Neural responses to binocular in-phase and anti-phase stimuli

Bruno Richard[1], Daniel H. Baker[2]

[1]Department of Math and Computer Sciences, Rutgers University, Newark, New Jersey, USA
[2]Department of Psychology, University of York, York, UK

Abstract

VSS Abstract - Binocular vision fuses similar stimuli into a single percept, yet incompatible stimuli result in other experiences such as rivalry, lustre, and diplopia. We measured neural responses to binocular stimuli with different phase relationships, intending to understand them using contemporary binocular models. Steady-State Visually Evoked Potentials (SSVEPs) were recorded from 15 observers in response to monocular and binocular stimulation at 3Hz, using either on-off or counterphase flicker. Across the eyes, binocular stimuli could be (i) in spatial and temporal phase, (ii) in temporal phase but spatial antiphase, (iii) in spatial phase but temporal antiphase, or (iv) in spatial and temporal antiphase (for counterphase flicker this is identical to condition (i)). Responses to monocular on-off flicker showed peaks at the fundamental frequency (3Hz) and its harmonics (integer multiples of 3Hz). In contrast, counterphase flicker produced responses only at twice the flicker frequency (6Hz) and its harmonics. Binocular in-phase stimulation resulted in a similar pattern of responses, consistent with 'ocularity invariance' – the observation that binocular and monocular stimuli appear equal at high contrasts. Changing the phase relationship modulated the harmonics pattern in complex ways. On-off flicker in temporal antiphase reduced the fundamental response, but there was no such effect for counterphase flicker. We modelled the data using a progression of binocular combination algorithms that increased in complexity from a simple linear sum to a two-stage binocular gain control model with parallel monocular and binocular phase-selective channels (the Lustre model). The most complex model (lustre) outperformed all other models in capturing the variance of our SSVEP data, although simpler phase-insensitive models performed similarly well in most experimental conditions. Simpler models struggled to capture the response magnitude to counterphase stimuli. Our findings suggest that explaining neural responses to binocular stimuli with different phase relationships requires parallel monocular and phase-selective channels.

The human visual system integrates input from both eyes to form a unified binocular representation of the world. Combining monocular inputs enhances sensitivity to the presented stimuli, particularly when the contrast is low or near detection thresholds (Baker et al., 2018; Campbell & Green, 1965; Meese et al., 2006). Contrast sensitivity can improve by

a factor of $\sqrt{2}$ or more when stimuli are presented binocularly versus monocularly (Baker et al., 2018; Blake & Wilson, 2011; Campbell & Green, 1965; Richard et al., 2018). Notably, the visual system also attempts to combine inputs even when the stimuli presented to each eye are markedly different (i.e., incompatible). In such cases, observers may perceive binocular rivalry (Blake, 1989; Wilson, 2003), diplopia, or visual lustre (Georgeson et al., 2016). Despite differences in perceptual outcomes, computationally, the underlying processes of binocular combination for compatible and incompatible inputs appear similar (Baker, Meese, & Georgeson, 2007; Legge, 1984). Human behavioural responses to compatible and incompatible stimuli can be effectively explained by a single psychophysical model that involves nonlinear transduction, followed by summation across monocular and binocular phase-selective channels (Baker, Meese, & Georgeson, 2007; Baker & Meese, 2007). Here we explore if the integrative processes defined over multiple behavioural tasks are also reflected at the neural level.

It has long been known that stimuli presented binocularly are summed across the eyes. In contrast detection tasks, where stimulus contrast is low, binocular presentation increases sensitivity by approximately $\sqrt{2}$ (Campbell & Green, 1965). This implies that observers require roughly 1.4 times more contrast to detect a monocular stimulus than a binocular one. The binocular improvement in sensitivity is consistent with a non-linearity operating before the signals from the two eyes (C_l and C_r) are combined (Legge, 1984):

$$R_B = C_l^m + C_r^m. \tag{1}$$

Here, the exponent m determines the degree of summation. When $m = 1$, summation is linear, yielding a doubling of sensitivity. When $m = 2$, summation is reduced to $\sqrt{2}$. Several studies have reported summation ratios over $\sqrt{2}$ with some approaching 1.8 (Meese et al., 2006; Simmons, 2005; Simmons & Kingdom, 1998). A recent meta-analysis of 65 studies (N = 716) found an average binocular summation ratio of 1.5 (Baker et al., 2018). Their work highlighted the challenges in accurately describing the binocular summation process (e.g., m) as individual variability and methodological differences can greatly impact binocular summation.

The contrast of stimuli, for example, is crucial in measuring binocular summation. Binocular summation can be very large when stimulus contrast is low; however, the binocular advantage is seldom observed in tasks that involve higher contrasts. In contrast discrimination tasks, where observers judge the contrast difference between otherwise identical stimuli, binocular presentation no longer confers a benefit in sensitivity (Legge, 1984; Maehara & Goryo, 2005; Meese et al., 2006): discrimination thresholds are identical whether stimuli are shown to one eye or both. This does not mean binocular summation is absent at higher contrasts (Meese et al., 2006; Meese & Baker, 2011). Instead, it is counteracted by normalization mechanisms that maintain consistency across viewing conditions. In contrast gain control models of early vision, normalization can stem from interocular and self-suppressive signals (Meese et al., 2006):

$$R_B = \frac{C_l^m}{S + C_l + C_r} + \frac{C_r^m}{S + C_r + C_l}, \tag{2}$$

When identical stimuli are presented to both eyes, doubling suppression offsets the enhanced excitatory signals, nullifying the binocular advantage.

A similar pattern of results is observed in neural measurements of binocular summation. In fMRI recordings, neural responses are significantly larger for binocular than monocular responses when stimulus contrast is low (Moradi & Heeger, 2009). At higher contrasts, observers no longer show the binocular advantage. These findings were well-explained by including interocular suppression and binocular contrast normalization, an identical mechanism to that used to explain behavioural data. The equivalency in binocular summation between psychophysics and neuroimaging findings means that both data types can be used to constrain binocular summation models. We, for example, have previously demonstrated that a popular model of binocular summation could be easily adapted to capture Steady-State Visually Evoked Potentials (SSVEPs) to monocular and binocular stimuli when one eye was occluded by a neutral density filter (Richard et al., 2018). Placing a neutral density filter in front of one eye darkens its input, reducing the amplitude and altering the phase of SSVEPs to stimuli presented to the filtered eye. Steady-State response amplitudes and phases to stimuli presented through different neutral density filter strengths were well described by a model that first used a biophysically plausible temporal filter on input stimuli, followed by self and interocular suppression, and finally, binocular contrast normalization. This model also captures psychophysically measured binocular summation in the same group of observers. A comprehensive description of the processes involved in binocular summation should be able to explain behavioural and neuroimaging findings under various experimental conditions of binocular summation.

