is increased). This means that choosing a higher value for β increases the saturation of the filled-in predicted color and also increases its intensity while preserving its hue. The α parameter affects the magnitude of the original gradient of the original stimulus. We arrived at the conclusion that the ratio between α and β determines the level of the filled-in predicted saturation. In all the simulations presented here $\alpha = 1$ and $\beta = 0.5$.

Comparison to Psychophysical Findings

In order to compare the predictions of the model to psychophysical findings we created sets of images that contain the same color values that have been used in previous psychophysical experiments (Devinck et al., 2005; Kimura and Kuroki, 2014b). Each color value used in the stimulus was converted from the CIE Lu'v' 1976 color space to the sRGB color space, in order to create the input images for the model. The model was then applied to each image stimulus, and the predicted colors were calculated and converted back to the CIE Lu'v' 1976 color space. These CIE Lu'v' 1976 color values are presented in the results section.

RESULTS

The results present the simulations of the model through its equations (according to the Methods section) implemented by MATLAB software. The model's equations were solved in a similar way to that reported in "Methods for Solving Equations" (Simchony et al., 1990) but another option was through "Poisson Image Editing" (Pérez et al., 2003).

Model's Simulation and Predictions

The model and simulation results (**Figure 1G**) are divided into three parts. The first part presents the model predictions for the assimilative (classic) watercolor effect. The second part presents the predictions of the model for the non-assimilative (non-classic) watercolor effect, while the third part presents the model predictions that relate to additional properties of the watercolor effect: the influence of the background luminance, and the effect of the inner color luminance on the perceived hue and the perceived brightness (Devinck et al., 2005, 2006; Cao et al., 2011; Kimura and Kuroki, 2014a,b).

Predictions—Assimilative (Classic) Watercolor Effect

The model simulations were tested on a large number of classic stimuli with a variety of chromatic thin polygonal curves (e.g., star shapes) that produce the watercolor effect. **Figure 3** shows that the model succeeded in predicting the correct coloration of the classic assimilative watercolor effect. Note that the most of the assimilative watercolor effects present the complementary colors of the IC and the OC (the IC and the OC color are complementary in these stimuli). Our model indeed predicts a strong filling-in color response to such stimuli, **Figures 3A–C**.

Figure 3 demonstrates that the filling-in perceived color is more prominent in the predicted result (right side), which represents the model prediction for the corresponding stimulus, i.e., the original stimuli (left side). The filling-in effect of the stimuli with orange and purple polygonal edges were obtained as expected, **Figure 3A**, as well as a reddish color and cyan, **Figure 3B**. The level of saturation in the simulation results can be controlled by the parameters α and β , Equation (8). We also tested our model with achromatic watercolor stimulus. **Figure 3C** shows that the model correctly predicts a perceived darker or lighter inner area, according to the luminance of the inner contour.

Comparison to psychophysical findings

We confronted our model predictions with quantitative psychophysical results (Devinck et al., 2005). **Figure 4** presents the predictions of the model in CIE Lu'v' (1976) coordinates instead of RGB images, see Methods. In order to enable the comparison between the model predictions and the psychophysical results, we applied the same set of colors as described in Devinck et al. (2005), as parameters to our model, see Methods.

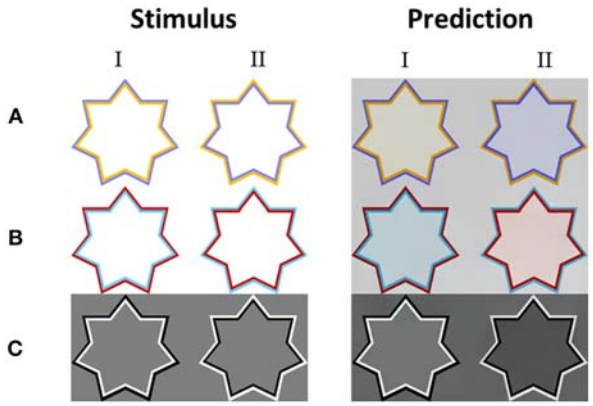


FIGURE 3 | The model's predictions for assimilative watercolor stimuli. **(A)** The classic watercolor stimulus (left) and the model's predictions (right). **(B)** Additional example of an assimilative watercolor stimulus (left), with different colors, and the model's predictions (right). **(C)** An example of achromatic watercolor stimulus (left) and the model's predictions (right). Our model predicts that in the assimilative watercolor stimuli, the inner contour color is spread to the inner area of the stars.

Figure 4 demonstrates the comparison of the model prediction with Devinck et al. (2005) findings, which tested the assimilative effect on three pairs of colors: Orange and Purple, Red and Green, and Blue and Yellow. Note that, the psychophysical findings are obtained from a hue cancellation test and therefore represent the complementary colors of the perceived colors; however, our results represent the predicted perceived colors. Most of the predicted colors, **Figure 4A**, are in agreement with the psychophysical findings, **Figure 4B**. Only in the orange and the purple stimuli pair the predicted color is slightly more yellowish then in the psychophysical findings for the IC: Orange OC: Purple stimulus (**Figure 4A** top left) and slightly more bluish then in the psychophysical findings for the IC: Purple OC: Orange stimulus (**Figure 4A** top right).

Predictions—Non-assimilative (Non-classic) Watercolor Effect

We also tested two known versions of the non-assimilative watercolor effect (Pinna, 2006; Kimura and Kuroki, 2014a). In

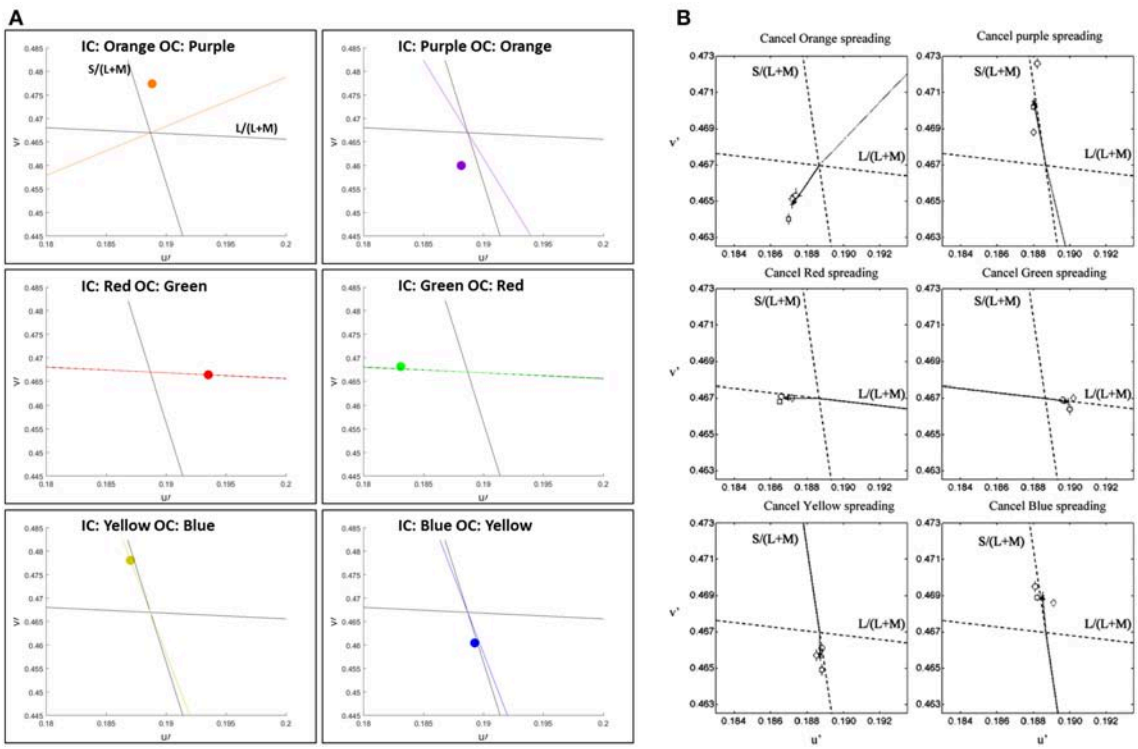


FIGURE 4 | Comparison between the predictions of the model and the psychophysical findings of the assimilative effect, both presented in u'v' (CIELu'v' 1976) color space. The prediction of the model **(A)** and the chromatic cancelation data **(B)** that are taken from Devinck et al. (2005). Each row **(A,B)** presents a pair of IC and OC colors, which are orange–purple, red–green, and blue–yellow, respectively. The colored dots **(A)** represent the predicted results. The colored lines **(A)** represent the hue line of the IC contour color that was used in each pair of contours.

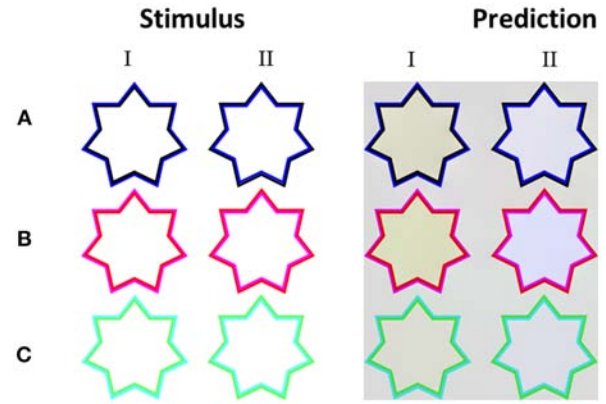


FIGURE 5 | The model's predictions for non-assimilative watercolor stimuli. Each row presents different color variation of the inner and outer contours **(A–C)**. Column I presents color configurations which produce a perceived non-assimilative effect, while column II presents color configurations that produce a perceived assimilative effect, even though the diffused perceived color does not reflect the color of the inner contour only. The colors predicted by the model are yellowish in the non-assimilative configurations (I), and blueish in the opposite assimilative configuration (II).

this case, we chose to test the three chromatic stimuli colors as tested originally by Kimura and Kuroki (2014a) for the non-assimilative watercolor effect. The stimuli in these versions have chromatic and achromatic edges/contours (**Figure 5A**) or specific pairs of colors (**Figures 5B,C**).

Kimura and Kuroki (2014a,b) psychophysically tested stimuli similar to those in **Figures 5A**,**B** and found that the induced colors were yellowish. The psychophysical results also demonstrated that a stimulus such as that in **Figure 5A** (left star), yielded a complementary color (yellowish) to the OC (bluish). Our model correctly predicts these complementary perceived coloration effects (filling-in effect), **Figure 5A** (left star).

Again, in accordance with psychophysical findings, our model could also correctly predict the influence exerted by the location of the chromatic contours, as to whether the same or complementary filling-in color is perceived in the inner area (Pinna, 2006; Kimura and Kuroki, 2014a), **Figure 5A**.

Kimura and Kuroki (2014a) observed that the perceived colors were not necessarily the "same" as or "complementary" to the IC/OC, but could be a combination of the IC and OC colors, **Figures 5B,C** (left stars). In agreement, the model results (**Figure 5**II) show indeed that the perceived color is determined by combination of the outer and the inner contours. In **Figure 5B** (left star), for example, the red IC contributes the same (red) color to the coloration effect, while the magenta OC contributes its complementary color (green). An additive combination of red and green colors yields a perceived yellowish coloration (Berns, 2000). These results are consistent with the model principles and Equations [Filling-in process; Equation (10)], such that both the IC and OC contours contribute as triggers to the filling-in process. The model correctly predicts the general trend that has been shown in previously reported experimental results (Pinna and Reeves, 2006) where the perceived chromatic filling-in color was determined by the combined influence of the chromatic and achromatic edges.

Comparison to psychophysical findings

Furthermore, we confronted our model predictions with quantitative psychophysical results (Kimura and Kuroki, 2014b). In order to enable the comparison between the model predictions and the non-assimilative watercolor effect experiment results, we applied the same set of colors as described in the results of Kimura and Kuroki (2014b), as parameters to our model, see Methods.

The psychophysical experiments of Kimura and Kuroki (2014b) investigate both the assimilative and the non-assimilative effects as well as the role of intensity in the perceived effect. **Figure 6** presents the model predictions and the results of Kimura and Kuroki (2014b) on a large repertoire of stimuli.

Figure 6 presents the predicted (A) and experimental results (B) of stimuli that share the same IC color at each sub-figure while the experiment tested 8 different OC colors. The top row presents the results for the red IC color and the bottom row presents the result for the achromatic IC color, while the outer color was presented with different chromatic colors. Left column presents the result when the IC color has a higher luminance level and the right column present the results when the IC color has a lower luminance level.

The stimuli with higher luminance of the red IC (**Figure 6B**) yielded perceived colors which were ranged from red to orange. Therefore, this trend of results shows an assimilative reddish color effect. The predicted result (**Figure 6A**) shows assimilate effects in adjustment to the red line. However, the perceived color is more reddish than orange as in the experimental results (**Figure 6B**). The stimuli with lower luminance of the red IC (**Figure 6B**) yielded an oval shape adjacent to the -S line. Our result also predicts an oval shape, but the shape is adjacent to the L line. It will be discussed in Discussion. The stimuli with higher luminance of the achromatic IC yielded a small magnitude of the perceived effects, in both the experimental (**Figure 6B**) and the predicted (**Figure 6A**) results. However, in the experimental results the effects slightly tend to be yellowish, while in the predicted results the effect is almost invisible (no filling-in effect). The stimuli with lower luminance of the achromatic IC also yielded a yellowish perceived color in the experimental results. In the predicted result the predicted colors are the complementary colors of the OC. It has to be noted that the achromatic configuration of the experimental result were tested also in additional studies such as Pinna (2006) and Hazenberg and van Lier (2013), and their trend of results are in better agreement with the prediction of the model (**Figure 6A**), see Discussion.

The Role of the Luminance Contrast Between the IC and the OC

Having discussed the model's predictions to highly saturated stimuli from the literature with different variations of chromatic properties (**Figures 3**, **5**) we then tested the model's predictions for stimuli with different luminance as well as different chromatic properties. Devinck et al. (2005) and Pinna et al. (2001) showed that the magnitude of the filling-in effect increases with

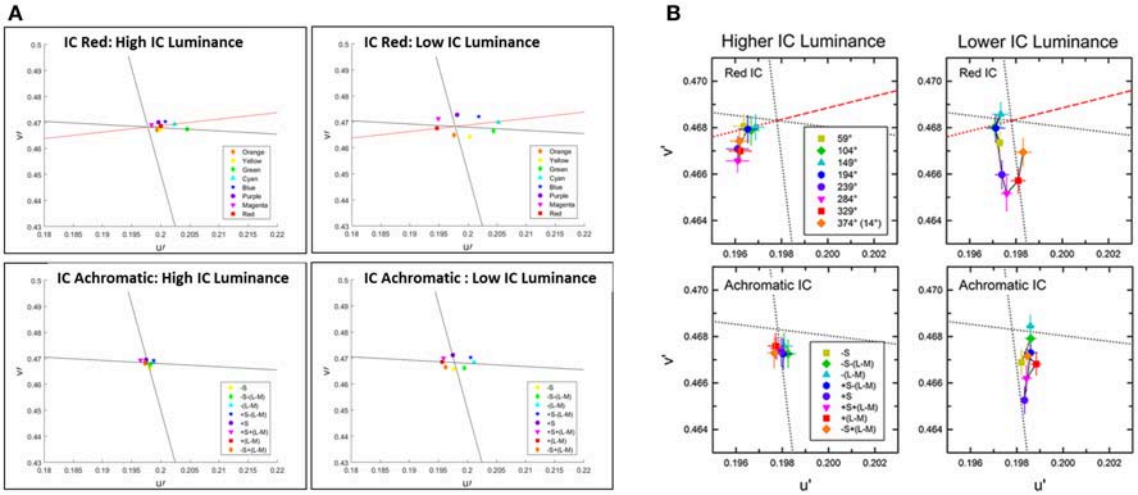


FIGURE 6 | Comparison between the predictions of the model and the psychophysical findings for the assimilative and non-assimilative effects. The prediction of the model **(A)** and the chromatic cancelation data **(B)** where done for 8 different colors of the OC, similarly as **Figure 4** Kimura and Kuroki (2014b). Top row **(A,B)** presents the experimental **(B)** and the predicted **(A)** results to stimuli with red IC. Bottom row present the experimental **(B)** and the predicted **(A)** results to stimuli with achromatic IC and the 8 different colors for the OC. In the left Column at each subfigure **(A,B)** the luminance of the IC is higher than the luminance of the OC. In the right column at each subfigure **(A,B)** the luminance of the IC is lower than the luminance of the OC as in Kimura and Kuroki (2014b).

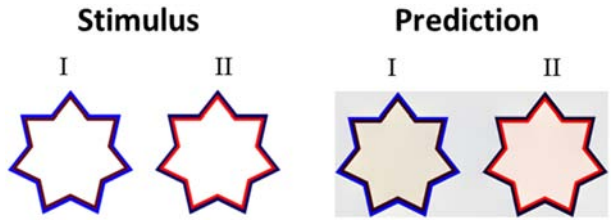


FIGURE 7 | The role of the luminance level of the IC and the OC. In both stimuli (I,II) the hue of the IC is reddish and the hue of the OC is bluish, but with a different level of luminance intensity. In stimulus I, the IC has a low luminance level (dark red), while the OC has a high level of intensity. The predicted color is yellowish (I right), thus the perceived effect is a non-assimilative effect. In stimulus II, the IC has a high luminance level, while the OC has a low intensity (dark blue). The predicted color is reddish (II right), thus the perceived result is due to an assimilative effect.

increasing luminance contrast between the relevant contours. Our model predicts this effect of luminance contrast between the IC and OC. **Figure 7** presents the model predictions to a "switching" effect (non-assimilative: **Figure 7**I vs. assimilative: **Figure 7**II) whereby the luminance contrast determines whether the perceived effect will be assimilative or non- assimilative (Kimura and Kuroki, 2014a). Even though the IC color in both stars is reddish and the OC color blueish, the predicted colors are different (pale yellowish in the left star and pale reddish in the right star), **Figure 7**. It should be noted that in this case, the model's prediction is in agreement with the experimental results of Kimura and Kuroki (2014a) that showed that the luminance condition suitable for the non-assimilative color spreading is the reverse (in their Weber contrast) of the assimilative color spreading. We argue that these experimental findings (Kimura and Kuroki, 2014a) shed a new light on the common assumption in the literature that assimilative and non-assimilative are different effects and are derived from different mechanisms (Kimura and Kuroki, 2014a,b). This topic will be discussed in more detail in the Discussion.

An additional important finding relates to the claim that only the assimilative type of watercolor effect is possible when the IC and the OC have the same level of luminance (Devinck et al., 2005). Accordingly, our model predicts that the assimilative effect should be perceived under such iso-luminance conditions and also predicts that the effect will be weaker than when the IC and the OC have different luminance values.

The Role of the Luminance Contrast Between the Background and the Contour

Several experimental studies that tested the role of background luminance on the perceived watercolor effect (Devinck et al., 2005; Cao et al., 2011; Kimura and Kuroki, 2014a) reported that the luminance contrast between the IC and the background, and between the OC and the background have a significant influence on the perceived effect.

Figure 8A presents the model's predictions for a response to the same stimuli used by Kimura and Kuroki (2014a), indicating that when the background is white (high luminance), the perceived color is yellowish. In contrast, when the background is darker (low luminance, **Figures 8A,B**), there is a tendency to a more greenish perceived color. This is because a change in the luminance of the background produces a change in the contrast between the contours (IC and OC) and the background, which in turn, influences the perceived effect. Importantly, the changes in

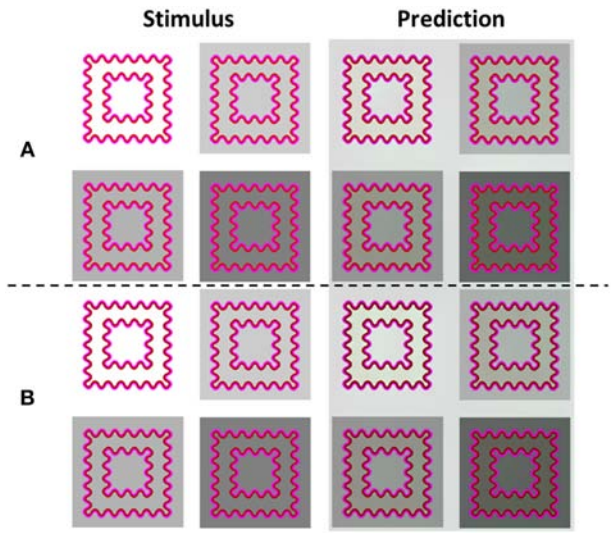


FIGURE 8 | The influences of the background luminance and the luminance ratio between the IC and the OC on the predicted filling-in colors. The left column (Stimulus) presents the original stimuli. The right column (Prediction) presents the model's predictions. The IC/OC ratio is higher in rows **(A)** than in rows **(B)**, because the IC is darker in **(B)** than in **(A)**. When the IC is darker **(B)**, the predicted color is greenish (more prominent in the predicted images, Prediction), while when the IC is lighter **(A)** the predicted color is yellowish, when the background luminance is high (**A**: upper left, Prediction) and greenish when the background is darker.

perceived color predicted by the model were in accordance with the experimental results (Kimura and Kuroki, 2014a).

Figure 8B demonstrates that there are three options for luminance contrast that play a role in the watercolor effect. The first one is the contrast between the IC and the OC, the second, the contrast between the IC and the background, and the third one is the contrast between the OC and the background. In **Figure 8B** the luminance of the IC is lower than in **Figure 8A**. As a result, the perceived filling-in color appears greenish in the stimulus with the white background (high background luminance). In contrast, the perceived filling-in color in **Figure 8A** appears yellowish. These perceived coloration effects were intensified in the model's simulation (**Figure 6** right) and support the suggestion that both the background and the luminance ratio between the IC and the OC contribute to the perceived effect. These predictions are in agreement with the psychophysical findings of Kimura and Kuroki (2014a).

DISCUSSION

We present here a generic computational model that describes the mechanisms of the visual system that activate the creation of chromatic surfaces from chromatic and achromatic edges. Our hypothesis was that these mechanisms can be revealed through a study of visual phenomena and illusions, such as the assimilative and non-assimilative watercolor effect and the Craik–O'Brien–Cornsweet (COC) illusions. The suggested model can be divided into two stages (or components). The first component determines the dominancy of the edges that trigger a diffusive filling-in process. The second component performs the diffusive filling-in process, which triggers the diffusion by heat sources. This process is modeled by the Poisson equation. The diffusion process is actually the same mechanism described for the afterimage effect (Cohen-Duwek and Spitzer, 2018).

In order to test the hypothesis, we developed a computational model that is able to predict both the assimilative and the non-assimilative watercolor effects. The model predictions, which are supported by psychophysical experiments (Pinna et al., 2001; Devinck et al., 2005, 2006; Pinna and Grossberg, 2005; Pinna and Reeves, 2006; Cao et al., 2011; Coia and Crognale, 2014; Kimura and Kuroki, 2014a,b), argue that both the assimilative and non-assimilative watercolor effects are derived from the same visual mechanism. In addition, the model can successfully predict quantitatively and qualitatively the psychophysical results reported by many researchers, such as the influence of the background luminance, contour intensities, contour saturations, and the relationship between them (Pinna et al., 2001; Devinck et al., 2005, 2006; Pinna and Grossberg, 2005; Pinna and Reeves, 2006; Cao et al., 2011; Coia and Crognale, 2014; Kimura and Kuroki, 2014a,b).

Comparison to Other Models

The only computational model in the literature, that is relevant to the watercolor effects, is the FACADE model (Pinna and Grossberg, 2005). In a more recent publication of Pinna and Grossberg (2005), the FACADE model was challenged by testing several stimulus parameters acting in the watercolor effect, such as the role of the contrast between the IC and the OC, the role of the background luminance, and different shape variations of the stimulus. While the FACADE model could predict the results of the stimuli on the assimilative watercolor effect it was not designed to, and indeed was unable to, predict the non-assimilative watercolor effect and its properties.

The FACADE model comprises two components. The first component, the BCS, detects the borders that block the diffusion process. The second component, the FCS, spreads the color to all directions until it is blocked by edges. The FACADE model is unable to predict the non-assimilative effect first because the spread of color is derived from the chromatic surface itself, and there is no mechanism that creates complementary colors. A second reason is that the border, which is detected by the BCS, prevents the OC color of the watercolor effect from spreading inside the inner area of the stimulus.

The ability of the FACADE model to predict only the assimilative effects (Pinna and Grossberg, 2005; Pinna, 2006; Cao et al., 2011; Kimura and Kuroki, 2014a,b) has contributed significantly to the general consensus in the literature that the assimilative and non-assimilative effects are derived from different mechanisms. In contrast, Kimura and Kuroki (2014a) found strong psychophysical evidence that assimilative and non-assimilative effects both share the same Weber contrast rule under specific psychophysical constraints. However, despite these Weber rules, they concluded that the effects might still involve different mechanisms.

Unlike FACADE, two factors allow our model to predict the non-assimilative watercolor effect. First, each edge in the stimulus triggers a diffusion process. Therefore, each edge contributes to the achromatic areas i.e., the inner area and the outer area. The color adjacent to the achromatic area contributes its color i.e., triggers a diffusion process of the same color, to this area; while the color in the other side of the edge contributes the complementary color to the same area. In other words, the color in the outer side of an edge triggers a diffusion process of its complementary color. The reason why the complementary color is obtained from the model is explained in the Model section. The exact colors that will be spread are calculated by the responses of the double-opponent RFs, Equations (8–10). The resultant colors, are therefore not necessarily exactly the "same" or "complementary" to the IC/OC, but rather a linear combination of the colors of the IC and the OC. In addition, the model assumes that the main role of the contours is to trigger the diffusion process as "heat sources," (Equation 10), and not as primarily designed to block the diffusion process.

It could be claimed that additional computational models that have been suggested for edge integration should be regarded here as competitors, which can explain this filling-in mechanism of chromatic and achromatic surfaces. Rudd (2014) summarized and discussed several computational models designed to perform the edge integration function in the visual system. He argued against the idea that the filling-in effect results from the activation of a low visual spatial frequency channel, due to the fact that the spatial extent of the filling-in effect is far larger than the area or distance spanned by the lowest spatial frequency filters in human vision (about 0.5 cycle/degree) (Wilson and Gelb, 1984). It should be noted that the watercolor effect has been shown to spread over 45° (Pinna et al., 2001), a spatial range that is not consistent with a low spatial frequency of the visual system.

Although Rudd (2014) also argued against the diffusive filling-in mechanism, we believe that his justification was based on the specific diffusive FACADE model suggested by Grossberg and his colleagues (Grossberg and Mingolla, 1987; Grossberg, 1997; Pinna and Grossberg, 2005). According to FACADE, the chromatic edges function as borders to block the diffusive process. If the watercolor stimulus is open (unclosed boundaries), the FACADE model predicts that the color would leak from the open ends, which, in reality, does not occur. In contrast, our diffusive computational model does not fail in such a case. **Figure 9** demonstrates that our model successfully predicts this effect, because the edges in our model are used as triggers, Equation (10), rather than borders for diffusion.

Rudd (2014) suggested a qualitative "Edge integration" model, through long range receptive fields in area V4 (Roe et al., 2012). Rudd suggested that lightness and darkness "edge integration" cells in V4 could integrate the responses of V1 simple receptive fields with a light or dark direction toward the center of the V4 receptive field. An additional neuron in the higher level of the visual pathway hierarchy then integrates these receptive fields, and performs a subtraction operation between the lightness and the darkness "edge integration" receptive fields. This model qualitatively predicts specific induction effects [Figures 2, 9 in Rudd (2014)] but fails to predict classic filling-in effects, such as the watercolor illusion that manifest filling-in in all directions and over very wide spatial regions.

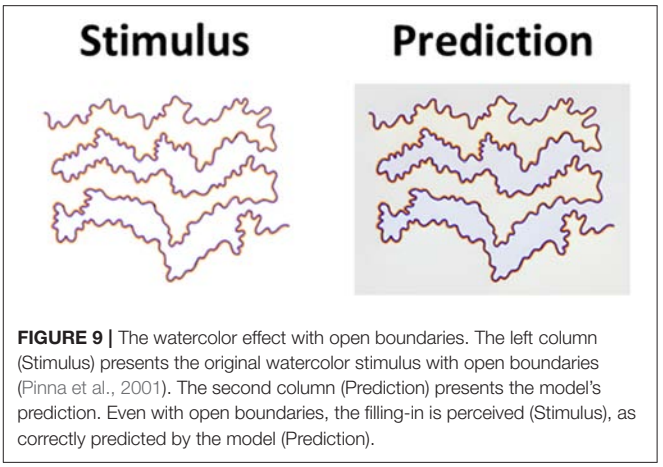


FIGURE 9 | The watercolor effect with open boundaries. The left column (Stimulus) presents the original watercolor stimulus with open boundaries (Pinna et al., 2001). The second column (Prediction) presents the model's prediction. Even with open boundaries, the filling-in is perceived (Stimulus), as correctly predicted by the model (Prediction).

Since Rudd (2014) related the induction effects to filling-in phenomena, he supplied an additional argument against the diffusive filling-in model, which is based on the model of Grossberg (Grossberg and Mingolla, 1987; Grossberg, 1997; Pinna and Grossberg, 2005). This argument is related to the FACADE model's failure to predict the specific induction effects, [Figure 2 in Rudd (2014)] and **Figure 9**.

There is currently a disagreement in the literature as to whether these specific induction effects are the result of a filling-in mechanism, an adaptation mechanism of the first order (Spitzer and Barkan, 2005), or a local or (remote) contrast mechanism (Blakeslee and McCourt, 1999, 2001, 2003, 2008). We argue that a visual effect may not necessarily be determined by a single dominant mechanism, and that several mechanisms could be involved. Different mechanisms could give rise to contradicting effects on one hand, or alternatively could work in synergy to enhance the perceived effect. An interesting question is whether this induction effect can also be predicted by our proposed model. **Figure 10** demonstrates that our filling-in model can predict the first order variation of the specific induction effect, [Figure 2 in (Rudd, 2014)]. Since this effect is predicted by our filling-in model, and also by an adaptation of the first order model (Spitzer and Barkan, 2005), we believe that the induction effect can be attributed to both mechanisms.

Experimental results show that the size of the inducer areas and the size of the induced area play a crucial role in the perceived induction effect (Shevell and Wei, 1998). The suggested filling-in model is based on edges that trigger a diffusion process, therefore the size of the induced area and the size of the inducer area do not play a role in our filling-in model. However, these two spatial factors do play a major role in the adaptation of the first order mechanism (Spitzer and Barkan, 2005).

We believe that there is a certain confusion in the literature regarding the source and the mechanisms of the induction and the filling-in effects. Kingdom (2011), for example, argued in his review that: "…'filling-in' of uniform regions is mediated by neural spreading has been seriously challenged by two sets

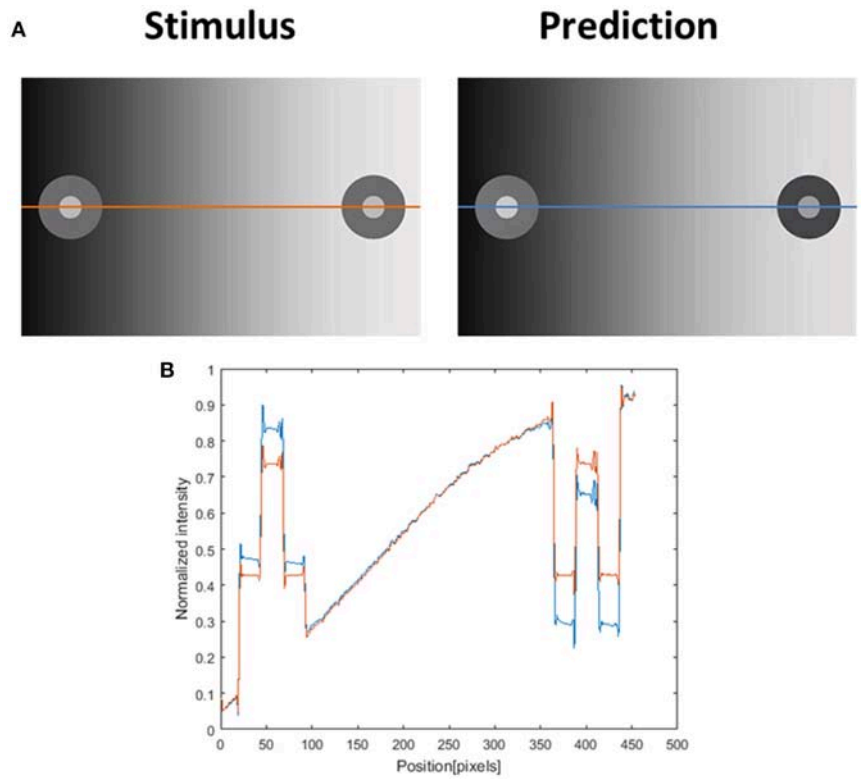


FIGURE 10 | Induction effect and the model's prediction. **(A)** The original induction stimulus from Rudd (2014) (Stimulus), and the model's prediction (Prediction). The second row **(B)** presents the luminance of the original image (orange line) and the predicted perceived luminance (at blue line) along the orange and blue axes in **(A)**. The predicted perceived luminance [demonstrated along the blue line in **(A)**] is higher than the original luminance [demonstrated along the orange line in **(A)**] in the left disk (including the inner circle and the outer ring of the disk), and shows a lower level of luminance than the original value in the right disk.

of findings: 1. That brightness induction is near-instantaneous and 2. That the Craik–Cornsweet–O'Brien illusion is dependent on the presence of residual low-frequency information and is not disrupted by the addition of luminance noise. 'Filling-in' should at best therefore be considered as a metaphor for the representation…". We argue that these claims are problematic, based on different psychophysical results (Pinna et al., 2001), and also query the feasibility of a mechanism, which is based on spatial filtering.

Kingdom (2011) assumed that these two effects of induction and other filling-in effects (the COC effect) derive from the same mechanism. For this reason, he argued against a diffusive filling-in mechanism, since a diffusive process requires more time. Kingdom (2011) also based his arguments on the findings reported by Blakeslee and McCourt (2008) that the temporal response of the induction effect (simultaneous contrast) lagged by <1 ms. In contrast, Pinna et al. (2001) found that the temporal response of the watercolor effect is about 100 ms. We believe that there is no contradiction between the two temporal results (Pinna et al., 2001; Blakeslee and McCourt, 2008), since they are associated with two different mechanisms, namely induction and the diffusive filling-in process. The first mechanism (induction of the first order) (Spitzer and Semo, 2002; Spitzer and Barkan, 2005; Tsofe et al., 2009; Kingdom, 2011) occurs in/at early visual areas, such as the retina, while the second mechanism (COC or watercolor, diffusive filling-in) occurs in a higher visual area. In addition, the spatial filling-in spread of 45°, reported for the watercolor illusion cannot be explained by any receptive field or low-spatial frequency channel of the visual system (Rudd, 2014).

In this context, we contend that positive and negative aftereffects (such as in "color dove illusion" and the "stars" illusion) (van Lier et al., 2009; Barkan and Spitzer, 2017), are perceived as a result of a diffusive filling-in process that cannot be explained by any spatial filtering mechanism. The reasons for this are: (1) The perceived color is obtained in an area that has not been stimulated by any color, at the time that the color is perceived [aftereffect with filling-in as in the "color dove illusion" and Van Lier "stars" (van Lier et al., 2009; Barkan and Spitzer, 2017)]. (2) The location of the achromatic reminder contour determines and triggers the perceived color. The filling-in model proposed here shares the same diffusion component, Equation (10), as suggested for the positive and the negative aftereffects (Cohen-Duwek and Spitzer, 2018). Although Kingdom (2011) supported the description of the filling-in and induction events by the filter models of Blakeslee and McCourt (2008), their model cannot predict the assimilative and the non-assimilative watercolor effects, or the aftereffects.

Predictions for Watercolor Properties

Having discussed the options of various alternative models for the "filling-in" phenomena, we were interested to test our model's predictions with studies that define general properties and rules for the watercolor effect, although without a computational model (Kimura and Kuroki, 2014a). We have already described the success of our model in correctly predicting experimental results (Kimura and Kuroki, 2014a) demonstrating crucial properties regarding the strength of the watercolor effect and its relation to the assimilative and non-assimilative effects. We explain below how the basic structure of the suggested model can explain these findings, without requiring any additional components.

Complementary colors: Several studies have demonstrated that a maximal filling-in response is perceived when the IC and the OC have complementary colors (Pinna et al., 2001; Devinck et al., 2006) and it should be noted that the model correctly predicts this trend, **Figure 6**. This can be explained by the model equations (Equations 3–10), through solving the Poisson equation. The IC triggers an assimilative filling-in (of the same color as the IC) toward the inner area, while the OC triggers a non-assimilative filling-in, with the opposite color to the IC contour (**Figure 2**, i.e., its complementary color), toward the inner area. According to the model, if the color of the OC is complementary to the color of the IC, the combination of colors that diffuse to the inner area will be the same as the color of the IC (assimilative color) and complementary to the color of the OC, which makes it the same color as the IC again. Consequently, the perceived color is enhanced.

Luminance contrast: Several studies have reported that the magnitude of the filling-in effect increases with increasing luminance contrast between the IC and OC contours (Pinna et al., 2001; Devinck et al., 2005; Devinck and Knoblauch, 2012). This property of the luminance contrast is treated similarly to the chromatic channels. The weights of the modified gradients calculation, Equations (7–8), gives greater dominancy to the gradients between the IC and the OC. It is therefore not surprising that the model correctly predicts the importance of the luminance contrast, between the IC and the OC, in the watercolor effect.

Saturation: Devinck et al. (2006) showed that increasing the saturation of the outer and inner contours increases the shift in chromaticity of the filling-in effect. This information is included in the model through the chromatic opponent channel, Equation (3). Higher color saturation is expressed as a higher response in the chromatic opponent channels. This property has been tested and the model predictions show good agreement with the results of experimental studies.

Weber rule – IC contrast/OC contrast: Kimura and Kuroki (2014a) reported that the ratio between the IC luminance contrast and the OC luminance contrast determines the perceived filling-in effect, **Figure 8**. The IC contrast is the Weber contrast of the chromatic IC luminance and the background luminance, while the OC contrast is the Weber contrast of the chromatic OC luminance and the background luminance, Equation (21). Note that since the background is achromatic, this Weber contrast is related only to the luminance domain. Kimura and Kuroki (2014a) argued that if the IC contrast is smaller than the OC contrast, an assimilative effect is perceived, Equation (21). In contrast, if the IC contrast is larger than the OC contrast, a non-assimilative effect is perceived, Equation (21).

$$\frac{|L_{IC}-L_{Bkg}|}{L_{Bkg}} < \frac{|L_{OC}-L_{Bkg}|}{L_{Bkg}} \rightarrow assimilative\ effect \qquad (21)$$
$$\frac{|L_{IC}-L_{Bkg}|}{L_{Bkg}} > \frac{|L_{OC}-L_{Bkg}|}{L_{Bkg}} \rightarrow non-assimilative\ effect$$

Where L_{IC}, L_{OC}, and L_{Bkg} are the luminances of the IC, OC, and the background, respectively.

Our model was tested with a variety of stimuli with different luminance backgrounds, different chromatic contours (**Figures 8A,B**), and different Weber ratios. **Figure 8** demonstrates the predictions of the Weber contrast rule with non-assimilative effect. Additional stimuli were tested, but showed a smaller perceived effect. Interestingly, the Weber contrast rule and the predictions of our model do not necessarily always yield the exact assimilative or non-assimilative colors, but rather a different color as found experimentally (Kimura and Kuroki, 2014a). For example, the stimuli in **Figures 8A,B** have the same colors (red and magenta), but because the IC in **Figure 8A** has a higher luminance than the IC in **Figure 8B**, this gives rise to a yellowish color in **Figure 8A** but a greenish color in **Figure 8B**. Note that despite the difference in luminance levels, both effects share the same trend of Weber contrast rule, and thus both appear as non-assimilative effects. The model's predictions are in agreement with the Weber contrast rules (Kimura and Kuroki, 2014a), **Figure 8**. This demonstrates that both the model and the Weber contrast rule can predict in which contrast configuration the perceived effect is assimilative or non-assimilative.

Let us explain how our model can predict this Weber contrast rule. If an IC has a high value of Weber contrast, the "heat source" located on the edge between the IC and the background has the highest value and the diffusion process from this edge has a strong influence on the perceived color. Accordingly, the color spreading from this "heat source" (the edge between the IC and the background) to the inner area has the same color as the color of the background (white **Figure 8A**), and the complementary color of the IC (cyan—the complementary color of the red IC), **Figure 8A**. The cyan color, which is a combination of green and blue, contributes to this bluish-greenish perceived effect (**Figure 8B**).

We were interested in whether the Weber contrast rule is applicable to the achromatic watercolor stimuli. Cao et al. (2011) conducted a psychophysical study in order to investigate the influence of the luminances of the IC, OC, and the background on the perceived achromatic watercolor effect. They found that the filling-in effect disappeared when the luminance of the OC was between the luminances of the IC and the background. Kimura and Kuroki (2014a) reported that the findings of Cao et al. (2011) are consistent with their psychophysical findings, and also with their suggestion for the role of the Weber contrast rule. The prediction of our model (**Figure 11**) is also in agreement

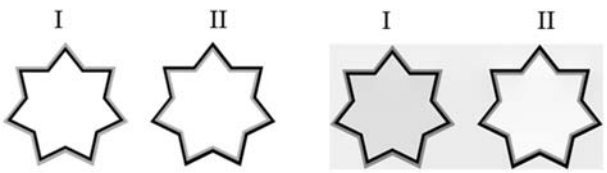


FIGURE 11 | The model's predictions for chromatic watercolor stimuli. In the left column (Stimulus I), the luminance of the OC (gray) lies between that of the IC (black) and the background (white). On the right side, (Stimulus II) the luminance of the IC lies between that of the OC and the background. The predicted filling-in luminance inside the left stimulus (Prediction I) is lower than the predicted filling-in luminance inside the right stimulus (Prediction II). Note that the Weber contrast of the OC is smaller than the Weber contrast of the IC in the left stimulus (I), and larger than the Weber contrast of the IC in the right stimulus (II). The topic is discussed in the Discussion section of the text.

with the Kimura and Kuroki (2014a) findings. In **Figure 11**, the luminance of the OC lies between that of the IC and the background. In terms of the Weber contrast rule, the Weber contrast of the OC is smaller than that of the IC. Therefore, such a configuration should lead to a non-assimilative perceived effect. However, since the perceived color inside the star is darker than the background (**Figure 8**I); this might be seen as a diffusive effect of the IC color ("assimilative" effect), which is black. According to our model, the perceived color is a combination of the same color as the IC (black) and the complementary color of the OC (gray, which is the complementary of gray), therefore, the model correctly predicts this effect. Accordingly, the terms "assimilative" and "non-assimilative" watercolor effects are not the precise terms regarding the perceived colors of the achromatic watercolor stimuli. It should be noted that there might be a dependency of the perceived effect on the stimulus size. This property should be further investigated experimentally.