Many have worked towards the development of a comprehensive description of the process of binocular summation in human vision using a variety of psychophysical and neuroimaging data (Baker, Meese, & Summers, 2007; Baker et al., 2008; Ding et al., 2013; Ding & Sperling, 2006; Legge, 1984; Lygo et al., 2021; Maehara & Goryo, 2005; Richard et al., 2018). Developing these models is not an insignificant challenge; they must account for multiple components of early vision, including identifying relevant signals, defining how these signals might interact, what non-linearities are present, and most importantly, how they are combined. One such model has proven very informative and capable of describing binocular summation under various experimental conditions in the two-stage contrast gain control model developed by Meese et al. (2006). The model captured detection and discrimination thresholds for monocular and binocular stimuli in addition to dichoptic stimuli, whereby the stimuli presented to both eyes are not identical in a two-stage process. First, the input is rectified by an excitatory non-linearity ($m \approx 1.3$) and normalized by self and interocular suppression (see Equation 2). The binocular (e.g., combined) input undergoes a second contrast normalization before the decision stage,

$$R = \frac{R_B^p}{Z + R_B^q}, \tag{3}$$

where the excitatory exponent p allows for greater increase in sensitivity than would be seen by the excitatory non-linearity of the first stage (m in Equation 2) alone.

Subsequent iterations of the two-stage contrast gain control model added channels for op-

posite contrast polarities (Baker & Meese, 2007), and monocular channels parallel the binocular summing channel (Georgeson et al., 2016). Polarity-specific channels were included to explain masking effects when the stimuli presented to each eye had opposing phase polarities (i.e., dichoptic presentation). The presentation of sinusoidal gratings with opposite polarities to each eye does not cancel them out: stimuli remain detectable and the two inputs sum, albeit weakly (Bacon, 1976; Baker & Meese, 2007; Simmons, 2005). Parallel monocular channels were added to account for adaptation after-effects that suggest monocular signals may be preserved and available for perception following binocular summation (Blake et al., 1981; Moulden, 1980). While the possibility of monocular channels had been considered (Legge, 1984), many assumed that only the binocularly summed signal contributed to perception. In a comprehensive study of binocular contrast perception, Georgeson et al. (2016) developed a specific experimental condition to assess the involvement of monocular channels. They devised a discrimination task where the target interval presented a contrast increment to one eye (e.g., 10% pedestal + 2%) and a contrast decrement to the other (e.g., 10% pedestal - 2%). If the only available signal is a binocularly summed one, the task would be nearly impossible to complete; the target interval would be perceptually identical to the pedestal-only interval. However, observers were able to complete the task. The two-stage contrast gain control model with parallel monocular channels was the only model able to capture observer thresholds from all experimental conditions, including the binocular increment and decrement condition.

The current architecture of the two-stage contrast gain control model has been rigorously evaluated on psychophysical data, providing a solid foundation for our understanding of binocular combination (Baker & Meese, 2007; Georgeson et al., 2016; Meese et al., 2006). While previous studies have utilized the two-stage contrast gain control model on neuroimaging data (Lygo et al., 2021; Richard et al., 2018), they only included experimental conditions where the phases of the sinusoidal gratings presented to each eye were identical. To accurately assess the current architecture of the two-stage contrast gain control model, neuroimaging data for binocular presentation of stimuli in opposite phase polarity are required. Here, we recorded SSVEPs to monocular and binocular stimuli with different spatial and temporal phase relationships to obtain the data needed to evaluate the two-stage contrast gain control model. By progressively increasing the complexity of the model, we demonstrate that many mechanisms of binocular combination, such as monocular non-linearities, interocular interactions, and parallel monocular channels, are required to explain neural responses to our set of experimental conditions. Thus, the two-stage contrast gain control model remains a powerful and flexible descriptor of the architecture of binocular combination for data collected across many experimental conditions and modalities.

Methods

Participants

Fifteen observers (2 authors: BR and DHB) with normal or corrected to normal visual acuity and binocular vision participated in our study. Written informed consent was obtained from all participants, and experimental procedures were approved by the ethics committee

of the Department of Psychology at the University of York.

Apparatus

All stimuli were presented using a gamma-corrected ViewPixx 3D display (VPixx Technologies, Canada) driven by a Mac Pro. Binocular separation with minimal crosstalk was achieved by synchronizing the display's refresh rate with the toggling of a pair of Nvidia stereo shutter goggles using an infrared signal. The monitor refresh rate was set to 120 Hz; each eye updated at 60 Hz (every 16.67 msec). The display resolution was set to 1920 X 1080 pixels. A single pixel subtended 0.027^o of visual angle (1.63 arc min) when viewed from 57 cm. The mean luminance of the display viewed through the shutter goggles was 26 cd/m^2.

EEG signals were recorded from 64 electrodes distributed across the scalp according to the 10/20 EEG system (Chatrian et al., 1985) in a WaveGuard cap (ANT Neuro, Netherlands). We monitored eye blinks with an electrooculogram consisting of bipolar electrodes placed above the eyebrow and the cheek on the left side of the participant's face. Stimulus-contingent triggers were sent from the ViewPixx display to the amplifier using a parallel cable. Signals were amplified and digitized using a PC with the ASAlab software (ANT Neuro, Netherlands). All EEG data were imported into MATLAB (Mathworks, MA, USA) using components of the *EEGlab* toolbox (Delorme & Makeig, 2004) and then exported for subsequent offline analysis using R.

Stimulus Creation

Observer SSVEPs were measured with a single horizontal sinusoidal grating that subtended 15^o of visual angle on the retina with a spatial frequency of 3 cycles/^{o} of visual angle (Figure 1 A). Our experimental conditions modulated the interocular spatial phase of stimuli (Figure 1 A). Under binocular viewing, the sinusoidal gratings could be presented in spatial phase or spatial anti-phase. When stimuli were presented in spatial phase, the aligned sinusoidal gratings were identical in both eyes ($\Delta\phi = 0$). The spatial anti-phase condition phase-shifted one of the sinusoidal gratings by 180^o ($\Delta\phi = \pi$). Stimuli were also modulated in their oscillatory pattern, which could be On/Off or counterphase flicker and also modulated in phase (Figure 1 B). On/Off and counterphase flicker had a temporal frequency of 3Hz. Under On/Off contrast flicker, the relative contrast of the gratings began at 0%, increased smoothly to 100% of the nominal maximum (100% Michelson contrast), and then returned to 0% over 333 ms (i.e., one cycle). On/Off flicker will generate SSVEPs at the fundamental frequency (3Hz) and its integer harmonics (2F, 3F, 4F, see Figure 1 C). Counterphase flicker reversed the phase of our gratings at a frequency of 3Hz. The contrast of the grating began at the relative maximum (100%), gradually decreased in contrast to 0% of the relative maximum, and then increased again to 100% but in the opposite phase. Unlike On/Off flicker, counterphase flicker generates two nearly identical transients per cycle and thus does not produce SSVEPs at the fundamental frequency (3Hz) but only its even harmonics (Wade & Baker, 2025).

To aid with binocular fusion, stimuli were surrounded by a static binocular texture presented beyond the central 19^o stimulus aperture. These textures were constructed by first

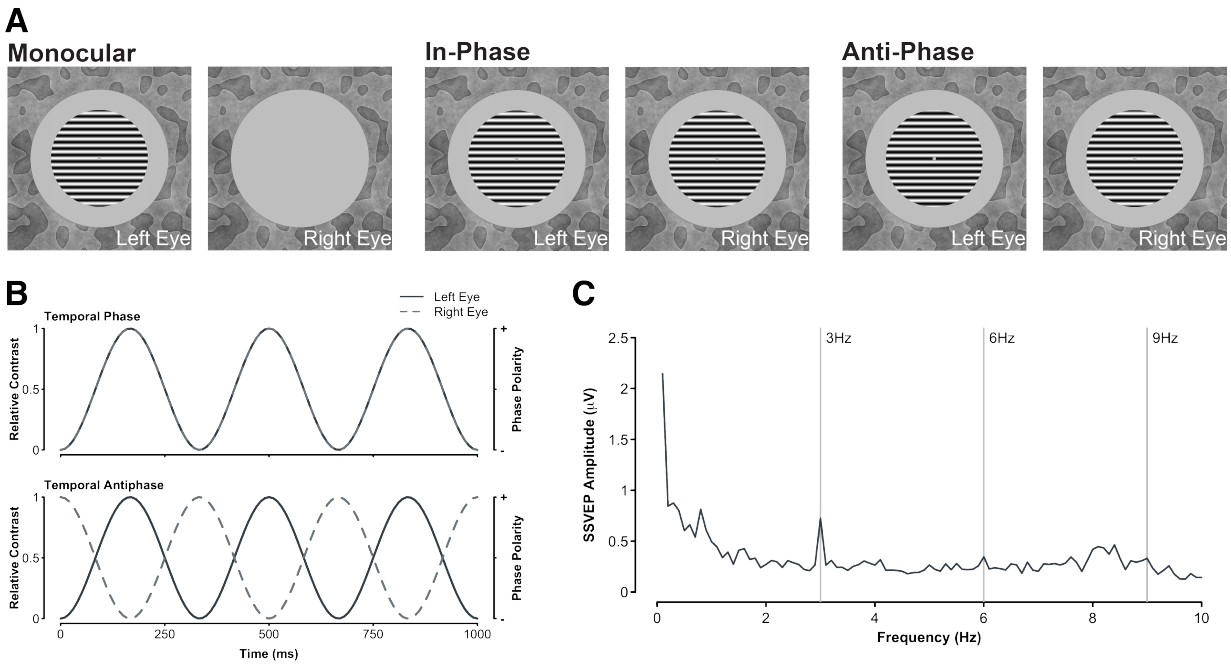

Figure 1: **A**. The spatial configuration of stimuli presented to observers in our experiment. Monocular conditions presented the sinusoidal grating to the left or right eye of observers (counterbalanced) while the other eye was presented with a gray screen set to mean luminance. Binocular conditions could be shown with stimuli in spatial phase, whereby the phase of both sinusoidal gratings was identical, or in spatial anti-phase, where the phase of the sinusoidal gratings presented to each eye was opposite. The background texture did not change throughout the trial to aid with binocular fusion. **B**. The temporal configuration of our stimuli. To generate SSVEPs, stimuli were contrast modulated in two ways: on/off(right Y axis) or counterphase (left Y axis). The oscillatory pattern could also be in phase, where both stimuli were modulated in the same manner, or in counterphase, where as one stimulus increased in contrast, the other decreased in contrast. **C**. For example, SSVEPs were generated under binocular spatial and temporal in-phase viewing of stimuli for one observer, with an average of four electrodes (Oz, Poz, O1, O2) and 12 repetitions.

low-pass filtering a white (amplitude $\propto 1/f^0$) noise pattern, dichotomizing its output into a binary image and taking its phase spectrum. A second flat (amplitude $\propto 1/f^0$) was adjusted by multiplying each spatial frequency's amplitude coefficient by f^{-1} to generate a pink amplitude spectrum (Hansen & Hess, 2006; Tadmor & Tolhurst, 1994). The pink amplitude spectrum and the phase spectrum of the binary image were rendered in the spatial domain by taking the inverse Fourier transform, resulting in the pattern shown in Figure 1 A.

Procedures

Steady-State Visually Evoked Potentials (SSVEPs) were recorded from monocular and binocular stimulation using either on-off or counterphase flicker at 3Hz. Across the eyes, binocular stimuli could be in spatial and temporal phases, in temporal phase but spatial anti-phase, in spatial phase but temporal anti-phase, or spatial and temporal anti-phase (on-off flicker only). Stimuli presented in spatial and temporal anti-phase under counterphase flicker would results in stimuli presented in spatial and temporal phase. Thus, this experiment was comprised of a total of nine conditions - two monocular and seven binocular - which were each repeated 12 times for a total of 108 trials. Stimulus presentation was separated into four experimental blocks, each containing 27 trials. A trial lasted 15 seconds; a grating stimulus flickered onscreen for 11 seconds, followed by a screen with its central 19^o set to mean luminance for 4 seconds. Participants completed all 27 trials of an experimental block in a single sequence (6.75 minutes) and were given breaks between experimental blocks. The trial order was randomized between each block and the participants. Participants did not receive explicit task instructions other than to fixate the marker in the center of the display and blink only during the blank period between stimulus presentations.

SSVEP Analysis

We used whole-head average referencing to normalize each electrode to the mean signal of all 64 electrodes (for each sample point). The EEG waveforms were Fourier transformed at each electrode for a 10-second window, one second after the stimulus onset to avoid onset transients. The Fourier spectra were coherently averaged (i.e., retaining the phase information) across four occipital electrodes (Oz, POz, O1 and O2) and trial repetitions (see Figure Figure 1 C). We then calculated signal-to-noise ratios (SNRs) by dividing the absolute amplitude in the signal bin (e.g., 3 Hz) by the mean of the absolute value of the ten adjacent bins ($\pm$0.5 Hz in steps of 0.1 Hz). Given that the distributions of SNRs are inherently skewed, the median SNR was taken across all participants. The median is a more robust descriptor of central tendency in skewed distributions.

Results

Figure 2 draws the cross-participant median SNR spectra for all experimental conditions. Responses for all On/Off flicker experimental conditions generated peaks at the fundamental frequency (3Hz) and its harmonics (integer multiples of 3Hz). Similarly, counterphase flicker produced responses at twice the flicker frequency (6 Hz) and its harmonics. We assessed