Not all experimental studies agree about the perceived color in the non-assimilative watercolor effect (Pinna, 2006; Kimura and Kuroki, 2014b). Kimura and Kuroki (2014b), for example, claim that if the luminance of the IC is low (very dark IC), the perceived filling-in effect is predominantly yellow, regardless of the OC color. Kimura reported this finding to be inconsistent with previous results reported by Pinna (2006), which described a complementary color filling-in effect with black IC and chromatic OC combinations. Additional experimental study supports the results of Pinna (2006) and the idea that complementary colors are perceived, when the IC color is dark (Hazenberg and van Lier, 2013). The model results predict that the perceived colors are predominantly complementary to the OC colors, when the IC is dark. Even though the predicted results, **Figure 6**, are predominantly complementary to the OC colors, when the IC color is dark red, the predicted colors are slightly shifted to the red IC color. When the IC is achromatic the predicted colors, **Figure 6**, are the complementary colors to the OC colors.

Our model, thus, supports the findings of Pinna (2006) and Hazenberg and van Lier (2013), **Figure 6**, and is not in agreement with Kimura and Kuroki (2014b) because the chromatic OC triggers a filling-in effect that is complementary to the inner area, and therefore the perceived color will be complementary to the OC (the IC is achromatic and so does not contribute any color to the effect).

Model's Predictions for the COC Effect

Although our model is mainly concerned with the predictions of the watercolor illusions, there are a number of other examples of filling-in effects, including the COC effect. We believe that the COC effect is driven solely by a diffusion mechanism, since the physical stimulus in this effect is only an edge. The model prediction for the COC effect, which is demonstrated in **Figure 12**, uses the same set of parameters as the watercolor illusions (**Figures 3**, **5**, **7**–**9**, **11**). Our suggestion that both phenomena (watercolor and COC) are related to the same visual mechanism, is in agreement with (Devinck et al., 2005; Todorovi, 2006; Cao et al., 2011) who showed that the watercolor stimulus profile is a discrete version of the COC stimulus profile. The success of the model prediction of the COC effect supports the suggestion that both effects (which are physically built only from edges) share the same "heat sources" diffusion mechanism, which is triggered by edges. The COC effect can actually be considered as a simpler case of the diffusive filling-in effect than the watercolor effects.

There are three main classes of computational models that have been used to investigate the COC effect. The first class is called the "Diffusive models" (Grossberg and Mingolla, 1987). Grossberg and Mingolla (1987) showed that the FACADE model can correctly predict the COC effect. Nevertheless, the FACADE model, in this case, can predict the COC effect when the stimulus contains open boundaries, but only through using an additional component that detects illusory contours, **Figure 12**. The illusory contours component will detect the illusory edges around the COC stimulus (**Figure 12**), and will prevent the color from spreading. However, this component is not necessary for the watercolor illusion, which can contain open boundaries. **Figure 9** presents, for example, open boundaries, and it can be seen that there is no perceived effect of edge completion (illusory contour). It has to be noted that the suggested model does not include a component that detects illusory contours, and therefore our model does not predict filling-in effects that involve illusory contours e.g., "Neon Color Spreading." Our model suggests that the illusory contours components are not necessary for the watercolor mechanism.

The second class of models is termed the "Spatial filtering models," where these models utilize low-frequencies spatial filters in order to predict the filling-in effects (Morrone et al., 1986; Burr, 1987; Morrone and Burr, 1988; Ross et al., 1989; Blakeslee and McCourt, 1999, 2001, 2003, 2004, 2005; Dakin and Bex, 2003; Blakeslee et al., 2005; Kingdom, 2011). We argue that the spatial filtering approach has limitations in predicting the COC effect because the filling-in can be spread to sizes which cannot be explained by the sizes of the receptive fields that exist in the LGN or V1–V2 cortical areas. In addition, the COC effect can be obtained from edges that are built only from ODOG (Oriented Difference of Gaussian) filters (Blakeslee and McCourt, 1999).

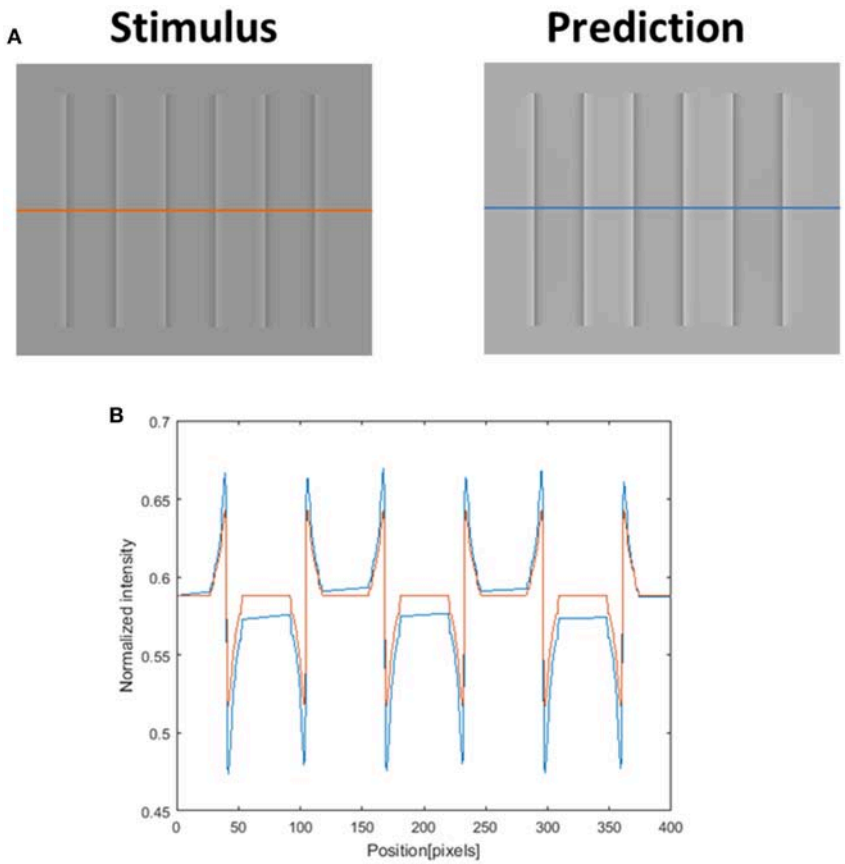


FIGURE 12 | The model's ability to predict the Craik–O'Brien–Cornsweet (COC) effect. The left image **(A)** represents the original COC effect (Stimulus, left), and the model's prediction (Prediction, right). The lower row **(B)** presents the luminance of the original stimulus (orange line) and predicted perceived luminance (blue line) along the orange and blue axes in **(A)**.

The third class of models is termed the "Empirical models." These models are designed to estimate the most likely reflectance values based on the pattern of the luminances observed in the image, together with learnt image statistics (Purves and Lotto, 2003; Brown and Friston, 2012). Typically, such an Empirical approach may explain why we perceive these visual effects, but cannot explain the neuronal mechanisms that lead to the perceived effects.

Neuronal Sources of the Filling-In Effect

Studies designed to identify the neuronal source of the filling-in effects that are triggered by edges, e.g., the watercolor and the COC effects, can shed additional light on the possible neuronal mechanisms. A recent fMRI study (Hong and Tong, 2017) compared the responses of the visual areas (V1–V4) to real colored surfaces and to illusory filled-in surfaces, such as occur in the afterimage effect of van Lier "stars"(van Lier et al., 2009). Hong and Tong (2017) found a high correlation between the two types of stimuli, the real and the illusory, only in areas V3 and V4. They, therefore concluded that the perception of filled-in surface color occurs in the higher areas of the visual cortex.

Rudd (2014) suggested an "edge integration" model that works through long range receptive fields in area V4 (Roe et al., 2012). Both the qualitative (Rudd, 2014) model and (Hong and Tong, 2017) experiments support the idea that the source of the filling-in mechanism is located in V4. It has to be noted that our computational model can be regarded as this diffusion process but also does not contradict a mechanism of edge integration that can be derived from long range receptive fields (Rudd, 2014). This "edge integration" mechanism can also be symbolic and appear as a diffusion process.

As already discussed, we argue that both the watercolor effect and the COC effect share the same visual mechanism; therefore, we would expect to identify a similar neuronal source for both effects. A literature survey of experimental studies that investigated these sources revealed a lack of consensus regarding the neuronal source of the COC effect. A few studies reported that the effect occurs in low visual areas: the LGN, V1 and V2 (MacEvoy and Paradiso, 2001; Roe et al., 2005; Cornelissen et al., 2006; Huang and Paradiso, 2008), while other studies showed evidence that the effect occurs in higher areas of the visual system such as the V3 and caudal intraparietal sulcus (Perna et al., 2005). It is possible that there is no

complete overlap between the cortical areas responsible for the COC effect and the watercolor effect, since the watercolor effect commonly involves color, while the COC effect involves achromatic stimuli.

Our model succeeds in predicting apparently conflicting perceived filling-in triggered-by-edges phenomena, e.g., the assimilative and the non-assimilative watercolor effects. The suggested mechanism is a filling-in process which is based on reconstruction of an image from its modified edges. The diffusion process, thus, is calculated by solving the heat equation with heat sources (Poisson equation). The edge of the trigger stimulus are modified by the model according to rules of dominancy, and computed as the heat sources in the Poisson equation. We therefore suggest that this model can predict all the filling-in-triggered-by-edges effect in both the spatial and temporal domains (Cohen-Duwek and Spitzer, 2018).

The challenge of "The interaction of the mechanisms underlying boundary and surface perception is an essential problem for vision scientists" has been presented previously (Cao et al., 2011). Here we introduce a new computational model that describes and predicts how any boundary can "create" surfaces by a filling-in process.

DATA AVAILABILITY

All datasets generated for this study are included in the manuscript and/or the supplementary files.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

REFERENCES

Adelson, E. H., Anderson, C. H., Bergen, J. R., Burt, P. J., and Ogden, J. M. (1984). Pyramid methods in image processing. *RCA Eng.* 29, 33–41.

Barkan, Y., and Spitzer, H. (2017). "The color dove illusion- chromatic filling in effect following a spatial-temporal edge," in *The Oxford Compendium of Visual Illusions*, eds A. G. Shapiro, and D. Todorovic (Oxford, NY: Oxford University Press), 752–755.

Berns, R. S. (2000). *Principles of Color Technology*. New York, NY: Wiley.

Bertalmio, M., Sapiro, G., Caselles, V., and Ballester, C. (2000). "Image inpainting," in *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques*, (New Orleans, LA: ACM Press/Addison-Wesley Publishing Co.), 417–424. doi: 10.1145/344779.344972

Blakeslee, B., and McCourt, M. E. (1999). A multiscale spatial filtering account of the white effect, simultaneous brightness contrast and grating induction. *Vision Res.* 39, 4361–4377. doi: 10.1016/S0042-6989(99)00119-4

Blakeslee, B., and McCourt, M. E. (2001). A multiscale spatial filtering account of the Wertheimer–Benary effect and the corrugated Mondrian. *Vision Res.* 41, 2487–2502. doi: 10.1016/S0042-6989(01)00138-9

Blakeslee, B., and McCourt, M. E. (2003). "A multiscale spatial filtering account of brightness phenomena," in *Levels of Perception*, eds L. Harris and M. Jenkin (New York, NY: Springer), 47–72. doi: 10.1007/0-387-22673-7_4

Blakeslee, B., and McCourt, M. E. (2004). A unified theory of brightness contrast and assimilation incorporating oriented multiscale spatial filtering and contrast normalization. *Vision Res.* 44, 2483–2503. doi: 10.1016/j.visres.2004.05.015

Blakeslee, B., and McCourt, M. E. (2005). A multiscale filtering explanation of gradient induction and remote brightness induction effects: a reply to Logvinenko (2003). *Perception* 34, 793–802. doi: 10.1068/p5303x

Blakeslee, B., and McCourt, M. E. (2008). Nearly instantaneous brightness induction. *J. Vis.* 8, 15–15. doi: 10.1167/8.2.15

Blakeslee, B., Pasieka, W., and McCourt, M. E. (2005). Oriented multiscale spatial filtering and contrast normalization: a parsimonious model of brightness induction in a continuum of stimuli including White, Howe and simultaneous brightness contrast. *Vision Res.* 45, 607–615. doi: 10.1016/j.visres.2004.09.027

Brown, H., and Friston, K. J. (2012). Free-energy and Illusions: the cornsweet effect. *Front. Psychol.* 3:43. doi: 10.3389/fpsyg.2012.00043

Burr, D. C. (1987). Implications of the craik-O'Brien illusion for brightness perception. *Vision Res.* 27, 1903–1913.

Cao, B., Yazdanbakhsh, A., and Mingolla, E. (2011). The effect of contrast intensity and polarity in the achromatic watercolor effect. *J. Vis.* 11, 18–18. doi: 10.1167/11.3.18

Cohen, M. A., and Grossberg, S. (1984). Neural dynamics of brightness perception: Features, boundaries, diffusion, and resonance. *Percept. Psychophys.* 36, 428–456. doi: 10.3758/BF03207497

Cohen-Duwek, H., and Spitzer, H. (2018). A Model for a filling-in process triggered by edges predicts "Conflicting" afterimage effects. *Front. Neurosci.* 12:559. doi: 10.3389/fnins.2018.00559

Coia, A. J., and Crognale, M. A. (2014). Asymmetric effects of luminance and chrominance in the watercolor illusion. *Front. Hum. Neurosci.* 8:723. doi: 10.3389/fnhum.2014.00723

Coia, A. J., Jones, C., Duncan, C. S., and Crognale, M. A. (2014). Physiological correlates of watercolor effect. *J. Opt. Soc. Am. A* 31:A15. doi: 10.1364/JOSAA.31.000A15

Conway, B. R. (2001). Spatial structure of cone inputs to color cells in alert macaque primary visual cortex (V-1). *J. Neurosci.* 21, 2768–2783. doi: 10.1523/JNEUROSCI.21-08-02768.2001

Conway, B. R., Chatterjee, S., Field, G. D., Horwitz, G. D., Johnson, E. N., Koida, K., et al. (2010). Advances in color science: from retina to behavior. *J. Neurosci.* 30, 14955–14963. doi: 10.1523/JNEUROSCI.4348-10.2010

Conway, B. R., and Livingstone, M. S. (2006). Spatial and temporal properties of cone signals in alert macaque primary visual cortex. *J. Neurosci.* 26, 10826–10846. doi: 10.1523/JNEUROSCI.2091-06.2006

Cornelissen, F. W., Wade, A. R., Vladusich, T., Dougherty, R. F., and Wandell, B. A. (2006). No functional magnetic resonance imaging evidence for brightness and color filling-in in early human visual cortex. *J. Neurosci.* 26, 3634–3641. doi: 10.1523/JNEUROSCI.4382-05.2006

Cornsweet, T. (1970). *Visual Perception*. New York, NY: Academic Press.

Dakin, S. C., and Bex, P. J. (2003). Natural image statistics mediate brightness "filling in." *Proc. R. Soc. Lond. B Biol. Sci.* 270, 2341–2348. doi: 10.1098%2Frspb.2003.2528

Devinck, F., Delahunt, P. B., Hardy, J. L., Spillmann, L., and Werner, J. S. (2005). The watercolor effect: quantitative evidence for luminance-dependent mechanisms of long-range color assimilation. *Vision Res.* 45, 1413–1424. doi: 10.1016/j.visres.2004.11.024

Devinck, F., Gerardin, P., Dojat, M., and Knoblauch, K. (2014). Quantifying the watercolor effect: from stimulus properties to neural models. *Front. Hum. Neurosci.* 8:805. doi: 10.3389/fnhum.2014.00805

Devinck, F., Hardy, J. L., Delahunt, P. B., Spillmann, L., and Werner, J. S. (2006). Illusory spreading of watercolor. *J. Vis.* 6, 7–7. doi: 10.1167/6.5.7

Devinck, F., and Knoblauch, K. (2012). A common signal detection model accounts for both perception and discrimination of the watercolor effect. *J. Vis.* 12, 19–19. doi: 10.1167/12.3.19

Devinck, F., and Spillmann, L. (2009). The watercolor effect: spacing constraints. *Vision Res.* 49, 2911–2917. doi: 10.1016/j.visres.2009.09.008

Gonzalez, R. C., and Woods, R. E. (2002). *Digital Image Processing*. Saddle River, NJ: Prentice hall Upper.

Grossberg, S. (1997). Cortical dynamics of three-dimensional figure–ground perception of two-dimensional pictures. *Psychol. Rev.* 104, 618–658. doi: 10.1037/0033-295X.104.3.618

Grossberg, S., Dasara, M., and Pinna, B. (2005). The problem of the perception of holes and figure-ground segregation in the watercolor illusion. *J. Vis*. 5, 54–54. doi: 10.1167/5.8.54

Grossberg, S., and Mingolla, E. (1985). Neural dynamics of form perception: Boundary completion, illusory figures, and neon color spreading. *Psychol. Rev*. 92, 173–211. doi: 10.1037/0033-295X.92.2.173

Grossberg, S., and Mingolla, E. (1987). "Neural dynamics of perceptual grouping: Textures, boundaries, and emergent segmentations," in *The Adaptive Brain II*, ed S. Grossberg (North Holland: Elsevier), 143–210. doi: 10.1016/B978-0-444-70414-6.50007-8

Hazenberg, S. J., and van Lier, R. (2013). Afterimage watercolors: an exploration of contour-based afterimage filling-in. *Front. Psychol*. 4:707. doi: 10.3389/fpsyg.2013.00707

Hong, S. W., and Tong, F. (2017). Neural representation of form-contingent color filling-in in the early visual cortex. *J. Vis*. 17, 10–10. doi: 10.1167/17.13.10

Horwitz, G. D., Chichilnisky, E. J., and Albright, T. D. (2007). Cone inputs to simple and complex cells in V1 of awake macaque. *J. Neurophysiol*. 97, 3070–3081. doi: 10.1152/jn.00965.2006

Huang, X., and Paradiso, M. A. (2008). V1 response timing and surface filling-in. *J. Neurophysiol*. 100, 539–547. doi: 10.1152/jn.00997.2007

Johnson, E. N., Hawken, M. J., and Shapley, R. (2001). The spatial transformation of color in the primary visual cortex of the macaque monkey. *Nat. Neurosci*. 4, 409–416. doi: 10.1038/86061

Johnson, E. N., Hawken, M. J., and Shapley, R. (2008). The orientation selectivity of color-responsive neurons in macaque V1. *J. Neurosci*. 28, 8096–8106. doi: 10.1523/JNEUROSCI.1404-08.2008

Kandel, E. R., Schwartz, J. H., Jessell, T. M., Siegelbaum, S. A., and Hudspeth, A. J. (2012). *Principles of Neural Science, 5th Edn.* New York, NY: McGraw-Hill Education / Medical.

Kanizsa, G. (1976). Subjective contours. *Sci. Am*. 234, 48–52. doi: 10.1038/scientificamerican0476-48

Kimura, E., and Kuroki, M. (2014a). Assimilative and non-assimilative color spreading in the watercolor configuration. *Front. Hum. Neurosci*. 8:722. doi: 10.3389/fnhum.2014.00722

Kimura, E., and Kuroki, M. (2014b). Contribution of a luminance-dependent S-cone mechanism to non-assimilative color spreading in the watercolor configuration. *Front. Hum. Neurosci*. 8:980. doi: 10.3389/fnhum.2014.00980

Kingdom, F. A. (2011). Lightness, brightness and transparency: a quarter century of new ideas, captivating demonstrations and unrelenting controversy. *Vision Res*. 51, 652–673. doi: 10.1016/j.visres.2010.09.012

Kitaoka, A. (2007). *Visual Completion*. Available online at: http://www.psy.ritsumei.ac.jp/~akitaoka/hokan-e.html

MacEvoy, S. P., and Paradiso, M. A. (2001). Lightness constancy in primary visual cortex. *Proc. Natl. Acad. Sci. U.S.A*. 98, 8827–8831. doi: 10.1073/pnas.161280398

Morrone, M. C., and Burr, D. C. (1988). Feature detection in human vision: A phase-dependent energy model. *Proc. R Soc. Lond. B* 235, 221–245. doi: 10.1098/rspb.1988.0073

Morrone, M. C., Ross, J., Burr, D. C., and Owens, R. (1986). Mach bands are phase dependent. *Nature* 324, 250–253. doi: 10.1038/324250a0

Movshon, J. A., Thompson, I. D., and Tolhurst, D. J. (1978). Spatial summation in the receptive fields of simple cells in the cat's striate cortex. *J. Physiol*. 283, 53–77.

Nicholls, J. G., Martin, A. R., Wallace, B. G., and Fuchs, P. A. (2001). *From Neuron to Brain*, Sunderland, MA: Sinauer Associates.

Pérez, P., Gangnet, M., and Blake, A. (2003). "Poisson image editing," in *ACM SIGGRAPH 2003 Papers* (New York, NY: ACM), 313–318. doi: 10.1145/1201775.882269

Perna, A., Tosetti, M., Montanaro, D., and Morrone, M. C. (2005). Neuronal mechanisms for illusory brightness perception in humans. *Neuron* 47, 645–651. doi: 10.1016/j.neuron.2005.07.012

Pinna, B. (2006). "The neon color spreading and the watercolor illusion: phenomenal links and neural mechanisms," in *Systemics of Emergence: Research and Development*, eds G. MinatiEliano and P. Abram (Boston, MA: Springer), 235–254. doi: 10.1007/0-387-28898-8_17

Pinna, B. (2008). Watercolor illusion. *Scholarpedia* 3:5352. doi: 10.4249/scholarpedia.5352

Pinna, B., Brelstaff, G., and Spillmann, L. (2001). Surface color from boundaries: a new "watercolor" illusion. *Vision Res*. 41, 2669–2676. doi: 10.1016/S0042-6989(01)00105-5

Pinna, B., and Grossberg, S. (2005). The watercolor illusion and neon color spreading: a unified analysis of new cases and neural mechanisms. *JOSA A* 22, 2207–2221. doi: 10.1364/JOSAA.22.002207

Pinna, B., and Reeves, A. (2006). Lighting, backlighting and watercolor illusions and the laws of figurality. *Spat. Vis*. 19, 341–373. doi: 10.1163/156856806776923434

Purves, D., and Lotto, R. B. (2003). *Why We See What We do: An Empirical Theory of Vision*. Sunderland, MA: Sinauer Associates.

Roe, A. W., Chelazzi, L., Connor, C. E., Conway, B. R., Fujita, I., Gallant, J. L., et al. (2012). Toward a unified theory of visual area V4. *Neuron* 74, 12–29. doi: 10.1016/j.neuron.2012.03.011

Roe, A. W., Lu, H. D., and Hung, C. P. (2005). Cortical processing of a brightness illusion. *Proc. Natl. Acad. Sci. U.S.A*. 102, 3869–3874. doi: 10.1073/pnas.05000971020

Ross, J., Morrone, M. C., and Burr, D. C. (1989). The conditions under which Mach bands are visible. *Vision Res*. 29, 699–715. doi: 10.1016/0042-6989(89)90033-3

Rudd, M. E. (2014). A cortical edge-integration model of object-based lightness computation that explains effects of spatial context and individual differences. *Front. Hum. Neurosci*. 8:640. doi: 10.3389/fnhum.2014.00640

Shapley, R., and Enroth-Cugell, C. (1984). Visual adaptation and retinal gain controls. *Prog. Retin. Res*. 3, 263–346. doi: 10.1016/0278-4327(84)90011-7

Shapley, R., and Hawken, M. (2011). Color in the Cortex—single- and double-opponent cells. *Vision Res*. 51, 701–717. doi: 10.1016/j.visres.2011.02.012

Shevell, S. K., and Wei, J. (1998). Chromatic induction: border contrast or adaptation to surrounding light? *Vision Res*. 38, 1561–1566. doi: 10.1016/S0042-6989(98)00006-6

Simchony, T., Chellappa, R., and Shao, M. (1990). Direct analytical methods for solving Poisson equations in computer vision problems. *IEEE Trans. Pattern Anal. Mach. Intell*. 12, 435–446. doi: 10.1109/34.55103

Solomon, S. G., and Lennie, P. (2007). The machinery of colour vision. *Nat. Rev. Neurosci*. 8, 276–286. doi: 10.1038/nrn2094

Spitzer, H., and Barkan, Y. (2005). Computational adaptation model and its predictions for color induction of first and second orders. *Vision Res*. 45, 3323–3342. doi: 10.1016/j.visres.2005.08.002

Spitzer, H., and Hochstein, S. (1985). A complex-cell receptive-field model. *J. Neurophysiol*. 53, 1266–1286. doi: 10.1152/jn.1985.53.5.1266

Spitzer, H., and Semo, S. (2002). Color constancy: a biological model and its application for still and video images. *Pattern Recognit*. 35, 1645–1659. doi: 10.1016/S0031-3203(01)00160-1

Tanca, M., Grossberg, S., and Pinna, B. (2010). Probing perceptual antinomies with the watercolor illusion and explaining how the brain resolves them. *Seeing Perce*. 23, 295–333. doi: 10.1163/187847510X532685

Thorell, L. G., de Valois, R. L., and Albrecht, D. G. (1984). Spatial mapping of monkey VI cells with pure color and luminance stimuli. *Vision Res*. 24, 751–769. doi: 10.1016/0042-6989(84)90216-5

Todorovi,ć, D. (2006). Lightness, illumination, and gradients. *Spat. Vis*. 19, 219–261. doi: 10.1163/156856806776923407

Tsofe, A., Spitzer, H., and Einav, S. (2009). Does the chromatic mach bands effect exist? *J. Vis*. 9, 20–20. doi: 10.1167/9.6.20

van de Sande, K., Gevers, T., and Snoek, C. G. (2010). Evaluating color descriptors for object and scene recognition. *IEEE Trans. Pattern Anal. Mach. Intell*. 32, 1582–1596. doi: 10.1109/TPAMI.2009.154

van Lier, R., Vergeer, M., and Anstis, S. (2009). Filling-in afterimage colors between the lines. *Curr. Biol*. 19, R323–R324. doi: 10.1016/j.cub.2009.03.010

van Tuijl, H. F., and Leeuwenberg, E. L. (1979). Neon color spreading and structural information measures. *Percept. Psychophys*. 25, 269–284. doi: 10.3758/BF03198806

von der Heydt, R., Friedman, H. S., and Hong, Z. (2003). "Searching for the neural mechanisms of color filling-in," in *Filling-In : From Perceptual Completion to Cortical Reorganization: From Perceptual Completion to Cortical Reorganization*, eds L. Pessoa and P. De Weerd (New York, NY: Oxford University Press), 106–127. doi: 10.1093/acprof:oso/9780195140132.003.0006

Wandell, B. A. (1995). *Foundations of Vision*. Sunderland, MA: Sinauer Associates.

Weickert, J. (1998). *Anisotropic Diffusion in Image Processing*. Stuttgart: Teubner.

Wilson, H. R., and Gelb, D. J. (1984). Modified line-element theory for spatial-frequency and width discrimination. *JOSA A* 1, 124–131. doi: 10.1364/JOSAA.1.000124

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 30 April 2019
doi: 10.3389/fnins.2019.00407

A Cross-Recurrence Analysis of the Pupil Size Fluctuations in Steady Scotopic Conditions

*Pietro Piu[1], Valeria Serchi[1], Francesca Rosini[1,2] and Alessandra Rufa[1,2]**

[1] Eye Tracking and Visual Application Lab, Department of Medicine, Surgery and Neuroscience, University of Siena, Siena, Italy, [2] Neurology and Neurometabolic Unit, Department of Medicine, Surgery and Neuroscience, University of Siena, Siena, Italy

Pupil size fluctuations during stationary scotopic conditions may convey information about the cortical state activity at rest. An important link between neuronal network state modulation and pupil fluctuations is the cholinergic and noradrenergic neuromodulatory tone, which is active at cortical level and in the peripheral terminals of the autonomic nervous system (ANS). This work aimed at studying the low- and high-frequency coupled oscillators in the autonomic spectrum (0–0.45 Hz) which, reportedly, drive the spontaneous pupillary fluctuations. To assess the interaction between the oscillators, we focused on the patterns of their trajectories in the phase-space. Firstly, the frequency spectrum of the pupil signal was determined by empirical mode decomposition. Secondly, cross-recurrence quantification analysis was used to unfold the non-linear dynamics. The global and local patterns of recurrence of the trajectories were estimated by two parameters: determinism and entropy. An elliptic region in the entropy-determinism plane (95% prediction area) yielded health-related values of entropy and determinism. We hypothesize that the data points inside the ellipse would likely represent balanced activity in the ANS. Interestingly, the Epworth Sleepiness Scale scores scaled up along with the entropy and determinism parameters. Although other non-linear methods like Short Time Fourier Transform and wavelets are usually applied for analyzing the pupillary oscillations, they rely on strong assumptions like the stationarity of the signal or the *a priori* knowledge of the shape of the single basis wave. Instead, the cross-recurrence analysis of the non-linear dynamics of the pupil size oscillations is an adaptable diagnostic tool for identifying the different weight of the autonomic nervous system components in the modulation of pupil size changes at rest in non-luminance conditions.

Keywords: pupil diameter, cross-recurrence quantification analysis, empirical mode decomposition, Epworth Sleepiness Scale, Gaussian-copula

OPEN ACCESS

Edited by:
Xavier Otazu, Autonomous University of Barcelona, Spain

Reviewed by:
Pablo De Gracia, Midwestern University, United States
Miriam Schwalm, Johannes Gutenberg University Mainz, Germany

***Correspondence:**
Alessandra Rufa
rufa@unisi.it

Specialty section:
This article was submitted to Perception Science, a section of the journal Frontiers in Neuroscience

Received: *22 September 2018*
Accepted: *10 April 2019*
Published: *30 April 2019*

Citation:
Piu P, Serchi V, Rosini F and Rufa A (2019) A Cross-Recurrence Analysis of the Pupil Size Fluctuations in Steady Scotopic Conditions. Front. Neurosci. 13:407. doi: 10.3389/fnins.2019.00407

Abbreviations: ANS, autonomic nervous system; CRQA, cross-recurrence quantification analysis; DET, percentage of determinism calculated from the cross-recurrence analysis; EmbDim, embedding dimension, a hyper-parameter to be estimated for the CRQA; EMD, empirical mode decomposition; ENT, entropy calculated from the cross-recurrence analysis; ESS, Epworth Sleepiness Scale; FAN, fixed amount of nearest neighbor; HF, high-frequency component in the ANS spectrum; IMFs, intrinsic mode functions extracted through the EMD; LF, low-frequency component in the ANS spectrum; MUSIC, multiple signal classification algorithm; R, neighborhood radius, a hyper-parameter to be estimated for the CRQA; RQA, recurrence quantification analysis; TD, time delay, a hyper-parameter to be estimated for the CRQA.

INTRODUCTION

The pupil controls the amount of light radiations reaching the retina, by modulating its diameter through the interaction of two muscles under sympathetic-parasympathetic control. The pupil constriction is regulated by the contraction of the iris sphincter muscle receiving parasympathetic innervation mainly through cholinergic fibers. The pupil dilatation is instead related to the contraction of the radial muscle of the iris, under sympathetic control (Loewenfeld and Lowenstein, 1993). Due to the well-known neuroanatomical substrate, the clinical examination of the pupillary light reflex is considered an indicator of the optic nerve conduction, brainstem integrity, vigilance and coma. In recent years, studies in rodents and non-human primates found a tight coupling between pupil size and cortical state even during quiet wakefulness, suggesting a non-luminance mediated system for pupil size variations, associated to neural network oscillations. Studies combining electrophysiology, optical imaging and neural networks modeling, indicated that the link between brain state activity and pupil size is related to the neuro-modulatory effect of the noradrenergic and cholinergic systems (Murphy et al., 2014; Costa and Rudebeck, 2016; Joshi et al., 2016; Eckstein et al., 2017). In this respect, a direct relationship between pupil size and moment-to-moment fluctuations in the activity of noradrenergic neurons of the brainstem locus coeruleus (LC) has been verified (Aston-Jones and Cohen, 2005; Nassar et al., 2012). Other forebrain nuclei and cortical areas connected to LC are activated during spontaneous and event driven pupil size changes (Wang and Munoz, 2015; Joshi et al., 2016) suggesting a circuit for pupil response, linked to arousal, attention and perception systems (Jones, 2004; Naber et al., 2013; Wang and Munoz, 2015; Fazlali et al., 2016; Reimer et al., 2016; Larsen and Waters, 2018). Overall, these studies outline a new role for the pupil size monitoring as a reliable and non-invasive peripheral marker of rapid brain state changes (Hartmann and Fischer, 2014; Schwalm and Jubal, 2017).

From a methodological point of view, a challenge in the analysis of the pupil size variations is the identification of specific patterns that may be representative of changes in the cortical state activity. Different methods have been proposed to assess the pupillary spontaneous oscillations in isoluminant–non-accommodation inducing conditions or in the dark (Lüdtke et al., 1998; Pong and Fuchs, 2000; Zénon et al., 2014; Zénon, 2017). According to the assumptions those methods meet, we distinguish: stationary and linear assumption meeting methods, non-linear assumption meeting methods and non-linear and non-stationary assumption meeting methods. Like other physiological non-stationary signals, under steady stimulation, the pupillary oscillatory signal is expected to show non-linear and chaotic patterns (Poon and Merrill, 1997; Morad et al., 2000; Wilhelm et al., 2001; Merritt et al., 2004; Muppidi et al., 2013; Regen et al., 2013). The non-linear methods assume that the dynamics of the pupil size follow the rules of deterministic chaos rather than a stochastic or linear process (Rosenberg and Kroll, 1999). Common non-linear methods for the analysis of pupillary oscillations imply the use of the Short Time Fourier Transform (Nowak et al., 2008) and wavelets transformations (Henson and Emuh, 2010; Nowak et al., 2013; Reiner and Gelfeld, 2014). These methods assume an underlying stationary signal or require an *a priori* knowledge of the shape of the single basis wave; assumptions that do not well reflect the pupillary dynamics (Onorati et al., 2016). Among the most recent proposed non-linear and non-stationary meeting methods for the analysis of the pupil oscillations, there are the Hilbert Huang Transform, the EMD (Ruiz-Pinales et al., 2016; Villalobos-Castaldi et al., 2016), and the recurrence plots (Mesin et al., 2013, 2014; Monaco et al., 2014). The Hilbert-Huang transform is a frequency domain transformation, with the advantage of maintaining a good temporal and frequency resolution. Through the EMD, the original signal is split into components with slowly varying amplitude and phase, also known as IMFs. By applying a Hilbert transform to the IMF, instantaneous frequencies are generated as functions of time that give sharp identifications of embedded structures (Barnhart, 2011; Ruiz-Pinales et al., 2016; Villalobos-Castaldi et al., 2016). The RQA consists in taking single physiological measurements, projecting them into multidimensional space by embedding procedures and in identifying correlations that are not apparent in one-dimensional time series. This method provides quantitative indexes related to the number and duration of recurrences of the trajectory of a dynamical system in the phase space (Marwan, 2008; Webber and Marwan, 2015). Then, by applying the cross-recurrence analysis (CRQA) which is a bivariate extension of the RQA, we can investigate the dynamic interactions among the systems modulating pupil size oscillations. The use of CRQA has the advantage to better capture the recurring properties of a dynamic system given by the interaction over time of streams of information (Marwan, 2008; Coco and Dale, 2014). For this purpose, the EMD and CRQA were applied in succession. The main goal of our analysis was the identification of specific frequency components of the oscillatory signal comprised in the range of ANS, that could be quantified by couples of DET and ENT lying within the 95% prediction ellipse. Our result suggests that, in awake healthy subjects at rest, pupils oscillate in darkness with high frequency (HF) and low frequency (LF) components that are in the range of ANS, suggesting a balance between noradrenergic/cholinergic tone. Moreover, the position of the points on the ENT-DET plane seems to be related to the ESS score, and therefore, could give insights into the sleepiness state.

MATERIALS AND METHODS

Participants

Twenty-six healthy subjects participated to the study (average age 36 $\pm$ 13 years old). The participants did not have neurological deficits or serious refractive problems. Moreover, the participants did not assume caffeine in the 2 h preceding the data collection (Wilhelm et al., 2014), and they reported to have slept more than 6 h in the night before the recording (average sleep hours 7.2 $\pm$ 0.1). The data collection was performed always between 3 and 6 pm. All subjects gave their written informed consent and the study respected the Declaration of Helsinki and was approved by the local Ethics Committee (Comitato Etico Locale

Azienda Ospedaliera Universitaria Senese, EVAlab protocol CEL no. 48/2010).

Experimental Setting

Pupil diameter recordings were performed monocularly with an ASL 504 eye-tracker device (Applied Science Laboratories, Bedford, MA, United States) sampling at a mean frequency of 240 Hz. The remote eye-tracker was placed at 650 mm far from the eye of the participant. The relative position of the subject's head with respect to the eye-tracker was kept still by mean of a chinrest.

Acquisition Protocol

Prior the data collection, the subjects were administered with an test ESS to investigate their vigilance state. ESS scores less than 11 are normally associated to subjects having normal sleepy state, while ESS scores greater than 11 suggested excessive daytime sleepiness (Parkes et al., 1998). ESS is a common used self-assessment questionnaire for the tiredness evaluation, hence it can turn to be a bias-prone measurement.

All the recordings were performed in a quiet light-controlled environment. To avoid the stimulation of pupillary light reflex, the subject was instructed to look straight for 15 min in a complete dark room (0 lux), similarly to the procedure adopted by Lüdtke et al. (1998). To reduce mental activity and cognitive load, subjects were instructed to try not to think to anything and to relax.

Data Processing

The flow chart of procedure employed to analyze the pupillary frequency balancing between the sympathetic and parasympathetic systems is shown in **Figure 1**. The pupil diameter data was exported in comma separated values format files and analyzed offline through Matlab (The Mathworks). The signal was de-blinked. Signal instances with the pupil diameter equal to zero were marked as blinks and removed from the signal. The remaining signal was then linearly interpolated. Moreover, machine artifacts introduced by the eye-tracker device due to failures to detect the pupil, were removed using Hampel filtering and low-pass filtered with a cut-off frequency (f_0) of 2 Hz. The Hampel function computes the median of the data within moving windows. The width of the filter window (w) was determined accordingly to the ratio of the sample frequency (f_s) over the cut-off frequency f_0 (Equation 1):

$$w = 0.44 \cdot f_s / f_0 \quad (1)$$

The variation of pupil size was computed with respect to a baseline value of the pupil estimated for each participant. Specifically, the baseline value of the pupil diameter signal was determined as the maximum value of the pupil size attained in the first 60 s of the signal in darkness (baseline), when the signal was expected to be more stable. The mean or the median of the pupil size were possible alternative reference values. However, taking the maximum value as reference enabled us to normalize the signal on the basis of a really observed value and to preserve the dynamics of the phenomenon. A baseline-corrected pupil diameter time series was then calculated as the diameter percentage change with respect to the value gathered in the basal

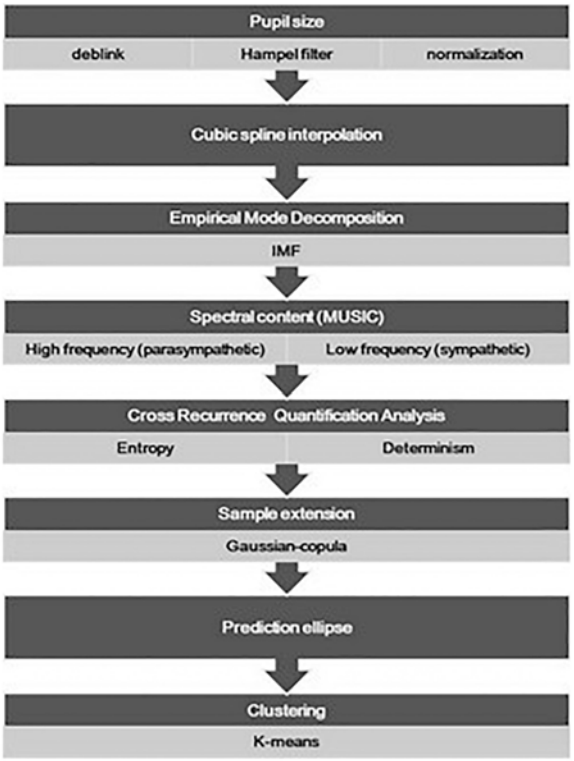


FIGURE 1 | The flow chart presents the major procedures adopted for the analysis of the pupil size oscillation, from the data pre-treatment (deblink and artifact removal) and normalization, to the final drawing of the prediction ellipse. Data points of the prediction region in the entropy-determination plane underwent a further classification analysis and a pairwise comparison of the identified clusters was also done.

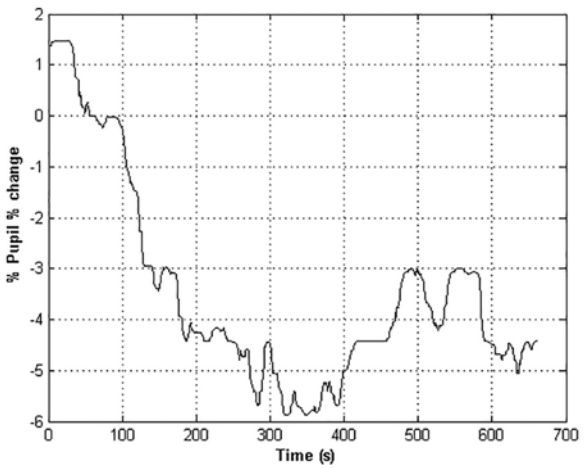


FIGURE 2 | The time course of the pupil % change of an healthy subject. Negative values indicate a restriction of the pupil size with respect to the baseline value.

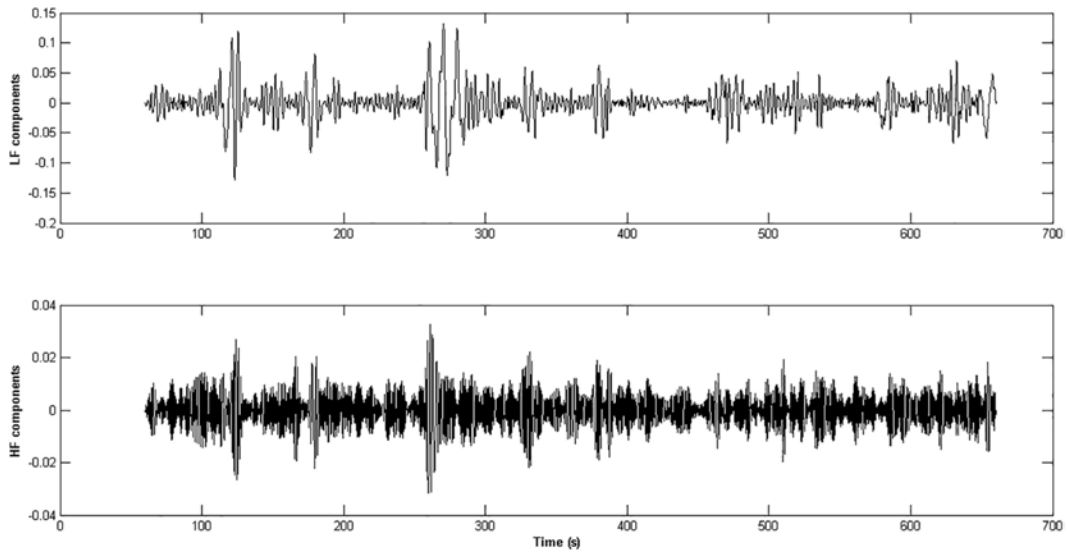


FIGURE 3 | The panels represent the HF and LF components extracted from the pupil % change time series of an healthy subject. The IMFs obtained through the application of the EMD technique whose frequency content was inside the range [0.15 – 0.45 Hz] were aggregated and form the HF component of the ANS activity (lower panel), while the IMFs in the range [0 – 0.15 Hz] gave rise to the LF component (upper panel).

condition (Equation 2).

$$\%change = \frac{X_t - Baseline}{Baseline} \cdot 100 \quad (2)$$

where X_t is the pupil diameter recorded at time t. The baseline correction provided the removal of inter-subject variability in pupil size percentage change of the pupil diameter signal (Lowenstein et al., 1963; **Figure 2**).