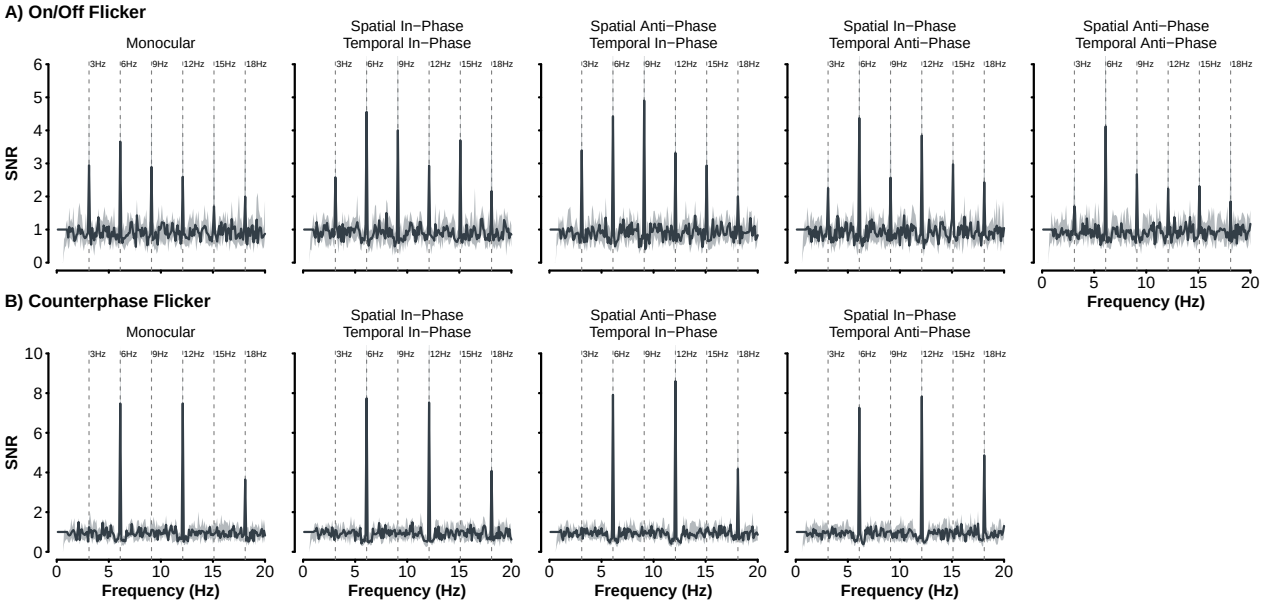

Figure 2: Cross-participant median SNRs for frequencies up to 20Hz. SNRs generated by On/Off flicker are shown in the top row (A), while those generated by counterphase flicker are shown in the bottom row (B). The light gray area represents 95% bootstrap confidence intervals that were calculated by resampling (with replacement) participant SNRs 2000 times.

differences in SNR magnitude across experimental conditions via a permutation test. A permutation test allows for a non-parametric comparison of a statistic between two experimental conditions. We first take the median difference between two experimental conditions (i.e., the real difference). The SNR values of both experimental conditions are then combined and randomly sampled without replacement to create two groups of sizes identical to their original, but with values that are not associated with a particular experimental condition. The median difference of the randomly sampled SNRs is then taken. This process is repeated multiple times (e.g., $N = 2000$) to build a distribution of median differences with no association of experimental condition (i.e., a null-hypothesis distribution). The original median difference is then compared to this distribution, and the proportion of scores greater (or less) than the real median difference represents the p value associated with the test. When comparing SNRs at the fundamental frequency (3Hz) for On/Off flicker, we find no statistically significant difference in median SNR magnitude between experimental conditions where stimuli were presented in temporal phase (see Figure 3). Monocular and binocular presentation for stimuli presented in the temporal phase resulted in a similar response pattern under both On/Off and counterphase flicker modulations. This is consistent with ocularity invariance; binocular and monocular stimuli are judged equal in magnitude at high contrast (Baker et al., 2018; Legge, 1984; Maehara & Goryo, 2005; Meese et al., 2006).

Changing the phase relationships of stimuli under On/Off flicker had some interesting impacts on the fundamental frequency (Figure 3 A). Stimuli presented in spatial phase and temporal anti-phase generated smaller SNRs ($\text{median}_{\text{SNR}} = 2.25$) than stimuli presented in spatial anti-phase and temporal phase ($\text{median}_{\text{SNR}} = 3.40$, $p = .047$). The reduction in the amplitude relative to stimuli in spatial anti-phase and temporal phase was also observed

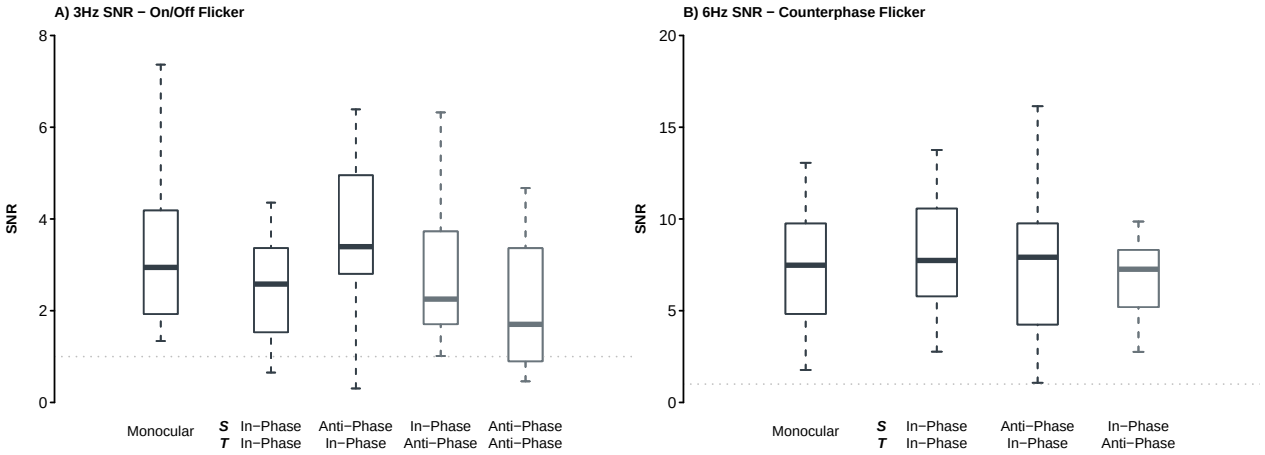

Figure 3: Boxplots of participant SNRs at the fundamental frequency for each experimental conditions in our study. A) Boxplots represent the participant SNR at 3Hz. The median SNR is shown in the thick line within the box, with the lower and upper border of the box represent the first (25%) and third (75%) quartile of the SNR distribution. Dashed lines show the lower and upper whisker limits, which are calculated as 1.5 times the interquartile range (distance between the third and first quartile). Boxplots for the binocular conditions have labels S for their spatial phase relationship and label T for their temporal phase relationship. Experimental conditions where stimuli are presented in temporal anti-phase are shown in a lighter gray. B) As in A, boxplots show participant SNRs at 6Hz, the fundamental frequency for counterphase flicker. In both graphs, the dashed line represents an SNR of 1.0.

for stimuli in temporal and spatial anti-phase ($\text{median}_{\text{SNR}} = 1.70$; $p = .023$). No other statistically significant difference in median SNRs were observed for all counterphase flicker conditions or the other harmonics for On/Off flicker (all p values were greater than .05). While the median SNRs under On/Off flicker shown in temporal anti-phase were reduced in comparison to other conditions, both the spatial phase temporal anti-phase condition ($p < .001$) and the spatial and temporal anti-phase conditions ($p = .004$) had median SNR values that were statistically significantly greater than 1.0. A 3Hz response for binocular stimuli presented in temporal anti-phase indicates monocular responses remain and contribute to the SSVEP, as these conditions generate two transients per cycle and a purely binocular signal would only generate responses at 6Hz (Blake et al., 1981; Georgeson et al., 2016; Moulden, 1980).

Modelling

The perception of stimulus contrast across eyes is well-explained by psychophysical models that process input contrast in two sequential contrast gain control stages interlaced by binocular summation (Baker, Meese, & Georgeson, 2007; Baker, Meese, & Summers, 2007; Baker et al., 2008; Baker & Meese, 2007; Meese et al., 2006). This simple, yet powerful, family of models not only captures behavioural data well, but can also explain neural responses to binocular and dichoptic stimuli (Baker & Wade, 2017; Lygo et al., 2021; Richard et al., 2018). Our SSVEP results show the expected pattern of binocular combination for stimuli presented at high contrast (i.e., ocularity invariance) but also intriguing effects that are

likely explainable by the most current version of the two-stage contrast gain control model, as defined in Georgeson et al. (2016). To explore the architecture required to describe our effects adequately, we progressively increase the complexity of binocular combination beginning with a wrong model (i.e., linear combination) and building up to a multi-channel model with monocular, binocular and phase-selective pathways (Figure 4).