Data Analysis

A cubic spline interpolation was used for compressing the percentage change time series with a resolution of five data points per second, which satisfied the Nyquist criterion (for the given 2 Hz cut-off frequency). The EMD was applied to the cubic spline interpolation of the percentage change time series in the autonomic frequency band ranging from 0 to 0.45 Hz (Huang et al., 1998). Since we were interested in a global spectral characterization of the IMFs derived from the EMD, the spectral content of the IMFs was estimated through MUSIC algorithms (Schmidt, 1986). The IMFs having most of the power in the autonomic frequency band were retained. The IMFs were then aggregated accordingly to the HF (0.15–0.45 Hz) and LF (0–0.15 Hz) ranges related to the parasympathetic and sympathetic systems activity (Cabrerizo et al., 2014; **Figure 3**). A CRQA was performed (Marwan and Kurths, 2002; Marwan, 2016) to assess the similarity between the dynamics of the parasympathetic and sympathetic processes by comparing the interaction between the LF and the HF components in the phase space. Three hyper-parameters must be set in the CRQA: EmbDim, TD, and the neighborhood radius (R). A symplectic geometry-based algorithm was used for estimating the EmbDim (Lei et al., 2002). The TD value was chosen as the one within the range (0: w/EmbDim) that maximized the sample entropy of the percentage change of pupil size. A FAN was taken as the neighborhood criterion, such that the cross-recurrence point density had a fixed predetermined value of 20%.

Two main parameters from CRQA were considered: the determinism (DET), which quantifies the fraction of periodic structures in the trajectories of the LF and HF dynamics in the phase space, and the entropy (ENT), which is the Shannon entropy of the diagonal line length distribution. Periodic signals are expected to yield high values of DET and small values of ENT (Marwan et al., 2007). To enlarge the sample size enough to apply clustering procedures on the DET-ENT plane and to investigate more carefully for possible highlights of this method on the analysis of the balancing of the sympathetic and parasympathetic systems thorough the analysis of the oscillations of the pupil diameter, we employed the Gaussian-copula simulation approach. Hence, firstly the association among age, ESS, ENT, and DET was measured by the Pearson's correlation matrix. Then, a Gaussian-copula which maintained the dependence structure was used to generate one-hundred correlated multivariate data of those variables.

The percentage of pupil size change of each of the simulated data was then represented as a point on the DET-ENT plane.

Statistical Analysis

On the simulated dataset the Doornik-Hansen multivariate normality test (Doornik and Hansen, 2008) was performed to verify the null hypothesis that the points in the DET-ENT plane were generated from a bivariate Gaussian distribution. The 95% prediction ellipse was calculated around the mean of observed points in the DET-ENT plane. Equations 3–4 indicate the formula

for determining the length of the two semi-axes:

$$a_x = 2 \cdot \sqrt{\lambda_2 \cdot \frac{(n_{obs}-1)\cdot n_{var}\cdot f(1-\alpha,\ n_{var},\ n_{obs}-n_{var})}{n_{obs}-n_{var}}} \quad (3)$$

$$a_y = 2 \cdot \sqrt{\lambda_2 \cdot \frac{(n_{obs}-1)\cdot n_{var}\cdot f(1-\alpha,\ n_{var},\ n_{obs}-n_{var})}{n_{obs}-n_{var}}} \quad (4)$$

where a_x and a_y are the major and minor semi-axes of the ellipse, n_{var} is the number of variables (=2), n_{obs} is the number of the observations (=100), λ_1 and λ_2 are the eigenvalues (in descending order) obtained from the spectral decomposition of the covariance matrix of ENT and DET, f is the pdf of the F distribution for the given significance α level and degrees of freedom (n_{var}, n_{obs}-n_{var}). The orientation of the ellipse is given (in radians) by the direction of the eigenvector associated to the largest eigenvalue:

$$\theta = atan\left(\frac{v_y}{v_x}\right) \quad (5)$$

where *atan* is the inverse tangent function, and v_x and v_y are the components of the eigenvector corresponding to the largest eigenvalue. The coordinates of the points $[P_x, P_y]$ laying on the ellipse contour are calculated as follows:

$$P_x = x_c + \left[\frac{a_x}{2}\cdot cos(t)\cdot cos(\vartheta) - \frac{a_y}{2}\cdot sin(t)\cdot sin(\vartheta)\right] \quad (6)$$

$$P_y = y_c + \left[\frac{a_x}{2}\cdot cos(t)\cdot sin(\vartheta) + \frac{a_y}{2}\cdot sin(t)\cdot cos(\vartheta)\right] \quad (7)$$

where x_c and y_c are the coordinates of the center of the ellipse, and t ranges in the interval $[0, 2\pi]$.

The prediction region can provide the regulatory reference points for assessing if the underlying slow oscillations in the autonomic band of a new observed pupil size time series have the characteristics of a normal pattern.

Afterward, unsupervised clustering through K-means method with two clusters and a L1-norm distance function was applied within the elliptic prediction area. The two clusters were compared in covariance matrices and means vectors. Accordingly, the Box's *M*-test was considered for verifying the homogeneity of the covariance matrices, and the Hotelling's T^2 test was used for testing the means. The variables age, ESS and *% change* associated to each cluster were separately compared through the Mann-Whitney unpaired test.

All statistical tests were two-sided and performed on Matlab with a 5% level of significance.

RESULTS

Self-organized adaptive systems like the brain generate complex signals which are inherently non-linear and non-stationary. Furthermore, unstable, weak, and state-dependent phase-locking characterizes the coupling between the biological oscillators (Shockley et al., 2002). Since the couplings between biological signals could also be predominantly transient, the canonical techniques of signal analysis, which basically rely on the assumption of stationary signals, are not appropriate. More importantly, the autonomic control of the spontaneous pupil fluctuations is expected to have non-linear/chaotic dynamics which can be well explored by recurrence analysis methods, whose domain is in the phase-space trajectories (Mesin et al., 2013). For these reasons, we chose the cross-recurrence method to analyze the spectral components of the ANS activity controlling the pupil fluctuations.

The EMD method was applied to the time series of pupil size variations to extract the low and high frequency components of the signal, which were found in the range of the ANS band. In fact, the EMD procedure, which is known to deal with non-linear and non-stationary signals like the pupil size oscillations, is a data driven method that overcomes the limitation of basis function shape typical of the wavelet decomposition method (Gonalves et al., 2007). The CRQA was then performed over the high and low frequency components and two parameters, i.e., entropy (ENT) and determinism (DET), were retained as the major features which quantified the non-linear dynamics of the high- and LF coupled oscillators in the autonomic band.

In **Table 1** the sampling distributions of age, ESS scores, ENT, DET and average pupil change are reported. The sample

TABLE 1 | Sampling distributions of age, Epworth Sleepiness Scale, entropy, determinism, and average pupil change.

Age	ESS	Entropy	% Determinism	% Change
24	3	1.08	52.34	−0.38
24	5	0.97	49.64	0.85
24	5	0.96	56.69	3.97
24	6	1.00	57.39	−2.36
24	13	1.03	46.94	−1.69
24	14	0.90	48.86	−2.00
25	6	0.86	37.91	−1.57
25	14	0.92	43.09	0.30
27	2	0.84	38.14	1.94
27	7	0.86	50.49	0.74
27	9	1.00	37.73	−3.35
28	3	0.89	41.85	−1.37
28	4	0.83	41.29	0.19
29	5	0.87	42.43	1.72
29	10	0.83	37.35	−1.37
29	13	0.97	56.99	1.39
31	5	0.72	29.79	1.43
33	8	0.83	44.9	−2.02
46	7	0.88	39.78	−2.60
46	9	0.99	46.22	−0.40
49	6	0.77	45.12	1.16
50	3	0.90	38.35	−3.46
51	4	0.95	35.6	0.33
56	7	1.00	51.47	2.34
62	14	0.94	56.19	2.07
63	8	1.04	50.69	−0.08

TABLE 2 | Values of age, Epworth Sleepiness Scale, entropy, determinism, and average pupil change generated from a Gaussian-copula.

Age	ESS	Entropy	% Determinism	% Change	Age	ESS	Entropy	% Determinism	% Change
24	1	0.76	33.90	−2.02	29	5	0.83	31.97	−1.37
24	7	0.91	51.04	−1.94	29	5	1.00	51.30	0.29
24	11	0.86	43.69	−1.37	29	6	0.96	37.97	−0.40
24	9	0.90	41.85	−1.37	29	4	0.89	37.76	−1.75
24	10	0.84	45.84	−3.04	29	6	0.91	43.87	−1.37
24	6	0.95	38.26	−3.55	29	7	0.83	41.42	−3.45
24	4	0.97	49.38	−1.37	29	3	0.88	42.35	0.42
24	5	0.93	41.80	−3.42	29	3	0.91	39.35	−2.02
24	8	0.89	44.10	−2.01	29	7	0.83	42.78	1.84
24	5	1.04	50.81	−2.01	29	5	0.87	38.07	0.48
24	9	0.84	41.36	−1.92	30	4	0.86	37.73	0.76
24	3	0.92	38.32	−3.42	31	6	0.83	36.60	−0.39
24	6	0.99	50.99	−1.37	31	14	0.97	56.78	2.23
24	14	0.81	39.87	−3.47	32	9	0.94	45.02	−0.39
24	4	0.83	35.69	−3.44	32	14	0.84	45.25	−1.55
24	3	0.77	36.14	−2.54	34	8	0.96	45.95	−0.39
24	14	0.97	51.16	−2.49	36	4	0.85	37.75	−3.36
24	12	0.97	57.02	−1.98	41	13	0.86	51.06	0.30
24	5	1.00	56.28	−1.37	41	6	0.83	35.96	−1.37
24	5	1.03	56.53	0.20	41	5	0.87	43.81	1.34
24	13	1.00	51.29	−2.28	42	2	0.90	38.30	2.15
24	5	1.00	50.54	1.71	43	6	0.89	46.47	1.52
25	5	0.87	41.70	−1.87	46	6	0.94	51.45	1.63
25	14	1.08	57.34	−2.48	46	3	0.68	37.74	1.39
25	14	0.83	41.76	−1.42	46	3	1.00	45.27	0.06
25	8	0.90	37.86	−2.92	46	7	1.00	52.01	0.31
26	3	0.93	38.25	−1.38	46	14	1.03	56.75	1.93
27	9	0.97	54.93	−2.01	46	14	1.00	52.63	−1.47
27	4	0.74	30.17	−1.67	46	6	0.89	50.60	1.42
27	2	0.88	49.28	1.39	46	14	0.83	37.84	−1.75
27	6	0.81	44.77	−1.04	46	5	1.00	54.86	−1.64
27	4	0.86	36.67	−2.27	47	3	0.92	38.32	0.33
27	3	0.75	33.77	0.74	47	8	0.90	56.60	1.76
27	5	0.88	39.46	0.31	48	4	0.89	46.61	1.42
27	3	1.00	41.90	−0.43	49	3	0.83	37.95	1.41
27	3	0.83	37.81	0.71	49	9	0.95	45.01	−2.25
27	6	0.83	37.51	−3.07	49	9	0.83	40.07	0.47
28	3	0.86	45.83	2.08	50	3	0.99	38.08	−2.01
28	7	0.95	43.95	−1.75	50	3	0.84	31.53	0.84
28	3	0.93	38.87	0.08	50	7	1.00	56.82	1.96
28	5	0.84	38.25	1.05	51	7	0.88	45.02	2.03
28	3	0.88	39.49	2.02	51	13	0.97	38.15	1.47
28	3	0.67	34.75	−0.67	51	7	0.73	37.41	2.03
28	5	0.83	28.51	−2.04	52	3	0.89	44.98	1.96
28	3	0.90	45.54	1.56	53	8	1.08	56.66	2.05
28	8	0.87	38.23	−2.39	56	14	1.05	57.08	1.15
28	7	0.89	38.19	0.03	56	7	0.96	41.81	−0.09
29	3	0.68	25.23	−1.04	60	5	0.92	42.09	−1.57
29	9	0.88	45.97	0.19	61	5	1.01	44.96	1.08
29	12	1.00	50.96	1.70	63	7	1.02	57.21	2.28

declared a normal level of diurnal drowsiness (ESS: mean = 7.3; *SD* = 3.7). Five subjects (four of age lower than 30, and one of age greater than 60) reported relatively high ESS scores (>10). The cross-recurrence analysis returned low values both for ENT (mean = 0.92; *SD* = 0.09) and for DET (mean = 45.28%; *SD* = 7.46%). The percentage of pupil change in the sample (%

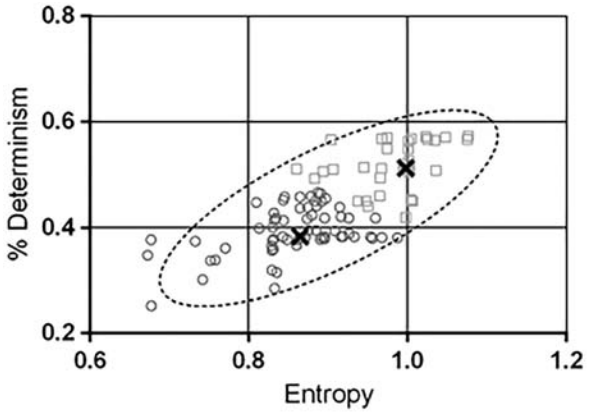


FIGURE 4 | The contour of the ellipse envelopes the region in the entropy-determinism space where new measured combinations of entropy (ENT) and determinism (DET) will fall with a 95% probability. The null hypothesis that the two variables are generated from a bivariate normal distribution was verified through the Doornik-Hansen test. Values of DET in the y-axis have been rescaled in the range [0 – 1]. A K-means algorithm was used to characterize sub-areas within the prediction ellipse and two distinct clusters were found out. The thick black crosses indicate the centres of the clusters. Cluster 2 (squares) exhibited significant higher combination of ENT and DET than cluster 1 (circles). In addition the ESS resulted greater in cluster 2 compared to cluster 1. This finding suggested that the ENT-DET bivariate distribution inherently conveys information about the drowsiness level of the subjects.

TABLE 3 | Parameters of the 95% prediction ellipse in the ENT-DET plane.

Center	(0.90, 0.44)
Semiaxis (x)	0.53
Semiaxis (y)	0.20
Angle (rad)	0.69
Angle (degree)	39°44'
Hyper-volume	0.08
Perimeter	0.12
Eccentricity	0.92

change) (mean = −0.16%; *SD* = 1.91%) indicated an overall slight loss of pupil size with respect to the baseline, but high variability of the fluctuations as well.

We firstly analyzed the possible association among the observed age, ESS, ENT, and DET. Based on the results, the DET and ENT variables were not significantly correlated to the age and the ESS score of the participants. Instead, a significant correlation between ENT and DET was found ($r = 0.58$, $p = 0.002$).

The bivariate distribution of ENT and DET obtained from Gaussian-copula simulated points (**Table 2**) is depicted in **Figure 4**, together with the 95% prediction ellipse. The simulated values of ENT (mean = 0.90; *SD* = 0.09) and DET (mean = 43.66%; *SD* = 7.43%) were consistent with the values observed in the sample.

The major parameters of the prediction ellipse are displayed in **Table 3**. The coordinates of the center of the ellipse are the means of the simulated ENT and DET vectors. The axes of the ellipse indicate the magnitude of the inertia along the directions of ENT and DET. The interval estimations of ENT and DET were obtained through Equations 6 and 7. **Table 4** displays the expected intervals of determinism for equally spaced intervals (0.05 bits) of entropy.

TABLE 4 | Normative intervals of determinism by ranges of entropy.

Entropy	% Determinism
0.70–0.75	24–40
0.75–0.80	24–46
0.80–0.85	24–51
0.85–0.90	26–55
0.90–0.95	29–58
0.95–1.00	33–60
1.00–1.05	38–60
1.05–1.10	45–60

TABLE 5 | Descriptive statistics of the clusters identified within the prediction ellipse.

	Cluster 1		Cluster 2	
	Mean	SD	Mean	SD
Age	33.1	10.2	36.3	12.4
ESS	5.7	3.2	8.2	3.7
Entropy	0.86	0.07	0.98	0.05
% Determinism	39	4	52	4
% Change	−0.69	1.77	−0.12	1.64

Through the K-means procedure, two clusters of points were identified within the prediction ellipse and their descriptive statistics is shown in **Table 5**.

The generated ENT-DET values underwent the Doornik-Hansen multinormality test. The hypothesis of bivariate normal distribution was not rejected (DH statistic = 6.97, $p = 0.14$).

The covariance matrices of the clusters were not significantly different (Box's *M*-test = 3.2; $p = 0.37$). The result of the Hotelling T^2 test indicated that the bivariate ENT-DET means vectors between the clusters were significantly different ($T^2 = 200.8$; $p < 0.0001$). The two clusters exhibited also significant different ESS scores (*U*-test = 701.5; $p = 0.002$), whilst they did not differ in age (*U*-test = 1073; $p = 0.64$), nor in *% change* (*U*-test = 933.5; $p = 0.14$).

DISCUSSION

The analysis of the pupil size oscillations is a promising diagnostic tool, enabling improvements in the identification of cortical state changes. Variations of cortical state activity during wakefulness have a strong influence on neural, perceptual and behavioral responses. Pupil diameter varies not only in response to variation of luminance and accommodation, but also during changes in alertness, attention, mental effort and decision making, suggesting a direct link between pupil size variation and cortical state changes (Preuschoff et al., 2011;

Nassar et al., 2012; Naber et al., 2013; Alnæs et al., 2014; de Gee et al., 2014). Changes in the cortical state are associated to well characterized variations of the cortical signal frequency. Specifically, in awake rodents the investigation of local field potentials demonstrated the prevalence of LF fluctuations during periods of quiet resting. However, the initiation of locomotion or whisking was related to the suppression of low frequency components and increased high frequency oscillations (Poulet et al., 2012; Eggermann et al., 2014, McGinley et al., 2015b). This transition between slow and fast cortical activity was also observed across cortical regions (Poulet and Crochet, 2019). Electrophysiological studies have revealed that pupil constriction is associated to slow and synchronous cortical responses and inattentive behavior. Conversely, the cortical activation during task engagement or locomotion shows a persistent desynchronized neuronal activity associated to the dilatation of the pupil (Reimer et al., 2014, 2016; McGinley et al., 2015b; Schwalm and Jubal, 2017). Pupil size fluctuations and cortical state variations are modulated by the central noradrenergic and cholinergic pathways. Thus, monitoring pupil dynamics could be a reliable proxy of the changes in cortical states (Reimer et al., 2014, 2016; McGinley et al., 2015a,b). More specifically, the release of acetylcholine (Ach) from the basal forebrain and noradrenaline (NA) from LC have been shown to drive both the state of cortical connectivity and the pattern of the pupil size oscillations also in resting conditions (Reimer et al., 2016; Schwalm and Jubal, 2017). At the peripheral level, both Ach and NA are neurotransmitters of the ANS (parasympathetic and sympathetic systems, respectively) also controlling the pupil diameter. Overall, these premises encourage exploring new and reliable techniques for pupil dynamics monitoring that allow the identification of parameters attributable to NA and Ach modulatory effect in various cortical state changes.

We propose here, a method that can be used as a quantitative measurement of the non-linear dynamics of the pupil fluctuations. We applied a cross-recurrence technique for estimating determinism (DET) and entropy (ENT) features and their distribution, in order to quantify the degree of coupling between the oscillators of the low (LF) and high frequency (HF) components of the pupillary signal. To the best of our knowledge this is the first study on the use of the ENT-DET plane for analyzing the dynamical systems associated to pupil size fluctuation during stationary scotopic visual conditions.

In our cohort of subjects, we observed low levels of determinism (<60%) and entropy (<1). This is consistent with spontaneous physiological signals recorded from healthy subjects, which are expected to be highly complex. Actually, low determinism can be associated to increase in the uncertainty of the signals, and hence to increase in the signal chaotic properties (i.e., complexity). In facts, complex systems are typically highly ordered. Therefore, they tend to preserve low entropy and counteract the second law of thermodynamics (*free energy principle*). A *de-complexification* process occurs when free-running physiological signals present sustained loss of complexity. The loss of complexity leads to less ordered states with higher entropy and with stronger coupling of the oscillators controlling the expression of the signal. This degradation in complexity is typically observed in pathological conditions or advanced aging. Therefore, the major result of this study is the identification of a normative elliptical region in the ENT-DET plane for the pupillary oscillators that could be compared with data from group of patients with neurodegenerative diseases. We hypothesize that the occurrence of points outside of the defined elliptical prediction region may signal potential pathological conditions related to alterations in the ANS. As secondary outcome, we observed that, within the elliptical region of confidence, clusters of points with different characteristics of ENT-DET highly differed also in their ESS scores. This finding suggests that the location of the points in the ENT-DET plane can also reveal alterations in the sleepiness state.

Our results indicate that in resting wakefulness conditions, without the influence of light and accommodation, pupil size oscillations are under the effect of a balanced cholinergic/noradrenergic tone. We believe that the employed CRQA-based method may help to lay the groundwork for studying the LF and HF components of the pupil, which may be related to neuronal network state of the brain at rest. Importantly, it consists in a non-invasive procedure that could be easily adopted in clinical context and for diagnostic assessment such as neurodegenerative conditions. Furthermore, this method is adaptable to different experimental conditions (e.g., variations of the visual stimulus, recording during cognitive tasks, etc) provided that the opportune frequency components are dug out from the signal. The joint recording of the pupil size fluctuations along with other physiological signals (e.g., heart rate variability, EEG, etc) would improve the method, since the study of possible time-dependent and/or frequency-related changes in autonomic functions would be facilitated by this integration.

ETHICS STATEMENT

The study was approved by the local Ethical Committee Comitato Etico Locale Azienda Ospedaliera Universitaria Senese, EVAlab protocol CEL no. 48/2010.

AUTHOR CONTRIBUTIONS

All authors conceived and designed the study, critically revised the manuscript, and approved the final version of the manuscript. FR and VS acquired the data. AR, PP, and VS involved in the analysis and interpretation of data, and drafted the manuscript. AR revised the scientific content of the study.

ACKNOWLEDGMENTS

We thank particularly Dr. Gemma Tumminelli for the help in the recruitment of the participants and for the data collection.

REFERENCES

Alnæs, D., Sneve, M. H., Espeseth, T., Endestad, T., van de Pavert, S. H. P., and Laeng, B. (2014). Pupil size signals mental effort deployed during multiple object tracking and predicts brain activity in the dorsal attention network and the locus coeruleus. *J. Vis.* 14:1. doi: 10.1167/14.4.1

Aston-Jones, G., and Cohen, J. D. (2005). An integrative theory of locus coeruleus-norepinephrine function: adaptive gain and optimal performance. *Annu. Rev. Neurosci.* 28, 403–450. doi: 10.1146/annurev.neuro.28.061604.135709

Barnhart, B. L. (2011). *The Hilbert-Huang Transform: Theory, Applications, Development*. Iowa: University of Iowa. doi: 10.1146/annurev.neuro.28.061604.135709

Cabrerizo, M., Cabrera, A., Perez, J. O., de la Rua, J., Rojas, N., Zhou, Q., et al. (2014). Induced effects of transcranial magnetic stimulation on the autonomic nervous system and the cardiac rhythm. *Sci. World J.* 2014:349718. doi: 10.1155/2014/349718

Coco, M. I., and Dale, R. (2014). Cross-recurrence quantification analysis of categorical and continuous time series: an R package. *Front. Psychol.* 5:510. doi: 10.3389/fpsyg.2014.00510

Costa, V. D., and Rudebeck, P. H. (2016). Previews more than meets the eye: the relationship between pupil size and locus coeruleus activity. *Neuron* 89, 8–10. doi: 10.1016/j.neuron.2015.12.031

de Gee, J. W., Knapen, T., and Donner, T. H. (2014). Decision-related pupil dilation reflects upcoming choice and individual bias. *Proc. Natl. Acad. Sci. U.S.A.* 111, E618–E625.

Doornik, J., and Hansen, H. (2008). An omnibus test for univariate and multivariate normality. *Oxf. Bull. Econ. Stat.* 70, 927–939. doi: 10.1111/j.1468-0084.2008.00537.x

Eckstein, M. K., Guerra-Carrillo, B., Singley, A. T. M., and Bunge, S. A. (2017). Beyond eye gaze: what else can eyetracking reveal about cognition and cognitive development? *Dev. Cogn. Neurosci.* 25, 69–91. doi: 10.1016/j.dcn.2016.11.001

Eggermann, E., Kremer, Y., Crochet, S., and Petersen, C. C. (2014). Cholinergic signals in mouse barrel cortex during active whisker sensing. *Cell Rep.* 9, 1654–1660. doi: 10.1016/j.celrep.2014.11.005

Fazlali, Z., Ranjbar-Slamloo, Y., Adibi, M., and Arabzadeh, E. (2016). Correlation between cortical state and locus coeruleus activity: implications for sensory coding in rat barrel cortex. *Front. Neural Circ.* 10:14. doi: 10.3389/fncir.2016.00014

Gonalves, P., Abry, P., Rilling, G., and Flandrin, P. (2007). "Fractal dimension estimation: empirical mode decomposition versus wavelets," in *Proceedings of 32nd IEEE International Conference on Acoustics, Speech and Signal Processing - ICASSP '07*, Honolulu.

Hartmann, M., and Fischer, M. H. (2014). Pupillometry: the eyes shed fresh light on the mind. *Curr. Biol.* 24, R281–R282. doi: 10.1016/j.cub.2014.02.028

Henson, D. B., and Emuh, T. (2010). Monitoring vigilance during perimetry by using pupillography. *Invest. Ophthalmol. Vis. Sci.* 51, 3540–3543. doi: 10.1167/iovs.09-4413

Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Shih, H. H., Yen, N., et al. (1998). The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc. R. Soc. A* 454, 903–995. doi: 10.1098/rspa.1998.0193

Jones, B. E. (2004). Activity, modulation and role of basal forebrain cholinergic neurons innervating the cerebral cortex. *Prog. Brain Res.* 145, 157–169. doi: 10.1016/s0079-6123(03)45011-5

Joshi, S., Li, Y., Kalwani, R. M., and Gold, J. I. (2016). Relationships between pupil diameter and neuronal activity in the locus coeruleus, colliculi, and cingulate cortex. *Neuron* 89, 221–234. doi: 10.1016/j.neuron.2015.11.028

Larsen, R. S., and Waters, J. (2018). Neuromodulatory correlates of pupil dilation. *Front. Neural Circ.* 12:21. doi: 10.3389/fncir.2018.00021

Lei, M., Wang, Z., and Feng, Z. (2002). A method of embedding dimension estimation based on symplectic geometry. *Phys. Lett. A* 303, 179–189. doi: 10.1016/s0375-9601(02)01164-7

Loewenfeld, I. E., and Lowenstein, O. (1993). *The Pupil: Anatomy, Physiology, and Clinical Applications*, Vol. 2. Hoboken, NJ: Wiley-Blackwell. doi: 10.1016/s0375-9601(02)01164-7

Lowenstein, O., Feinberg, R., and Loewenfeld, I. E. (1963). Pupillary movements during acute and chronic fatigue a new test for the objective evaluation of tiredness. *Invest. Ophthalmol. Vis. Sci.* 2, 138–158.

Lüdtke, H., Wilhelm, B., Adler, M., Schaeffel, F., and Wilhelm, H. (1998). Mathematical procedures in data recording and processing of pupillary fatigue waves. *Vis. Res.* 38, 2889–2896. doi: 10.1016/s0042-6989(98)00081-9

Marwan, N. (2008). A historical review of recurrence plots. *Eur. Phys. J. Spec. Top.* 164:3. doi: 10.1140/epjst/e2008-00829-1

Marwan, N. (2016). *CRP Toolbox For Matlab Toolbox*. Available at: https://www.academia.edu/14066129/CRP_Toolbox_for_MATLAB (accessed July 16, 2018).

Marwan, N., and Kurths, J. (2002). Nonlinear analysis of bivariate data with cross-recurrence plots. *Phys. Lett. A* 302, 299–307. doi: 10.1016/s0375-9601(02)01170-2

Marwan, N., Romano, M. C., Thiel, M., and Kurths, J. (2007). Recurrence plots for analysis of complex systems. *Phys. Rep.* 438, 237–329. doi: 10.1016/j.physrep.2006.11.001

McGinley, M. J., David, S. V., and McCormick, D. A. (2015a). Cortical membrane potential signature of optimal states for sensory signal detection. *Neuron* 87, 179–192. doi: 10.1016/j.neuron.2015.05.038

McGinley, M. J., Vinck, M., Reimer, J., Batista-Brito, R., Zagha, E., Cadwell, C. R., et al. (2015b). Waking state: rapid variations modulate neural and behavioral responses. *Neuron* 87, 1143–1161. doi: 10.1016/j.neuron.2015.09.012

Merritt, S. L., Schnyders, H. C., Patel, M., and Basner, R. C. (2004). Pupil staging and EEG measurement of sleepiness. *Int. J. Psychophysiol.* 52, 97–112. doi: 10.1016/j.ijpsycho.2003.12.007

Mesin, L., Cattaneo, R., Monaco, A., and Pasero, E. (2014). "Pupillometric Study of the Dysregulation of the Autonomous Nervous System by SVM Networks," in *Recent Advances of Neural Network Models and Applications. Smart Innovation, Systems and Technologies*, Vol. 26, eds S. Bassis, A. Esposito, and F. Morabito (Cham: Springer), 107–115. doi: 10.1007/978-3-319-04129-2_11

Mesin, L., Monaco, A., and Cattaneo, R. (2013). Investigation of nonlinear pupil dynamics by recurrence quantification analysis. *BioMed. Res. Int.* 2013:420509. doi: 10.1155/2013/420509

Monaco, A., Cattaneo, R., Mesin, L., Fiorucci, E., and Pietropaoli, D. (2014). Evaluation of autonomic nervous system in sleep apnea patients using pupillometry under occlusal stress: a pilot study. *Cranio* 32, 139–147. doi: 10.1179/0886963413z.00000000022

Morad, Y., Lemberg, H., Yofe, N., and Dagan, Y. (2000). Pupillography as an objective indicator of fatigue. *Curr. Eye Res.* 21, 535–542. doi: 10.1076/0271-3683(200007)2111-zft535

Muppidi, S., Adams-Huet, B., Tajzoy, E., Scribner, M., Blazek, P., Spaeth, E. B., et al. (2013). Dynamic pupillometry as an autonomic testing tool. *Clin. Auton. Res.* 23, 297–303. doi: 10.1007/s10286-013-0209-7

Murphy, P. R., O'Connell, R. G., O'Sullivan, M., Robertson, I. H., and Balsters, J. H. (2014). Pupil diameter covaries with BOLD activity in human locus coeruleus. *Hum. Brain. Mapp.* 35, 4140–4154. doi: 10.1002/hbm.22466

Naber, M., Alvarez, G. A., and Nakayama, K. (2013). Tracking the allocation of attention using human pupillary oscillations. *Front. Psychol.* 4:919. doi: 10.3389/fpsyg.2013.00919

Nassar, M. R., Rumsey, K. M., Wilson, R. C., Parikh, K., Heasly, B., and Gold, J. I. (2012). Rational regulation of learning dynamics by pupil-linked arousal systems. *Nat. Neurosci.* 15:1040. doi: 10.1038/nn.3130

Nowak, W., Hachol, A., and Kasprzak, H. (2008). Time-frequency analysis of spontaneous fluctuation of the pupil size of the human eye. *Opt. Appl.* 38, 469–480.

Nowak, W., Szul-Pietrzak, E., and Hachol, A. (2013). "Wavelet Energy and Wavelet Entropy as a New Analysis Approach in Spontaneous Fluctuations of Pupil Size Study – Preliminary Research," in *Proceedings XIII Mediterranean Conference on Medical and Biological Engineering and Computing*, Seville, 807–810. doi: 10.1007/978-3-319-00846-2_200

Onorati, F., Mainardi, L. T., Sirca, F., Russo, V., and Barbieri, R. (2016). Nonlinear analysis of pupillary dynamics. *Biomed. Tech.* 61, 95–106. doi: 10.1515/bmt-2015-0027

Parkes, J. D., Chen, S. Y., Clift, S. J., Dahlitz, M. T., and Dunn, G. (1998). The clinical diagnosis of the narcoleptic syndrome. *J. Sleep Res.* 7, 41–52. doi: 10.1046/j.1365-2869.1998.00093.x

Pong, M., and Fuchs, A. F. (2000). Characteristics of the pupillary light reflex in the macaque monkey: discharge patterns of pretectal neurons. *J. Neurophysiol.* 84, 964–974. doi: 10.1152/jn.2000.84.2.964

Poon, C. S., and Merrill, C. K. (1997). Decrease of cardiac chaos in congestive heart failure. *Nature* 389, 492–495. doi: 10.1038/39043

Poulet, J. F., and Crochet, S. (2019). The cortical states of wakefulness. *Front Syst. Neurosci.* 8:64. doi: 10.1037/11149-006

Poulet, J. F., Fernandez, L. M., Crochet, S., and Petersen, C. C. (2012). Thalamic control of cortical states. *Nat. Neurosci.* 15:370. doi: 10.1038/nn.3035

Preuschoff, K., t Hart, B. M., and Einhauser, W. (2011). Pupil dilation signals surprise: Evidence for noradrenaline's role in decision making. *Front. Neurosci.* 5:115. doi: 10.3389/fnins.2011.00115

Regen, F., Dorn, H., and Danker-Hopfe, H. (2013). Association between pupillary unrest index and waking electroencephalogram activity in sleep-deprived healthy adults. *Sleep Med.* 14, 902–912. doi: 10.1016/j.sleep.2013.02.003

Reimer, J., Froudarakis, E., Cadwell, C. R., Yatsenko, D., Denfield, G. H., and Tolias, A. S. (2014). Pupil fluctuations track fast switching of cortical states during quiet wakefulness. *Neuron* 84, 355–362. doi: 10.1016/j.neuron.2014.09.033

Reimer, J., McGinley, M. J., Liu, Y., Rodenkirch, C., Wang, Q., McCormick, D. A., et al. (2016). Pupil fluctuations track rapid changes in adrenergic and cholinergic activity in cortex. *Nat. Commun.* 7:13289. doi: 10.1038/ncomms13289

Reiner, M., and Gelfeld, T. M. (2014). Estimating mental workload through event-related fluctuations of pupil area during a task in a virtual world. *Int. J. Psychophysiol.* 93, 38–44. doi: 10.1016/j.ijpsycho.2013.11.002

Rosenberg, M. L., and Kroll, M. H. (1999). Pupillary hippus: an unrecognized example of biologic chaos. *J. Biol. Syst.* 7, 85–94. doi: 10.1142/s0218339099000085

Ruiz-Pinales, J., Salamanca, C., and De Santiago, V. (2016). "Pupillometric-based analysis of central autonomic levels using HHT," in *Proccedings of 2016 International Conference on Mechatronics, Electronics and Automotive Engineering (ICMEAE)*, Cuernavaca, 14–19.

Schmidt, R. (1986). Multiple emitter location and signal parameter estimation. *IEEE Trans. Antennas Propag.* 34, 276–280. doi: 10.1109/tap.1986.1143830

Schwalm, M., and Jubal, E. R. (2017). Back to pupillometry: How cortical network state fluctuations tracked by pupil dynamics could explain neural signal variability in human cognitive neuroscience. *eNeuro* 4:ENEURO.0293-16.2017. doi: 10.1523/ENEURO.0293-16.2017

Shockley, K., Butwill, M., Zbilut, J. P., and Webber, C. L. Jr. (2002). Cross recurrence quantification of coupled oscillators. *Phys. Lett. A* 305, 59–69. doi: 10.1016/s0375-9601(02)01411-1

Villalobos-Castaldi, F. M., Ruiz-Pinales, J., Kemper, N. C., and Flores, M. (2016). Biomedical signal processing and control time-frequency analysis of spontaneous pupillary oscillation signals using the Hilbert-Huang transform. *Biomed. Signal Process Control* 30, 106–116. doi: 10.1016/j.bspc.2016.06.002

Wang, C. A., and Munoz, D. P. (2015). A circuit for pupil orienting responses: implications for cognitive modulation of pupil size. *Curr. Opin. Neurobiol.* 33, 134–140. doi: 10.1016/j.conb.2015.03.018

Webber, C. L., and Marwan, N. (2015). *Recurrence Quantification Analysis: Theory and best practices*, eds L. W. Charles Jr. and M. Norbert (Berlin: Springer). doi: 10.1016/j.conb.2015.03.018

Wilhelm, B., Körner, A., Heldmaier, K., Moll, K., Wilhelm, H., and Lüdtke, H. (2001). Normwerte des pupillographischen schafrigkeitstests fur frauen und manner zwischen 20 und 60 jahren. *Somnologie* 5, 115–120. doi: 10.1046/j.1439-054x.2001.01156.x

Wilhelm, H., Wilhelm, B., Stuiber, G., and Holger, L. (2014). The effect of caffeine on spontaneous pupillary oscillations. *Ophthal. Physiol. Opt.* 34, 73–81. doi: 10.1111/opo.12094

Zénon, A. (2017). Time-domain analysis for extracting fast-paced pupil responses. *Sci. Rep.* 7:41484. doi: 10.1038/srep41484

Zénon, A., Sidibé, M., and Olivier, E. (2014). Pupil size variations correlate with physical effort perception. *Front. Behav. Neurosci.* 8:286. doi: 10.3389/fnbeh.2014.00286

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Computational Neuroscience

ORIGINAL RESEARCH
published: 28 May 2019
doi: 10.3389/fncom.2019.00034

Bio-Inspired Presentation Attack Detection for Face Biometrics

Aristeidis Tsitiridis, Cristina Conde, Beatriz Gomez Ayllon and Enrique Cabello**

Computer Science and Statistics, King Juan Carlos University, Móstoles, Spain

Today, face biometric systems are becoming widely accepted as a standard method for identity authentication in many security settings. For example, their deployment in automated border control gates plays a crucial role in accurate document authentication and reduced traveler flow rates in congested border zones. The proliferation of such systems is further spurred by the advent of portable devices. On the one hand, modern smartphone and tablet cameras have in-built user authentication applications while on the other hand, their displays are being consistently exploited for face spoofing. Similar to biometric systems of other physiological biometric identifiers, face biometric systems have their own unique set of potential vulnerabilities. In this work, these vulnerabilities (presentation attacks) are being explored via a biologically-inspired presentation attack detection model which is termed "BIOPAD." Our model employs Gabor features in a feedforward hierarchical structure of layers that progressively process and train from visual information of people's faces, along with their presentation attacks, in the visible and near-infrared spectral regions. BIOPAD's performance is directly compared with other popular biologically-inspired layered models such as the "Hierarchical Model And X" (HMAX) that applies similar handcrafted features, and Convolutional Neural Networks (CNN) that discover low-level features through stochastic descent training. BIOPAD shows superior performance to both HMAX and CNN in all of the three presentation attack databases examined and these results were consistent in two different classifiers (Support Vector Machine and *k*-nearest neighbor). In certain cases, our findings have shown that BIOPAD can produce authentication rates with 99% accuracy. Finally, we further introduce a new presentation attack database with visible and near-infrared information for direct comparisons. Overall, BIOPAD's operation, which is to fuse information from different spectral bands at both feature and score levels for the purpose of face presentation attack detection, has never been attempted before with a biologically-inspired algorithm. Obtained detection rates are promising and confirm that near-infrared visual information significantly assists in overcoming presentation attacks.

Keywords: face biometrics, presentation attack detection, anti-spoofing, multiple sensor fusion, biologically-inspired biometrics

OPEN ACCESS

Edited by:
Hagit Hel-Or,
University of Haifa, Israel

Reviewed by:
Alejandro Linares-Barranco,
University of Seville, Spain
Manuel Jesus Dominguez-Morales,
University of Seville, Spain

****Correspondence:***
Aristeidis Tsitiridis
aristeidis.tsitiridis@urjc.es
Enrique Cabello
enrique.cabello@urjc.es

Received: *17 July 2018*
Accepted: *09 May 2019*
Published: *28 May 2019*

Citation:
Tsitiridis A, Conde C, Gomez Ayllon B and Cabello E (2019) Bio-Inspired Presentation Attack Detection for Face Biometrics.
Front. Comput. Neurosci. 13:34.
doi: 10.3389/fncom.2019.00034

INTRODUCTION

Biometrics have a long history of existence and usage in various security environments. Modern biometric systems utilize a variety of physiological characteristics also known as "biological identifiers." For example, non-intrusive biometric patterns extracted from a finger, palm, iris, voice, gait (and their fusion in multimodal biometric systems), can provide a wealth of identity

information about a person. Face biometrics in particular, pose a challenging practical problem in computer vision due to dynamic changes in their settings such as fluctuations in illumination, pose, facial expressions, aging, clothing accessories, and other facial feature changes such as tattoos, scars, wrinkles and piercings. The main advantage of face biometric applications is that they can be deployed in diverse environments at low cost (in many cases, a simple RGB camera is sufficient) without necessitating substantial participation and inconvenience from the public. Public acceptance of face biometrics is also the highest amongst all other biological identifiers. Modern day applications making extensive use of face biometric systems include, mobile phone authentication, border or customs control, visual surveillance, police work, and human-computer interaction. Regardless of the numerous practical challenges in this field, face biometrics still remain a heavily researched topic in security systems.