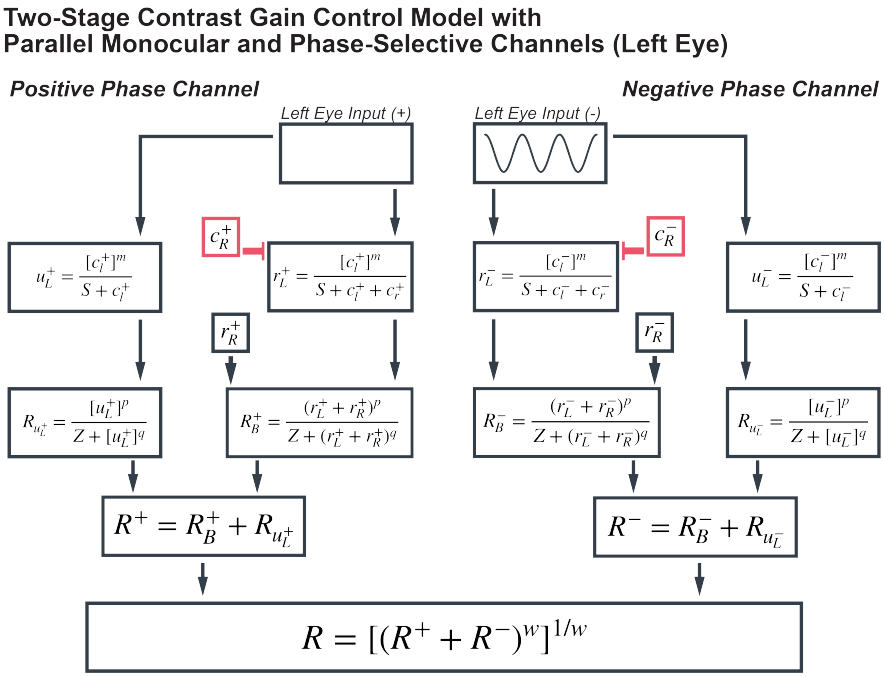

Figure 4: The multi-channel variant of the two-stage contrast gain control model. This diagram shows the channels for the left eye. Contributions from the right eye (c_R^+ and r_R^+) to the binocular channels are shown in small boxes. Unlike the model defined by Georgeson et al. (2016), responses from the parallel monocular channels are added to those of the binocular channels before signal selection for phase-selective channels. This change accounts for methodological differences when fitting neuroimaging data. To fit the final response of the model, R, was Fast Fourier Transformed, and a pink noise spectrum was added to allow for a comparable calculation of model SNRs as is done with human data.

The architecture of the models explored differs, but all received the same input and had their final outputs processed identically. The input to all models was a 3 Hz sine wave, adjusted to accurately represent the various experimental conditions of this study (see Figure 1). For example, stimuli presented with On/Off flicker in temporal anti-phase had the left eye input generated by the following equation,

$$c_l = A * (\cos(2\pi f t) + 1)/2 \tag{4}$$

while the right eye input would be defined as,

$$c_r = A * (-\cos(2\pi f t) + 1)/2. \tag{5}$$

A represents stimulus contrast (amplitude), f the temporal flicker frequency (i.e., 3 Hz), and t time in milliseconds. The input to the other eye (c_r) is phase shifted by 180^o, which can be

accomplished using the negative cosine function ($-cos$). Finally, sine waves are rectified to range between 0 and 1 to represent the relative contrast presented to observers. The same experimental condition with counterphase flicker has the following sinusoidal profile for the left eye,

$$c_l = A * [\cos(2\pi ft)]_+ \tag{6}$$

and for the right eye,

$$c_r = A * [-\cos(2\pi ft)]_+. \tag{7}$$

These profiles are identical to the On/Off flicker, but the sine waves are half-wave rectified to represent the counterphase oscillation. To fit model outputs (rectified sine waves) to observer data, the final response of the models was Fast Fourier Transformed, and a pink noise spectrum was added to the Fourier amplitude, $|FFT(R_{\text{model}})| + 1/f$, before calculating model SNRs. All models developed in this study were fit by minimizing the sum of squared errors between the model output and the observer median SNRs for the first 6 SSVEP components (3Hz, 6Hz, 9Hz, 12Hz, 15Hz, and 18Hz).

Evidently wrong models

As a first step in defining the necessary architecture to capture our results, we built wrong models with no monocular stage or phase selectivity. The first is a purely linear summation model of binocular combination,

$$R_B = c_l + c_r, \tag{8}$$

the binocular response (R_B) is a sum of the monocular inputs. The fits of the linear summation model are shown in Figure 5 and its performance metrics in Table 1. For On/Off flicker, the linear summation model only generates responses at the fundamental frequency (3Hz) that grossly overestimate observer SNRs. This is expected as this model lacks the rectification and non-linearities required to generate responses at the harmonics (Wade & Baker, 2025). In a linear sum, stimuli presented under On/Off flicker in temporal anti-phase cancel each other, and thus the model generates no response. The model does generate responses at the fundamental and harmonics of the counterphase flicker condition (Figure 5 B), but this is attributable to the input's rectification (Equation 6, Equation 7) and not the model architecture.

Model	R^2	$\text{RMS}_{\text{error}}$	AIC
Linear sum	-	3.234	281.99
Linear sum, with Rectification	0.648	1.278	187.76
Two-Stage, no interocular interactions	0.587	1.385	200.44
Two-Stage, with interocular interactions	0.698	1.184	183.47
Two-Stage with parallel monocular channels	0.900	0.681	123.81
Two-Stage with phase-selective channels & monocular channels	0.908	0.654	121.40

Table 1: Goodness-of-fit metrics for all models developed in this study. Errors in predictions for the linear sum model were too large to calculate R^2. $\text{RMS}_{\text{error}}$ is the Root Mean Square error and AIC is the Aikaike Information Criterion.

Adding a rectification to the inputs and normalization aids in capturing responses to harmonic frequencies,

$$R_B = \frac{(c_L + c_R)^p}{Z + (c_L + c_R)^q}. \tag{9}$$

In this slightly improved iteration of a linear summation model, the binocular response of the model is defined as the sum of monocular inputs raised to the power p, normalized by the sum of monocular inputs raised to the power q. The parameter Z prevents division by zero. The model slightly improves the fits to human SNRs as it can now generate responses at the harmonics (see Figure 5 and Table 1). It struggles, however, at capturing SNR values for the fundamental frequency (3Hz) and is, as with the linear summation model, incapable of describing SNRs to stimuli presented with On/Off flicker in temporal anti-phase.