Face biometric systems are susceptible to intentional changes in facial appearance or falsification of photos in official documents known as, "presentation attacks." For example, impostors may acquire a high quality face image of an individual and manipulate it either printed on paper, on a mask or even on a smartphone display to deceive security camera checkpoints. The significant reduction in high-definition portable camera size also means that impostors have easy access to tiny digital cameras that discretely or secretively capture face images of unsuspecting individuals. Moreover, with the vast online availability of face images in public or social media, it is relatively easy to acquire and reproduce a person's image without their consent. "Presentation Attack Detection (PAD)" or less formally known "anti-spoofing," engulfs the detection of all spoofing attempts made on biometric systems. Therefore, accurate and fast PAD is an important problem for authentication systems across many platforms and applications (Galbally et al., 2015) in the fight against malicious security system attacks. Basic face presentation attacks often are: (a) printed face on a paper sheet. Sometimes a printed face is shown with eyes cropped out so that the impostor's eyes blink underneath. (b) Digital face displayed on a screen from digital devices such as tablets, smartphones, and laptops. This kind of face presentation attacks can be static or video. In video attacks facial movements, eye blinking, mouth/lip movements or expressions are usually simulated through a short video sequence. (c) A 3D mask (paper, silicon, cast, rubber etc.) specifically molded for a targeted face. In addition, impostors may also try identity spoofing by using more sophisticated appearance alteration techniques or their combinations: (1) Glasses corrective or otherwise and/or contact lenses with possible color change. (2) Hairstyle, change in color, cut/trim, hair extensions etc. (3) Make-up or fake facial scars. (4) Real and/or fake facial hair. (5) Facial prosthetics and/or plastic surgery.

Presentation attacks in images can be detected from anomalies in image characteristics such as liveness, reflectance, texture, quality, and spectral information. Sensor-based approaches are considered efficient strategies to investigate such image characteristics and naturally involve the usage (and fusion) of various camera sensors that capture minute discrepancies. A sensor-based method that uses a light field camera sensor with 26 different focus measures together with image descriptors (Raghavendra et al., 2015) reported promising PAD scores. With the aid of infrared sensors authors in Prokoski and Riedel (2002) analyzed facial thermograms for rapid, and varied illumination environments. Similar thermography methods were presented in Hermosilla et al. (2012) and Seal et al. (2013). Motion-based techniques are mostly employed in video sequences to detect motion anomalies between frames. Some representative methods of this type of PAD algorithms used Eulerian Video Motion Magnification (Wu et al., 2012), Optical Flow (Anjos et al., 2014), and non-rigid motion with face-background fusion analysis (Yan et al., 2012). Liveness-based approaches extract image features that focus on the liveness phenomena of a particular subject. Using this approach, algorithms scan liveness patterns in certain facial parts such as facial expressions, mouth or head movements, eye blinking, and facial vein maps (Pan et al., 2008; Chakraborty and Das, 2014). Texture based methods investigate texture, structure and overall shape information of faces. In conventional terms, commonly used texture-based methods rely on Local Binary Patterns (Maatta et al., 2011; Chingovska et al., 2012; Kose et al., 2015), Difference of Gaussians (Zhang et al., 2012) and Fourier frequency analysis (Li et al., 2004). For quality characteristics, a notable image quality method in Galbally et al. (2014) proposed 25 different image quality metrics as extracted between real and fake images in order to train classifiers which are then used for the detection of potential attacks.

In today's society, face perception is extremely important. In the distant past, our very survival in the wild depended on our ability to collaborate collectively as species. As a consequence, the human brain over the millennia has evolved to perform facial perception in an effortless, rapid and efficient manner (Ramon et al., 2011). The ever increasing requirements in complexity, power and processing speed, have motivated the biometric research community to explore new ways of optimizing facial biometric systems. Therefore, it should not come as a surprise that biology has recently become a valuable source of inspiration for fast, power efficient and alternative methods (Meyers and Wolf, 2008; Wang et al., 2013).

The fundamental biologically-motivated vision architecture consists of alternating hierarchical layers mimicking the early processing stages of the primary visual cortex (Hubel and Wiesel, 1967). It is established from past research that as visual stimuli are transmitted up the cortical layers (from V1–V4), visual information progressively exhibits a combination of selectivity and invariance to object translations such as size, position, rotation, depth etc. In the past, there have been many vision models and variants inspired from this approach such as the "Neocognitron" (Fukushima et al., 1980), "Convolutional neural network" (LeCun et al., 1998), and "Hierarchical model and X" (Riesenhuber and Poggio, 2000). Over the years, these models have performed incredibly well in many object perception tasks and today are recognized as equal alternatives to statistical techniques. In face perception, biologically-inspired methodologies have been applied successfully for some years and have proven reliable as well as accurate (Lyons et al., 1998; Wang and Chua, 2005; Perlibakas, 2006; Rose, 2006; Meyers and Wolf,

2008; Pisharady and Martin, 2012; Li et al., 2013; Slavkovic et al., 2013; Wang et al., 2013).

There are many common characteristics in biologically-motivated algorithms and perhaps the most important aspect is the extensive use of texture-based features in either 2D or 3D images. Reasons for designing a biologically-inspired model would be its projected efficiency, parallelization and speed in extremely demanding biometric situations. Contemporary state-of-the-art methods are efficient in selected environments with high availability of data but sifting each frame with laborious and lengthy CNN training, sliding windows or pixel-by-pixel approaches requires an incredible amount of available resources such as storage capacity, processing speed and power. Nevertheless, biologically-inspired systems have almost entirely been expressed by deep learning CNN architectures. In Lakshminarayana et al. (2017), spatio-temporal mappings of faces extraction is followed by a CNN schema, and discriminative features for liveness detection were subsequently acquired. This approach produced impressive results on the databases examined but their setup relied solely on video sequences which penalize processing speed and are not always available in the real world, especially in border control areas where a single image should suffice. Other CNN models (Alotaibi and Mahmood, 2017; Atoum et al., 2017; Wang et al., 2017) explored depth perception prior to application of a CNN that distinguished original vs. impostor access attempts. In Alotaibi and Mahmood (2017), depth information was produced with a non-linear diffusion method based on an additive operator splitting scheme. Even though only a single image was required in this work, the use of only one database (and the high error rates in the Replay-Attack database) did not entirely reveal the potential of this approach. Another CNN approach was presented in Atoum et al. (2017) where a two-stream CNN setup for face anti-spoofing was employed by extracting local image features and holistic depth maps from face frames of video sequences. Experimentation with this CNN setup showed reliable results with a significant cost on practicality i.e., training two separate CNNs along with all intermediate processing steps. In Wang et al. (2017), a representation joining together 2D textual information and depth information for face anti-spoofing was presented. Texture features were learned from facial image regions using a CNN and face depth representation was extracted from Kinect images. The high error rates and limited experimentation procedure made their findings rather questionable. Finally, in Liu et al. (2018) a CNN-RNN (Recursive Neural Network) model was used to acquire face depth information with pixel-wise supervision, by estimating remote photoplethysmography signals together with sequence-wise supervision. The accuracy of this method relied heavily on the number of frames per video which makes this approach computationally heavy.

Overall, Convolutional Neural Network approaches and the manner in which they are executed or accelerated in hardware is a big subject of debate in our world today. They require large amounts of resources in hardware, software and energy to be effectively trained. However, since end-users have different hardware/software configurations, no particular effort was given to hardware optimization or software acceleration. The investigation of a biologically-inspired PAD secure system was developed as part of two funded projects, the European project ABC4EU and the Spanish national project BIOINPAD. End-users in both projects (i.e., the Spanish national police, Estonian police, Rumanian Border Guard) were interested in a new approach to the PAD problem.

Over the years, bio-inspired systems have received significant interest from the computer vision community because their solutions can relate to real-world human experiences. Thus, the main research contribution of this work has been the introduction of a system that handles video presentation attack detection from a biologically-inspired perspective. A system that has a straightforward and simple architecture able to cope with visual information from a single frame at high precision rates. Our design focus has been the development of a bio-inspired system with a clear structure and relatively little effort. In addition, this paper summarizes precision rate results obtained during our research and compares them against other known models to enhance the comparative scope and understanding. The system has been evaluated with different databases in the visible, and near-infrared (and their fusion) spectral regions. This is illustrated over several sections of this article which is organized in the following way. In section Methodology and BIOPAD's structure, definitions and methodology that have led us to the development of the BIOPAD model are discussed, followed by a detailed explanation of the model's structure. Furthermore, in that section, we demonstrate the biologically-inspired techniques used, the model's general layout, and individual layer functionality. Section Experiments describes all databases used (section Databases), explains our biometric evaluation procedures (section Presentation attack results) and analyses all experiments conducted for the BIOPAD, Hierarchical Model And X (HMAX) and CNN (AlexNet) models. Section Experiments is further divided into visible (section Visible spectrum experiments) and near-infrared (section Near-infrared experiments and cross-spectral fusion) experiments for a better comparison between the two approaches explored. Finally, the last section summarizes all of our conclusions in this research work.

METHODOLOGY AND BIOPAD'S STRUCTURE

In the first part of this section, the overall layered structure is described, followed by the biologically-inspired concepts that have been used as core mechanisms in BIOPAD. In the last section, each layer is individually explored, along a full explanation of its operation in a pseudo-like manner.

Center-Surround and Infrared Channels

Mammals perceive incoming photons through the retina in their eyes. The number of individual photoreceptors in the retina of the human eye varies from person to person and in the same person from time to time, but on average each eye consists of ∼5 million cones, 120 million rods and 100 thousand photosensitive retinal ganglion cells (Goldstein, 2010).

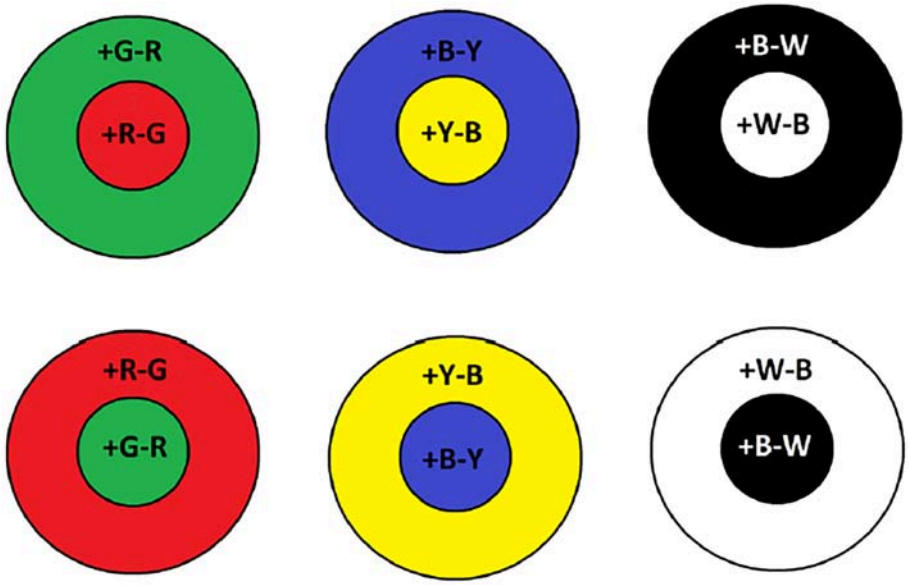


FIGURE 1 | Examples of on-center and off-center receptive fields for color opponency channels. Plus sign indicates whether the particular color is on and the minus off.

In the human retinae, rod photoreceptors peak at ∼500 nm, they are slow response receptors, come in small numbers, possess large receptive fields, and are suitable for dark environments i.e., night time. However, cone receptive fields are narrower and are tuned to different wavelengths of light. They are considerably greater in numbers than rods and hence, are responsible for visual acuity. Bipolar retinal cells bear the task of unifying incoming visual information from cones and rods (Engel et al., 1997). Furthermore, on-center and off-center bipolar cells operate in a center-surround process between red-green and blue-yellow wavelengths. For example, on-center Green-Red (RG) bipolar cells are going to maximally respond when red hits the center of their receptive field only and are inhibited when green is at their surrounding region. Vice versa, this operation is reversed for an off-center RG bipolar cell where excitation only occurs when the detectable green wavelength is incident in the surrounding region. As shown in **Figure 1**, this can be further applied for the blue-yellow and lightness channels. The color opponent space is defined by the following equations (Van De Sande et al., 2010):

$$O1 = (R - G)/\sqrt{2} \quad (1)$$

$$O2 = (R + G - 2B)/\sqrt{6} \quad (2)$$

$$O3 = (R + G + B)/\sqrt{3} \quad (3)$$

The O3 opponent channel is the intensity channel and color information is conveyed by channels O1 and O2. In BIOPAD, when the input image is in RGB, all three opponent channels are processed simultaneously and in order to make use of the available infrared information, an additional channel NIR is added in the fourth channel dimension.

The use of infrared or thermal imaging alongside the visible spectrum, has been the subject of investigation many times in the past (Kong et al., 2005) and Gabor filters with near-infrared data have been applied together with computer vision algorithms (Prokoski and Riedel, 2002; Singh et al., 2009; Zhang et al., 2010; Chen and Ross, 2013; Shoja Ghiass et al., 2014). However, the use of infrared spectra in presentation attack detection using a biologically-motivated model, to our knowledge, is a first with this research work.

The actual infrared range of wavelengths can be huge, spanning from 7 microns all the way up to 300 microns and generally these bands, are undetectable to the human eye. However, there is evidence that infrared wavelengths up to 10 microns under certain circumstances are detectable by humans as visible light (Palczewska et al., 2014). From a biological perspective, the exact mechanism of near-infrared perception in the visual cortex is unknown. In BIOPAD and at low feature level, it is treated as an additional channel input from the retina, with a range of normalized pixel values as provided by the sensor (**Figure 2**). Infrared data acquisition and sensor information is shown in section Presentation attack results.

Area V1—Edge Detection

As visual signals travel to the primary visual cortex through the lateral geniculate nucleus, area V1 orientation selective simple cells process incoming information (Hubel and Wiesel, 1967) from the retinae and perform basic edge detection operations for all subsequent visual tasks. They serve as the building block units of biological vision. It is already well-established from literature that orientation selectivity in V1 simple cells can be precisely matched by Gabor filters (Marcelja, 1980; Daugman, 1985; Webster and De Valois, 1985).

A Gabor filter is a linear filter which is defined as the product of a sinusoid with a 2D Gaussian envelope and for values in pixel

coordinates *(x, y)*, it is expressed as:

$$G(x,y) = \exp\left(-\frac{X^2+\gamma^2Y^2}{2\sigma^2}\right)\cos\left(\frac{2\pi}{\lambda}\right)X \quad (4)$$

$$X = x\cos\theta - y\sin\theta \quad (5)$$

$$Y = -x\sin\theta + y\cos\theta \quad (6)$$

In Equation 5, γ is the aspect ratio and in this work is set to 0.3. Parameter λ is known as the wavelength of the cosine factor and together with the effective width, parameter σ, specify the spatial tuning accuracy of the Gabor filter. Ideally, to optimize the extraction of contour features from V1 units for a particular set of objects, some form of learning is necessary to isolate an optimum range of filters. However, this process adds complexity and it is time-consuming since it requires a huge number of samples, as experiments on convolutional neural networks have shown in literature. In order to avoid this step, Gabor filter parameters are hardcoded directly into our model following parameterization sets that have been identified from past studies. Two different parameterization settings have been considered (Serre and Riesenhuber, 2004; Lei et al., 2007; Serrano et al., 2011). Our preliminary experiments have shown that the two particular Gabor filter parameterization ranges, have no noticeable effect on PAD results. Thus, we chose the parameterization values given (Serrano et al., 2011).

Additionally, it is known that V1 cell receptive field sizes vary considerably (McAdams and Reid, 2005; Rust et al., 2005; Serre et al., 2007) to provide a range of thin to coarse spatial frequencies. Similarly, four different receptive field sizes were used here with pixel dimensions 3 × 3, 5 × 5, 7 × 7, and 9 × 9. Coarser features are handled by area V2, explained in the next section.

Area V2—Texture Features

In general, the significance of textural information is sometimes neglected or even downplayed in past biologically-inspired vision models. In face biometrics, as explained previously in the introductory section, there is a long list of texture-based presentation attack detection models and texture information is considered a crucial feature against attacks.

The role of cortical area V2 in basic shape and texture perception is essential. V2 cells share many of the edge properties found in V1. Nevertheless, V2 cell selectivity has broader receptive fields and is attuned to more complex features compared with V1 cells (Hegdé and Van Essen, 2000; Schmid et al., 2014). In addition to broader spatial features, this layer processes textural information and is therefore capable of expressing the different nature of surfaces. This is a crucial advantage in face presentation attack detection where there is a wealth of information hidden within the texture of faces, facial features or face attacks. For example, texture of beards, skin, and glasses can prove a valuable feature against spoofing attacks mimicking their nature.

V2 cells are effectively expressed by a sinusoidal grating cell operator though other shape characteristics also correspond well (Hegdé and Van Essen, 2000). The grating cell operator has not only shown great biological plausibility with respect to actual V2 texture processes but has also proven superior to Gabor filters in texture related tasks (Grigorescu et al., 2002). Its response is relatively weak to single bars but in contrast, it responds maximally to periodic patterns.

The approach used here (Petkov and Kruizinga, 1997) consists of two stages. In the first stage grating subunits generate on-center and off-center cells responding to periodicity much like retina cells. In the following stage, grating cell responses of a particular orientation and periodicity are added together, a process also known in neurons as spatial summation (Movshon et al., 1978).

A certain response *Gr* of a grating subunit at position *(x, y),* with orientation θ and periodicity λ is given by Petkov and Kruizinga (1997):

$$\mathrm{Gr}(x,y)_{\theta,\lambda} = \begin{cases} 1, \ if\ \forall\, n,\ \mathrm{M}(\mathbf{x},\mathbf{y})_{\theta,\lambda,n} \geq \rho \mathrm{M}(\mathbf{x},\mathbf{y})_{\theta,\lambda} \\ 0, \ if\ \exists\, n,\ \mathrm{M}(\mathbf{x},\mathbf{y})_{\theta,\lambda,n} < \rho \mathrm{M}(\mathbf{x},\mathbf{y})_{\theta,\lambda} \end{cases} \quad (7)$$

where $n \in \{-3 \ldots 2\}$, ρ is the threshold parameter between 0 and 1 (typically 0.9). The maximum activities of M at a given location *(x, y)* and for a particular selection of θ, λ, n, are calculated as followed (Petkov and Kruizinga, 1997):

$$M(\mathbf{x},\mathbf{y})_{\theta,\lambda,n} = \mathbf{max}\begin{cases} \mathbf{s}(\mathbf{x}',\mathbf{y}')_{\theta,\lambda,\varphi_\mathbf{n}} \mid \\ \mathbf{n}\frac{\lambda}{2}\boldsymbol{cos}\theta \leq \mathbf{x}'-\mathbf{x} < (\mathbf{n}+\mathbf{1})\frac{\lambda}{2}\cos\theta \\ \mathbf{n}\frac{\lambda}{2}\boldsymbol{sin}\theta \leq \mathbf{y}'-\mathbf{y} < (\mathbf{n}+\mathbf{1})\frac{\lambda}{2}\sin\theta \end{cases} \quad (8)$$

$$\phi_n = \begin{cases} 0, \ n = -3, -1, 1 \\ \pi, \ n = -2, 0, 2 \end{cases} \quad (9)$$

and

$$\mathrm{M}(\mathbf{x},\mathbf{y})_{\theta,\lambda,n} = \mathbf{max}\left(\mathrm{M}(\mathbf{x},\mathbf{y})_{\theta,\lambda,n}\right) \quad (10)$$

The responses at $M(x, y)_{\theta,\lambda,n}$ in Equation 9, are simple cell responses with symmetric receptive fields along a line segment 3λ. Essentially this means that there are three peak responses for each grating subunit at point *(x, y)* at a given orientation θ. This line segment is split in $\lambda/2$ intervals. The particular position of each interval defines the response of on-center and off-center cells. In other words, a grating cell subunit is maximally activated when on-center and off-center cells of the same orientation and spatial frequency are activated at point *(x, y)*. In Equation 10, ϕ_n is the phase offset and for values between 0 and π, it corresponds to symmetric center-on and center-off operations, respectively.

In the second part of V2 grating cell design, a response w of grating cell centered on *(x, y)* along orientation θ and periodicity λ, is the weighted summation of grating subunits with orientations θ and $\theta + \pi$, as given below:

$$\boldsymbol{w}(\mathbf{x},\mathbf{y})_{\lambda,\theta} = \int \exp\left(-\frac{(\boldsymbol{x}-\boldsymbol{x}')^2+(\boldsymbol{y}-\boldsymbol{y}')^2}{\mathbf{2}(\beta\sigma)^2}\right) \left(\mathrm{Gr}(x',y')_{\theta,\lambda}+\mathrm{Gr}(x,y)_{\theta+\pi,\lambda}\right)\boldsymbol{dx}'\boldsymbol{dy}', \ \theta\in[\mathbf{0},\pi) \quad (11)$$

Parameter β is the summation area size with a typical value of 5. In our experiments the number of simple cells were empirically chosen at 3 and all other parameter values were set at default values according to Petkov and Kruizinga (1997).

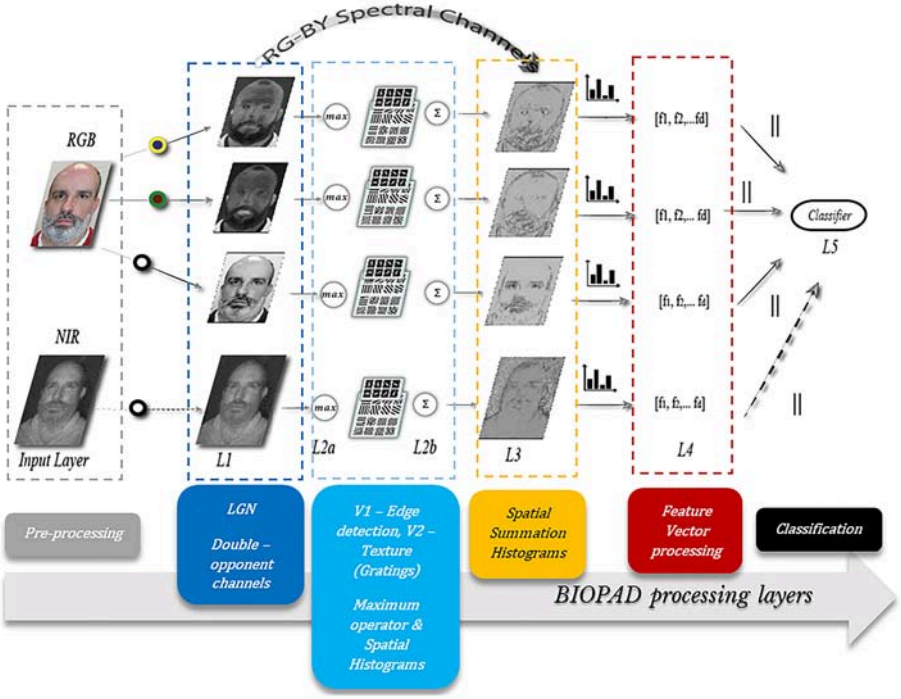


FIGURE 2 | The proposed model structure. Several layers L1 to L5 progressively process spatial and spectral facial features. All participants gave written informed consent for the publication of this manuscript.

BIOPAD Structure

Light waves are being continuously perceived by our eyes and every generated electrical impulse passes via the lateral geniculate nucleus of our brain to arrive at the first neurons in the striate cortex (Hubel and Wiesel, 1967). Countless neurons organized in progressive layers then process this information through cascades of cerebral layer modules each intended for a specific operation. Broadly, visual areas in the human brain after visual area V2 follow the dorsal and ventral visual pathways, the "where" and "what" pathways (Schneider, 1969; Ungerleider and Mishkin, 1982). The two streams are layers along two distinct cerebral paths that localize and analyse meaningful information in constant neuronal communication.

BIOPAD's structure mimics the basic visual areas V1 and V2 in the primary visual cortex in a bottom-up fashion (**Figure 2**). Its operation relies on the early stages of biological visual cognition, without any external biases or influences. The design successively processes extracted biologically-inspired features reducing their dimensionality to an extent that they can be used with classifiers that determine original from fake access attempts. Furthermore, through successive biologically-motivated filtering BIOPAD's main strength lies in its ability to transform extracted features into higher dimensional vectors in a simple way that maximizes the separation between them. For example, an important difference between BIOPAD and HMAX is that the latter model's main focus is view-invariant representation of objects irrespective of their size, position, rotation and illumination. Conversely, BIOPAD's purpose is the detection of face spoofing attempts and to this end, invariance properties such as size and position could be valuable with future extensions. Even though invariance properties are generally meaningful in face recognition (Yokono and Poggio, 2004; Perlibakas, 2006; Rolls, 2012), in this particular scenario of face presentation attack detection they add unnecessary complexity or processing delays and are therefore not explored further. More specifically, BIOPAD's proposed structure is separated in the following layers (**Figure 2**):

Input Layer: The purpose of the input layer is to prepare image information by scaling down all input RGB images to a minimum of 300 pixels for the shortest edge in order to preserve the image's aspect ratio. This particular image size was chosen as a good compromise between speed/time and computational cost.

Layer L1: This layer plays the role of the lateral geniculate nucleus and separates visual stimuli in the appropriate double-opponency channels (bipolar cells) as given in section Area V1—Edge detection while scaling all pixel values to the same range between 0 and 1.

Layer L2a: Gabor filter operations perform edge detection according to parameterization values given in section Area V2—Texture features producing feature maps for each channel. It is important to note that after obtaining filtered outputs from all Gabor filters (in total 192) for each double-opponency channel, a maximum operator is applied so that a particular maximum response of L2a vectors $(x_1 \ \ldots \ x_m)$ in a neighborhood j is

given by:

$$\mathbf{r} = \mathbf{arg}\max_{\mathbf{j}}(\mathbf{x_j}) \tag{12}$$

The maximum operator is a well-known non-linear biological property exhibited by certain visual cells at low levels of visual cognition that assists in pooling visual inputs from previous layers (Riesenhuber and Poggio, 1999; Lampl et al., 2004) to greater receptive fields. This hierarchical process gradually projects meaningful visuospatial information to higher cortical layers in the mammalian brain (**Figures 3a,b**).

Layer L2b: In this layer grating cell operations are performed according to the settings given in section BIOPAD structure. Subsequently, grating outputs are spatially summed with outputs from L2a, in order to form a single L2 output for each of the three double-opponency channels. Spatial summation is another property of the visual cortex and like the maximum operator it is intended to linearly combine presynaptic inputs into outputs for higher layers (Movshon et al., 1978). Spatial summation is used in this layer in order to preserve the spatial integrity and sensitive texture information in faces (**Figure 3c**).

Layer L3: The three double-opponency channels after spatial summation (**Figure 3d**), contain both edge and texture features. The information of these channels along with the RG-BY spectral channels from L1 that contain the spectral differences of each image, are aggregated into spatial histograms with a window size of 20 units and bin size of 10. These values were empirically selected after experimentation as ideal for the particular layer dimensions. These spatial histograms have been used before in the context of face recognition but with lower level features at L1 (Zhang et al., 2005). Here, they are employed at an intermediate level of feature processing and with various types of biological-like features. It is further important to note here that since all these spatio-spectral channels carry different types of visual information, they are never mixed together.

Layer L4: In this layer all L3 information from the previous layer is simply concatenated and sorted in a multidimensional vector for either the training or testing phase, without any further processing. Vector dimensions vary according to the size of the dataset and choice of parameters within the model. For example, if from the previous L3 settings spatial histograms are performed over larger regions or if the input image layer of the image is set to smaller dimensions (for faster processing speeds), then the total number of vectors extracted will be smaller. Moreover, if the total number of images in the dataset changes, so does the vector dimension size, i.e., $m_d \times n_p$, where m are the vectors extracted from previous layers with length d and n are the columns of vectors per image p.

Layer L5: Supervised classification takes place in this layer and any classifiers used can be trained with the extracted feature vector from L4. Training data are selected by following the 10-fold cross-validation technique. The supervised classifiers chosen for this work were a Support Vector Machine (SVM) with a linear kernel and k-Nearest Neighbor (KNN) with Euclidean distance.

BIOPAD's overall operation is further demonstrated with a pseudo-code approach below:

RGB Data Setup

Each PAD database consists of single **RGB** frame samples for a particular person's authentic video sequence and their presentation attacks. The PAD image database is then split in 70% training samples (T_r) 30% samples for testing (T_s) with cross-validation in *10-folds*.

if *RGB case train* ***then***,
for each *random* T_r *sample of each fold* ***do,***

(1) **Input:** Load a $m \times n$ T_r sample and scale to *300 pixels* for the shortest edge.

(2) **Center-surround:** Convert **RGB** space to **O1, O2, O3** channel opponent space using Equations (2–4) thus obtain **opponency** frame O_r of the same dimensions.
for *each opponency channel O1 (red –green differences), O2 (blue–yellow) and O3(lightness)* ***do***,

(3) **Process V1:** Load *3x3, 5x5, 7x7, 9x9* ***Gabor filters*** (G_f) parameterised with σ *=1, and* $\lambda = 4, 5.6, 7.9, 11.31, 15.99, 22.61$ in total *192* filters **then**.
- $L1_{Tr} = O_r \cdot G_f$, where $L1_{Tr}$ is a multidimensional array of $m \times n \times 192$ convolved versions of the T_r frame with V1-Gabor like filters.
- Extract the maximum response using Equation (12) at every position along the dimension of convolutions to obtain a new matrix $L1_M$
- Normalize $L1_M$ with zero mean and unit variance.

(4) **Process V2**: Load grating filters (G_r) using $\theta = 0$–360° in 45° steps, $\lambda = 5.42$, $\rho = 0.9$, and $\beta = 5$.
- $L2_{Tr} = O_r \cdot G_r$, where $L2_{Tr}$ is a multidimensional array of $m \times n \times \theta$ convolved versions of the T_r frame with V2 -grating filters.
- Extract the maximum response using Equations (10–12) at every position along the dimension of convolutions to obtain a new matrix $L2_M$.
- Normalize $L2_M$ with zero mean and unit variance.

(5) **Spatial summation** of $L1_M$ *and* $L2_M$ features yielding an array of the same size as the input.

(6) **Spatial histograms** on summation output from step *5*, with a fixed window size of *20x20* L3 units and bin size of *10*, then concatenate histograms into a column of *5920* L4 vectors for each sample

(7) Train classifier after all T_r have been processed through steps (1–6).

else if *RGB case test* ***then***,
for each *random* T_S *sample of each fold* ***do,***

repeat steps (1-6) as above and use *5920* column vectors of T_S to extract predictions from the trained classifier

RGB and NIR Data Setup

The FRAV database consists of **RGB** and **NIR** single samples for a particular person's authentic video sequence and their presentation attacks. The PAD image database is then split in 70% training samples (T_r) 30% samples for testing (T_s) with cross-validation in *10-folds,* maintaining RGB and NIR original sample ratios.

if *RGB and NIR case train* ***then,***
for each *random* T_r *sample of each fold,* ***do***
repeat steps (1-2) and (3-6). At L1 for each opponency channel O1 (red –green differences), O2 (blue – yellow), O3(lightness), NIR (near-infrared) extract *7100* L4 column vectors for each T_r sample during classifier training.
else if *RGB and NIR case test* ***then***,
for each *random* T_S *sample of each fold* ***do,***
repeat steps (1-2) and (3-6). At L1 for each opponency channel O1 (red –green differences), O2 (blue – yellow), O3(lightness), NIR (near-infrared) extract *7100* L4 column vectors of T_S for predictions obtained from the trained classifier.

EXPERIMENTS

It is important to note that in all experiments for both the genuine access and impostor attacks, only one photo per person was used from the entire video sequences. The databases employed in this work and their different spoofing attacks are explained in section Databases. Section Presentation attack results presents the obtained results in conventional biometric evaluation measures. The remaining part of this section is further divided into experiments in the visible and near-infrared spectrum. In this subsection, the different spectra are examined individually and subsequently, their cross-spectral fusion at feature, and score levels. Since our model currently does not perform any liveness detection method, successive video frames are not being considered. For the purpose of homogeneity and statistical accuracy between datasets, train and test data were divided with the cross-validation technique, bypassing the original train/test data split of some databases as has been explained in the previous section in more detail.

Databases

The Facial Recognition and Artificial Vision (FRAV) group's "attack" database addresses several critical issues compared to other available face PAD databases. The number and type of attacks can vary significantly in each facial presentation attack database and by large, databases of the past never included a large sample of known threats. In addition to the sample of individuals examined being relatively small, little attention was paid in the multitude of human characteristics often occurring within human populations e.g., beards, glasses, eye color, haircuts etc. At the same time, sensor equipment is often limited and out-dated to contemporary technology products found in the market today. These shortcomings necessitated the creation of an up-to-date PAD facial database according to ISO/IEC and ICAO standards with a larger statistical sample, multi-sensor information and inclusion of all basic attacks. This database serves as a simulation stepping stone for experimentation ahead for any real-world situation and supplements the list of existing databases found publicly. The introduction of this new database from our group offers the following main characteristics and contributions:

- The largest PAD-ready facial database to date with 185 different individuals of both genders and various age groups.
- The largest collection of sensor data aimed at PAD algorithms. Four different types of sensors namely Intel's

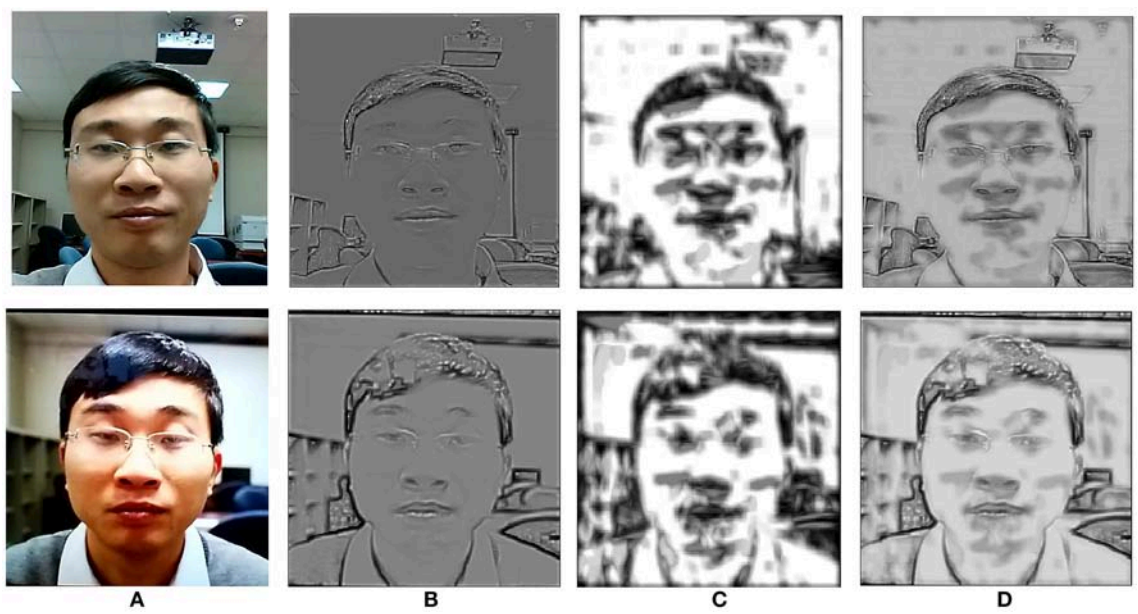


FIGURE 3 | A genuine access attempt vs. a photo-print attack. Top row shows the progressive process of a genuine photo attempt. Bottom row shows the printed photo attack. Column **(A)** shows the input layer images. Column **(B)** the L2a layer as processed from edge detection Gabor filters, column **(C)** the L2b layer processed from texture grating cells and column **(D)** the combined layers L2a and L2b after spatial summation. The richness and depth of edge-texture information in the original image (top row) is apparent. All participants gave written informed consent for the publication of this manuscript.

Realsense F200, FLIR ONE mobile phone thermal sensor, Sony A6000 ILCE-A6000 and a HIKVISION surveillance camera and therefore covering a range of spectral bands in the visible, near-infrared (at 860 nm) and infrared (800–1500 nm).

• Various spoofing attack scenarios examined, which include the following types of spoofing attacks:

1. Printed photo attacks with high resolution A4 paper.
2. Mask attacks from printed paper.
3. Mask attacks from printer paper with eye areas exposed an eye blinking effect.
4. Video attack with a tablet electronic device.
5. 3D Mask attack (to this day limited but will be expanded in the future)

Lastly, particular attention was paid at uniformly illuminating all faces using artificial lighting. Two T4 fluorescent tubes operating at 6,000 K−12 Watts each, evenly distributing multi-directional light to all subjects. **Figure 4** illustrates all of the presentation attack types explored in the FRAV "attack" database for a given subject using RGB and NIR sensor information.

The CASIA Face Anti-Spoofing (Zhang et al., 2012) database is a database from the Chinese Academy of Sciences (CASIA) Center for Biometrics and Security Research (CASIA-CBSR). This database contains videos at 10 s of real-access and spoofing attacks of 50 different subjects, divided into train and test sets with no overlap. All samples were captured with three devices at different resolutions: (a) low resolution with an old 640 × 480 webcam, (b) normal resolution with a more up-to-date 640 × 480 webcam and c) high resolution with a 1920x 1080 Sony NEX-5 camera. Three different attacks were considered, (a) warped, spoofing attacks are performed with curved copper paper hardcopies of high-resolution digital photographs from genuine users, (b) cut, attacks are performed using hardcopies of high-resolution digital photographs from genuine users, with the eye areas cut out to simulate eye blinking, c) video, genuine user videos are replayed in front of the capturing device using a tablet.

The MSU Mobile Face Spoofing Database or MFSD (Wen et al., 2015) for face spoof attacks, consists of 280 video clips of photo and video attack attempts of 35 different users. This database was produced at the Michigan State University Pattern Recognition and Image Processing (PRIP) Lab, in East Lansing, US. The MSU database has the following properties, (a) mobile phones were used to acquire both genuine faces and spoofing attacks, (b) printed photos were generated as high-definition prints and their authors claim that these have much better quality than printed photos in other databases of this kind. Two types of cameras were used in this database, (a) built-in camera in MacBook Air at a resolution of 640 × 480, and (b) front-facing camera in the Google Nexus 5 Android phone at a resolution of 720 × 480. Spoofing attacks were generated using a Canon SLR camera, recording at 18.0 M pixel photographs and 1,080

FIGURE 4 | An example of a subject from the FRAV "attack" database. Top row left to right: Genuine access RGB photo, RGB Printed photo attack, RGB printed mask attack, RGB printed mask with eyes exposed attack, RGB tablet attack. Bottom row left to right: Genuine access NIR photo, NIR printed photo attack, NIR printed mask attack, NIR printed mask with eyes exposed attack, NIR tablet attack. All participants gave written informed consent for the publication of this manuscript.

p high-definition video clips and iPhone 5S back-facing camera, recording 1,080 p video clips.

Presentation Attack Results

BIOPAD was evaluated with three different databases, FRAV-attack, CASIA, and MFSD. The main concern of our experiments was the detection success rate of spoofing attacks made by potential impostors. In simple terms, the system was required to effectively differentiate between fake and genuine access attempts. This was treated as a two-class classification problem. The applied biometric evaluation procedures are defined for the spoofing False Acceptance Rate (sFAR) and False Rejection Rate (FRR) as:

$$sFAR = \frac{Impostor\ attacks\ seen\ as\ genuine}{Total\ number\ of\ attacks} \tag{13}$$

$$FRR = \frac{Rejected\ genuine\ access\ attempts}{Total\ number\ of\ genuine\ access\ attempts} \tag{14}$$

Moreover, presentation attack detection is further presented according to SC37ISO/IEC JTC1 Biometrics (2014) with an additional measure, Average Classification Error Rate (ACER). The average of impostor attacks incorrectly classified as genuine attempts and normal presentation incorrectly classified as impostor attacks is given by:

$$ACER = \frac{sFAR + FRR}{2} \tag{15}$$

Train and test data were partitioned using the k-fold cross validation technique. All scores were obtained using 10-folds and in order to further testify performance scores, and L4 feature vectors were essentially classified using two different schemas. A Support Vector Machine (SVM) classifier with two different kernels linear, Radial Basis Function (RBF) and a k-nearest neighbor (KNN) classifier of $n = 2$ nearest neighbors with Euclidean distance as a distance measure. In reality, the number of neighbors varies according to the dataset but for the two class problem here out of all n values examined, two produced the best average on all datasets as found through cross-validation. In the beginning, BIOPAD was examined only on the RGB images of all three databases and then on both RGB/Near-Infrared (NIR) images at feature-score levels for the FRAV attack database only since infrared data is unavailable for the other databases.

Visible Spectrum Experiments

Accuracy rates are defined as the number of images for each database correctly classified as genuine or fake, i.e., true positives and true negatives. The average classification accuracy scores and standard deviation values from all trials in **Tables 1**, **2**, respectively, highlight the large differences between datasets and classifiers. From **Table 1** it can be deduced that BIOPAD analyses presentation threats better than HMAX under all of the examined databases. Depending on the choice of training and testing data as provided by cross-validation, significant deviations in results may occur. This is largely due to the relatively small sample sizes in databases, especially in CASIA and MFSD, leading to significant statistical variance. This has an obvious effect on the KNN classifier which portrays an unstable and low performance with respect to SVM. The CASIA presentation attack database produced the worst overall results in terms of PAD.

TABLE 1 | The average detection percentages (%) of 10 trials with cross-validation.

Dataset	BIOPAD			HMAX		
	SVM linear	SVM RBF	KNN	SVM linear	SVM RBF	KNN
CASIA	92.75	90.13	57.37	90.25	88.63	63.50
MFSD	97.08	86.04	82.08	90	87.08	70.42
FRAV	98.91	98.71	94.71	96.57	93.91	81.23

TABLE 2 | The average standard deviation values (σ^2) of 10 trials with cross-validation.

Dataset	BIOPAD			HMAX		
	SVM linear	SVM RBF	KNN	SVM linear	SVM RBF	KNN
CASIA	5.06	5.96	10.18	6.06	5.6	17.17
MFSD	3.82	3.68	9.97	7.84	9.86	11.23
FRAV	1.14	1.4	1.99	2.18	3.18	4.98

The highest performance has been achieved with the FRAV "attack" database closely followed by the performance achieved with the MFSD database. This is not entirely surprising since both datasets consist of good quality images and high resolution print attacks. The worst performance has been noticed when operating with CASIA photos. The total average performance from all datasets in the BIOPAD SVM linear case is at 96.24% while for HMAX at 92.27%. HMAX is not a dedicated PAD algorithm, nor has it been ever designed for such a purpose. Nevertheless, it can be seen from **Table 1** that HMAX has performed remarkably well which beyond doubt proves the adaptability and capacity that bio-inspired computer vision models have.

In **Table 2**, standard deviation values further paint a picture of relationships between models and datasets. The highest performance was observed in BIOPAD with SVM using the FRAV database and the worst in HMAX KNN using CASIA. Between them there is a sizeable difference of 16% indicating the impact of choosing a particular scenario and classifier in PAD performance. It is further noticeable from this table that BIOPAD provides a more consistent set of results with SVM linear being the overall winner in performance. The detection accuracy rates in **Table 1** provide an insight into the overall ability of the PAD model to detect spoofing attacks. From these results it is seen that the model can achieve a high detection rate at almost 99% with a consistent standard deviation value of 1.14 for the SVM linear kernel case in the FRAV database. Overall, the KNN classifier with the CASIA database has shown the worst performance. While conclusions from **Tables 1**, **2** are useful, biometric evaluation becomes more meaningful when measured in terms of sFAR and FRR which can effectively capture the nature of error.