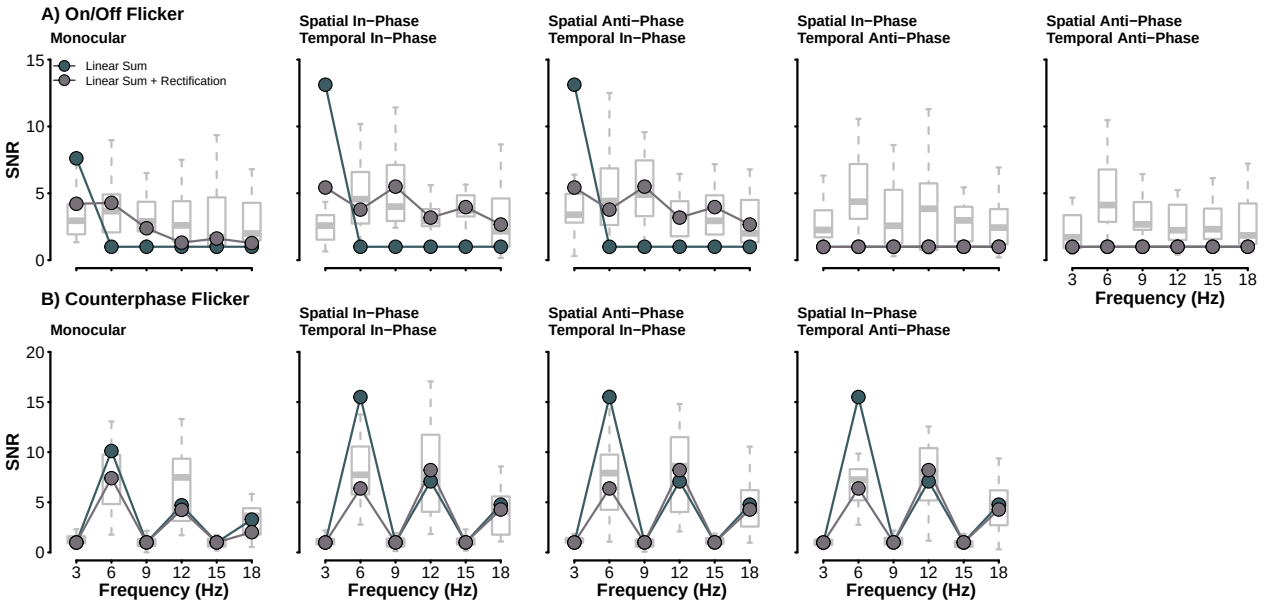

Figure 5: Fits of the linear sum (green) and the rectified linear sum (brown) models. Boxplots behind model responses show the distribution of observer SNRs. Model SNRs were fit to the median SNR of observers, which is represented by the thicker line within the box.

The Two-Stage Contrast Gain Control Model

The simple models described above could not accurately represent our study's SSVEPs. They overestimated SNRs at the fundamental frequency and failed to generate responses for stimuli presented in temporal anti-phase with On/Off flicker. We can account for the magnitude of responses by adding a monocular stage, with its rectification and normalization, for the left eye,

$$r_L = \frac{c_L^m}{S + c_L} \tag{10}$$

with the same equation written for the right eye. The two rectified and normalized monocular responses are then fed to the binocular stage (Meese et al., 2006; Meese & Baker, 2011),

$$R_B = \frac{(r_L + r_R)^p}{Z + (r_L + r_R)^q}. \tag{11}$$

In this variant, m is the monocular stage's excitatory component and determines the summation degree. The monocular input also undergoes self-suppression before being fed into the binocular stage. Unlike the original two-stage contrast gain control model, this iteration does not include an inter-ocular suppression term, which impacts the model's performance (see Table 1). The suppression at the monocular stage is not large enough to reduce responses at the fundamental frequency (Figure 6). While it does generate responses for stimuli presented in temporal anti-phase, as the rectification prevents their sum from cancelling, they are too small in magnitude to capture observer SNRs.

Strengthening the normalization of the monocular stage by adding an interocular suppression term,

$$r_L = \frac{c_L^m}{S + c_L + c_R}, \tag{12}$$

improves the fits at the fundamental frequency by reducing the SNR magnitude of the model (see Figure 6) and generates larger responses for stimuli presented in temporal anti-phase (see Table 1). As in the original two-stage contrast gain control model, while the excitatory component of the monocular stage may be large, the suppression received at both the monocular stage and at the subsequent binocular stage keeps responses low, better representing human SNRs (Baker et al., 2018; Meese et al., 2006). This model does struggle to fit SNRs at the fundamental frequency and odd-harmonics (3F, 5F) for stimuli presented in temporal anti-phase under On/Off stimulation. As discussed above, these responses are likely driven by monocular signals. The two-stage contrast gain control model does not preserve the monocular response after the first stage.

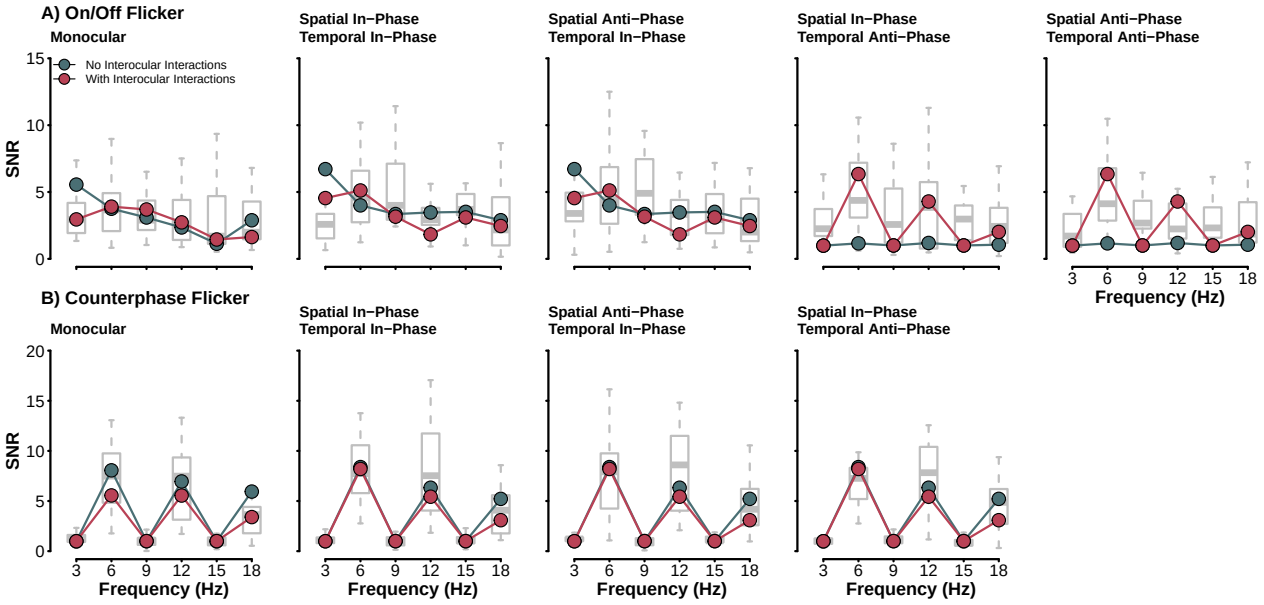

Figure 6: Fits of the two-stage contrast gain control model without (green) and with (red) interocular suppression. Boxplots behind model responses show the distribution of observer SNRs. Model SNRs were fit to the median SNR of observers, which is represented by the thicker line within the box.