In addition to HMAX and for a more complete comparison with BIOPAD, the selected databases were analyzed using Convolutional Neural Network. Multiple lines of research have been explored for CNN architectures in last two decades and a huge number of different methods are proposed in references (Canziani et al., 2016; Ramachandram and Taylor, 2017). In this part of the experiments, the objective is to compare the proposed bio-inspired method with a base line CNN model. The architecture selected was based on the well-known LeNet method (LeCun et al., 1998) with the improvements implemented in AlexNet (Krizhevsky et al., 2012). AlexNet has been tested for detecting presentation attacks using faces (Yang et al., 2014; Xu et al., 2016; Lucena et al., 2017). The architecture of the net is formed by eight layers, five convolutional and three fully-connected. All results provided in **Table 3** are the average of 10 trials.

Table 3 shows that error percentages are relatively small and comparable with another state-of-the-art algorithm like CNN that have been used in the past. The sFAR percentages for the CASIA and MFSD databases are comparable but there is a significant difference between the two databases in their FRR percentages. Naturally, this is also reflected onto the ACER percentages. The significant difference in FRR percentages indicates the difficulty of distinguishing attacks from genuine access attempts in the CASIA database. The error percentages for the best classifier choice (SVM linear) appear particularly improved for the FRAV attack database. In effect, this proves the importance of image quality in terms of both verification and presentation attack cases. Image quality is a consequence of various reasons and is also reflected in PAD results seen in **Table 1**. We further wanted to investigate the impact V1 and V2 edge and texture operations have on the overall performance of presentation attack detection. These tests were only performed for the SVM linear kernel case. It is worthwhile therefore to examine the separate and combined effect of V1 and V2 operations which can be seen in **Table 4** below in terms of classification percentages. PAD scores rise when V1 and V2 feature vectors are combined together and standard deviation values across all trials indicate better performance. While these values are indicative in these early stages of experimentation, a separate study on optimum parameterization for each layer may yet reveal a more important relationship between edge and texture features in presentation attack detection.

In order to better understand the intrinsic quality difference of the databases used in this work, various metrics were explored. There are numerous image quality metrics that have been developed over the years such as mean square error, maximum difference, normalized cross-correlation and peak signal-to-noise ratio amongst many others. Some of these metrics in fact have been successfully used as a separate PAD algorithm (Galbally et al., 2014). The majority of quality metrics requires the examined image to be subtracted from a reference image. This produces accurate error results only when the images are identical i.e., when the image content is identical. However, in practice face databases are a collection of images from various sensors at different angles. So in this particular case, sharpness metrics capable of measuring the content quality from a single image would be more suitable and useful. Likewise as before with quality metrics, there is a huge list of sharpness metrics being used in literature today, e.g., absolute central moment, image contrast and curvature, histogram entropy, steerable filters, energy gradients etc. An in-depth database quality analysis is beyond the scope of this work, and we have experimented with several sharpness metrics noting similar responses from all. **Table 5**, shows indicative sharpness results by using the spatial frequency quality (Eskicioglu and Fisher, 1995) metric which has been representatively chosen.

TABLE 3 | AlexNet and BIOPAD average sFAR and FRR scores over 10 trials.

Dataset	AlexNet			BIOPAD		
	sFAR	FRR	ACER	sFAR	FRR	ACER
CASIA	2.857	13.9	8.37	2.77	14.58	8.67
FRAV	2.98	17.34	10.16	0.85	2.43	1.64
MFSD	9.64	39.07	24.34	3.44	5	4.22

TABLE 4 | The average classification percentages (%) and standard deviation values of 10 trials with cross-validation for V1 and V2 operations.

Dataset	μ		σ^2	
	V1	V1 and V2	V1	V1 and V2
CASIA	90	92.75	8.6	5.06
MFSD	95.63	97.08	6.25	3.82
FRAV	97.73	98.91	2.48	1.14

It is evident from the mean values (μ) in **Table 5** that the CASIA dataset on average does not possess the high quality of spatial features seen in the MFSD and FRAV databases. Furthermore, the MFSD dataset has produced the best scores, however it should be highlighted that it does not have the same variety of presentation attacks found in the FRAV "attack" database nor the abundance of test subjects. The "Smartphone" and "Tablet" attacks are a similar type of electronic device attack and there is no provision of mask attack data. To further understand the importance of the aforementioned better, we employ the t-Distributed Stochastic Neighbor Embedding (t-SNE) (Van Der Maaten and Hinton, 2008) technique to visualize and compare presentation attacks in each dataset. L4 vectors as extracted from BIOPAD are used with t-SNE technique at "default" value settings, i.e., 30 dimensions for its principal component analysis part and 30 for the Gaussian kernel perplexity factor, and shown in **Figure 5**.

In **Figures 5A,C,E**, real access attempts vs. impostor attacks are visualized within the same space. These illustrations help understand how genuine users distance from their attacks. It can be easily observed in **Figure 5A** that for the CASIA dataset real access attempts are scattered across the same space as presentation attacks, making the classification process complex and difficult to achieve. This is also confirmed by its reduced detection rates. Different patterns are exhibited from results in **Figure 5B**, where real access attempts occupy a denser area

TABLE 5 | Direct comparison of spatial frequency quality index values for three datasets and for each of their presentation attacks.

Dataset	Printed photo	Printed mask	Printed Photo/Mask with Eye blinking	Smartphone	Tablet	Real users	μ
CASIA	0.803	–	0.8957	–	1.0221	1.094	0.9538
MFSD	2.4191	–	–	2.7054	2.9603	2.754	2.7097
FRAV	1.8275	1.6544	1.5081	–	1.4906	1.831	1.6623

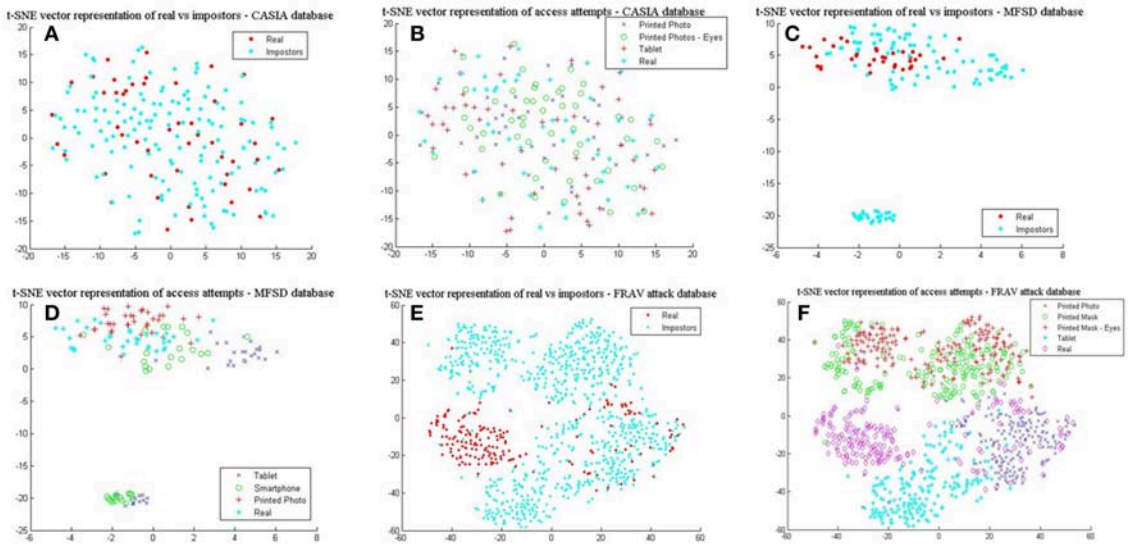


FIGURE 5 | L4 vectors visualized with t-SNE for the three datasets. **(A)** real vs. impostors–CASIA database, **(B)** presentation attacks—CASIA database, **(C)** real vs. impostors –MFSD database, **(D)** presentation attacks—MFSD database **(E)** real vs. impostors—FRAV "attack" database, and **(F)** presentation attacks—FRAV "attack" database.

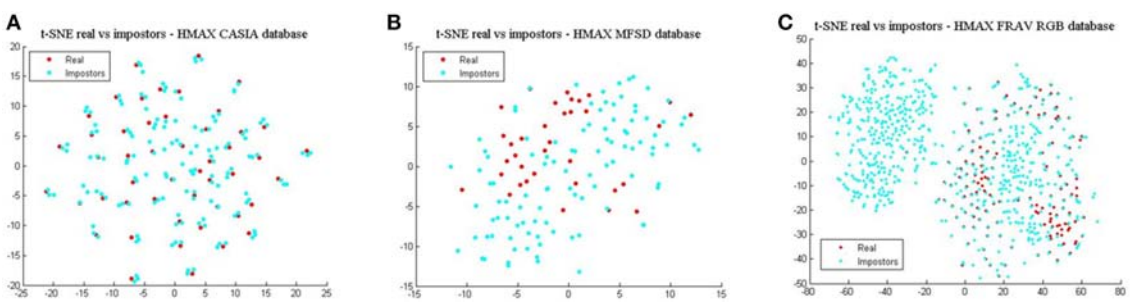


FIGURE 6 | HMAX vectors visualized with t-SNE for the three datasets in terms of real access attempts vs. impostors. **(A)** t-SNE for the CASIA dataset, **(B)** t-SNE for the MFSD dataset, and **(C)** t-SNE for the FRAV "attack" dataset.

within the impostor attack zone and finally in **Figure 5C**, in which real access attempts fall within a separate space. Looking at the presentation attack images in all datasets closely, it is not surprising to understand why these patterns occur. In **Figure 5B**, mainly due to the low image sharpness in CASIA (**Table 5**) and the nature of attack experiments, L4 vectors cover almost the same range of values and dimensional space. As the separation of presentation attacks and real access attempts improve in **Figures 5D,F** so do the results in **Table 1**. Finally, in **Figure 5F**, some real access attempts exhibit a noticeable overlap with their respective presentation attacks, particularly within the printed photo space, which is the main source of sFAR and FRR errors for the FRAV database. Arguably, the presentation attack that, in general, best matches genuine user information is the "printed photo" attack which can be efficiently faced in the NIR spectrum (section Near-infrared experiments and cross-spectral fusion).

Finally, comparing BIOPAD L4 vectors with HMAX vectors using t-SNE (**Figure 6**), it can be noted that HMAX vectors do not display the same amount of consistency in distinct areas but rather vectors from all attacks appear merged and scattered

TABLE 6 | BIOPAD detection rates and their standard deviation values over 10 trials.

Dataset	SVM linear	SVM RBF	KNN	σ2-SVM linear	σ2-SVM RBF	σ2-KNN
FRAV RGB	96.13	94.58	85.95	2.26	3.21	3.91
FRAV NIR	97.81	97.17	92.28	1.72	2.16	3.2
FRAV(RGB + NIR) Feature level	96.33	95.71	86.49	3.08	2.93	3.07
FRAV(RGB + NIR) Score level	96.97	95.87	89.11	1.99	2.68	3.55

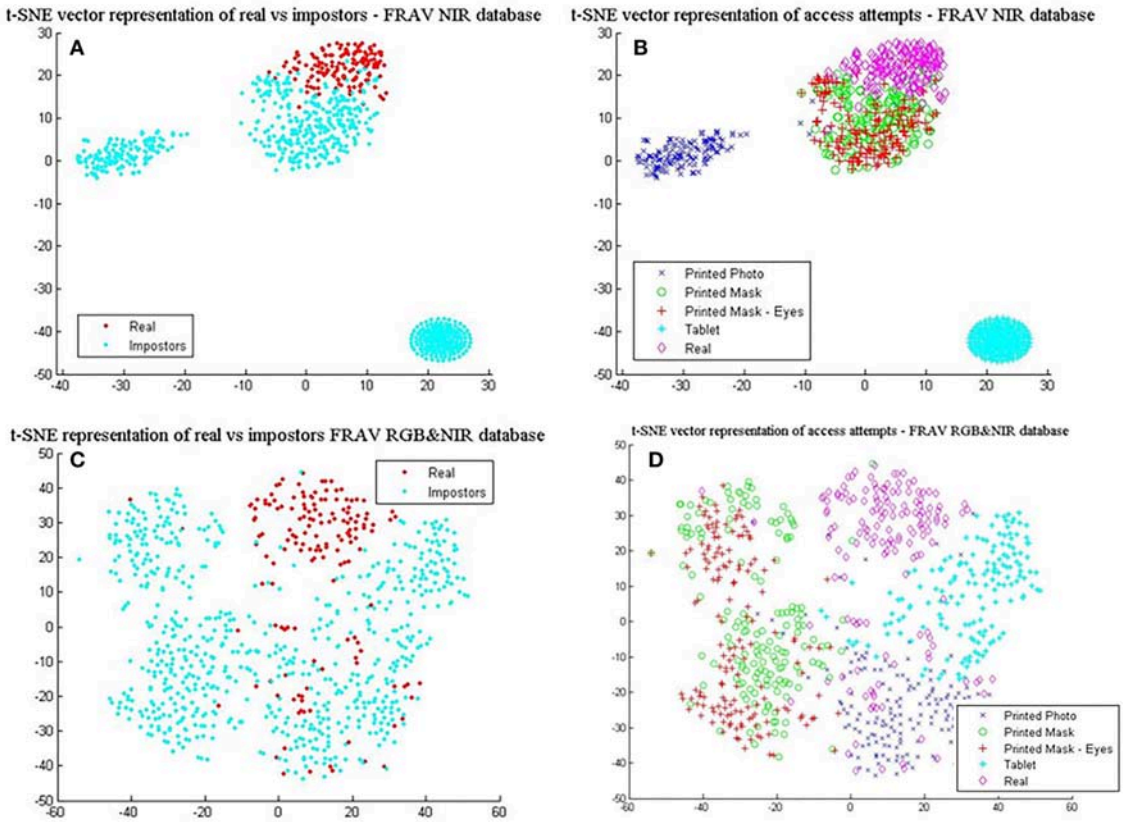


FIGURE 7 | L4 vectors visualized with t-SNE for the FRAV "attack" database and its NIR information. **(A)** real vs. impostors—FRAV "attack" database with NIR information only, and **(B)** presentation attacks—FRAV "attack" database with NIR information only, **(C)** real vs. impostors—FRAV "attack" database with RGB&NIR information fused at feature level, and **(D)** presentation attacks—FRAV "attack" database with RGB&NIR information fused at feature level.

across the same area. HMAX lack of bio-inspired features capable of processing texture and color information, leads to hardly distinguishable classes. In effect, this has a toll in presentation attack detection results (**Table 1**).

Near-Infrared Experiments and Cross-Spectral Fusion

BIOPAD experiments in the previous section have centered on the visible spectral bands and have shown great promise. Nonetheless, there were noticeable overlaps with certain presentation attacks and so we wanted to further expand BIOPAD's capacity to cope with these attacks and minimize the contribution of errors either directly from the subjects or their ambience. For this reason, our experiments in this section present a direct comparison between the performance for each spectral band, then their fusion at feature and score levels i.e., fusion between the visible and NIR band. At feature level, NIR is treated like an additional channel (**Figure 2**) and L4 vectors from all bands are equally processed in the model. Conversely, at score level visible—NIR bands are processed and classified separately. However, after classification, vectors for each subject are examined over all trials using the weighted sum score level fusion technique in order make a decision on whether the subject is genuine or not.

For this round of experiments, we only process the FRAV "attack" dataset since NIR data is unavailable in other datasets and to our knowledge the FRAV "attack" database is the only face presentation attack dataset in literature. Originally, the FRAV "attack" dataset consists of 185 different subjects and experiments in the previous section were conducted under this sample. In these experiments, available data for different subjects is changed

to 157 individuals since there were failure-to-acquire instances during database acquisition. All other setup parameters remain unchanged as before.

In **Table 6**, the best results with the least standard deviation values for BIOPAD across all classifiers were obtained by using NIR images. The drop in performance in the visible spectrum is nearly 1.5% for the SVM linear classifier case and this pattern trend is consistent with other classifier settings. NIR superiority in this type of presentation attack experiments can be further viewed from their t-SNE results in **Figures 7A,B**, where it is apparent that classes are well-separated. These representations can be directly compared with the visible spectrum case (**Figures 5E,F**) where there was a clear overlap between genuine and impostor attacks leading to errors being introduced in sFAR and FFR. The overlap between genuine access attempts and printed photo attacks does not exist in the NIR case and the "tablet" is completely neutralized since there isn't any useful attack information being projected at NIR. Fusing visual information between the visible and NIR at feature level, caused BIOPAD to lose slightly in detection rate performance with respect to NIR only by ∼1.5%, also noticeable in standard deviation values. Moreover, when visualized at feature level and with the visible spectrum analyzed (**Figures 7C,D**), attack patterns appear slightly improved to **Figures 5E,F** but otherwise similar patterns are noticeable.

Furthermore, the performance between the different visual information can be viewed from the Detection Error Tradeoff (DET) curve as shown in **Figure 8**. The DET curve for the FRAV "attack" illustrates the relationship within sFAR and FRR. Naturally, sFAR and FRR confirm the same behavior seen in the percentages, also presented in **Table 6**. As expected the best curve is obtained by BIOPAD with NIR followed by RGB + NIR (feature level) and RGB. Equal error rate or Attack Presentation Equal Error Rate (APEER) is a biometric security system indicator that determines the threshold values for sFAR and FRR. When these rates are equal, their common value is known as the "equal error rate." This value specifies the proportion of false acceptances to false rejections. Low equal error rates mean higher accuracy. In **Figure 8**, the difference between APEERs in BIOPAD's case is 4.15% and undoubtedly shows that for the types of attacks present in the FRAV "attack" database, the best acquisition method for PAD is with the use of a NIR sensor.

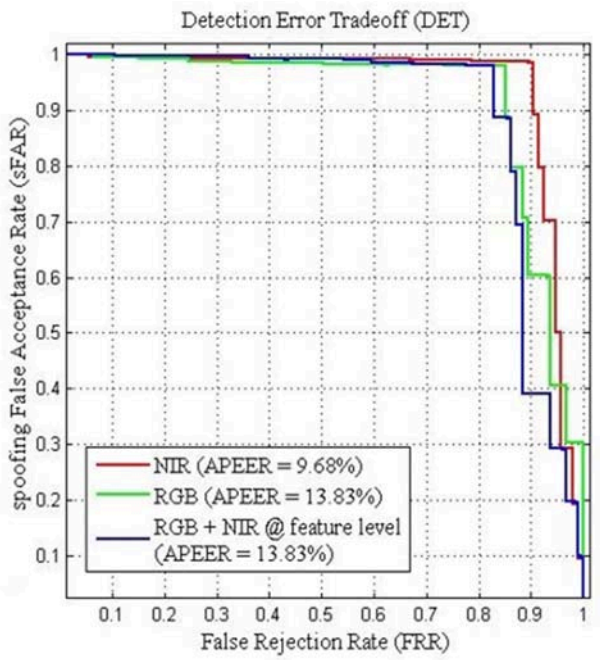


FIGURE 8 | BIOPAD Detection Error Tradeoff curves of SVM linear classifier for the FRAV "attack" database in NIR(red), RGB + NIR at feature level (blue) and RGB (green). Attack Presentation Error Rate—APER.

CONCLUSIONS

In this article we presented a novel presentation attack detection algorithm that relies on the extraction of edge and texture biologically-inspired features, by mimicking biological processes found in areas V1 and V2 of the human visual cortex. This model termed as "BIOPAD," reproduced impressive presentation attack detection rates of up to 99% in certain cases by only utilizing one photo per person and for all attacks examined in the three datasets that were investigated. The main contributions of this research work were to (a) Present a novel biologically-inspired PAD algorithm which behaves comparably to other state-of-the-art algorithms. (b) Introduce a new PAD database called FRAV- "attack," and (c) Introduce near-infrared band information for PAD experimentation at feature and score levels.

BIOPAD has been successful in surpassing other standard biological-like techniques such as HMAX and CNN which are considered state-of-the-art and benchmark models in biologically-inspired vision research. In addition, the creation, introduction and implementation of a new face presentation attack database by our group termed as "FRAV attack," extended our investigation conclusions with high definition samples and diverse scenarios for the most commonly used spoofing attacks. The "FRAV attack" dataset which encompasses visual data that span from visible to infrared, is expected to set future standards for all new databases in face biometrics.

For the first time in literature, a biologically-inspired algorithm has been directly applied with near-infrared information, specifically for the purposes of face presentation attack detection. As observed from the experimental analysis in section Presentation attack results, BIOPAD features maximize the separation between attacks and as a consequence increase attack detection performance. The sFAR and FRR indicate that BIOPAD error performance falls within acceptable limits and it was further evident from our experiments that the nature of data were better separated in classification by a SVM linear classifier. However, future research in classification might reveal classification schema more effective in dealing with incoming data from multiple sensors.

Our results have also shown that near infrared sensor information is of extreme value and importance for presentation attack detection, significantly outperforming visible spectrum data. In our case, an increase in detection rate of almost 6% was observed between the near-infrared and visible scenarios. While the usefulness of near infrared information appears indisputable, we have proposed data fusion from multiple sensors to minimize errors from future elaborate attack methods that have not yet been investigated. To this end, data fusion at feature and score level indicate enhanced detection rates with respect to rates obtained from the visible spectrum.

Overall, results were promising and BIOPAD can serve as a foundation for further enhancements. Future work will include refinement of the biological-like operations to significantly increase performance and speed, optimization of presentation attack detection for video, and real time processes by incorporating biologically-inspired liveness detection algorithms, experimentation with multiple sensors, different types of novel and sophisticated presentation attacks, and experimentation in dynamic—real world situations.

ETHICS STATEMENT

This study was carried out in accordance with the recommendations of the European Union, Spanish police, Spanish government, and University of Rey Juan Carlos with written informed consent from all subjects. All subjects gave written informed consent in accordance with the Declaration of Helsinki. The protocol was approved by the University of Rey Juan Carlos in Spain.

AUTHOR CONTRIBUTIONS

AT is the principal author, main contributor, and researcher of this work. CC helped in the following sections: original research, experiments, and text revision. BG helped during experiments. EC supervised this work and helped in the following sections: original research, during experiments, and text revision.

FUNDING

This research work has been partly funded by ABC4EU project (European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement No 312797) and by BIOinPAD project (funded by Spanish national research agency with reference TIN2016-80644-P).

ACKNOWLEDGMENTS

Preliminary stages of this work were presented in our work titled Face Presentation Attack Detection using Biologically-inspired Features (Tsitiridis et al., 2017). The authors would like to specially thank David Ortega del Campo for his significant contribution and effort in acquiring the new FRAV attack database.

REFERENCES

Alotaibi, A., and Mahmood, A. (2017). Deep face liveness detection based on nonlinear diffusion using convolution neural network. *Signal Image Video Process.* 11, 713–720. doi: 10.1007/s11760-016-1014-2

Anjos, A., Chakka, M. M., and Marcel, S. (2014). Motion-based counter-measures to photo attacks in face recognition. *IET Biometrics* 3, 147–158. doi: 10.1049/iet-bmt.2012.0071

Atoum, Y., Liu, Y., Jourabloo, A., and Liu, X. (2017). "Face anti-spoofing using patch and depth-based CNNs," in *2017 IEEE International Joint Conference on Biometrics (IJCB)* (Denver, CO). doi: 10.1109/BTAS.2017.8272713

Canziani, A., Paszke, A., and Culurciello, E. (2016). An analysis of deep neural network models for practical applications. *arXiv. arXiv:1605.07678v4.*

Chakraborty, S., and Das, D. (2014). An overview of Face Liveness Detection. *Int. J. Inf. Theory* 3, 11–25. doi: 10.5121/ijit.2014.3202

Chen, C., and Ross, A. (2013). "Local gradient Gabor pattern (LGGP) with applications in face recognition, cross-spectral matching, and soft biometrics," in *SPIE Defense, Security, and Sensing.* (Baltimore, MD). doi: 10.1117/12.2018230

Chingovska, I., Anjos, A., and Marcel, E. (2012). "On the effectiveness of local binary patterns in face anti-spoofing," in *2012 BIOSIG - Proceedings of the International Conference of Biometrics Special Interest Group (BIOSIG) (*Darmstadt). Available online at: http://ieeexplore.ieee.org/xpl/login.jsp?tp=andarnumber=6313548andurl=http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6313548

Daugman, J. G. (1985). Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *J. Opt. Soc. Am.* 2, 1160–1169. doi: 10.1364/JOSAA.2.001160

Engel, S., Zhang, X., and Wandell, B. (1997). Colour tuning in human visual cortex measured with functional magnetic resonance imaging. *Nature* 388, 68–71. doi: 10.1038/40398

Eskicioglu, a,. M., and Fisher, P. S. (1995). Image quality measures and their performance. *IEEE Trans. Commun.* 43, 2959–2965. doi: 10.1109/26.477498

Fukushima, K., Miyake, S., and Ito, T. (1980). Neocognitron: a neural network model for a mechanism of visual pattern recognition. *IEEE Trans. Syst Man Cybernet.* SMC-13, 826—834. doi: 10.1109/TSMC.1983.6313076

Galbally, J., Marcel, S., and Fierrez, J. (2014). Image quality assessment for fake biometric detection: application to Iris, fingerprint, and face recognition. *IEEE Trans. Image Process.* 23, 710–724. doi: 10.1109/TIP.2013.2292332

Galbally, J., Marcel, S., and Fierrez, J. (2015). Biometric antispoofing methods: a survey in face recognition. *IEEE Access* 2, 1530–1552. doi: 10.1109/ACCESS.2014.2381273

Goldstein, B. E. (2010). *Sensation and Perception*. Belmont, CA: Wadsworth.

Grigorescu, S. E., Petkov, N., and Kruizinga, P. (2002). Comparison of texture features based on Gabor filters. *IEEE Trans. Image Process.* 11, 1160–1167. doi: 10.1109/TIP.2002.804262

Hegdé, J., and Van Essen, D. C. (2000). Selectivity for complex shapes in primate visual area V2. *J. Neurosci.* 20:RC61. doi: 10.1523/JNEUROSCI.20-05-j0001.2000

Hermosilla, G., Ruiz-Del-Solar, J., Verschae, R., and Correa, M. (2012). A comparative study of thermal face recognition methods in unconstrained environments. *Pattern Recognit.* 45, 2445–2459. doi: 10.1016/j.patcog.2012.01.001

Hubel, D. H., and Wiesel, T. N. (1967). Receptive fields and functional architecture of monkey striate cortex. *J. Physiol.* 195, 215–243. doi: 10.1113/jphysiol.1968.sp008455

Kong, S. G., Heo, J., Abidi, B. R., Paik, J., and Abidi, M. A. (2005). Recent advances in visual and infrared face recognition - A review. *Comput. Vis. Image Underst.* 97, 103–135. doi: 10.1016/j.cviu.2004.04.001

Kose, N., Apvrille, L., and Dugelay, J.-L. (2015). "Facial makeup detection technique based on texture and shape analysis," in *2015 11th IEEE International*

Conference and Workshops on Automatic Face and Gesture Recognition (FG) (Ljubljana: IEEE), 1–7. doi: 10.1109/FG.2015.7163104

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* 60, 84–90. doi: 10.1145/3065386

Lakshminarayana, N. N., Narayan, N., Napp, N., Setlur, S., and Govindaraju, V. (2017). "A discriminative spatio-temporal mapping of face for liveness detection," in *2017 IEEE International Conference on Identity, Security and Behavior Analysis, ISBA 2017*. (New Delhi). doi: 10.1109/ISBA.2017.7947707

Lampl, L., Ferster, D., Poggio, T., and Riesenhuber, M. (2004). Intracellular measurements of spatial integration and the MAX operation in complex cells of the cat primary visual cortex. *J. Neurophysiol.* 92, 2704–2713. doi: 10.1152/jn.00060.2004

LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proc. IEEE* 86, 2278–2324. doi: 10.1109/5.726791

Lei, Z., Li, S. Z., Chu, R., and Zhu, X. (2007). Face recognition with local gabor textons. *Adv. Biometrics* 49–57. doi: 10.1007/978-3-540-74549-5_6

Li, J., Wang, Y., Tan, T., and Jain, A. K. (2004). "Live face detection based on the analysis of fourier spectra," in *Defense and Security*, 296–303. doi: 10.1117/12.541955

Li, M., Bao, S., Qian, W., and Su, Z. (2013). "Face recognition using early biologically inspired features," in *2013 IEEE Sixth International Conference on Biometrics: Theory, Applications and Systems (BTAS),* (Arlington, VA), 1–6. doi: 10.1109/BTAS.2013.6712711

Liu, Y., Jourabloo, A., and Xiaoming, L. (2018). "Learning deep models for face anti-spoofing: binary or auxiliary supervision," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition* (Salt Lake City, UT), 389–398. doi: 10.1109/CVPR.2018.00048

Lucena, O., Junior, A., Moia, V., Souza, R., Valle, E., and Lotufo, R. (2017). "Transfer learning using convolutional neural networks for face anti-spoofing," in *Lect. Notes Comput. Sci (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*. (Montreal, qc), doi: 10.1007/978-3-319-59876-5_4

Lyons, M., Akamatsu, S., Kamachi, M., and Gyoba, J. (1998). "Coding facial expressions with Gabor wavelets," in *Proceedings Third IEEE International Conference on Automatic Face and Gesture Recognition* (Nara), 200–205. doi: 10.1109/AFGR.1998.670949

Maatta, J., Hadid, A., and Pietikäinen, M. (2011). "Face spoofing detection from single images using micro-texture analysis," in *2011 International Joint Conference on Biometrics (IJCB)* (Washington, DC), 1–7. doi: 10.1109/IJCB.2011.6117510

Marcelja, S. (1980). Mathematical description of the responses of simple cortical cells. *J. Opt. Soc. Am.* 70, 1297–1300. doi: 10.1364/JOSA.70.001297

McAdams, C. J., and Reid, R. C. (2005). Attention modulates the responses of simple cells in monkey primary visual cortex. *J. Neurosci.* 25, 11023–11033. doi: 10.1523/JNEUROSCI.2904-05.2005

Meyers, E., and Wolf, L. (2008). Using biologically inspired features for face processing. *Int. J. Comput. Vis.* 76, 93–104. doi: 10.1007/s11263-007-0058-8

Movshon, J. A., Thompson, I. D., and Tolhurst, D. J. (1978). Spatial and temporal contrast sensitivity of neurones in areas 17 and 18 of the cat's visual cortex. *J. Physiol.* 283, 101–120. doi: 10.1113/jphysiol.1978.sp012490

Palczewska, G., Vinberg, F., Stremplewski, P., Bircher, M. P., Salom, D., Komar, K., et al. (2014). Human infrared vision is triggered by two-photon chromophore isomerization. *Proc Natl Acad Sci U.S.A.* 111, E5445–E5454. doi: 10.1073/pnas.1410162111

Pan, G., Wu, Z., and Sun, L. (2008). Liveness detection for face recognition. *Recent Adv. Face Recognit.* 236, 109–124. doi: 10.5772/6397

Perlibakas, V. (2006). Face recognition using principal component analysis and log-gabor filters. *Analysis* 3:23. arXiv:cs/0605025.

Petkov, N., and Kruizinga, P. (1997). Computational models of visual neurons specialised in the detection of periodic and aperiodic oriented visual stimuli: bar and grating cells. *Biol. Cybern.* 76, 83–96. doi: 10.1007/s004220050323

Pisharady, P. K., and Martin, S. (2012). Pose invariant face recognition using neuro-biologically inspired features. *Int. J. Futur. Comput. Commun.* 1, 316–320. doi: 10.7763/IJFCC.2012.V1.85

Prokoski, F. J., and Riedel, R. B. (2002). "Infrared identification of faces and body parts," in *Biometrics*, 191–212. doi: 10.1007/0-306-47044-6_9. Available online at: http://www.springerlink.com/index/x442p40qv2734757.pdf

Raghavendra, R., Raja, K. B., and Busch, C. (2015). "Presentation attack detection for face recognition using light field camera," in *IEEE Transactions on Image Processing,* Vol. 24, 1060–1075. doi: 10.1109/TIP.2015.2395951

Ramachandram, D., and Taylor, G. W. (2017). "Deep multimodal learning: a survey on recent advances and trends," in *IEEE Signal Processing Magazine*. doi: 10.1109/MSP.2017.2738401

Ramon, M., Caharel, S., and Rossion, B. (2011). The speed of recognition of personally familiar faces. *Perception* 40, 437–449. doi: 10.1068/p6794

Riesenhuber, M., and Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nat. Neurosci.*2, 1019–1025. doi: 10.1038/14819

Riesenhuber, M., and Poggio, T. (2000). Models of object recognition. *Nat. Neurosci.* 3, 1199–1204. doi: 10.1038/81479

Rolls, E. T. (2012). Invariant visual object and face recognition: neural and computational bases, and a model, VisNet. *Front. Comp. Neurosci.* 6:35. doi: 10.3389/fncom.2012.00035

Rose, N. (2006). "Facial expression classification using gabor and log-gabor filters," in *7th International Conference on Automatic Face and Gesture Recognition, 2006* (Southampton), 346–350.

Rust, N. C., Schwartz, O., Movshon, J. A., and Simoncelli, E. P. (2005). Spatiotemporal elements of macaque V1 receptive fields. *Neuron* 46, 945–956. doi: 10.1016/j.neuron,.2005.05.021

SC37ISO/IEC JTC1 and Biometrics (2014). *Information Technology—Presentation Attack Detection—Part 3: Testing, Reporting and Classification of Attacks*. SC37ISO/IEC JTC1 and Biometrics

Schmid, A. M., Purpura, K. P., and Victor, J. D. (2014). Responses to orientation discontinuities in V1 and V2: physiological dissociations and functional implications. *J. Neurosci.* 34, 3559–3578. doi: 10.1523/JNEUROSCI.2293-13.2014

Schneider, G. E. (1969). Two visual systems. *Science* 163, 895–902. doi: 10.1126/science.163.3870.895

Seal, A., Ganguly, S., Bhattacharjee, D., Nasipuri, M., and Basu, D. K. (2013). Automated thermal face recognition based on minutiae extraction. *Int. J. Comput. Intell. Stud.* 2, 133–156. doi: 10.1504/IJCISTUDIES.2013.055220

Serrano, Á., Martín De Diego, I., Conde, C., and Cabello, E. (2011). Analysis of variance of Gabor filter banks parameters for optimal face recognition. *Pattern Recognit. Lett.* 32, 1998–2008. doi: 10.1016/j.patrec.2011.09.013

Serre, T., and Riesenhuber, M. (2004). Realistic modeling of simple and complex cell tuning in the HMAX model, and implications for invariant object recognition in cortex. *Methods* 17, 1–12. doi: 10.21236/ADA459692

Serre, T., Wolf, L., Bileschi, S., and Riesenhuber, M. (2007). Robust Object Recognition with Cortex-like mechanisms. *IEEE Trans. Pattern Anal. Mach. Intell.* 29, 411–426. doi: 10.1109/TPAMI.2007.56

Shoja Ghiass, R., Arandjelović, O., Bendada, A., and Maldague, X. (2014). Infrared face recognition: A comprehensive review of methodologies and databases. *Pattern Recognit.* 47, 2807–2824. doi: 10.1016/j.patcog.2014.03.015

Singh, R., Vatsa, M., and Noore, A. (2009). Face recognition with disguise and single gallery images. *Image Vis. Comput.* 27, 245–257. doi: 10.1016/j.imavis.2007.06.010

Slavkovic, M., Reljin, B., Gavrovska, A., and Milivojevic, M. (2013). "Face recognition using Gabor filters, PCA and neural networks," in *2013 20th International Conference on Systems, Signals and Image Processing (IWSSIP)* (Bucharest), 35–38. doi: 10.1109/IWSSIP.2013.6623443

Tsitiridis, A., Conde, C., De Diego, I. M., and Cabello, E. (2017). "Face presentation attack detection using biologically-inspired features," in *VISIGRAPP 2017 - Proceedings of the 12th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*. (Killarney). doi: 10.5220/0006124603600370

Ungerleider, L. G., and Mishkin, M. (1982). Two cortical visual systems. *Anal. Vis. Behav.* 549–586.

Van De Sande, K., Gevers, T., and Snoek, C. (2010). "Evaluating color descriptors for object and scene recognition," in *IEEE Transactions on Pattern Analysis and Machine Intelligence,* Vol. 32, 1582–1596. doi: 10.1109/TPAMI.2009.154

Van Der Maaten, L. J. P., and Hinton, G. E. (2008). Visualizing high-dimensional data using t-sne. *J. Mach. Learn. Res.* 9, 2579–2605.

Wang, S., Xia, X., Qing, Z., Wang, H., and Le, J. (2013). "Aging face identification using biologically inspired features," in *2013 IEEE International Conference on Signal Processing, Communication and Computing (ICSPCC* 2013), (Kunming), 1–5. doi: 10.1109/ICSPCC.2013.6664116

Wang, Y., and Chua, C. (2005). Face recognition from 2D and 3D images using 3D Gabor filters. *Image Vis. Comput.* 23, 1018–1028. doi: 10.1016/j.imavis.2005.07.005

Wang, Y., Nian, F., Li, T., Meng, Z., and Wang, K. (2017). Robust face anti-spoofing with depth information. *J. Vis. Commun. Image Represent.* 49, 332–337. doi: 10.1016/j.jvcir.2017.09.002

Webster, M. A., and De Valois, R. L. (1985). Relationship between spatial-frequency and orientation tuning of striate-cortex cells. *J. Opt. Soc. Am. A.* 2, 1124–1132. doi: 10.1364/JOSAA.2.001124

Wen, D., Han, H., and Jain, A. K. (2015). "Face spoof detection with distortion analysis," in *IEEE Transactions on Information Forensics and Security*, Vol. 10, 746–761. doi: 10.1109/TIFS.2015.2400395

Wu, H.-Y., Rubinstein, M., Shih, E., Guttag, J., Durand, F., and Freeman, W. (2012). "Eulerian video magnification for revealing subtle changes in the world," in *ACM Transactiona on Graphics* (New York, NY), 31, 1–8. doi: 10.1145/2185520.2185561

Xu, Z., Li, S., and Deng, W. (2016). "Learning temporal features using LSTM-CNN architecture for face anti-spoofing," in *Proceedings - 3rd IAPR Asian Conference on Pattern Recognition, ACPR 2015.* (Kuala Lumpur), doi: 10.1109/ACPR.2015.7486482

Yan, J., Zhang, Z., Lei, Z., Yi, D., and Li, S. Z. (2012). "Face liveness detection by exploring multiple scenic clues," in *12th International Conference on Control Automation Robotics and Vision (ICARCV)* (Guangzhou), 188–193. doi: 10.1109/ICARCV.2012.6485156

Yang, J., Lei, Z., and Li, S. Z. (2014). Learn convolutional neural network for face anti-spoofing. *arXiv:1408.5601*.

Yokono, J. J., and Poggio, T. (2004). *Rotation Invariant Object Recognition from One Training Example*. Available online at: http://cbcl.mit.edu/publications/ai-publications/2005/AIM-2005-023.pdf

Zhang, B., Zhang, L., Zhang, D., and Shen, L. (2010). Directional binary code with application to PolyU near-infrared face database. *Pattern Recogn. Lett.* 31, 2337–2344. doi: 10.1016/j.patrec.2010.07.006

Zhang, W., Shan, S., Gao, W., Chen, X., and Zhang, H. (2005). "Local Gabor Binary Pattern Histogram Sequence (LGBPHS): a novel non-statistical model for face representation and recognition," in *Tenth IEEE International Conference on Computer Vision (ICCV'05)* (Beijing), 786–791.

Zhang, Z., Yan, J., Liu, S., Lei, Z., Yi, D., and Li, S. Z. (2012). "A face antispoofing database with diverse attacks," in *IEEE Biometrics CompendiumIEEE RFIC Virtual JournalIEEE RFID Virtual Journal* (New Delhi), 26–31. doi: 10.1109/ICB.2012.6199754

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 28 May 2019
doi: 10.3389/fnins.2019.00523

Scene Regularity Interacts With Individual Biases to Modulate Perceptual Stability

Qinglin Li[1,2,3,4*†], *Andrew Isaac Meso*[5†], *Nikos K. Logothetis*[1,6] *and Georgios A. Keliris*[1,3,4*]

[1] Department of Physiology of Cognitive Processes, Max Planck Institute for Biological Cybernetics, Tübingen, Germany, [2] IMPRS for Cognitive and Systems Neuroscience, University Tuebingen, Tübingen, Germany, [3] Bernstein Center for Computational Neuroscience, Tübingen, Germany, [4] Department of Biomedical Sciences, University of Antwerp, Wilrijk, Belgium, [5] Psychology and Interdisciplinary Neurosciences Research Group, Faculty of Science and Technology, Bournemouth University, Poole, United Kingdom, [6] Division of Imaging Science and Biomedical Engineering, University of Manchester, Manchester, United Kingdom

OPEN ACCESS

Edited by:
Hedva Spitzer,
Tel Aviv University, Israel

Reviewed by:
Szonya Durant,
Royal Holloway, University of London, United Kingdom
Huseyin Boyaci,
Bilkent University, Turkey

***Correspondence:**
Qinglin Li
qinglin.li@tuebingen.mpg.de
Georgios A. Keliris
georgios.keliris@uantwerpen.be

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Perception Science, a section of the journal Frontiers in Neuroscience

Received: *16 October 2018*
Accepted: *06 May 2019*
Published: *28 May 2019*

Citation:
Li Q, Meso AI, Logothetis NK and Keliris GA (2019) Scene Regularity Interacts With Individual Biases to Modulate Perceptual Stability. Front. Neurosci. 13:523. doi: 10.3389/fnins.2019.00523

Sensory input is inherently ambiguous but our brains achieve remarkable perceptual stability. Prior experience and knowledge of the statistical properties of the world are thought to play a key role in the stabilization process. Individual differences in responses to ambiguous input and biases toward one or the other interpretation could modulate the decision mechanism for perception. However, the role of perceptual bias and its interaction with stimulus spatial properties such as regularity and element density remain to be understood. To this end, we developed novel bi-stable moving visual stimuli in which perception could be parametrically manipulated between two possible mutually exclusive interpretations: transparently or coherently moving. We probed perceptual stability across three composite stimulus element density levels with normal or degraded regularity using a factorial design. We found that increased density led to the amplification of individual biases and consequently to a stabilization of one interpretation over the alternative. This effect was reduced for degraded regularity, demonstrating an interaction between density and regularity. To understand how prior knowledge could be used by the brain in this task, we compared the data with simulations coming from four different hierarchical models of causal inference. These models made different assumptions about the use of prior information by including conditional priors that either facilitated or inhibited motion direction integration. An architecture that included a prior inhibiting motion direction integration consistently outperformed the others. Our results support the hypothesis that direction integration based on sensory likelihoods maybe the default processing mode with conditional priors inhibiting integration employed in order to help motion segmentation and transparency perception.

Keywords: visual perception, bias, bayesian, computational modeling, regularity, psychophysics, human perception, motion perception

INTRODUCTION

Our brains are subjected to ambiguous sensory inputs from a variety of sources, yet the world that we perceive appears stable and coherent. To constantly maintain such a percept, dynamic sensory inputs are thought to be combined with our prior knowledge and experience to form what should be consistent neural representations (Knill and Richards, 1996; Rao et al., 2002). Alternative percepts compete dynamically, continuously resulting in changes to the dominant representation driven by interactions taking place at several stages of the cortical hierarchy. Perception can thus vary between multiple outcomes by a myriad of possible mechanisms (Desimone and Duncan, 1995; Beck and Kastner, 2009; Meso et al., 2016b). Biased competition theory suggested that objects simultaneously presented in the visual field compete for neural representation and attention can bias this competition (Desimone and Duncan, 1995; Desimone, 1998; Beck and Kastner, 2009). When stimuli are inherently more ambiguous, such internal processes become more critical in perceptual selection and could govern the outcome of the competition. However, the role of observer bias and how that might interact with key visual stimulus properties which may often control signal strength, remains unexplored. Questions arise following evidence recently found that the human visual system possesses internal templates for regular patterns, indicating that regularity is a coded feature in human vision (Morgan et al., 2012; Ouhnana et al., 2013).

Here, we developed novel bi-stable visual stimuli (**Figure 1**) that exploited the significant role of plaid local elements such as intersections (Stoner et al., 1990), to parametrically manipulate perception between two possible interpretations, coherent and transparently moving. We then probed perceptual stability during the resulting ambiguous motion perception across three stimulus density levels with normal or degraded regularity using a factorial design. Further, a set of Bayesian observer models based on the causal inference frame work (Shams and Beierholm, 2010) were developed to perform a perceptual task analogous to the experiments carried out in order to support the investigation of the underlying mechanism. Causal inference has been demonstrated to model perceptual judgements of multisensory integration (Körding et al., 2007; Sato et al., 2007) and fine motion direction judgments done using discrimination (Stocker and Simoncelli, 2007). The approach tackles the problem of having to decide whether two sensory signals come from the same source (in which case they should be integrated) or come from different sources (in which case they should be segregated). These models typically have just four parameters which correspond to the observer's *individual bias* toward one or the other of the of the alternatives; two parameters capturing the *sensory noise* associated with the representation of each competing alternative and finally a *prior width* parameter which defines the extent of the influence the prior has across the measurement space when it is applied. We implement the models in the current experimental context to explore whether performance changes across the density and regularity conditions measured during the tasks are better explained by shifts in one or both sensory likelihood parameters or in prior parameters.

MATERIALS AND METHODS

Participants and Apparatus

Five subjects (college students, four females) participated in all the experiments, four of whom were naïve to the aims of the study. All had normal or corrected-to-normal vision. The study was approved by the ethical committee of the University of Tuebingen. Before data collection, a written participant informed consent was obtained from each subject.

The experiments were performed in a dimly lit room. The stimuli were programmed using Matlab Psychophysics toolbox (Brainard, 1997) and presented on a 17-inch CRT monitor (iiyama, 21sd017) with a resolution of 1,280 × 1,024 and a refresh rate of 100 Hz. The monitor was gamma corrected with a mean luminance of 15.6 cd/m^2. The distance from the eyes of the subject to the monitor was 43 cm. Responses from subjects were acquired by using a bespoke 2-button response box (see Procedures). Eye movements were monitored continuously using an infrared video eye tracker (iView X^{TM} Hi-speed, SMI).

Stimuli

The novel plaid stimuli in this study were designed to mimic and manipulate the local elements—lines and intersections—that are carrying the motion signals within the square line plaid stimuli that have been used extensively in psychophysics (Stoner et al., 1990). To achieve this, we decomposed the original plaids into two different types of stimulus patches (see **Figures 1A–C**; **Supplementary Movies 1, 2**): separated lines (SL) and line intersections (LI). Although in what follows we refer to these patches as apertures, it should be noted that their dynamic content remained always the same (SL or LI) independent of the position they were plotted. Thus, this allowed us to manipulate the locations of these motion signals to be either consistent with an underlying plaid or jittered in space. The mimicked plaid from which these apertures were created, consisted of two identical superimposed asymmetric line gratings (Hupé and Rubin, 2003; Takahashi, 2004; Moreno-Bote et al., 2010) with a directional difference of 120° (±60 with respect to vertical). Stimulus directions were fixed with respect to the vertical rather than being randomized during the task to avoid previously reported idiosyncratic anisotropies in participant representations of direction (Rauber and Treue, 1999) and to simplify simulated categorical perceptual decisions during the modeling. The spatial frequency of each narrow line grating was 1 cycle per degree, with a duty cycle of 1 pixel or 0.03° and a speed of 2° per second. In order to minimize the luminance effect of the intersection for plaid stimuli (Stoner et al., 1990; Thiele and Stoner, 2003), the luminance of the small intersections remained the same as that of the line. The color of the lines was black (0.9 cd/m^2) and the background was gray (15.6 cd/m^2). In Experiment 1 (Regular; **Figure 1B**) their positions were selected based on a regular grid of locations where either intersections or single lines would be expected in the classic plaid (see positions of red and green dotted circles in **Figure 1A**). In Experiment 2 (Irregular; **Figure 1C**), the possible positions of apertures were dynamically jittered vertically from the grid locations (±0.025° of visual-angle) and SL and LI could be located in any of the locations

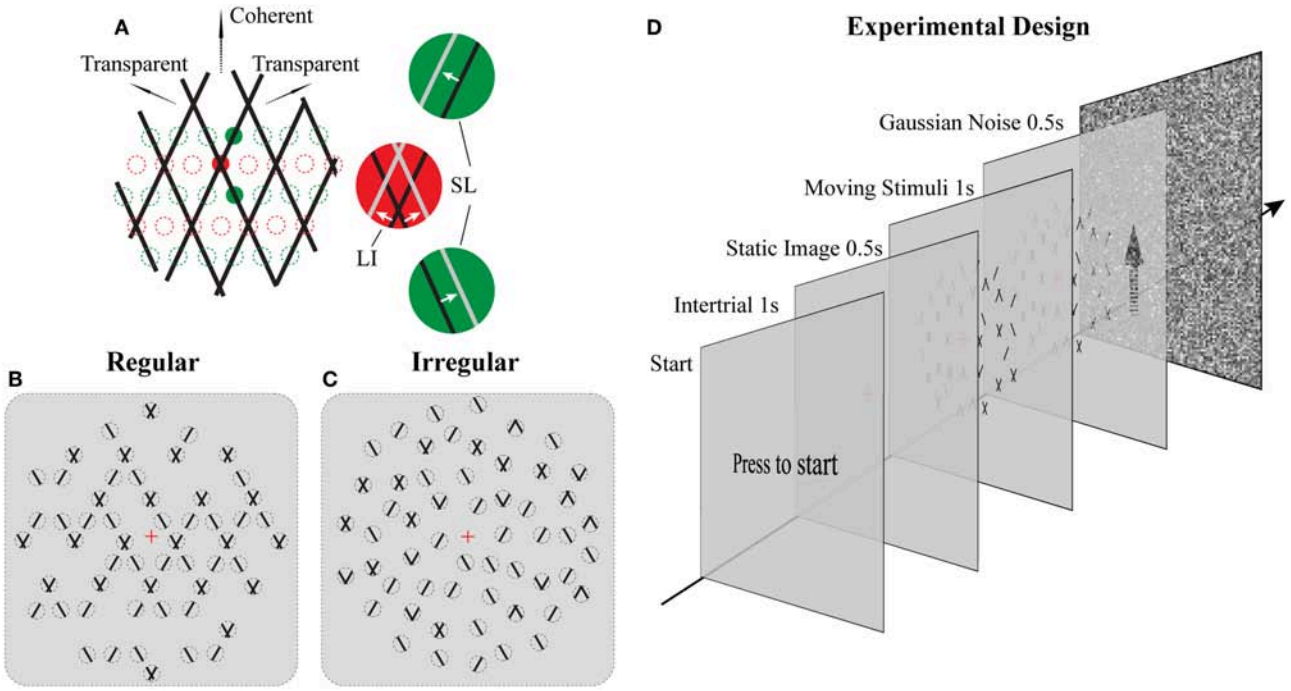


FIGURE 1 | Stimuli and Experimental design. **(A)** Illustration of moving stimuli. The line-plaid (composed of two overlaid drifting line gratings) can be analyzed as containing two different local inputs and namely LI (red) and SL (green). Locations of lines at time t and t + Δt were plotted in black and gray, respectively, in each aperture. Red and green dotted lines indicate exemplary positions of LI and SL apertures respectively. **(B,C)** Cartoon versions of regular and irregular stimuli. In the regular condition **(B)** as the pattern moves up only SI or nothing can appear in SI apertures and likewise for the LI. In the irregular condition **(C)** the aperture locations including their contents were jittered. Dotted lines indicate the locations of the apertures but were not visible **(D)** Experimental design. Subjects had to press a key on the response box to start a trial. After that, a red fixation cross was shown on the center of the monitor for 1 s. The luminance is the same as the mean luminance of the following trial to exclude the influences of luminance changing. A static image of the following trial was presented for 0.5 s to avoid transitional eye movements. After that, the stimulus was shown for 1 s, and subjects had to report their perception by a button press. The trial ended with Gaussian noise presented for 0.5 s to mask potential effects of previous stimuli in subsequent trials.

on the underlying grid abolishing the regularity of Experiment 1. The diameter of each aperture was 0.2° of viewing-angle and 720 potential locations were used with no overlap over a stimulus area with a 23°diameter. A rhombus-shaped mask was applied upon each aperture so that no terminators leading to the perception of circular apertures would be seen (Pack et al., 2003). The vertical and horizontal distance between the centers of adjacent apertures was 0.5° and 0.28° of view-angle, respectively. A red fixation cross (0.2° of visual-angle) was shown at the center of the stimuli. No apertures were located within a circular area (2° of visual-angle diameter) where the fixation was centered. The stimuli shared some similarities with previously used multi-aperture stimuli but also had some critical differences (Amano et al., 2009, 2012): (a) within the apertures we used moving lines instead of drifting Gabors, (b) in the regular condition aperture locations for lines and intersections were selected according to the underlying plaid pattern (Experiment 1), (c) the number of apertures was systematically manipulated, and (d) the proportion of different aperture types was used to parametrically change perception.

The total number of apertures was chosen based on three density conditions: low, medium, and high; with 180, 340, and 680, apertures, respectively. New random positions were selected according to these numbers for each trial. In addition, we parametrically manipulated the ratio between SL and LI along 11 homogeneously spaced proportions within the range of 0% to 100%.

Procedures

For both Experiments 1 and 2, subjects were instructed to press a key on the response box to start a trial (see **Figure 1D**). After that, a red fixation cross was shown on the center of the monitor for 1 s. Before trial onset, background luminance was slightly adjusted to the mean luminance depending on the density condition to have a homogeneous mean luminance across conditions and trials. First, a static image was presented for 0.5 s to control for transitional eye movements. Then, the stimulus started moving for 1 s, and subjects had to report their perception (either coherent or transparent) during this period by pressing one of two keys. They were instructed to do so as fast as possible and according to their first impression. In order to avoid potential adaptation effects, each trial was followed with a 0.5 s full field Gaussian noise pattern with mean luminance equal to the average of all trials. A method of constant stimuli was used and each psychometric point came from 30 measurements for each of the 11 points along the parametric manipulation of the ratio of the

different types of apertures for each subject. All conditions were presented in a pseudo-randomized fashion.

At the beginning of each block, a standard nine-point eye tracking calibration was performed. Subjects took a break after each block. For training, subjects performed 4 blocks of 15 trials before each experiment. They were instructed to fixate the center of the screen and use a chin-rest to avoid head movements.

Theory and Models

Modeling transparent motion perception presents a challenge of separating unlabeled signals which can come from one source or from multiple sources, posing a computational problem similar to that previously studied with vowel sounds (Sato et al., 2007; Feldman et al., 2009). Here, we used the causal inference framework which originates in multisensory perception and considered the problem to be solved as an explicit two-step hierarchical process with an initial unity vs. separation choice and subsequent direction perception made subject to the influence of the initial decision as a conditional estimate (Stocker and Simoncelli, 2007; Zamboni et al., 2016). This class of models typically has four parameters (Körding et al., 2007; Stocker and Simoncelli, 2007): a participant bias parameter—which we did not use in the current work for reasons explained later, two sensory likelihood parameters corresponding to each alternative sensory representation and a prior width parameter which determines the extent to which the likelihoods can be shifted along the measurement space.

An optimal Bayesian model would average over the probability of both hypotheses (Körding et al., 2007; Sato et al., 2007), which in this case would be, coherent dominated by components given by $H = h_c$ and transparent dominated by the plaid pattern given by $H = h_p$, making a decision by reading out from the averaged probability distribution. For a difficult categorical perceptual decision associated with a global percept with mutually exclusive alternatives like ambiguous global motion, we followed previous work (Sato et al., 2007; Stocker and Simoncelli, 2007; Zamboni et al., 2016), and used an implementation in which the optimality of averaging was sacrificed for a quick and self-consistent decision. In other words, a categorical decision is made and this adjusts the shape of the prior probabilities to influence the refined estimate of the second stage. The visual stimulus contains a superimposed distribution of multiple directions of components θ_s, from which a sensory measurement of the perceived direction distribution θ_m, is made by the visual system; an estimate contaminated by Gaussian noise. Given the task at hand in which the alternatives, h_c (components dominate) and h_p (single pattern dominates) cannot mutually exist, we impose an assumption that ambiguity resolution forces the system to commit to one alternative, and its corresponding posterior distribution only, which is either $P(\theta|h_c)$ or $P(\theta|h_p)$, illustrated in **Figure 2** (Sato et al., 2007).

Three model variants made the following assumptions about the prior: M1 assumed no additional hypothesis about the direction space, i.e., a flat prior with all directions equally likely, then estimation of maximum likelihood $P(\theta_m)$ and then categorization of direction; M2 selectively applied a prior on trials where an initial hierarchical step suggested motion integration of the input was needed, consistent with the use of a slow speed prior which has been shown to explain some cases of motion perception (Weiss et al., 2002); The categorical decision in the second step was based on the estimated maximum posterior direction after multiplication with the excitatory prior (h_p). M3 similarly computes a categorical decision from the maximum posterior after multiplication with an inhibitory prior (h_c) but in contrast on trials which could not be selected by M2, where component separation is suggested by early noisy computations, which supports motion segregation. This novel configuration implements a prior distribution centered diametrically opposite to the average stimulus direction in the circular direction space so that the average direction is inhibited. This is a viable probability distribution configuration in a circular space. Note that for simulations of configuration M2, no segregate priors (i.e., M3) were applied on trials where integrate was chosen and similarly, for the separate simulations under M3 prior no integrate prior (i.e., M2) was applied to any trials. M4 is a control condition which uses either prior (h_c or h_p) on each individual trial following the initial estimate, a biologically implausible architecture which we used to allow us to contrast conditions.

The probability of the alternative categorical hypotheses H, is given by Equation (1) which includes all the respective likelihoods and priors,

$$P(H|\theta_m) = P(\theta_m|H)P(H)/P(\theta_m) \tag{1}$$

Applying model averaging over the posterior distribution (Stocker and Simoncelli, 2007) of each model results in Equation (2):

$$\int P(\theta_s|\theta_m)\, d\theta = 1, \tag{2}$$

$$\begin{aligned} P(\theta_s|\theta_m) = {} & P\left(\theta_s|\theta_m\,,\, H = h_c\right) P\left(H = h_c|\theta_m\right) \\ & + P\left(\theta_s|\theta_m\,,\, H = h_p\right) P\left(H = h_p|\theta_m\right), \end{aligned} \tag{3}$$

where the composite posterior in Equation (3) is obtained by adding both alternative posterior probabilities corresponding to each perceptual alternative. We simplify Equation (3) which includes the two separate posterior terms by using model selection to propose an initial fast binary variable computation $\chi_{(1,2)}$, (see simulations) corresponding to hypotheses $H = h_c$ and $H = h_p$, respectively, to hierarchically separate the early discrimination and the estimation tasks (Luu and Stocker, 2018). In each case, one alternative is selected and the remaining term is set to a probability of zero (Stocker and Simoncelli, 2007). We do not seek an optimal solution to Equations (3) and instead following the lead from previous work sacrifice optimality for consistency (Stocker and Simoncelli, 2007; Luu and Stocker, 2018). During simulations, we assign a decision value of $\chi = 1$, if the MLE is closer to the average (pattern direction) than the component direction, and $\chi = 2$ if the MLE is closer to the transparent component direction (see **Figure 4**). This heuristic crudely solves the "one vs. two" component problem and reduces the number of free parameters used in this type of experiments from four to three by avoiding the inclusion of a parameter for bias. While individual differences in participant biases have been

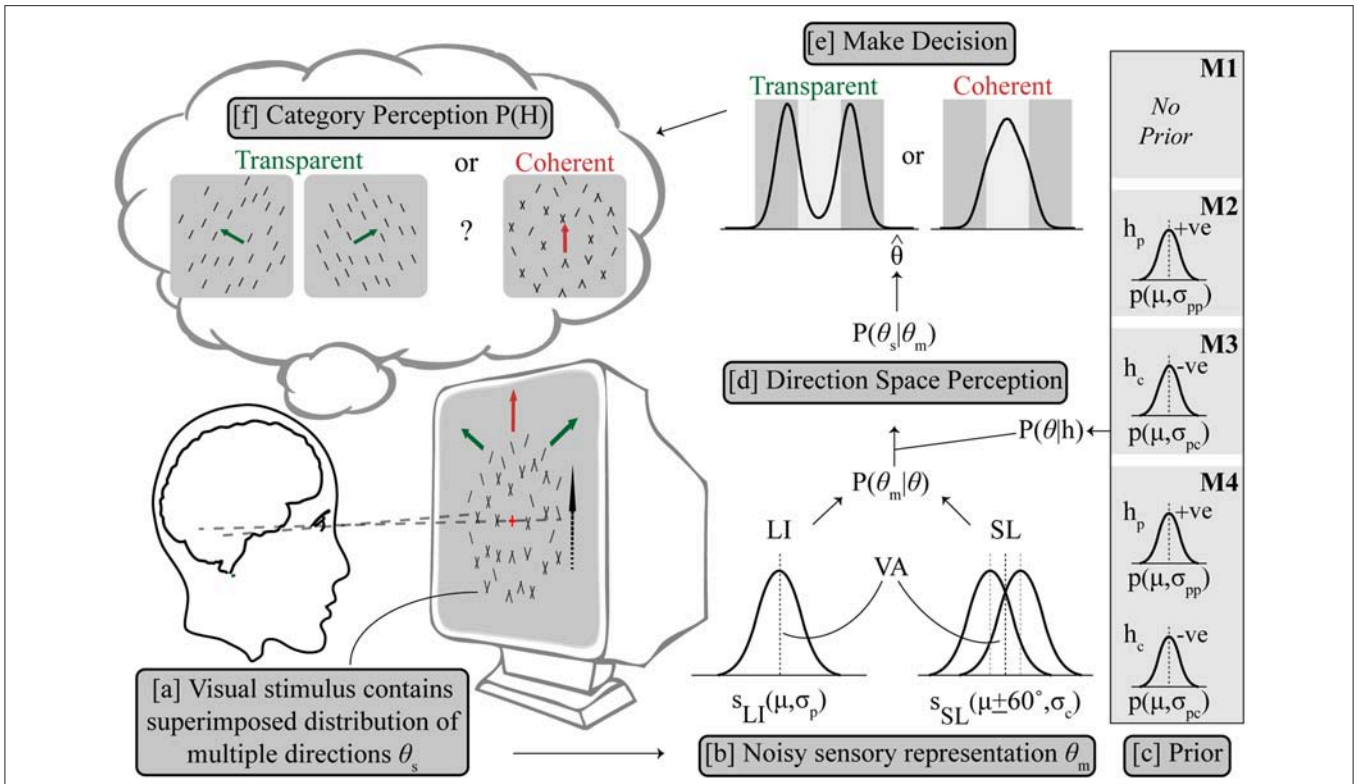


FIGURE 2 | Outline of Bayesian observer model. **(a)** The visual stimulus contains multiple directions of components θ_S, from which a sensory estimation was made as θ_m with uncertainty. **(b)** LI is represented as a Gaussian probability density function (S_{LI}) centered on the vector averaged (VA) direction (μ) in the direction space with variance σ_p, while SL is similarly modeled as two Gaussian probability density functions (S_{SL}) centered on $\mu + 60°$ and $\mu - 60°$ respectively with same variance σ_c. The likelihood $P(\theta_m|\theta)$ contains S_{LI} and S_{SL}, combining with the respective prior term $P(\theta|h)$ **(c)** to get the posterior distribution of $P(\theta_s|\theta_m)$ **(d)**. Prior settings are different for M1–M4, see text for details. The prior terms $P(\theta|h_p)$ and $P(\theta|h_c)$ are also both Gaussian terms centered on the VA direction which either enhance (h_p) or inhibit (h_c) the pattern to support integration or segregation, respectively. **(e)** Decision is made based on a final direction using MAP estimation leading to categorical perception **(f)**.

previously found and modeled (Odegaard and Shams, 2016), in the current work we expected there might be differences within participants across our scene structure conditions and so focused on the interaction between the role of sensory representations and the strength of prior biases. Our heuristic computation of χ similarly constrained all the participants' categorical estimation.

The conditional inference is therefore computed on a given trial according to either,

$$P(\theta|\theta_m, \chi = 1) = P(\theta_m|\theta)P\left(\theta\middle|h_p\right)/P(\theta_m), \quad (4)$$

in the coherent case where pattern motion is reported or,

$$P(\theta|\theta_m, \chi = 2) = P(\theta_m|\theta)P\left(\theta\middle|h_c\right)/P(\theta_m), \quad (5)$$

in the case of the transparent choice where the two components are simultaneously perceived. In both Equations (4) and (5), the likelihood term $P(\theta_m|\theta)$ is identical and contains Gaussian functions of two components and one pattern term whose width captures the sensory noise, and these are shown together as Equation (6).

$$P(\theta_m|\theta) = \frac{A_S}{\sqrt{2\pi}}\exp\left(-\frac{(\theta-\theta_S)^2}{2\sigma_S^2}\right) + \frac{A_S}{\sqrt{2\pi}}\exp\left(-\frac{(\theta+\theta_S)^2}{2\sigma_S^2}\right) + \frac{A_L}{3\sqrt{2\pi}}\exp\left(-\frac{(\theta)^2}{2\sigma_L^2}\right) \quad (6)$$

The average direction of the distribution in Equation (6) is also the pattern direction, $\theta_L = 0$. The relative scaling of the Gaussian terms corresponding to the alternative percepts is related by $A_S = 1\text{-}A_L$. The respective prior terms $P(\theta|h_p)$ and $P(\theta|h_c)$ are both Gaussian terms centered on the average direction θ=0 which either enhance (h_p) or inhibit (h_c) the pattern to support integration or segregation, respectively. These are given by Equations (7) and (8) and illustrated in **Figure 2**.

$$P\left(\theta\middle|h_p\right) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{(\theta)^2}{2\sigma_P^2}\right) \quad (7)$$

$$P\left(\theta\middle|h_c\right) = 1 - \left(\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{(\theta)^2}{2\sigma_C^2}\right)\right) \quad (8)$$

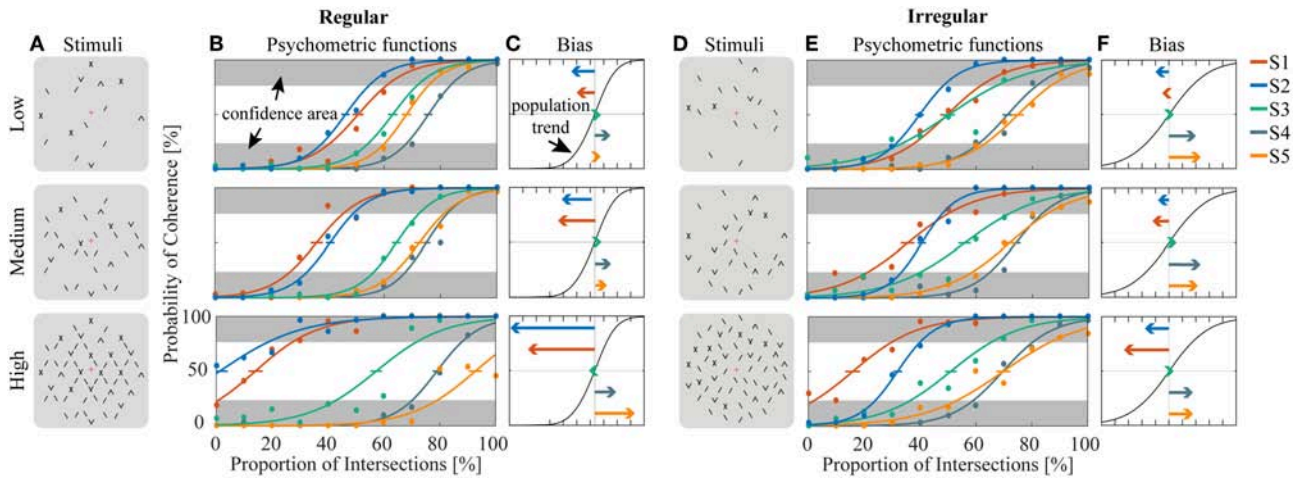


FIGURE 3 | Estimation of bias and stability for regular **(A–C)** and irregular **(D–F)** experiments. **(A,D)** Cartoons of the stimuli across density conditions for the regular **(A)** and irregular **(D)** experiments. **(B,E)** Fitted psychometric functions for each subject across density conditions for the regular **(B)** and irregular **(E)** experiments. The error bar on each psychometric function is the standard error of mean estimated by bootstrapping processing by resampling 400 times. Confidence area (in gray) was defined as where the probability of coherent or transparent perception was higher than 75%. **(C,F)** The direction and amplitudes of bias for each subject corresponding to the conditions of **(B)** and **(E),** respectively.

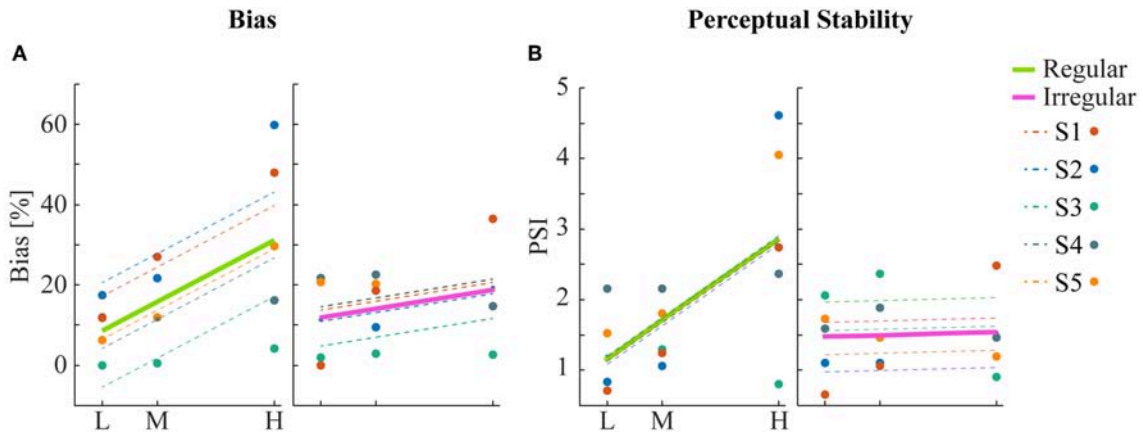


FIGURE 4 | Statistical analysis of bias and perceptual stability. **(A)** Mean bias across subjects for each condition. Linear regression analysis shows a significant correlation between bias and density for the regular but not for the irregular condition. **(B)** Perceptual stability index (PSI, see text) across subjects for each condition. Significant linear correlation between PSI and density was found only for the regular but not for the irregular condition.

The prior which acts to enhance the vector average direction of Equation (7) is consistent with a previously proposed slow speed prior which has been demonstrated to explain illusory perception for a range of ambiguous motion stimuli (Weiss et al., 2002). The prior inhibiting the part of the direction space where the average lies is a novel contribution in the current work and is consistent with observations of motion repulsion effects which push direction estimates away from the averages of transparent component directions (Mahani et al., 2005; Meso et al., 2016a). Simulated trials are used to generate psychometric data to study the interaction of sensory motion representations and prior distributions that is most consistent with each participant's performance.

Simulations

In each trial, assuming a two-step hierarchical process, an MLE estimate based on reduced draws of direction samples of Equation (6) (i.e., 20% of 5,000 used for the full simulation) was used to compute χ based on the distance between the peak of the direction distribution θ_{MAX} and the pattern/zero direction. We note that we adopted the convention of making the vertical direction the zero direction so that the component directions flanked this on either side as $\pm 60^{\circ}$. Having fixed directions rather than fully randomizing stimulus directions across space over trials simplifies the process of computing the thresholds of Equation (10). The initial estimation of χ varied with a logistic type non-linear probability as the percentage of LI apertures went

from 0 to 100. Slope depended on the likelihood parameters and the PSE ($P = 0.5$) was influenced by the relative widths of the pair of likelihoods. This step captures an implicit categorical decision taken when the stimulus is interpreted at onset using the formulation

$$\theta_{MAX} = argmax(P(\theta_m)) \tag{9}$$

$$\chi = \begin{cases} 1, \; if \; -\frac{\theta_L}{2} < \theta_{MAX} < \frac{\theta_L}{2} \\ 2, \; if \; |\theta_{MAX}| > \frac{\theta_L}{2} \end{cases} \tag{10}$$

With χ determined, the posterior of Equation (3) is then simulated using the model selection estimates of Equation (4) or (5) which eliminate the redundant term. Five-thousands draws of direction samples are then used for each trial, binned into a discrete probability distribution with a 0.5° bin resolution. A MAP estimation computes a direction θ_i for each single trial i, from which a second forced choice decision for the simulated trial is made. Transparent or coherent is selected based on the maximum direction (T: $\theta_S/2 < |\theta_i|$ or C: $\theta_S/2 > |\theta_i|$) in a similar way to Equation (10). The estimates used to make the categorical decisions assume symmetry across the direction space for simplicity and therefore search for one peak which could be near the pattern direction or within either transparent component, both left and right.

Each simulated trial had a fixed set of stimulus parameters, $\theta_S = 60°$ and $\theta_L = 0°$. The two sensory likelihood parameters σ_S and σ_L along with the relevant prior parameters σ_P or σ_C [for M2 or M3] were used to generate psychometric functions for comparison to the empirical psychometric functions for each participant under all six conditions. The best fitting parameters [σ_S, σ_L *and* σ_P/σ_C] were obtained using an iterative Kullback-Leibler minimization to search the simulated parameter space. Fits to the data were compared across models using Akaike information criterion (Akaike, 1981).

RESULTS

Human psychophysics experiments were performed using novel bi-stable line-plaid stimuli (**Figures 1B,C**). Subjects were instructed to report their perception of either a coherent pattern moving upward, or two transparent surfaces sliding over each other in leftward and rightward oblique directions (see Methods). Inspired by the geometric properties of typically used moving line-plaids (**Figure 1A**) (Adelson and Movshon, 1982; Pack et al., 2003) and the architecture of the visual system with very small receptive fields (RFs) in early visual areas, we developed this novel stimulus by decomposing the plaid into two types of local stimulus elements we refer to as apertures: separated lines (*SL*) and line intersections (*LI*). In this way, the stimuli could mimic two basic inputs that the visual system could experience locally: 1D- or 2D-motion (green/red apertures, respectively, in **Figure 1A**) based on the dimensions of the features within the aperture. We performed two experiments with the only difference being the positioning of apertures: in Experiment 1 (regular, R) the structure of the mimicked plaid was maintained (**Figure 1B**), whereas in Experiment 2 (irregular, I) the element apertures were spatially jittered (**Figure 1C**). All subjects could consistently fixate within a circular window with radius 0.4 degrees of visual angle (**Figure S1**). For each subject, we first estimated the relative bias toward one of the two possible percepts (transparent or coherent), by calculating the difference between the 50% coherence threshold taken from its fitted psychometric function and the same threshold calculated from the low-density population trend that was used as a reference (**Figure 3**). Interestingly, for higher stimulus densities we observed gradual increases in the bias and this effect was more pronounced in Experiment 1 (Regular) in comparison to Experiment 2 (Irregular). Statistical analysis was performed using a linear mixed effects model approach with the bias as independent variable and density and regularity as fixed effects. Subjects were considered as a random effect thus allowing for different intercepts in the model (**Figure 4A**). Statistical significance was evaluated after parameter estimation using an F-test for the fixed effects with density being significant ($F_{(22)} = 11.83$, $P = 0.0023$) while the interaction between density and regularity remained a trend ($F_{(22)} = 3.32$, $P = 0.0822$). Regularity as a main effect was not significant ($F_{(22)} = 1.11$, $P = 0.3$) indicating that on average the two experiments showed comparable biases.

To obtain a quantitative estimate of the stability of the two percepts for each condition, a perceptual stability index (PSI, **Figure 4B**) was calculated for each subject as follows: first, we defined as perceptually stable the stimuli that resulted in either coherent or transparent perception with probability over 75% (i.e., see the shaded areas in either side of the psychometric curve with $P_{\text{coherent}} < 25\%$ or $P_{\text{coherent}} > 75\%$ in **Figure 3**). Then, the PSI was calculated as the fraction of fitted data-points within the side of the confidence area corresponding to the dominant percept, and the rest of the points (**Figure 4B**). Similar linear mixed effects modeling analysis as for the bias was then performed with the PSI as independent variable. The results showed a significant main effect of density ($F_{(22)} = 6.38$, $P = 0.0193$) as well as significant interaction between density and regularity ($F_{(22)} = 5.55$, $P = 0.0278$). Regularity as a main effect was not significant ($F_{(22)} = 1.88$, $P = 0.18$).

To study the relative contribution of prior experience and sensory representation to the processing of the ambiguous motion direction, we modeled the underlying motion perception task using a Bayesian causal inference framework (Sato et al., 2007; Stocker and Simoncelli, 2007; Shams and Beierholm, 2010). To this end, we used models of increasing complexity (no prior, a transparent prior or a coherent prior, and as a control a model with the use of both priors). In the simplest model architecture (M1, no prior), the maximum likelihood was estimated and categorized depending on whether it was closer to the coherent or transparent direction. For models M2 and M3, a hierarchical sequential computation was assumed and on each simulated trial an initial noisy direction estimate χ, was used to determine whether to apply an excitatory (M2, run as a separate independent simulation from M3) or an inhibitory (M3, run separate from M2) prior, each of which required a single additional Gaussian width parameter centered on the average direction. These would have an effect of shifting posterior probabilities to bias perception either toward coherent (M2) or transparent (M3).

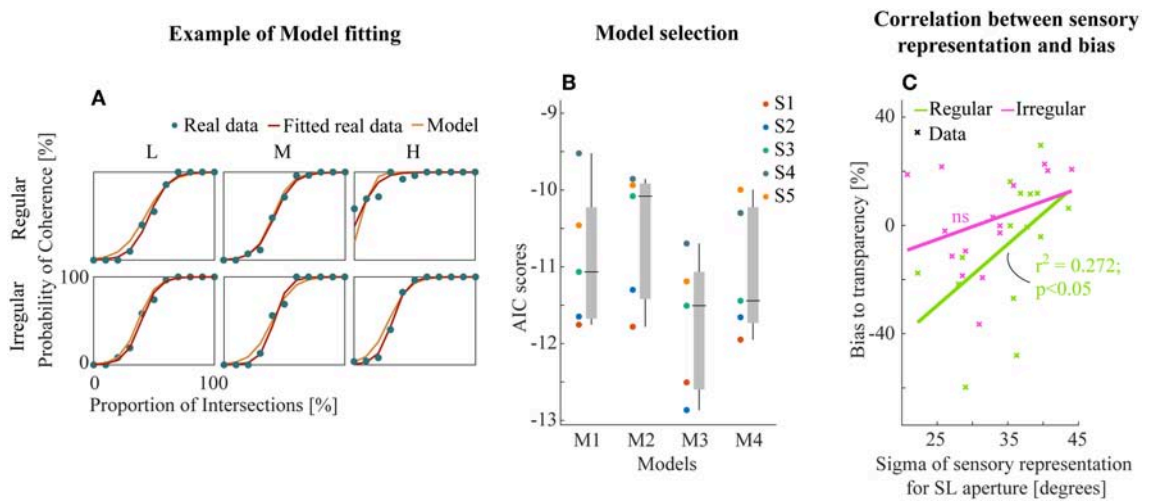


FIGURE 5 | (A) Example model fitting results from a representative subject. Empirical and model simulated psychometric functions were plotted for each experiment and conditions. **(B)** The raw AIC scores of all models (see text). Each subjects (left) and box plot (median ± s.e.m., right). **(C)** Linear correlation between the sensory representation of SL aperture and the amount of bias to transparent perception. Data points were collapsed across all density conditions of regular/irregular experiments.

Last, in a control condition, a model M4 was simulated by using the best fitting M2/3 parameters and therefore included separate optimal priors for separation and integration. Motion direction was represented as a linear combination of Gaussian probability density functions representing the LI and SL aperture direction and variance (**Figure 2**; also see Methods). The set of models, M1–M4 were tasked with a forced choice decision on whether each simulated trial corresponded to transparent or coherent, over a number of conditions recreating Experiments 1 and 2.

Example model-fitting results for a representative subject are shown in **Figure 5A** (results for all subjects in **Figure S2**). We then performed model comparison based on the Akaike criterion measures (AIC, Akaike, 1981) to identify the optimal model architecture. The AIC measurements use likelihoods from the fitting residuals to determine which model provides the best explanation for the data, giving a lower score for better fits but penalizing models with more parameters. M3 (transparent prior) was found to be the most appropriate model for the data set based on AIC scores (**Figure 5B**). This suggests a general tendency within the visual system toward separating motion components unless there is strong sensory evidence for integration into a single object (here provided by the line intersections (LI) apertures).

Further, we analyzed the relationship between the best model parameters of M3 and perceptual bias from empirical data to investigate the potential insights into sensory mechanisms of subjective biases. We found a significant linear correlation between the bias and the variability of sensory representation (Gaussian likelihoods) for SL apertures ($r^2 = 0.272$, $p < 0.05$, **Figure 5C**) only for the regular experiment suggesting that regularity influences the effectiveness of the sensory representation by decreasing variance. There were no similar trends in the fitted parameters for LI sensory likelihoods and the prior (**Figure S3**).

DISCUSSION

In this study, we used bi-stable motion perception as a tool to understand processes of perceptual stabilization in the human brain. We used a Bayesian causal inference framework (Sato et al., 2007; Stocker and Simoncelli, 2007; Shams and Beierholm, 2010) to model the internal decision process leading to one of the two alternative interpretations with the aim to understand the relative role of priors and sensory evidence in the selection process. We found, counter-intuitively, that adding more motion information by increasing the number of apertures increased response biases in the task. Individuals' tendencies to either one or the other of the percepts were amplified substantially when we increased the density of stimulus apertures. This led to an increased inter-subject variability, with each subject diverging from the population trend with a magnitude and direction that was related to their original bias (**Figure 4A**). Interestingly, this effect was largely abolished in the irregular condition when the position of elements was jittered with respect to their original location, indicating that this form of contextual organization created by spatial regularity played a major role in the amplification of the bias. As a measure of the effect of bias amplification, we computed a perceptual stability index and found that it linearly increased for higher element density.

To further understand the brain processes leading to this result, we adapted hierarchical motion perception models that posit sequential stages of brain processing including local motion detection, global combination of these local signals and then an interpretation of the representation to support

categorical/qualitative decisions. This broad mechanistic view is widely supported by evidence in the literature for both psychophysics and physiology (Burr and Thompson, 2011; Nishida, 2011). In the context of our work, the representation of the local motion information can be reflected directly in the neural responses in directionally selective areas such as MT/MST, however, one of the classic difficulties of motion transparency perception is how such a local representation can be transformed into the qualitative percept (e.g., see Qian et al., 1994; Treue et al., 2000; Meso and Zanker, 2009). To this end, and in particular with respect to prior information encoded in the brain of each participant, we built a battery of Bayesian models (M1–M4; see Methods) with the task to probabilistically select one of the two percepts on a trial-by-trial basis simulating the experiments. These modeled the sensory representations of the 1D- and 2D-motion input-signals as Gaussian processes each with separate sigma likelihood parameters and, in addition, one of four different prior probability configurations. M3 (which included a segregation prior) provided the best model, suggesting that the visual system selectively applies an inhibition within the direction space to help separate components. Importantly, it should be noted that M3 was the better model even in subjects that were biased toward coherent percepts. We conjecture, that the brain when faced with such tasks applies a conditional implementation of separating priors on some critical trials (Zamboni et al., 2016) and not an integrating one because integration might arise naturally from overlapping signal distributions (Mahani et al., 2005). The proposed hierarchical computation extends recent findings in which participants performed an orientation discrimination followed by an orientation estimation task, with the discrimination found to influence the estimation task (Luu and Stocker, 2018). A similar effect had been found for motion stimuli (Zamboni et al., 2016) with a need for self-consistency proposed as an explanation. We argue that this hierarchical two-step computation might occur during our task, with an implicit early categorical decision needed to resolve the ambiguity resolution known to occur early in motion stimuli (Meso et al., 2016a). In the implemented model, for simplicity, fixed directions were explicitly associated with the categorical decisions. Similar models could be implemented in the future in which, the decision need not be based on the absolute directions but reached based on the distribution of global motion directions after pooling (i.e., a bimodal distribution would signify transparency and a unimodal coherence). In that case, the future tested priors could be adjusted and made independent of direction for example by acting broadly as an attractor or repellant of nearby directions.

Bias stands at the core of signal detection theory (SDT) when applied to both living organisms and machines. In fact, (Green and Swets, 1966), being the first to develop SDT approaches in psychophysics, directly criticized previously used methods for not being able to separate the sensitivity of subjects from their potential biases. In addition to the principle problem of detecting signal within noise, our brains also face the problem of inherently ambiguous sensory inputs. Thus, to make veridical interpretations of the outside world, the brain needs to employ additional mechanisms such as attention and prior experience (Knill and Richards, 1996; Desimone, 1998; Rao et al., 2002; Beck and Kastner, 2009; Meso et al., 2016a). One theory suggested that objects simultaneously presented in the visual field compete and attention can bias the outcome of this competition (Desimone and Duncan, 1995; Desimone, 1998; Beck and Kastner, 2009). Our results are consistent with the general framework of the biased competition hypothesis; however, attention does not seem to be the primary source of the observed biases as there is no reason to expect attention to vary systematically across the different density or regularity conditions. The subjects had to continuously perform the task of reporting their percepts in randomized trials within blocks so attention should have remained largely constant. Moreover, individual bias directions were independent of the stimulus configuration (which was the same for all subjects) precluding bottom-up stimulus driven attention effects. The subject specific results suggested a strong influence of prior experience or assumptions and thus we expected our modeling results might reveal that some subjects would use a "coherence" prior (M2) while others a "transparency" prior (M3). To our surprise, M3 (in comparison to M2; **Figure 5B**) was a better model for all our subjects, including those with biases toward coherence. This suggests that the sensitivity of the visual system of each participant to the two motion signals (sensory σ) was more important for determining bias direction in comparison to the integration prior. We conjecture that motion direction integration based on sensory likelihoods maybe the default processing mode with conditional priors inhibiting integration employed in order to help motion segmentation and transparency perception.