Phase-Selective and Parallel Monocular Channels

To maintain the monocular responses until the final stage of the model, we add parallel monocular channels as was done by Georgeson et al. (2016). These channels are fully monocular and therefore have no interocular suppression term,

$$\mu_L = \frac{C_L^m}{S + C_L}, \mu_R = \frac{C_R^m}{S + C_R}, \tag{13}$$

where μ_L is the output of the first stage of the monocular channel, while μ_R is that of the right eye. The excitatory exponent m is identical to that of the binocular channel. The output of the monocular channel undergoes a second rectification and normalization, identical to that of the binocular channel,

$$R_{\mu_L} = \frac{\mu_L^p}{Z + \mu_L^q}, R_{\mu_R} = \frac{\mu_R^p}{Z + \mu_R^q}, \tag{14}$$

where R_L and R_R represent the final responses of the left and right monocular channels. The parameters m, p, q, S and Z are identical to those of the binocular channel and thus no additional free parameters are required to define the parallel monocular channels. The addition of these channels poses an interesting problem regarding cue selection. The final stage of this model has three channel responses: monocular left (R_L), monocular right (R_R), and binocular (R_B). In behavioural variants of this model Georgeson et al. (2016), cue selection is implemented as a MAX rule, as a Minkowski sum with a very large (≈ 30) exponent. This method of signal combination was not appropriate for our data. The binocular response is always larger than the monocular and thus dominates in the final signal, preventing the model from capturing the fundamental and odd-harmonics responses for anti-phase stimuli. Instead, we found that a linear combination of the binocular signal and the monocular signal,

$$R = R_B + R_{\mu_L}, \tag{15}$$

generated better fits to our data. In our model fits, we use the monocular response from the left eye (R_{ _L}) to add to the binocular channel. Adding parallel monocular channels to the two-stage contrast gain control model significantly improved the fit of our observer data (Figure 7). With the monocular response preserved, the model can now generate SSVEPs at the fundamental frequency and odd-harmonics necessary to capture observer data for stimuli presented in temporal anti-phase under On/Off flicker (see Table 1). Thus, monocular responses to binocular stimuli are preserved along the processing pipeline and can be recorded in the SSVEPs of observers.

Our results do not show a strong effect of stimulus spatial phase on observer SSVEPs. Still, it is important to determine whether adding phase-selective channels to the two-stage contrast gain control model highlights a subtle impact of phase-selective responses. We add phase-selective channels to the binocular and monocular channels of the two-stage contrast gain control model. Phase-selectivity required replicating the equations from both stages to have binocular and monocular channels selective for either positive or negative phase (see Figure 4). The phase-selective channels in this model are completely independent; they do not interact with each other. This model iteration, which we refer to as the full model, replicates the two-stage contrast gain control model developed by Georgeson et al. (2016).

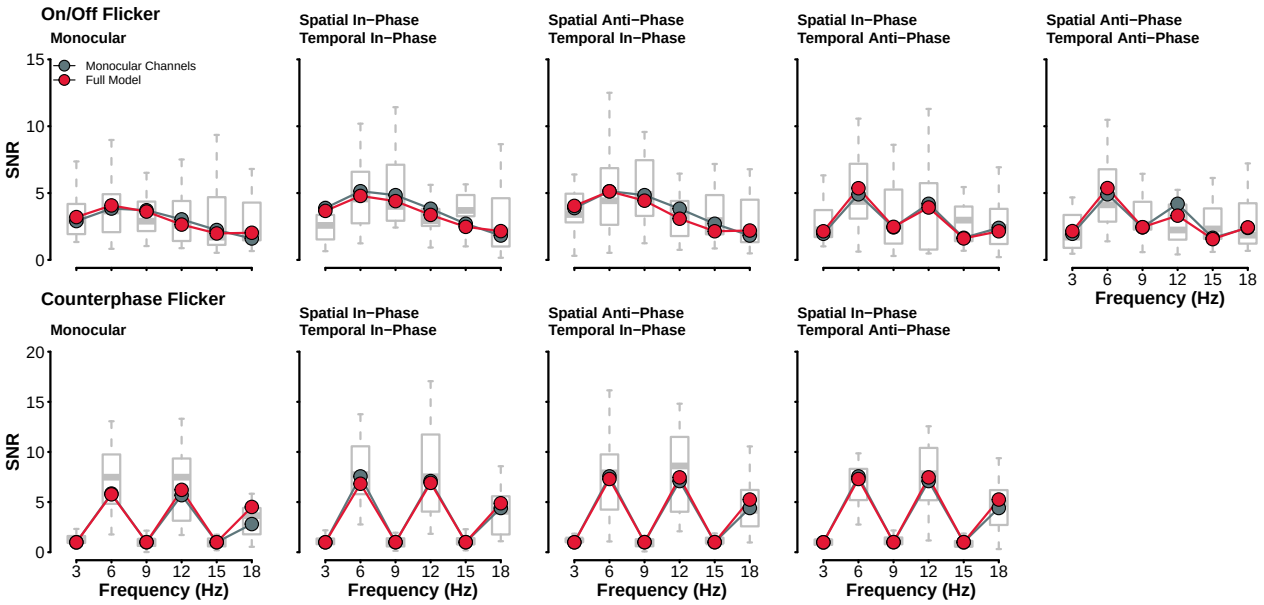

Figure 7: Fits of the two best performing models to our observer data. Both models, that including parallel monocular channels and the full model with the added phase-selective channels perform quite similarly. The two stage contrast gain control model with parallel monocular channels is sufficient to explain the effects in our data. The best fitting parameters for the monocular channels model were: m = 2.68, p = 8.72, q = 2.03, S = 0.25, Z = -0.02. These parameters suggest strong binocular combination at the monocular stage.

As we encountered above with the addition of monocular channels, the addition of phase-selectivity means that we now have two final signals following the combination of the binocular and monocular channels (Equation 15), a positive phase and a negative phase signal. To generate the final model SSVEP, we utilize a MAX rule implemented as a Minkowski sum, as was done in Georgeson et al. (2016),

$$R_{\text{MAX}} = [(R^{+})^{w} + (R^{-})^{w}]^{\frac{1}{w}}. \tag{16}$$

We introduce the new parameter w, which is expected to be large (> 20) to select the phase channel, either positive or negative, that generated the largest response. The addition of phase-selectivity to our model improved the fits slightly as the overall R^2 was increased. The Aikaike Information Criterion (AIC) of the phase selective model (AIC = 121.4) is slightly smaller than that of the two-stage contrast gain control model with parallel monocular channels (AIC = 123.81). While the effects of spatial phase on observer SSVEPs are small, accounting for its impact on our model by adding phase-selective channels offers a better description of our results.

While the whole model outperformed the others, the simpler models are nested variants and thus difficult to compare using R^2 and AIC values alone. We use an F test for nested models to determine if the difference in the quality of fits between our smaller models and the full model is statistically significant. The full results of the F tests are reported in Table 2. We do not find a statistically significant difference between the fits of the two-stage contrast gain control with parallel monocular channels and the full model, $F(1, 47) = 4.00, p = .051$.

While the full model does improve model fits, its contributions to explaining our findings are not sufficiently large to pass the F test for nested models, at an α level of .05.

Model	df	SS	MS	F	p
Linear sum	6	541.54	90.26	183.61	<.001
Linear sum, with Rectification	3	65.12	21.71	44.16	<.001
Two-Stage, no interocular interactions	1	80.52	80.52	163.8	<.001
Two-Stage, with interocular interactions	1	52.58	52.58	106.95	<.001
Two-Stage with parallel monocular channels	1	1.97	1.97	4	0.051

Table 2: Results of the F test for nested models. The full model - two-stage contrast gain control model with phase selective and parallel monocular channels - outperformed all but the two-stage model with parallel monocular channels at an α level of .05.