Furthermore, bias in our experiments was increased with stimulus element density. This was also an unexpected finding, as previous studies have shown that increases in the density of random-dot-kinematograms (RDKs) result in coherence thresholds also decreasing (Barlow and Tripathy, 1997) or being unaffected (Eagle and Rogers, 1997; Talcott et al., 2000; Welchman and Harris, 2000). We note, however, that RDK experiments are closer to the foundations of SDT (i.e., detecting signal within noise). We propose that in our scenario, competition between the two motion representations may be enhanced by density increments resulting in the observed increase of the bias toward a preferred representation which would act like a perceptual attractor, an area within the direction space where probability increases at higher densities. This is consistent with reports in previous literature where contrast-based motion signal increases resulted in stronger 2D motion attractors compared to 1D directions in a tri-stable ambiguous motion stimulus (Meso et al., 2016b). In addition, research with RDKs demonstrated that coherence thresholds in 5–6-year olds were (a) much higher, and (b) decreased with dot density in comparison to adults (Narasimhan and Giaschi, 2012). In our view, this provides evidence for coherent perception or integration as the earliest unelaborated default computation and with perhaps the connectivity of the underlying neural circuitry prone to changes by experience during development. This could explain the different directions of the biases in different subjects.

Interestingly, the bias-amplification and the increases in the perceptual stability index with density were largely abolished in the irregular stimuli with jittered aperture positions. This is consistent with previous work demonstrating the importance of regularity (Morgan et al., 2012; Ouhnana et al., 2013) which appears to play a role in the selection of stable neural representations. Another interpretation is that reduction of regularity eliminates in parallel the correspondence of the single stimulus elements to the underlying patterns or "objects," interfering with their spatial integration. This is consistent with studies that have demonstrated a precedence of global features in visual perception (Beck and Kastner, 2005; Phillips et al., 2015; Ding et al., 2017). Moreover, the profound influence of position jitter on the bias indicates that the scale of the integration cannot be completely local nor global as in that case the regular/irregular conditions should not elicit an effect. These results directly indicate that the motion integration mechanisms contributing to individual biases are of "meso-scale" i.e., go beyond single-neuron receptive fields (RFs) in V1 to scales more typical for area V5/MT but not the very large RFs found in size-invariant object selective areas like inferotemporal cortex (IT).

Previous research has found strong evidence for active perceptual stabilization mechanisms in the visual system, such as reorganization of sensory representation during intermittent viewing (Leopold et al., 2002); top-down modulation of beta-band synchronization (Kloosterman et al., 2015); feedforward inhibition (Bollimunta and Ditterich, 2012) arousal (Mather and Sutherland, 2011; de Gee et al., 2014); and memory (Wimmer and Shohamy, 2012). Our study suggests that bias serves as an additional factor our brains actively use to stabilize our perception of the world.

ETHICS STATEMENT

This study was carried out in accordance with the recommendations of the ethical guidelines, University of Tuebingen with written informed consent from all subjects. All subjects gave written informed consent in accordance with the Declaration of Helsinki. The protocol was approved by the ethical committee of the University of Tuebingen.

AUTHOR CONTRIBUTIONS

QL and GK conceived and designed the psychophysics experiments. QL performed psychophysics experiments and analyzed all data. AM developed the models. QL, AM, and GK run model simulations and wrote the manuscript. NL supported the study and provided experimental equipment. GK supervised the study. All authors interpreted the experimental results and contributed to the final manuscript and gave final approval for publication.

FUNDING

This work was supported by the Max Planck Society, the German Federal Ministry of Education and Research (BMBF; FKZ: 01GQ1002), and a BOF DOCPRO1 (FFB150293) to GK and QL from the University of Antwerp.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2019.00523/full#supplementary-material

Supplementary Figure S1 | Eye movement results. Each subplot shows the averaged eye movement results of each subject from regular/irregular conditions.

Supplementary Figure S2 | Psychometric functions. Empirical and simulated psychometric functions were plotted for each experiment and condition.

Supplementary Figure S3 | The amount of bias of transparent perception is not correlated with sensory representation of LI aperture (regular condition: r: −0.43, p: 0.10; irregular condition: r: 0.10, p: 0.70), nor with prior (regular condition: r: −0.14, p: 0.60; irregular condition: r: 0.06, p: 0.81)

Supplementary Movie 1 | Regular stimuli with representative three density conditions (100, 50 and 0% of LI from in total 500 apertures each, the same as Movie 2).

Supplementary Movie 2 | Irregular stimuli. Note that the demo movies were not used for real experiments.

REFERENCES

Adelson, E. H., and Movshon, J. A. (1982). Phenomenal coherence of moving visual patterns. *Nature* 300, 523–5. doi: 10.1038/300523a0

Akaike, H. (1981). Likelihood of a model and information criteria. *J. Econom.* 16, 3–14. doi: 10.1016/0304-4076(81)90071-3

Amano, K., Edwards, M., Badcock, D. R., and Nishida, S. (2009). Adaptive pooling of visual motion signals by the human visual system revealed with a novel multi-element stimulus. *J. Vis.* 9, 4.1–25. doi: 10.1167/9.3.4

Amano, K., Takeda, T., Haji, T., Terao, M., Maruya, K., Matsumoto, K., et al. (2012). Human neural responses involved in spatial pooling of locally ambiguous motion signals. *J. Neurophysiol.* 107, 3493–3508. doi: 10.1152/jn.00821.2011

Barlow, H., and Tripathy, S. P. (1997). Correspondence noise and signal pooling in the detection of coherent visual motion. *J. Neurosci. 17*, 7954–66. doi: 10.1523/JNEUROSCI.17-20-07954.1997

Beck, D. M., and Kastner, S. (2005). Stimulus context modulates competition in human extrastriate cortex. *Nat. Neurosci.* 8, 1110–1116. doi: 10.1038/nn1501

Beck, D. M., and Kastner, S. (2009). Top-down and bottom-up mechanisms in biasing competition in the human brain. *Vis. Res.* 49, 1154–1165. doi: 10.1016/j.visres.2008.07.012

Bollimunta, A., and Ditterich, J. (2012). Local computation of decision-relevant net sensory evidence in parietal cortex. *Cereb. Cortex* 22, 903–17. doi: 10.1093/cercor/bhr165

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vis.* 10, 433–6. doi: 10.1163/156856897X00357

Burr, D. C., and Thompson, P. G. (2011). Motion psychophysics: 1985-2010. *Vis. Res.* 51, 1431–1456. doi: 10.1016/j.visres.2011.02.008

de Gee, J. W., Knapen, T., and Donner, T. H. (2014). Decision-related pupil dilation reflects upcoming choice and individual bias. *Proc. Natl. Acad. Sci. U.S.A.* 111, E618–E625. doi: 10.1073/pnas.1317557111

Desimone, R. (1998). Visual attention mediated by biased competition in extrastriate visual cortex. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 353, 1245–1255. doi: 10.1098/rstb.1998.0280

Desimone, R., and Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annu. Rev. Neurosci.* 18, 193–222. doi: 10.1146/annurev.ne.18.030195.001205

Ding, S., Cueva, C. J., Tsodyks, M., and Qian, N. (2017). Visual perception as retrospective Bayesian decoding from high- to low-level features. *Proc. Natl. Acad. Sci. U.S.A.* 114, E9115–E9124. doi: 10.1073/pnas.1706906114

Eagle, R. A., and Rogers, B. J. (1997). Effects of dot density, patch size and contrast on the upper spatial limit for direction discrimination in random-dot kinematograms. *Vis. Res.* 37, 2091–102. doi: 10.1016/S0042-6989(96)00153-8

Feldman, N. H., Griffiths, T. L., and Morgan, J. L. (2009). The influence of categories on perception: explaining the perceptual magnet effect as optimal statistical inference. *Psychol. Rev.* 116, 752–782. doi: 10.1037/a0017196

Green, D. M., and Swets, J. A. (1966). *Signal Detection Theory and Psychophysics*. Wiley. Available online at: https://books.google.de/books?id=Ykt9AAAAMAAJ

Hupé, J.-M., and Rubin, N. (2003). The dynamics of bi-stable alternation in ambiguous motion displays: a fresh look at plaids. *Vis. Res.* 43, 531–48. doi: 10.1016/S0042-6989(02)00593-X

Kloosterman, N. A., Meindertsma, T., Hillebrand, A., van Dijk, B. W., Lamme, V. A. F., and Donner, T. H. (2015). Top-down modulation in human visual cortex predicts the stability of a perceptual illusion. *J. Neurophysiol.* 113, 1063–1076. doi: 10.1152/jn.00338.2014

Knill, D. C., and Richards, W. (1996). *Perception as Bayesian Inference*. Cambridge University Press. Available online at: https://books.google.de/books?id=_cTLCgAAQBAJ

Körding, K. P., Beierholm, U., Ma, W. J., Quartz, S., Tenenbaum, J. B., and Shams, L. (2007). Causal inference in multisensory perception. *PLoS ONE* 2:e943. doi: 10.1371/journal.pone.0000943

Leopold, D. A., Wilke, M., Maier, A., and Logothetis, N. K. (2002). Stable perception of visually ambiguous patterns. *Nat. Neurosci.* 5, 605–609. doi: 10.1038/nn0602-851

Luu, L., and Stocker, A. A. (2018). Post-decision biases reveal a self-consistency principle in perceptual inference. *Elife* 7:e33334. doi: 10.7554/eLife.33334

Mahani, A. S., Carlsson, A. E., and Wessel, R. (2005). Motion repulsion arises from stimulus statistics when analyzed with a clustering algorithm. *Biol. Cybern.* 92, 288–291. doi: 10.1007/s00422-005-0556-0

Mather, M., and Sutherland, M. R. (2011). Arousal-Biased Competition in Perception and Memory. *Perspect. Psychol. Sci.* 6, 114–133. doi: 10.1177/1745691611400234

Meso, A. I., Haruhana, K., Masson, G. S., and Gardner, J. L. (2016a). "Repulsion of perceived visual motion direction as an emergent property of deciding to unify or segregate sources," in *Program No. 54.20/AA8. 2016 Neuroscience Meeting Planner* (San Diego, CA: Society for Neuroscience).

Meso, A. I., Rankin, J., Faugeras, O., Kornprobst, P., and Masson, G. S. (2016b). The relative contribution of noise and adaptation to competition during tri-stable motion perception. *J. Vis.* 16:6. doi: 10.1167/16.15.6

Meso, A. I., and Zanker, J. M. (2009). Perceiving motion transparency in the absence of component direction differences. *Vis. Res.* 49, 2187–2200. doi: 10.1016/j.visres.2009.06.011

Moreno-Bote, R., Shpiro, A., Rinzel, J., and Rubin, N. (2010). Alternation rate in perceptual bistability is maximal at and symmetric around equi-dominance. *J. Vis.* 10:1. doi: 10.1167/10.11.1

Morgan, M. J., Mareschal, I., Chubb, C., and Solomon, J. A. (2012). Perceived pattern regularity computed as a summary statistic: implications for camouflage. *Proc. Biol. Sci.* 279, 2754–2760. doi: 10.1098/rspb.2011.2645

Narasimhan, S., and Giaschi, D. (2012). The effect of dot speed and density on the development of global motion perception. *Vis. Res.* 62, 102–107. doi: 10.1016/j.visres.2012.02.016

Nishida, S. (2011). Advancement of motion psychophysics: review 2001-2010. *J. Vis.* 11, 11. doi: 10.1167/11.5.11

Odegaard, B., and Shams, L. (2016). The brain's tendency to bind audiovisual signals is stable but not general. *Psychol. Sci.* 27, 583–591. doi: 10.1177/0956797616628860

Ouhnana, M., Bell, J., Solomon, J. A., and Kingdom, F. A. A. (2013). Aftereffect of perceived regularity. *J. Vis.* 13:18. doi: 10.1167/13.8.18

Pack, C. C., Livingstone, M. S., Duffy, K. R., and Born, R. T. (2003). End-stopping and the aperture problem: two-dimensional motion signals in macaque V1. *Neuron* 39, 671–80. doi: 10.1016/S0896-6273(03)00439-2

Phillips, W. A., Clark, A., and Silverstein, S. M. (2015). On the functions, mechanisms, and malfunctions of intracortical contextual modulation. *Neurosci. Biobehav. Rev.* 52, 1–20. doi: 10.1016/j.neubiorev.2015.02.010

Qian, N., Andersen, R. A., and Adelson, E. H. (1994). Transparent motion perception as detection of unbalanced motion signals. III. Modeling. *J. Neurosci.* 14, 7381–7392. doi: 10.1523/JNEUROSCI.14-12-07381.1994

Rao, R. P. N., Olshausen, B. A., Lewicki, M. S., Jordan, M. I., and Dietterich, T. G. (2002). *Probabilistic Models of the Brain: Perception and Neural Function*. MIT Press. Available online at: https://books.google.de/books?id=mzBlvComcqwC

Rauber, H.-J., and Treue, S. (1999). Revisiting motion repulsion: evidence for a general phenomenon? *Vis. Res.* 39, 3187–3196. doi: 10.1016/S0042-6989(99)00025-5

Sato, Y., Toyoizumi, T., and Aihara, K. (2007). Bayesian inference explains perception of unity and ventriloquism aftereffect: identification of common sources of audiovisual stimuli. *Neural Comput.* 19, 3335–3355. doi: 10.1162/neco.2007.19.12.3335

Shams, L., and Beierholm, U. R. (2010). Causal inference in perception. *Trends Cogn. Sci.* 14, 425–432. doi: 10.1016/j.tics.2010.07.001

Stocker, A. A., and Simoncelli, E. P. (2007). A bayesian model of conditioned perception. *Adv. Neural Informat. Process. Syst.* 2007, 1409–1416.

Stoner, G. R., Albright, T. D., and Ramachandran, V. S. (1990). Transparency and coherence in human motion perception. *Nature* 344, 153–155. doi: 10.1038/344153a0

Takahashi, N. (2004). Effect of spatial configuration of motion signals on motion integration across space. *Swiss J. Psychol.* 63, 173–182. doi: 10.1024/1421-0185.63.3.173

Talcott, J. B., Hansen, P. C., Assoku, E. L., and Stein, J. F. (2000). Visual motion sensitivity in dyslexia: evidence for temporal and energy integration deficits. *Neuropsychologia* 38, 935–43. doi: 10.1016/S0028-3932(00)00020-8

Thiele, A., and Stoner, G. (2003). Neuronal synchrony does not correlate with motion coherence in cortical area MT. *Nature* 421, 366–370. doi: 10.1038/nature01285

Treue, S., Hol, K., and Rauber, H. J. (2000). Seeing multiple directions of motion-physiology and psychophysics. *Nat. Neurosci.* 3, 270–276. doi: 10.1038/72985

Weiss, Y., Simoncelli, E. P., and Adelson, E. H. (2002). Motion illusions as optimal percepts. *Nat. Neurosci.* 5, 598–604. doi: 10.1038/nn0602-858

Welchman, A. E., and Harris, J. M. (2000). The effects of dot density and motion coherence on perceptual fading of a target in noise. *Spatial Vis.* 14, 45–58. doi: 10.1163/156856801741350

Wimmer, G. E., and Shohamy, D. (2012). Preference by association: how memory mechanisms in the hippocampus bias decisions. *Science* 338, 270–273. doi: 10.1126/science.1223252

Zamboni, E., Ledgeway, T., McGraw, P. V., and Schluppeck, D. (2016). Do perceptual biases emerge early or late in visual processing? Decision-biases in motion perception. *Proc. R. Soc. B Biol. Sci.* 283:20160263. doi: 10.1098/rspb.2016.0263

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Neuroscience

REVIEW
published: 16 August 2019
doi: 10.3389/fnins.2019.00865

Reconciling Color Vision Models With Midget Ganglion Cell Receptive Fields

*Sara S. Patterson[1,2], Maureen Neitz[1] and Jay Neitz[1]**

[1] Department of Ophthalmology, University of Washington, Seattle, WA, United States, [2] Neuroscience Graduate Program, University of Washington, Seattle, WA, United States

Midget retinal ganglion cells (RGCs) make up the majority of foveal RGCs in the primate retina. The receptive fields of midget RGCs exhibit both spectral and spatial opponency and are implicated in both color and achromatic form vision, yet the exact mechanisms linking their responses to visual perception remain unclear. Efforts to develop color vision models that accurately predict all the features of human color and form vision based on midget RGCs provide a case study connecting experimental and theoretical neuroscience, drawing on diverse research areas such as anatomy, physiology, psychophysics, and computer vision. Recent technological advances have allowed researchers to test some predictions of color vision models in new and precise ways, producing results that challenge traditional views. Here, we review the progress in developing models of color-coding receptive fields that are consistent with human psychophysics, the biology of the primate visual system and the response properties of midget RGCs.

Keywords: primate retina, color vision, color perception, computational vision, linking hypotheses, cone photoreceptor, retinal ganglion cells

OPEN ACCESS

Edited by:
Misha Vorobyev,
The University of Auckland,
New Zealand

Reviewed by:
Pablo De Gracia,
Midwestern University, United States
Jihyun Yeonan-Kim,
National Institutes of Health (NIH),
United States
Andrew B. Metha,
The University of Melbourne, Australia

***Correspondence:**
Jay Neitz
jneitz@uw.edu

Specialty section:
This article was submitted to Perception Science, a section of the journal Frontiers in Neuroscience

Received: *01 January 2019*
Accepted: *02 August 2019*
Published: *16 August 2019*

Citation:
Patterson SS, Neitz M and Neitz J (2019) Reconciling Color Vision Models With Midget Ganglion Cell Receptive Fields. Front. Neurosci. 13:865. doi: 10.3389/fnins.2019.00865

INTRODUCTION

The first stage of visual processing occurs in the retina, an outpost of the brain located at the back of the eye. Under photopic conditions, photons of light are absorbed by three types of cone photoreceptor (**Figure 1A**), processed by five main classes of retinal neuron, then visual signals are conveyed to the brain by the axons of retinal ganglion cells (RGCs; Wässle, 2004). Midget RGCs make up a large majority of all RGCs in the central retina, where each L- and M-cone provides the sole direct input to an ON and OFF midget RGC circuit (**Figure 1C**; Wässle et al., 1990, 1998; Kolb and Marshak, 2003).

The midget RGC receptive field has a center-surround organization (Kuffler, 1953). In the central retina, this receptive field compares the photon catch in the single L- or M-cone center to the photon catch in neighboring L/M-cones in the surround (**Figure 1C**). Since this configuration compares the activity of cones that differ in both spatial location and spectral sensitivity, midget RGCs have been implicated in both color and spatial vision (Schiller et al., 1990; Martin et al., 2011). Mammalian RGCs have been described as acting as feature detectors, with different types showing specificity for motion, form or color conferred by the spatial, spectral, and temporal characteristics of their receptive field (Field and Chichilnisky, 2007; Gollisch and Meister, 2010; Baden et al., 2016). Here, we review evidence for the role of midget RGC receptive fields as the first step for detection of two elementary visual features, (1) hue detectors which encode information about spectral reflectances of surfaces as red, green, blue and yellow percepts, (2) high acuity edge detectors which encode the boundaries of objects as required for form vision.

Because their receptive fields exhibit both spectral and spatial opponency, midget RGCs respond to both chromatic and achromatic edges and thus confound the two (Wiesel and Hubel, 1966). Like all RGCs, midget RGCs encode and transmit information to the brain in binary, as all-or-nothing action potentials. A downstream neuron has no way of knowing, from an individual midget RGC's response, whether the midget RGC responses represent the chromatic or spatial structure of a stimulus. At the level of perception, however, we can distinguish between achromatic and equiluminant chromatic edges, even though individual midget RGCs cannot. How and where the spectral and spatial information encoded by midget RGCs is extracted remains one of the most important unanswered questions of primate vision.

Midget RGCs provide, arguably, the best model for linking low-level receptive fields to perception. Understanding how color and spatial information are encoded may provide insight into general organizational principles employed by neural circuits to parse specific features of a stimulus. Furthermore, restoration of color and spatial vision are an important goal for retinal prosthetics, some of which must replace the upstream circuitry that defines the midget RGC receptive field (Yue et al., 2016). Efforts to restore these fundamental aspects of visual perception may benefit from a better understanding of how they are computed in normal vision.

RECEPTIVE FIELDS

All receptive fields are built from the photoreceptor outputs (**Figure 1A**). The photoreceptors' output encodes a single variable: the number of photons absorbed (Rushton, 1972; Baylor et al., 1987). An important implication is that wavelength and intensity are interchangeable and, under the right conditions, any two lights differing in wavelength can be "substituted silently" for each other (Estevez and Spekreijse, 1982). For example, the probability of photon absorption by an M-cone is the same for 467 and 582 nm lights, thus the response of the M-cone shown in **Figure 1B** to the two lights will be indistinguishable. Meanwhile, a 535 nm light with twice the probability of photon absorption can be matched by doubling the intensity of the 467 nm light.

The visual system extracts information about wavelength and spatial contrast by virtue of receptive fields that compare the outputs of multiple cones. The basic computation for extracting wavelength is a comparison between cones of different spectral types, while spatial contrast requires comparing neighboring cones at different spatial locations, regardless of type (Calkins and Sterling, 1999). The characteristics of receptive fields form the foundation of each color vision model discussed here.

WHAT IS THE OPTIMAL RECEPTIVE FIELD FOR SPATIAL VISION?

Because midget RGCs are implicated in high acuity form vision, any discussion of their color-coding role must also include their role in spatial-coding. The first step of spatial vision requires delineating the boundaries of objects, essentially performing an edge detection task.

Spatial Opponency

By comparing the relative activity of cones at different locations, spatially opponent receptive fields signal spatial contrast rather than raw quantal catch (Srinivasan et al., 1982). For low-level edge detectors, circularly symmetric center and surround receptive fields are optimal and will provide sensitivity to all edges, regardless of their orientation (Marr and Hildreth, 1980).

Spectral Opponency

While spatial vision is sometimes assumed to operate only on light intensity (Marr, 1982; Billock et al., 1996), equiluminant edges are also common in natural scenes (Hansen and Gegenfurtner, 2009). Accordingly, an optimal edge detector would be sensitive to all edges regardless of whether the edge is defined by a change in wavelength or intensity. Thus, an optimal edge-detecting receptive field might not just be spatially opponent, but also spectrally opponent. In this case, the purpose of spectral opponency is not to signal the hue of a surface but rather an edge defined by spectral contrast.

WHAT IS THE OPTIMAL COLOR-CODING RECEPTIVE FIELD FOR HUE PERCEPTION?

In the natural world, most colors we perceive are from lights reflected from objects. The purpose of hue perception is to provide information about the surface reflectance of objects, which, in turn, tells us about their internal contents or state. For example, we know the ripeness of fruit and when children are getting sunburned from their surface colors. However, there are significant challenges to this task. Individual cones themselves are not selective for the *distribution* of wavelengths reflected from a surface. If L-cones are active, light could be coming from a red surface reflecting only long wavelengths, a yellow surface reflecting both middle and long wavelengths, a violet surface reflecting both short and long wavelengths or a white surface reflecting all wavelengths. In addition, information from any individual cone will be further confounded by the spectral characteristics of the illuminant. For example, the amount of illumination from blue sky light relative to direct sunlight changes throughout the day. As a result, the illuminant color can vary from blue to yellow (Foster, 2011; Pauers et al., 2012; Spitschan et al., 2016; Woelders et al., 2018). The ideal receptive fields for serving hue perception would be designed to help extract surface spectral reflectance independent of the illuminant. Here we discuss the features of theoretical receptive fields optimized to overcome the challenges associated with consistently signaling hue, independent of any underlying neural substrates.

Spectrally Opponent

Color vision is the ability to discriminate between different wavelengths, independent of intensity (Jacobs, 2018). Receptive

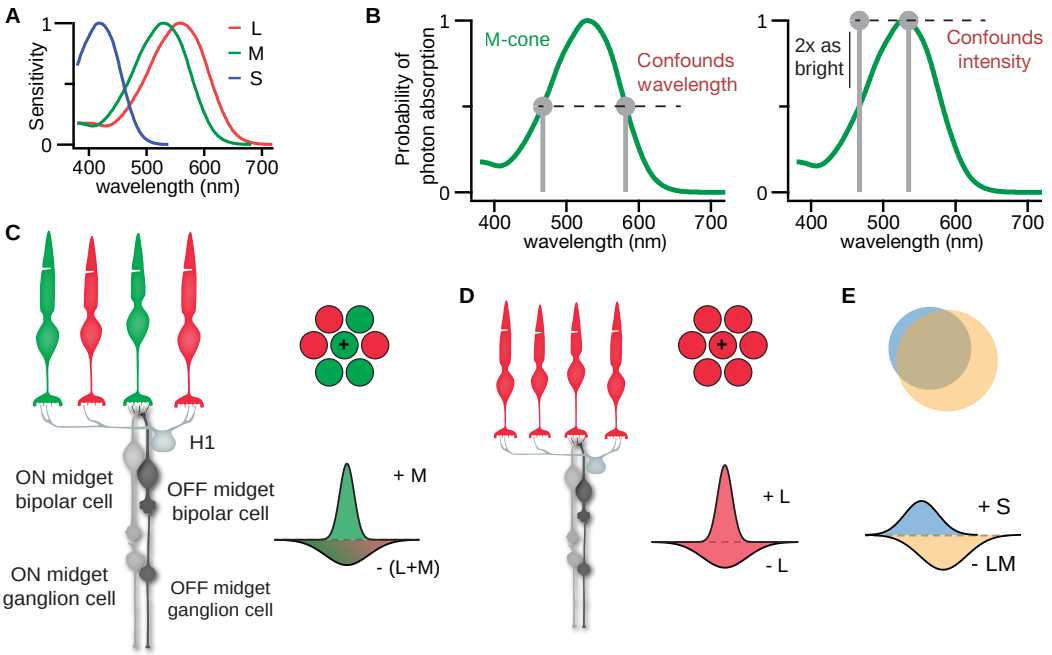


FIGURE 1 | Color-coding receptive fields of the primate retina. **(A)** The normalized spectral sensitivities of human L-, M-, and S-cone photoreceptors. **(B)** Demonstration of monochromatic lights that elicit the same probability of photon absorption, and thus elicit the same response from M-cones. On left, 467 and 582 nm lights are indistinguishable, despite a 115 nm difference. On the right, a 467 and 535 nm lights of different intensities can be confounded by adjusting the intensity of the 467 nm light. **(C)** Circuit diagram of the upstream input to ON and OFF midget RGCs. **(D)** Same as in **(C)**, but for the midget RGCs of a dichromat. **(E)** Small bistratified RGC receptive field diagram illustrating the lack of perfectly coincident spectrally opponent receptive fields as required for pure color cells.

fields with spectrally opponent interactions can extract wavelength information and thus *carry color information* (Paulus and Kroger-Paulus, 1983; Neitz and Neitz, 2011; Chang et al., 2013). However, cone opponent receptive fields are not necessarily optimized for hue perception.

Spatially Coextensive

The first receptive field proposed to create a "pure color cell," was the single opponent receptive field, which exhibits spectral opponency without any spatial opponency (**Figure 2A**). Also called spatially co-extensive or Type II (Wiesel and Hubel, 1966; Crook et al., 2009), this receptive field provides *color selectivity*, the ability to extract spectral information unconfounded by spatial information. Spatially co-extensive, spectrally opponent receptive fields like **Figure 2A** would be theoretically color selective in that they respond to chromatic stimuli, but not achromatic patterns. However, these receptive fields act as simple wavelength detectors and cannot compensate for the changes in illuminant discussed above.

Double Opponency

To consistently signal hue, an optimal color-coding receptive field must compensate for the changes in illuminant discussed above. Double-opponent receptive fields, superimposing two opposing, spectrally and spatially opponent receptive fields (**Figure 2A**) have been proposed to help provide this *color constancy* (Daw, 1973; D'Zmura and Lennie, 1986). Double opponent receptive fields exploit the fact that, in the natural world, hue typically changes abruptly at object boundaries while illumination changes slowly across a visual scene. When the center receives light from the edge of an object surface, some light falling in the surround is reflected from other objects in the scene under the same illuminant. If the illuminant changes to have more long-wavelength light, the increased L-cone stimulation in the center is opposed by greater L-cone stimulation in the surround, and ideally, the change in illumination is removed from the visual signal. Thus, double opponent receptive fields confer sensitivity to chromatic contrast at the edges of objects while remaining relatively insensitive to global changes in illumination.

Trichromatic

Normal humans are trichromats and a special requirement of optimal color coding for trichromats is that the receptive fields must compare all three cone types. This is because for neurons comparing only two out of the three cone types, a change in activity in the unsampled cone will not change the hue signaled by that neuron. For example, an L vs. M opponent neuron without S-cone input, as in **Figure 2A**, cannot discriminate between a red surface reflecting only long wavelengths and a violet surface reflecting both long and short wavelengths (Fuld et al., 1981).

Low Spatial Resolution

If the ideal retina is composed of multiple types of feature detectors, spatial constraints must be considered, and the relative

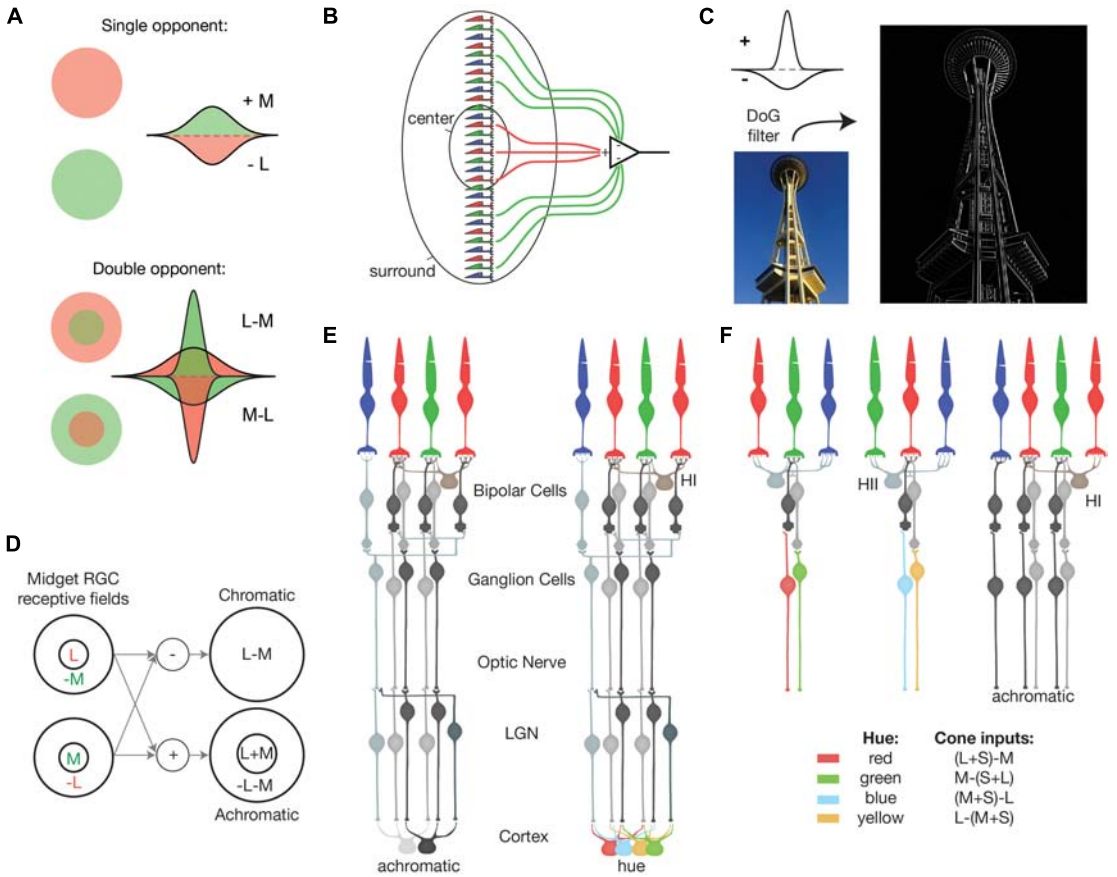


FIGURE 2 | Models of receptive fields encoding form and color vision. **(A)** Diagrams of single and double opponent receptive fields. Adapted from Daw (1973). **(B)** Putative receptive field formed by selective wiring. Adapted from Wiesel and Hubel (1966). **(C)** Edge detection performed by convolution of an image with an achromatic center-surround receptive field, or Difference of Gaussians filter. **(D)** Circuit for "de-multiplexing" chromatic and achromatic information from midget RGC receptive fields. Adapted from Derrico and Buchsbaum (1991). **(E)** Circuit proposed by the multi-stage color model (De Valois and De Valois, 1993). **(F)** Circuit proposed by the parallel processing model (Neitz and Neitz, 2016).

density of any one type should be no higher than required to serve its specific function. The color of a surface tends to be consistent all across it. Thus, in contrast to spatial vision, that requires a high density of detectors to capture the fine details of the shape of objects, hue detectors can accurately capture surface colors using a much lower resolution array of detectors. In summary, the ideal trichromatic hue-encoding system is a relatively sparse array of receptive fields with structures that are double-opponent and receive input from all three types of cones.

INTERPRETING MIDGET RGC RECEPTIVE FIELDS

Early models linking L vs. M midget RGCs to visual perception focused on either spatial or spectral opponency in isolation. Models focusing on their spectral opponency emphasized their potential role in encoding red and green hues. In contrast, models accounting only for achromatic spatial opponency lead to the perspective that spectral opponency is an unintended consequence of trichromacy and may be considered "poor engineering" (Marr, 1982).

Are Midget RGC Receptive Fields Optimal for Hue Perception?

The earliest models followed the first parvocellular LGN (P cell) recordings (De Valois et al., 1966; Wiesel and Hubel, 1966), which have similar receptive field properties as their L vs. M midget RGC inputs. At the time, opponent process theory was still highly controversial (Hurvich and Jameson, 1957) and the discovery of color-opponent neurons in the visual system was groundbreaking. The resulting hypothesis that the parvocellular LGN projections of midget RGCs are responsible for red-green hue perception arguably played a large role in shaping later research. Further, spatial opponency and the resulting responses to achromatic and spatially-structured stimuli were overlooked in many accounts of the physiological basis of hue perception.

In emphasizing, the proposed role of midget RGCs in mediating red-green hue percepts, it was argued that the optimal color-coding receptive field, was one in which an L-cone is

surrounded entirely by M-cones, or vice versa. This receptive field, which would seem to require some cone-specific selective wiring, maximizes the spectral difference between the center and surround to maximally decorrelate the outputs of L- and M-cones' overlapping spectral sensitivities (**Figure 1A**; Buchsbaum and Gottschalk, 1983; Párraga et al., 2002; Sun et al., 2006). The "selective-wiring" model in **Figure 2B** was challenged by theoretical studies demonstrating that mixed L/M-cone receptive fields could generate sufficient spectral opponency (Paulus and Kroger-Paulus, 1983; Lennie et al., 1991). Though still debated by some (Lee, 1996; Wool et al., 2018), there is, at most, only a slight functional bias toward selective wiring (Buzás et al., 2006; Field et al., 2010).

A lack of selective wiring may be one argument against the idea that midget RGCs are optimized for hue perception. However, more importantly, from above, the ideal trichromatic hue-encoding system is a relatively sparse array of receptive fields with structures that are double-opponent and receive input from all three types of cones. The common L vs. M midget RGCs do not conform to any of these theoretical features of hue-encoding neurons. While our theoretical discussion cannot rule out a contribution to hue, we can conclude L vs. M midget RGCs, by themselves, are "non-optimal" for hue perception.

Are Midget RGC Receptive Fields Optimal for Spatial Vision?

Near the fovea, the midget RGC's receptive field center represents the cone providing direct input to the midget bipolar cell, while the surround is formed by feedback from horizontal cells contacting neighboring cones (**Figure 1C**; Verweij et al., 2003). This feedback weights each cone's response by the quantal catch in neighboring cones, essentially subtracting out the mean light level and allowing each individual cone feeding the center of midget RGCs to encode spatial contrast (Jadzinsky and Baccus, 2013). In the central retina, midget RGCs set the limits of human visual acuity (Rossi and Roorda, 2010).

Indeed, theoretical attempts to derive an optimal receptive field for the first step of spatial vision have all converged on the same circularly symmetric center-surround organization (Marr and Hildreth, 1980; Srinivasan et al., 1982; Atick et al., 1992), often modeled as a Difference of Gaussians (Enroth-Cugell and Robson, 1966; Croner and Kaplan, 1995; Dacey et al., 2000). As **Figure 2C** demonstrates, center-surround receptive fields are ideal edge detectors for encoding spatial contrast.

In contrast to early ideas emphasizing their putative role in color perception, more recent research into the evolution of the primate visual system provides a useful context for a modern understanding of L vs. M midget RGC function. Though sometimes compared to the X-cells of the mammalian retina, there is no true homolog to the midget circuit prior to prosimians (Peng et al., 2019). The midget RGC circuitry evolved before uniform trichromacy (Nathans, 1999). In dichromats, for example, with only S- and L-cones, the midget RGC's antagonistic center-surround receptive field functions as an achromatic edge detector by comparing the input of a single L-cone to surrounding L-cones (**Figure 1D**).

Interim Conclusions

The receptive field structure of L vs. M midget RGCs is consistent with a role in edge detection. Their ability to respond to equiluminant edges defined only by wavelength differences makes visible forms that would be otherwise invisible. Spectral opponency can also increase the signal-to-noise ratio for edges defined by both intensity and wavelength. The idea that spectral opponency in L vs. M midget RGCs could enhance edge detection rather than contribute to color perception raises an important point. A response to wavelength changes does not imply a causal role in hue perception. As introduced above, hue perception requires detectors that will not respond to black-white edges.

In conclusion, while it may be arguable whether or not midget L vs. M RGCs are ideal achromatic encoders, it is indisputable that they are far from ideal for red-green hue encoding. This leaves two major unanswered questions: what is the physiological basis for hue perception and what role do midget RGCs play? Several different theories involving both the spectral and spatial aspects of midget RGC receptive fields have been proposed as tentative answers to this question. We next review the two main classes of explanation: multiplexing and parallel processing.

MULTIPLEXING MODELS

The first class of models share the idea that each individual midget RGC does "double duty," carrying information for both color vision and achromatic spatial vision, which are extracted by circuitry at higher levels of processing in the geniculostriate pathway. It has been said that red-green and black-white percepts are "de-multiplexed" by downstream circuits (Boycott and Wässle, 1999; Lennie and Movshon, 2005). The idea of multiplexing originated as an analog to attempts to efficiently compress chromatic and spatial information for color televisions (Ingling and Martinez-Uriegas, 1983; Derrico and Buchsbaum, 1991).

The most common models, summarized in **Figure 2D**, combine the outputs of midget RGCs to perform two main transforms: one to extract spectral information by removing spatial correlations and another to extract achromatic spatial information by removing spectral information. The achromatic channels (L + M) sum L- and M-center midget RGC signals to serve as intensity contrast detectors. The putative chromatic channels (L vs. M) difference L-ON-center with M-ON center receptive fields to produce spatially coincident spectrally opponent receptive fields, as discussed above (**Figure 2A**). Accordingly, achromatic spatial structure will be absent in the chromatic channel, resulting in a low-pass chromatic filter, while the achromatic channel will retain the band-pass spatial tuning necessary for spatial vision.

A separate aspect of one of the best-known versions, the De Valois and De Valois (1993) multi-stage color model, was the need to reconcile the difference in cone inputs measured for L vs. M cone-opponent neurons and the opponent receptive fields required to account for hue perception, illustrated in **Figure 3A**. The four fundamental hue sensations are often assumed to

represent the responses of four groups of hue-encoding neurons. Over the last 50 years, there have been different ideas about the exact nature of the cone inputs to the four fundamental hues. However, a convergence of modern evidence from experiments directly measuring hue perception indicate that all three cone types contribute to each hue in the following combinations: L + S vs. M for red-green and M + S vs. L for blue-yellow, respectively (**Figure 3A**; Wooten and Werner, 1979; Drum, 1989; Webster et al., 2000a; Schmidt et al., 2016).

One of the great insights of the DeValois and DeValois model was that hue perception requires S-cone inputs to L vs. M opponent pathways (Wooten and Werner, 1979; Drum, 1989; Webster et al., 2000b). As an *ad hoc* solution to the discrepancy between L vs. M midget RGCs and the receptive fields required for hue perception, their multi-stage color model proposed that the necessary S-cone input to an L vs. M channel is accomplished by mixing in the outputs of S-cone opponent neurons (**Figure 2E**).

Evaluating the Double Duty Hypothesis

The DeValois and DeValois model was firmly based on the most recent anatomical, psychophysical and physiological results of the time, yet a number of assumptions were necessary where open questions remained. We can now revisit these assumptions in light of the research published in the 25 years since the multi-stage model was first proposed. One example is their explanation of how the required S-cone inputs from small bistratified RGCs are added in the process of building cortical receptive fields for hue perception. More recently, the classification of small bistratified RGCs as single opponent "pure color cells" has been called into question [compare **Figures 1E**, **2A** (Field et al., 2007; Tailby et al., 2010); but see Crook et al. (2009)]. Thus, small bistratified RGCs and their S-ON projections may also confound spatial and spectral information. Moreover, the S-cone ON neurons were later identified as a part of the functionally distinct koniocellular pathway (Martin et al., 1997) and there is no direct evidence for specific circuits combining signals from the koniocellular and parvocellular pathways.

While the theoretical L-M and L + M channels would decorrelate the outputs of midget RGCs, it has been argued that not all decorrelations are created equal (Pitkow and Meister, 2012) and the benefits depend on how these channels are implemented by neural circuitry. In general, however, asking a neuron to perform two jobs simultaneously has been said to ensure that both are done poorly (Sterling and Laughlin, 2017). Moreover, there don't appear to be any true modern examples of multiplexing RGCs involving two functions performed simultaneously. Perhaps the closest parallel is the fact that the same RGCs serve both photopic and scotopic vision, however, these functions are primarily performed separately under different conditions (Field et al., 2009; Grimes et al., 2014). Other examples of multiplexing RGCs involve one stimulus dimension modulating the encoding of another (Deny et al., 2017), however, this is different from two functions being encoded simultaneously.

The "de-multiplexing" multi-stage models are the result of speculation about the type of computation that would be required to produce selective detectors for wavelength and spatial contrast from combinations of spectrally opponent center-surround neurons, however, they lack firm experimental evidence from cortical physiology (Lee, 2008). They have also been criticized from an image compression standpoint, with the argument that decorrelation of chromatic and spatial information is best done early, ideally before transmission through the optic nerve (Derrico and Buchsbaum, 1991). In contrast, an effort to test de-multiplexing models concluded the two dimensions cannot be disentangled in the early visual system (Kingdom and Mullen, 1995). Moreover, the most successful models based on the "double duty" hypothesis do not make predictions about both spatial and spectral responses (Rider et al., 2018).