Discussion

References

Bacon, J. H. (1976). The interaction of dichoptically presented spatial gratings [Journal Article]. *Vision Res*, *16*(4), 337–344. https://doi.org/10.1016/0042-6989(76)90193-0

Baker, D. H., Lygo, F. A., Meese, T. S., & Georgeson, M. A. (2018). Binocular summation revisited: Beyond $\sqrt{2}$. *Psychol Bull*, *144*(11), 1186–1199. https://doi.org/10.1037/bul0000163

Baker, D. H., & Meese, T. S. (2007). Binocular contrast interactions: Dichoptic masking is not a single process [Journal Article]. *Vision Res*, *47*(24), 3096–3107. https://doi.org/10.1016/j.visres.2007.08.013

Baker, D. H., Meese, T. S., & Georgeson, M. A. (2007). Binocular interaction: Contrast matching and contrast discrimination are predicted by the same model [Journal Article]. *Spat Vis*, *20*(5), 397–413. https://doi.org/10.1163/156856807781503622

Baker, D. H., Meese, T. S., & Hess, R. F. (2008). Contrast masking in strabismic amblyopia: Attenuation, noise, interocular suppression and binocular summation [Journal Article]. *Vision Res*, *48*(15), 1625–1640. https://doi.org/10.1016/j.visres.2008.04.017

Baker, D. H., Meese, T. S., & Summers, R. J. (2007). Psychophysical evidence for two routes to suppression before binocular summation of signals in human vision. *Neuroscience*, *146*(1), 435–448. https://doi.org/https://doi.org/10.1016/j.neuroscience.2007.01.030

Baker, D. H., & Wade, A. R. (2017). Evidence for an optimal algorithm underlying signal combination in human visual cortex [Journal Article]. *Cereb Cortex*, *27*(1), 254–264. https://doi.org/10.1093/cercor/bhw395

Blake, R. (1989). A neural theory of binocular rivalry [Journal Article]. *Psychol Rev*, *96*(1), 145–167. https://doi.org/10.1037/0033-295x.96.1.145

Blake, R., Overton, R., & Lema-Stern, S. (1981). Interocular transfer of visual aftereffects [Journal Article]. *J Exp Psychol Hum Percept Perform*, *7*(2), 367–381. https://doi.org/10.1037//0096-1523.7.2.367

Blake, R., & Wilson, H. (2011). Binocular vision. *Vision Research*, *51*(7), 754–770. https://doi.org/https://doi.org/10.1016/j.visres.2010.10.009

Campbell, F. W., & Green, D. G. (1965). Monocular versus binocular visual acuity [Journal Article]. *Nature*, *208*(5006), 191–192. https://doi.org/10.1038/208191a0

Chatrian, G. E., Lettich, E., & Nelson, P. L. (1985). Ten percent electrode system for topographic studies of spontaneous and evoked EEG activities [Journal Article]. *American Journal of EEG Technology*, *25*(2), 83–92. https://doi.org/10.1080/00029238.1985.11080163

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*, 9–21. https://doi.org/10.1016/J.JNEUMETH.2003.10.009

Ding, J., Klein, S. A., & Levi, D. M. (2013). Binocular combination of phase and contrast explained by a gain-control and gain-enhancement model [Journal Article]. *J Vis*, *13*(2), 13. https://doi.org/10.1167/13.2.13

Ding, J., & Sperling, G. (2006). A gain-control theory of binocular combination [Journal Article]. *Proc Natl Acad Sci U S A*, *103*(4), 1141–1146. https://doi.org/10.1073/pnas.0509629103

Georgeson, M. A., Wallis, S. A., Meese, T. S., & Baker, D. H. (2016). Contrast and lustre: A model that accounts for eleven different forms of contrast discrimination in binocular vision. *Vision Research*, *129*, 98–118. https://doi.org/10.1016/j.visres.2016.08.001

Hansen, B. C., & Hess, R. F. (2006). Discrimination of amplitude spectrum slope in the fovea and parafovea and the local amplitude distributions of natural scene imagery [Journal Article]. *J Vis*, *6*(7), 696–711. https://doi.org/10.1167/6.7.3

Legge, G. E. (1984). Binocular contrast summation–i. Detection and discrimination [Journal Article]. *Vision Res*, *24*(4), 373–383. https://doi.org/10.1016/0042-6989(84)90063-4

Lygo, F. A., Richard, B., Wade, A. R., Morland, A. B., & Baker, D. H. (2021). Neural markers of suppression in impaired binocular vision [Journal Article]. *Neuroimage*, *230*(October 2020), 117780. https://doi.org/10.1016/j.neuroimage.2021.117780

Maehara, G., & Goryo, K. (2005). Binocular, monocular and dichoptic pattern masking [Journal Article]. *Optical Review*, *12*(2), 76–82. https://doi.org/DOI 10.1007/s10043-004-0076-5

Meese, T. S., & Baker, D. H. (2011). Contrast summation across eyes and space is revealed along the entire dipper function by a "swiss cheese" stimulus [Journal Article]. *J Vis*, *11*(1), 1–23. https://doi.org/10.1167/11.1.23

Meese, T. S., Georgeson, M. A., & Baker, D. H. (2006). Binocular contrast vision at and above threshold. *Journal of Vision*, *6*, 1224–1243. https://doi.org/10.1167/6.11.7

Moradi, F., & Heeger, D. J. (2009). Inter-ocular contrast normalization in human visual cortex [Journal Article]. *J Vis*, *9*(3), 13 1–22. https://doi.org/10.1167/9.3.13

Moulden, B. (1980). After-effects and the integration of patterns of neural activity within a channel [Journal Article]. *Philos Trans R Soc Lond B Biol Sci*, *290*(1038), 39–55. https://doi.org/10.1098/rstb.1980.0081

Richard, B., Chadnova, E., & Baker, D. H. (2018). Binocular vision adaptively suppresses

delayed monocular signals [Journal Article]. *Neuroimage*, *172*, 753–765. https://doi.org/10.1016/j.neuroimage.2018.02.021

Simmons, D. R. (2005). The binocular combination of chromatic contrast [Journal Article]. *Perception*, *34*(8), 1035–1042. https://doi.org/10.1068/p5279

Simmons, D. R., & Kingdom, F. A. A. (1998). On the binocular summation of chromatic contrast [Journal Article]. *Vision Research*, *38*(8), 1063–1071. https://doi.org/10.1016/S0042-6989(97)00272-1

Tadmor, Y., & Tolhurst, D. J. (1994). Discrimination of changes in the second-order statistics of natural and synthetic images [Journal Article]. *Vision Research*, *34*(4), 541–554. https://doi.org/10.1016/0042-6989(94)90167-8

Wade, A. R., & Baker, D. H. (2025). Measuring contrast processing in the visual system using the steady state visually evoked potential (SSVEP). *Vision Research*, *231*, 108614. https://doi.org/https://doi.org/10.1016/j.visres.2025.108614

Wilson, H. R. (2003). Computational evidence for a rivalry hierarchy in vision [Journal Article]. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(24), 14499–14503. https://doi.org/10.1073/pnas.2333622100