The assumption that different aspects of color vision are all based on the same underlying neural substrates (e.g., L vs. M midget RGCs) has resulted in a tendency to expect the visual system to somehow extract hue information from the midget RGCs' receptive field output. However, the computational complexity required to separate chromatic from spatial information at subsequent stages of visual processing should not be underestimated. One higher stage is proposed to decorrelate spatial and spectral information, a second higher stage to add the required S-cone input (**Figure 2E**) and yet an additional stage, that has not been incorporated into current de-multiplexing models, to generate the double opponent receptive field structure required to create neurons that are able to contribute to invariant hue-encoding of spectral reflectance.

Multiplexing in the Light of Information Coding in the Retina

The need to compress RGC axons down to a 2 mm cable is often referred to as an "information bottleneck" within the visual system. Proponents of multiplexing models might claim superiority on this account: combining color and spatial information into one RGC could reduce the number axons in the optic nerve without reducing the transmission of information. Indeed, there are about six to seven million cones in a human eye and only about a million optic nerve fibers (Sterling and Laughlin, 2017). However, this represents the situation in the peripheral retina where convergent input from a large number of cones to each RGC results in a huge reduction in visual acuity relative to what could be supported by the cone mosaic. The loss of spatial information from this convergence is never recovered at higher levels in the visual pathway.

At the time multiplexing models were first proposed, a dominant view on the purpose of retinal function was to reduce redundancy and compress visual information to fit through the optic nerve, with the computations defining visual perception occurring in the cortex (Barlow, 1961). However, contrary to the idea of information compression, in the fovea there is a divergence from cones to RGCs such that the ratio is about 3:1 RGCs:cone. Recent work in non-primate animal models has contributed to a growing appreciation for the diversity of RGC types (Wässle, 2004; Baden et al., 2016) and the sophisticated computations occurring within the retina (Gollisch and Meister, 2010; Wienbar and Schwartz, 2018). Even near the primate fovea,

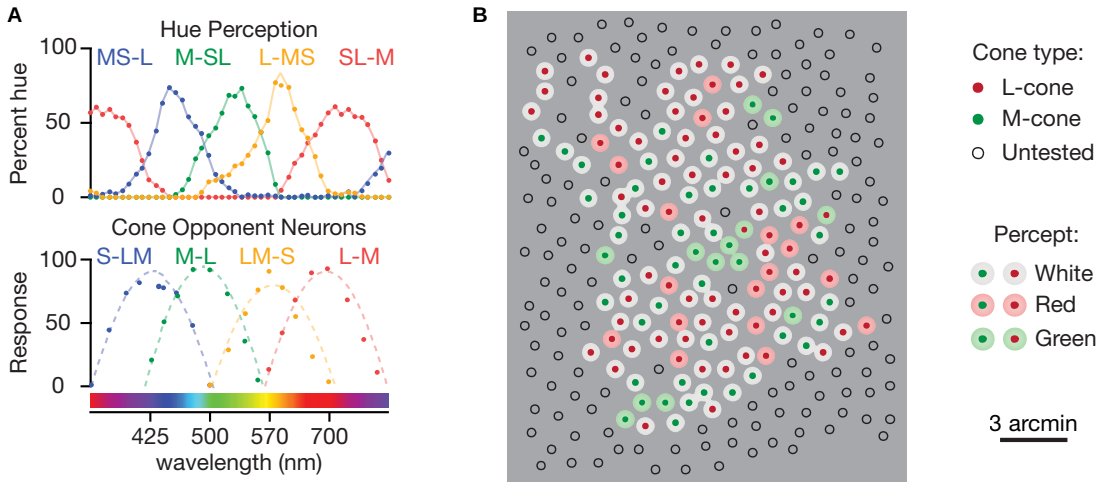


FIGURE 3 | Experiments inspiring the revision of existing color models. **(A)** Spectral sensitivities and the corresponding cone inputs of the mechanisms responsible for red, green, blue, and yellow hues. The data was obtained from a hue scaling experiment, where participants report the percentage of red, green, blue, and yellow. Bottom: Averaged responses of color-opponent LGN neurons, which reflect their color-opponent RGC inputs. Both panels are replotted with wavelength units from De Valois (2004). **(B)** Percepts associated with stimulating individual L- and M-cones in isolation may represent the responses of two types of individual midget RGCs, a larger group of achromatic contrast detectors and a smaller group that function as hue detectors. Adapted from Sabesan et al. (2016).

many of the at least 20 different RGC types are represented (Percival et al., 2013; Peng et al., 2019). What failed to be appreciated in the early work on the primate retina is that, with the exception of the midget RGCs, for which there are two for every cone (one ON and one OFF), each of the twenty or more RGC types represents a small percentage of the total. Thus, the retina is a massively parallel processing machine with many different types of RGCs carrying out diverse functions most of which operate at low spatial acuity and require only sparse representations. Thus, as discussed below, it seems plausible that, consistent with the current understanding of the plan of the retina, hue perception could be mediated by a relatively sparse set of RGCs that serve as hue detectors.

Recent considerations of the metabolic cost of information transmission have also questioned the efficiency of compressing information into a smaller set of RGCs, and revealed a more nuanced set of constraints defined not by the number of axons, but by their diameter. RGC axon diameters scale linearly with average firing rate (Perge et al., 2009, 2012). This relationship forms the basis of a law of diminishing returns – metabolic cost increases supralinearly with axon diameter while the information per spike falls as spike rate increases (Rieke et al., 1997; Koch et al., 2006).

A population of parallel neurons, each carrying as much information as possible, is the most efficient coding scheme (Laughlin, 2001). The midget RGC circuit, acting as an edge contrast detector, is already a model of energy-efficient parallel processing – each cone in the central retina contacts a single ON and OFF midget bipolar cell (**Figure 1C**). This allows baseline activity to remain low while the response ranges of each ON and OFF cell are devoted to signaling increments or decrements, respectively, in parallel (Berry et al., 1997). Theoretically, multiplexing increments and decrements would double the information per axon, thus halving the number of axons while increasing axon diameter (and thus energetic cost) fourfold (Sterling and Laughlin, 2017). Taking these costs into account creates a strong pressure for more types of RGCs with thinner axons and lower spike rates, consistent with a parallel processing model.

PARALLEL PROCESSING MODELS

L vs. M midget RGCs receptive fields are near optimal for high acuity spatial vision and are poorly suited for encoding hue. These facts plus the computational complexity required to separate hue from spatial information from L vs. M midget RGCs and a newer understanding of information processing in the retina has led to the suggestion of an alternative hypothesis: that the L vs. M midget RGCs' only serve spatial vision – the function for which they are optimized – and they do not contribute to red-green hue perception. According to this idea, the front-end computations for hue perception are served, in parallel, by a second population of RGCs that have receptive field properties that are specifically optimized as hue detectors (Rodieck, 1991; Calkins and Sterling, 1999; Schmidt et al., 2014; Neitz and Neitz, 2016). The "pixel density" of the L vs. M midget RGCs is high to serve high spatial acuity but, as introduced above, the proposed parallel set of hue detectors need to be only relatively sparse to recover surface reflectance with much lower spatial acuity.

Separate Subtypes of Midget RGCs for Hue and Spatial Vision

If L vs. M midget RGCs mediate spatial vision, which RGCs encode color? To match the acuity of our hue perception, an undiscovered RGC type would need roughly the sampling density

of the S-cone mosaic (Mullen, 1985; Calkins and Sterling, 1999). The lack of alternative hue encoders makes midget RGCs an obvious candidate. We have proposed that the four fundamental hues are encoded by a small subset of L vs. M midget RGCs receiving input from neighboring S-cones (**Figure 2F**; Schmidt et al., 2014). The resulting L + S vs. M and M + S vs. L midget RGCs match the cone inputs for the four fundamental hues, as well as a population of rare RGCs (De Monasterio and Gouras, 1975; De Monasterio et al., 1975) and LGN neurons (Derrington et al., 1984; Tailby et al., 2008). These rare RGCs should not be ignored, as a potential hue-encoding RGC type needs to be only ∼5–10% of foveal RGCs to match color acuity (Calkins and Sterling, 1999).

Each S-cone has a surround created by S-cone-preferring HII horizontal cells. Hue-encoding receptive fields are proposed to arise from the superposition of the S-cone center-surround receptive field with the L vs. M cone center-surround. These two are predicted to be combined by feedforward synapses (Puller et al., 2014) from HII horizontal cells to L vs. M midget bipolar cells. The result simultaneously creates the S-cone input to L vs. M opponent cells and double opponency required to create nearly ideal hue-encoding RGCs (discussed in detail in Neitz and Neitz, 2016). Indeed, computational models of such color-coding midget RGCs can account for previously unexplained color phenomena, such as unique hues and variations in hue perception with L/M-cone ratios (Schmidt et al., 2016).

Key strengths of this parallel processing hypothesis are its simplicity and specificity. All the key features of ideal hue-encoding neurons are proposed to be created in the retina simply by feed-forward from HII horizontal cells at the level of the bipolar cells in a single step as opposed to the idea of multiple stages at unspecified higher levels. The predicted mechanism for a parallel set of double opponent neurons includes specific cell types, neurotransmitters, and biophysical mechanisms (Puller et al., 2014). While this level of detail may invite additional criticism, it also generates testable predictions that can be addressed by experiment. In contrast, the DeValois and DeValois model specified the computations for their "de-multiplexing" neurons, but not the underlying neural substrates.

Recent Research Supporting Parallel Processing Models

The parallel processing approach draws from the idea that each RGC's receptive field acts as a feature detector, tuned to extract a specific type of visual information, such as direction, defocus, edges or hue. From this perspective, L vs. M midget RGCs that respond equally to red–green and black–white edges are not multiplexing, nor even confounding, red–green and black–white signals. Rather, they are reliably signaling a particular feature – the presence or absence of an edge. Accordingly, hue-encoding RGCs are signaling a different feature – the detection of a specific spectral reflectance distribution (**Figure 3A**). Importantly, these RGCs would not be directly responsible for percepts of hue and edges, but instead we are proposing that they serve as front-end mechanisms for making these computations.

A particularly influential line of evidence has been provided by high-precision psychophysics experiments enabled by the development of adaptive optics systems capable of delivering small spots of light while simultaneously imaging the underlying mosaic of cones (Harmening et al., 2014). Early experiments investigating spatial acuity found individual midget RGCs set the limit for spatial resolution (Rossi and Roorda, 2010). These results are inconsistent with models proposing midget RGC outputs are combined to "de-multiplex" color and spatial information. The loss of spatial information from the convergence in **Figure 2D** can never be recovered at higher levels in the visual pathway.

The unprecedented precision provided by adaptive optics imaging systems combined with recent advances in eye tracking and cone type classification (Sabesan et al., 2015) have enabled highly precise psychophysics experiments investigating the percepts resulting from single cones (reviewed by Kling et al., 2019). The responses are highly consistent and reflect activity in the midget RGCs with single cone centers (Schmidt et al., 2019). Consistent with parallel processing of hue and spatial information by separate types of midget RGCs, stimulation of most L/M-cones in the central retina results in percepts of white, with only a small subset eliciting color percepts (**Figure 3B**; Sabesan et al., 2016; Schmidt et al., 2018a,b). Further, the homogeneity of the surrounding cone type had no effect on which cones were associated with a perceived color, arguing against the idea that midget RGCs with strong L vs. M opponency serve hue perception. These experiments were the first to target stimuli to single cones of a known type and represent a major advance in linking perception to underlying neural substrates in awake, behaving humans and the results will undoubtedly continue to challenge long-held assumptions.

HOW DOES THE CORTEX USE WAVELENGTH INFORMATION?

Hue perception is just one of many functions that uses wavelength information. For example, the retina contains photopigments such as melanopsin and neuropsin, which carry additional wavelength information, but have no impact on the dimensionality of color vision (Horiguchi et al., 2013; Buhr et al., 2015). There are many examples of neurons carrying temporal, spatial or spectral information that is not extracted for visual perception, including color-opponent V1 neurons responding to chromatic stimuli that are not perceived (Gur and Snodderly, 1997; Jiang et al., 2007).

In fact, many RGCs do not contribute to conscious perception at all, but instead mediate functions such as visually guided movements or circadian photoentrainment (for review, see Neitz and Neitz, 2016). Wavelength information is extracted by several types of spectrally opponent RGCs for many functions other than color vision. For example, circadian rhythm photoentrainment and the pupillary light reflex are mediated by intrinsically photosensitive RGCs (reviewed in Do and Yau, 2010). Their receptive fields match the wavelength-encoding, single opponent receptive fields discussed above (**Figure 2A**;

Dacey et al., 2005) – ideal for measuring the changes in chromaticity of ambient light throughout the day (Pauers et al., 2012; Spitschan et al., 2017) but they do not contribute to hue perception.

Several lines of evidence indicate that the ability to detect red-green edges is a distinct feature encoded separately from the ability to classify the appearance of lights as red or green. For example, patients with cerebral achromatopsia who suffer a total loss of hue perception, but still can detect chromatic borders, perceive shape from color and discriminate the direction in which colored patterns move (Cowey and Heywood, 1997). The existence of multiple mechanisms and uses for wavelength information also seems evident when comparing the cone inputs mediating color detection and color appearance. The studies identifying L + S vs. M and M + S vs. L as the cone inputs to hue perception measured color appearance (Wooten and Werner, 1979; Drum, 1989; Webster et al., 2000a; Schmidt et al., 2016). However, the classic psychophysical experiments that identified L vs. M and S vs. L + M as the "cardinal directions of color space" (Krauskopf et al., 1982), measured detection. Krauskopf et al. (1982) noted the disparity between their cardinal directions and the red-green (L + S vs. M) and blue-yellow (M + S vs. L) hue axes of color appearance and later questioned the evidence for cardinal mechanisms (Krauskopf, 1997).

There is common ground between multiplexing and parallel processing models. In discussing the abundance of chromatic cortical neurons, DeValois and DeValois argue that only a few are responsible for the specification of color, while the majority instead use color information to specify the spatial (or other) characteristics of stimuli. A problem was a lack of agreement on which cells were relevant for hue perception. Though their proposed color transformations were not consistent with the majority of published cortical color tuning studies, DeValois and DeValois pointed out inconsistencies in the literature and claimed one could "cite some cortical study in support of (or against) almost any suggestion about cortical color processing" (De Valois and De Valois, 1993) We argue a similar situation exists today in the retina where different studies can be sited in support or against the existence of S-cone inputs to midget RGCs [for example, compare the cone opponency reported by De Monasterio and Gouras (1975), Sun et al. (2006), and Field et al. (2010)].

FUTURE DIRECTIONS

Both the parallel processing and multiplexing models would benefit from experiments linking the theories to their underlying neural substrates. However, an overarching difficulty for resolving the controversy over parallel vs. multiplexing theories is that each point of view reflects a deep-seated theoretical conviction. For those preferring the multiplexing view of L vs. M midget RGCs, "If the color signal is extractable, it makes little sense not to use it" (Billock et al., 1996). From a parallel processing standpoint, encoding color and spatial vision, two of the most fundamental aspects of visual perception, in a single binary channel makes little sense (Calkins and Sterling, 1999) and the information gained must outweigh the cost of extracting a color signal (Laughlin et al., 1998).

Thus, further experiments to characterize the response properties of visual neurons alone are not going to settle the controversy. Initial surveys of cone inputs to neurons in the retinal and LGN reported S-cone input to a subset of L vs. M neurons (De Monasterio and Gouras, 1975; De Monasterio et al., 1975; Derrington et al., 1984) and later surveys confirmed these findings (Tailby et al., 2008; Field et al., 2010). However, skeptics of the parallel processing models favor a study by Sun et al. (2006) in which the authors recorded from a large population of midget RGCs and concluded S-cone input was unlikely (Sun et al., 2006). An underlying problem is that the answers depend on how you ask the question. Results from receptive field measurements are a function of stimulus choice. For example, a full-field stimulus (Lee et al., 1998) may have reduced S-cone responses by driving the antagonistic S-cone surround receptive field mediated by HII horizontal cell feedback (Dacey et al., 1996). Indeed, the Sun et al. (2006) experiments did not detect S-OFF midget RGCs, despite a growing consensus that these neurons make up 5–10% of OFF midget RGCs in the macaque central retina (Klug et al., 2003; Field et al., 2010; Tsukamoto and Omi, 2015; Patterson et al., 2019). Taken together, these results further demonstrate the need to account for both the spatial and spectral dimensions of midget RGC receptive fields.

Consideration of underlying theoretical perspectives and stimulus biases will be essential for designing future experiments linking color vision models to their underlying neural substrates. Also, a broader perspective may help answer the larger questions about how our eye and brain process visual information. Hopefully, future research using cutting-edge technologies will provide satisfying explanations for long unanswered mysteries of vision.

AUTHOR CONTRIBUTIONS

SP wrote the manuscript. MN and JN edited the final version of the manuscript.

FUNDING

This work was supported by NIH grants R01EY027859 (JN), T32EY07031 (SP), T32NS099578 (SP), P30EY001730 (Core Grant for Vision Research), and Research to Prevent Blindness. The National Institute of Health contributed to the salaries of the authors and the institutional facilities. A research to prevent blindness unrestricted grant supports the authors research efforts in the Department of Ophthalmology.

ACKNOWLEDGMENTS

We thank Steve Buck and Ram Sabesan for helpful discussions.

REFERENCES

Atick, J. J., Li, Z., and Redlich, A. N. (1992). Understanding retinal color coding from first principles. *Neural Comput.* 4, 559–572. doi: 10.1162/neco.1992.4.4.559

Baden, T., Berens, P., Franke, K., Román Rosón, M., Bethge, M., Euler, T., et al. (2016). The functional diversity of retinal ganglion cells in the mouse. *Nature* 529, 345–350. doi: 10.1038/nature16468

Barlow, H. (1961). "Possible principles underlying the transformations of sensory messages," in *Sensory Communication*, ed. W. A. Rosenbith (Cambridge, MA: MIT Press).

Baylor, D. A., Nunn, B. J., and Schnapf, J. L. (1987). Spectral sensitivity of cones of the monkey Macaca fascicularis. *J. Physiol.* 390, 145–160. doi: 10.1113/jphysiol.1987.sp016691

Berry, M. J., Warland, D. K., and Meister, M. (1997). The structure and precision of retinal spike trains. *Proc. Natl. Acad. Sci. U.S.A.* 94, 5411–5416. doi: 10.1073/pnas.94.10.5411

Billock, V. A., Dacey, D. M., and Masland, R. H. (1996). Consequences of retinal color coding for cortical color decoding. *Science* 274, 2118–2120.

Boycott, B., and Wässle, H. (1999). Parallel processing in the mammalian retina: the proctor lecture. *Investig. Ophthalmol. Vis. Sci.* 40, 1313–1327.

Buchsbaum, G., and Gottschalk, A. (1983). Trichromacy, opponent colours coding and optimum colour information in the retina. *Proc. R. Soc. Lond. B Biol. Sci.* 220, 89–113. doi: 10.1098/rspb.1983.0090

Buhr, E. D., Yue, W. W., Ren, X., Jiang, Z., Liao, H. W., Mei, X., et al. (2015). Neuropsin (OPN5)-mediated photoentrainment of local circadian oscillators in mammalian retina and cornea. *Proc. Natl. Acad. Sci. U.S.A.* 112, 13093–13098. doi: 10.1073/pnas.1516259112

Buzás, P., Blessing, E. M., Szmajda, B. A., and Martin, P. R. (2006). Specificity of M and L cone inputs to receptive fields in the parvocellular pathway: random wiring with functional bias. *J. Neurosci.* 26, 11148–11161. doi: 10.1523/jneurosci.3237-06.2006

Calkins, D. J., and Sterling, P. (1999). Evidence that circuits for spatial and color vision segregate at the first retinal synapse. *Neuron* 24, 313–321. doi: 10.1016/s0896-6273(00)80846-6

Chang, L., Breuninger, T., and Euler, T. (2013). Chromatic coding from cone-type unselective circuits in the mouse retina. *Neuron* 77, 559–571. doi: 10.1016/j.neuron.2012.12.012

Cowey, A., and Heywood, C. A. (1997). Cerebral achromatopsia: colour blindness despite wavelength processing. *Trends Cogn. Sci.* 1, 133–139. doi: 10.1016/S1364-6613(97)01043-7

Croner, L. J., and Kaplan, E. (1995). Receptive fields of P and M ganglion cells across the primate retina. *Vision Res.* 35, 7–24. doi: 10.1016/0042-6989(94)e0066-t

Crook, J. D., Davenport, C. M., Peterson, B. B., Packer, O. S., Detwiler, P. B., Dacey, D. M., et al. (2009). Parallel ON and OFF cone bipolar inputs establish spatially coextensive receptive field structure of blue-yellow ganglion cells in primate retina. *J. Neurosci.* 29, 8372–8387. doi: 10.1523/JNEUROSCI.1218-09.2009

Dacey, D., Packer, O. S., Diller, L., Brainard, D., Peterson, B., Lee, B., et al. (2000). Center surround receptive field structure of cone bipolar cells in primate retina. *Vision Res.* 40, 1801–1811. doi: 10.1016/s0042-6989(00)00039-0

Dacey, D. M., Lee, B. B., Stafford, D., Polkorny, J., and Smith, V. C. (1996). Horizontal cells of the primate retina: cone specificity without spectral opponency. *Science* 271, 656–659. doi: 10.1126/science.271.5249.656

Dacey, D. M., Liao, H. W., Peterson, B. B., Robinson, F. R., Smith, V. C., Pokorny, J., et al. (2005). Melanopsin-expressing ganglion cells in primate retina signal colour and irradiance and project to the LGN. *Nature* 433, 749–754. doi: 10.1038/nature03387

Daw, N. W. (1973). Neurophysiology of color vision. *Physiol. Rev.* 53, 571–603.

De Monasterio, F. M., and Gouras, P. (1975). Functional properties of ganglion cells of the rhesus monkey retina. *J. Physiol.* 251, 167–195. doi: 10.1113/jphysiol.1975.sp011086

De Monasterio, F. M., Gouras, P., and Tolhurst, D. J. (1975). Trichromatic colour opponency in ganglion cells of the rhesus monkey retina. *J. Physiol.* 251, 197–216. doi: 10.1113/jphysiol.1975.sp011087

De Valois, R. L. (2004). "Neural coding of color," in *The New Visual Neurosciences*, eds J. S. Werner and L. M. Chalupa (Cambridge, MA: MIT Press), 1003–1016.

De Valois, R. L., Abramov, I., and Jacobs, G. H. (1966). Analysis of response patters of LGN cells. *J. Opt. Soc. Am.* 56, 966–977.

De Valois, R. L., and De Valois, K. K. (1993). A multi-stage color model. *Vision Res.* 33, 1053–1065. doi: 10.1016/0042-6989(93)90240-w

Deny, S., Ferrari, U., Macé, E., Yger, P., Caplette, R., Picaud, S., et al. (2017). Multiplexed computations in retinal ganglion cells of a single type. *Nat. Commun.* 8:1964. doi: 10.1038/s41467-017-02159-y

Derrico, J. B., and Buchsbaum, G. A. (1991). computational model of spatiochromatic image coding in early vision. *J. Vis. Commun. Image Represent.* 2, 31–38. doi: 10.1016/1047-3203(91)90033-c

Derrington, A. M., Krauskopf, J., and Lennie, P. (1984). Chromatic mechanisms in lateral geniculate nucleus of macaque. *J. Physiol.* 357, 241–265. doi: 10.1113/jphysiol.1984.sp015499

Do, M. T. H., and Yau, K.-W. (2010). Intrinsically photosensitive retinal ganglion cells. *Physiol. Rev.* 90, 1547–1581. doi: 10.1152/physrev.00013.2010

Drum, B. (1989). Hue signals from short- and middle-wavelength-sensitive cones. *J. Opt. Soc. Am. A* 6, 153–157.

D'Zmura, M., and Lennie, P. (1986). Mechanisms of color constancy. *J. Opt. Soc. Am. A* 3:1662.

Enroth-Cugell, C., and Robson, J. G. (1966). The contrast sensitivity of retinal ganglion cells of the cat. *J. Physiol.* 187, 517–552. doi: 10.1113/jphysiol.1966.sp008107

Estevez, O., and Spekreijse, H. (1982). The 'silent substitution' method in visual research. *Vision Res.* 22, 681–691. doi: 10.1016/0042-6989(82)90104-3

Field, G. D., and Chichilnisky, E. J. (2007). Information processing in the primate retina: circuitry and coding. *Annu. Rev. Neurosci.* 30, 1–30. doi: 10.1146/annurev.neuro.30.051606.094252

Field, G. D., Gauthier, J. L., Sher, A., Greschner, M., Machado, T., Jepson, L. H., et al. (2010). Functional connectivity in the retina at the resolution of photoreceptors. *Nature* 467, 673–677. doi: 10.1038/nature09424

Field, G. D., Greschner, M., Gauthier, J. L., Rangel, C., Shlens, J., Sher, A., et al. (2009). High-sensitivity rod photoreceptor input to the blue-yellow color opponent pathway in macaque retina. *Nat. Neurosci.* 12, 1159–1164. doi: 10.1038/nn.2353

Field, G. D., Sher, A., Gauthier, J. L., Greschner, M., Shlens, J., Litke, A. M., et al. (2007). Spatial properties and functional organization of small bistratified ganglion cells in primate retina. *J. Neurosci.* 27, 13261–13272. doi: 10.1523/jneurosci.3437-07.2007

Foster, D. H. (2011). Color constancy. *Vision Res.* 51, 674–700. doi: 10.1016/j.visres.2010.09.006

Fuld, K., Wooten, B. R., and Whalen, J. J. (1981). The elemental hues of short-wave and extraspectral lights. *Percept. Psychophys.* 29, 317–322. doi: 10.3758/bf03207340

Gollisch, T., and Meister, M. (2010). Eye smarter than scientists believed: neural computations in circuits of the retina. *Neuron* 65, 150–164. doi: 10.1016/j.neuron.2009.12.009

Grimes, W. N., Schwartz, G. W., and Rieke, F. (2014). The synaptic and circuit mechanisms underlying a change in spatial encoding in the retina. *Neuron* 82, 460–473. doi: 10.1016/j.neuron.2014.02.037

Gur, M., and Snodderly, D. M. (1997). A dissociation between brain activity and perception: chromatically opponent cortical neurons signal chromatic flicker that is not perceived. *Vision Res.* 37, 377–382. doi: 10.1016/s0042-6989(96)00183-6

Hansen, T., and Gegenfurtner, K. R. (2009). Independence of color and luminance edges in natural scenes. *Vis. Neurosci.* 26, 35–49. doi: 10.1017/S0952523808080796

Harmening, W. M., Tuten, W. S., Roorda, A., and Sincich, L. C. (2014). Mapping the perceptual grain of the human retina. *J. Neurosci.* 34, 5667–5677. doi: 10.1523/JNEUROSCI.5191-13.2014

Horiguchi, H., Winawer, J., Dougherty, R. F., and Wandell, B. A. (2013). Human trichromacy revisited. *Proc. Natl. Acad. Sci. U.S.A.* 110, E260–E269. doi: 10.1073/pnas.1214240110

Hurvich, L. M., and Jameson, D. (1957). An opponent-process theory of color vision. *Psychol. Rev.* 64, 384–404. doi: 10.1037/h0041403

Ingling, C. R., and Martinez-Uriegas, E. (1983). The relationship between spectral sensitivity and spatial sensitivity for the primate r-g X-channel. *Vision Res.* 23, 1495–1500. doi: 10.1016/0042-6989(83)90161-x

Jacobs, G. H. (2018). Photopigments and the dimensionality of animal color vision. *Neurosci. Biobehav. Rev.* 86, 108–130. doi: 10.1016/j.neubiorev.2017.12.006

Jadzinsky, P. D., and Baccus, S. A. (2013). Transformation of visual signals by inhibitory interneurons in retinal circuits. *Annu. Rev. Neurosci.* 36, 403–428. doi: 10.1146/annurev-neuro-062012-170315

Jiang, Y., Zhou, K., and He, S. (2007). Human visual cortex responds to invisible chromatic flicker. *Nat. Neurosci.* 10, 657–662. doi: 10.1038/nn1879

Kingdom, F. A. A., and Mullen, K. T. (1995). Separating colour and luminance information in the visual system. *Spat. Vis.* 9, 191–219. doi: 10.1163/156856895x00188

Kling, A., Field, G. D., Brainard, D. H., and Chichilnisky, E. J. (2019). Probing computation in the primate visual system at single-cone resolution. *Annu. Rev. Neurosci.* 42, 169–186. doi: 10.1146/annurev-neuro-070918-050233

Klug, K., Herr, S., Ngo, I. T., Sterling, P., and Schein, S. (2003). Macaque retina contains an S-cone OFF midget pathway. *J. Neurosci.* 23, 9881–9887. doi: 10.1523/jneurosci.23-30-09881.2003

Koch, K., McLean, J., Segev, R., Freed, M. A., Berry, M. J., Balasubramanian, V., et al. (2006). How much the eye tells the brain. *Curr. Biol.* 16, 1428–1434. doi: 10.1016/j.cub.2006.05.056

Kolb, H., and Marshak, D. (2003). The midget pathways of the primate retina. *Doc. Ophthalmol.* 106, 67–81.

Krauskopf, J. (1997). "Paucity of evidence for cardinal mechanisms," in *John Dalton's Color Vision Legacy*, eds C. Dickinson, I. Murray, and D. Carden (Routledge: Taylor & Francis).

Krauskopf, J., Williams, D. R., and Heeley, D. W. (1982). Cardinal directions of color space. *Vision Res.* 22, 1123–1131. doi: 10.1016/0042-6989(82)90077-3

Kuffler, S. W. (1953). Discharge patterns and functional organization of mammalian retina. *J. Neurophysiol.* 16, 37–68. doi: 10.1152/jn.1953.16.1.37

Laughlin, S. B. (2001). Energy as a constraint on the coding and processing of sensory information. *Curr. Opin. Neurobiol.* 11, 475–480. doi: 10.1016/s0959-4388(00)00237-3

Laughlin, S. B., De Ruyter Van Steveninck, R. R., and Anderson, J. C. (1998). The metabolic cost of neural information. *Nat. Neurosci.* 1, 36–41. doi: 10.1038/236

Lee, B. B. (1996). Receptive field structure in the primate retina. *Vision Res.* 36, 631–644. doi: 10.1016/0042-6989(95)00167-0

Lee, B. B. (2008). Neural models and physiological reality. *Vis. Neurosci.* 25, 231–241. doi: 10.1017/S0952523808080140

Lee, B. B., Kremers, J., and Yeh, T. (1998). Receptive fields of primate retinal ganglion cells studied with a novel technique. *Vis. Neurosci.* 15, 161–175. doi: 10.1017/s095252389815112x

Lennie, P., Haake, P. W., and Williams, D. R. (1991). "The design of chromatically opponent receptive fields," in *Computational Models of Visual Processing*, eds M. S. Landy and J. A. Movshon (Cambridge, MA: MIT Press), 71–82.

Lennie, P., and Movshon, J. A. (2005). Coding of color and form in the geniculostriate visual pathway (invited review). *J. Opt. Soc. Am. A* 22, 2013–2033.

Marr, D. (1982). *Vision: A Computational Investigation into The Human Representation and Processing of Visual Information*. Cambridge, MA: MIT Press.

Marr, D., and Hildreth, E. (1980). Theory of edge detection. *Proc. R. Soc. Lond. Ser. B* 207, 187–217.

Martin, P. R., Blessing, E. M., Buzás, P., Szmajda, B. A., and Forte, J. D. (2011). Transmission of colour and acuity signals by parvocellular cells in marmoset monkeys. *J. Physiol.* 589, 2795–2812. doi: 10.1113/jphysiol.2010.194076

Martin, P. R., White, A. J. R., Goodchild, A. K., Wilder, H. D., and Sefton, A. E. (1997). Evidence that blue-on cells are part of the third geniculocortical pathway in primates. *Eur. J. Neurosci.* 9, 1536–1541. doi: 10.1111/j.1460-9568.1997.tb01509.x

Mullen, K. T. (1985). The contrast sensitivity of human colour vision to red-green and blue-yellow chromatic gratings. *J. Physiol.* 359, 381–400. doi: 10.1113/jphysiol.1985.sp015591

Nathans, J. (1999). The evolution and physiology of human color vision: insights from molecular genetic studies of visual pigments. *Neuron* 24, 299–312. doi: 10.1016/s0896-6273(00)80845-4

Neitz, J., and Neitz, M. (2011). The genetics of normal and defective color vision. *Vision Res.* 51, 633–651. doi: 10.1016/j.visres.2010.12.002

Neitz, J., and Neitz, M. (2016). Evolution of the circuitry for conscious color vision in primates. *Eye* 31, 286–300. doi: 10.1038/eye.2016.257

Párraga, C. A. C. A., Troscianko, T., and Tolhurst, D. J. D. J. (2002). Spatiochromatic properties of natural images and human vision. *Curr. Biol.* 12, 483–487. doi: 10.1016/s0960-9822(02)00718-2

Patterson, S. S., Kuchenbecker, J. A., Anderson, J. R., Bordt, A. S., Marshak, D. W., Neitz, M. M., et al. (2019). An S-cone circuit for edge detection in the primate retina. *Sci. Rep.*, doi: 10.1101/667204

Pauers, M. J., Kuchenbecker, J. A., Neitz, M., and Neitz, J. (2012). Changes in the colour of light cue circadian activity. *Anim. Behav.* 83, 1143–1151. doi: 10.1016/j.anbehav.2012.01.035

Paulus, W., and Kroger-Paulus, A. (1983). A new concept of retinal colour coding. *Vision Res.* 23, 529–540. doi: 10.1016/0042-6989(83)90128-1

Peng, Y. R., Shekhar, K., Yan, W., Herrmann, D., Sappington, A., Bryman, G. S., et al. (2019). Molecular classification and comparative taxonomics of foveal and peripheral cells in primate retina. *Cell* 176, 1222–1237. doi: 10.1016/j.cell.2019.01.004

Percival, K. A., Martin, P. R., and Grünert, U. (2013). Organisation of koniocellular-projecting ganglion cells and diffuse bipolar cells in the primate fovea. *Eur. J. Neurosci.* 37, 1072–1089. doi: 10.1111/ejn.12117

Perge, J. A., Koch, K., Miller, R., Sterling, P., and Balasubramanian, V. (2009). How the optic nerve allocates space, energy capacity, and information. *J. Neurosci.* 29, 7917–7928. doi: 10.1523/JNEUROSCI.5200-08.2009

Perge, J. A., Niven, J. E., Sterling, P., Mugnaini, E., and Balasubramanian, V. (2012). Why do axons differ in caliber? *J. Neurosci.* 32, 626–638. doi: 10.1523/JNEUROSCI.4254-11.2012

Pitkow, X., and Meister, M. (2012). Decorrelation and efficient coding by retinal ganglion cells. *Nat. Neurosci.* 15, 628–635. doi: 10.1038/nn.3064

Puller, C., Haverkamp, S., Neitz, M., and Neitz, J. (2014). Synaptic elements for GABAergic feed-forward signaling between HII horizontal cells and blue cone bipolar cells are enriched beneath primate S-cones. *PLoS One* 9:e88963. doi: 10.1371/journal.pone.0088963

Rider, A. T., Henning, G. B., Eskew, R. T., and Stockman, A. (2018). Harmonics added to a flickering light can upset the balance between ON and OFF pathways to produce illusory colors. *Proc. Natl. Acad. Sci. U.S.A* 115, E4081–E4090. doi: 10.1073/pnas.1717356115

Rieke, F., Warland, D., de Ruyter van Steveninck, R. R., and Bialek, W. (1997). *Spikes: Exploring the Neural Code*. Cambridge, MA: MIT Press.

Rodieck, R. (1991). "Which cells code for color?," in *From Pigments to Perception: Advances in Understanding Visual Processes*, eds A. Valberg and B. Lee (New York, NY: Plenum Press), 83–93. doi: 10.1007/978-1-4615-3718-2_10

Rossi, E. A., and Roorda, A. (2010). The relationship between visual resolution and cone spacing in the human fovea. *Nat. Neurosci.* 13, 156–157. doi: 10.1038/nn.2465

Rushton, W. A. H. (1972). Review lecture. Pigments and signals in colour vision. *J. Physiol.* 220, 1–31. doi: 10.1113/jphysiol.1972.sp009719

Sabesan, R., Hofer, H., and Roorda, A. (2015). Characterizing the human cone photoreceptor mosaic via dynamic photopigment densitometry. *PLoS One* 10:e1003652. doi: 10.1371/journal.pone.0144891

Sabesan, R., Schmidt, B. P., Tuten, W. S., and Roorda, A. (2016). The elementary representation of spatial and color vision in the human retina. *Sci. Adv.* 2:e1600797. doi: 10.1126/sciadv.1600797

Schiller, P. H., Logothetis, N., and Charles, E. R. (1990). Functions of the colour-opponent and broad-band channels of the visual system. *Nature* 343, 68–70. doi: 10.1038/343068a0

Schmidt, B. P., Boehm, A. E., Foote, K. G., and Roorda, A. (2018b). The spectral identity of foveal cones is preserved in hue perception. *J. Vis.* 18:19. doi: 10.1167/18.11.19

Schmidt, B. P., Boehm, A. E., Tuten, W. S., and Roorda, A. (2019). Spatial summation of individual cones in human color vision. *bioRxiv*

Schmidt, B. P., Neitz, M., and Neitz, J. (2014). Neurobiological hypothesis of color appearance and hue perception. *J. Opt. Soc. Am. A. Opt. Image Sci. Vis.* 31, 195–207. doi: 10.1364/JOSAA.31.00A195

Schmidt, B. P., Sabesan, R., Tuten, W. S., Neitz, J., and Roorda, A. (2018a). Sensations from a single M-cone depend on the activity of surrounding S-cones. *Sci. Rep.* 8:8561. doi: 10.1038/s41598-018-26754-1

Schmidt, B. P., Touch, P., Neitz, M., and Neitz, J. (2016). Circuitry to explain how the relative number of L and M cones shapes color experience. *J. Vis.* 16, 1–17. doi: 10.1167/16.8.18

Spitschan, M., Aguirre, G. K., Brainard, D. H., and Sweeney, A. M. (2016). Variation of outdoor illumination as a function of solar elevation and light pollution. *Sci. Rep.* 6, 1–14. doi: 10.1038/srep26756

Spitschan, M., Lucas, R. J., and Brown, T. M. (2017). Chromatic clocks: color opponency in non-image-forming visual function. *Neurosci. Biobehav. Rev.* 78, 24–33. doi: 10.1016/j.neubiorev.2017.04.016

Srinivasan, M. V., Laughlin, S. B., and Dubs, A. (1982). Predictive coding: a fresh view of inhibition in the retina. *Proc. R. Soc. B Biol. Sci.* 216, 427–459. doi: 10.1098/rspb.1982.0085

Sterling, P., and Laughlin, S. (2017). *Principles of Neural Design*. Cambridge, MA: MIT Press.

Sun, H., Smithson, H. E., Zaidi, Q., and Lee, B. B. (2006). Specificity of cone inputs to macaque retinal ganglion cells. *J. Neurophysiol.* 95, 837–849. doi: 10.1152/jn.00714.2005

Tailby, C., Dobbie, W. J., Solomon, S. G., Szmajda, B. A., Hashemi-Nezhad, M., Forte, J. D., et al. (2010). Receptive field asymmetries produce color-dependent direction selectivity in primate lateral geniculate nucleus. *J. Vis.* 10, 1–18. doi: 10.1167/10.8.1

Tailby, C., Solomon, S. G., and Lennie, P. (2008). Functional asymmetries in visual pathways carrying S-cone signals in macaque. *J. Neurosci.* 28, 4078–4087. doi: 10.1523/JNEUROSCI.5338-07.2008

Tsukamoto, Y., and Omi, N. (2015). OFF bipolar cells in macaque retina: type-specific connectivity in the outer and inner synaptic layers. *Front. Neuroanat.* 9:122. doi: 10.3389/fnana.2015.00122

Verweij, J., Hornstein, E. P., and Schnapf, J. L. (2003). Surround antagonism in macaque cone photoreceptors. *J. Neurosci.* 23, 10249–10257. doi: 10.1523/jneurosci.23-32-10249.2003

Wässle, H. (2004). Parallel processing in the mammalian retina. *Nat. Rev. Neurosci.* 5, 747–757. doi: 10.1038/nrn1497

Wässle, H., Grünert, U., Röhrenbeck, J., and Boycott, B. B. (1990). Retinal ganglion magnification cell density and cortical magnification in the primate. *Vision Res.* 30, 1897–1911. doi: 10.1016/0042-6989(90)90166-i

Wässle, H., Grünert, U., Rohrenbeck, J., and Boycott, B. B. (1998). Cortical magnification factor and the ganglion cell density of the primate retina. *Nature* 341, 643–645.

Webster, M. A., Miyahara, E., Malkoc, G., and Raker, V. E. (2000a). Variations in normal color vision. I. Cone-opponent axes. *J. Opt. Soc. Am. A* 17, 1535–1544.

Webster, M. A., Miyahara, E., Malkoc, G., and Raker, V. E. (2000b). Variations in normal color vision II Unique hues. *J. Opt. Soc. Am. A Opt. Image Sci. Vis.* 17, 1545–1555.

Wienbar, S., and Schwartz, G. W. (2018). The dynamic receptive fields of retinal ganglion cells. *Prog. Retin. Eye Res.* 67, 102–117. doi: 10.1016/j.preteyeres.2018.06.003

Wiesel, T. N., and Hubel, D. H. (1966). Spatial and chromatic interactions in the lateral geniculate body of the rhesus monkey. *J. Neurophysiol.* 29, 1115–1156. doi: 10.1152/jn.1966.29.6.1115

Woelders, T., Wams, E. J., Gordijn, M. C. M., Beersma, D. G. M., and Hut, R. A. (2018). Integration of color and intensity increases time signal stability for the human circadian system when sunlight is obscured by clouds. *Sci. Rep.* 8, 1–10. doi: 10.1038/s41598-018-33606-5

Wool, L. E., Crook, J., Droy, J. B., Packer, O. S., Zaidi, Q., Dacey, D. M., et al. (2018). Nonselective wiring accounts for red-green opponency in midget ganglion cells of the primate retina. *J. Neurosci.* 38, 1520–1540. doi: 10.1523/JNEUROSCI.1688-17.2017

Wooten, B. R., and Werner, J. S. (1979). Short-wave cone input to the red-green opponent channel. *Vision Res.* 19, 1053–1054. doi: 10.1016/0042-6989(79)90231-1

Yue, L., Weiland, J. D., Roska, B., and Humayun, M. S. (2016). Retinal stimulation strategies to restore vision: fundamentals and systems. *Prog. Retin. Eye Res.* 53, 21–47. doi: 10.1016/j.preteyeres.2016.05.002

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: info@frontiersin.org | +41 21 510 17 00

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership